A multi-criteria assessment to determine the customers' technology preference in the context of apparel e-commerce

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Ruchika Kalpoe

Student number: 4192001

To be defended in public July 9th 2020

Graduation committee

Chairperson	: Prof. dr. G.P.B. Van Wee, Section Transport & Logistics
First Supervisor	: Dr. J. Rezaei, Section Transport & Logistics
Second Supervisor	: H. Asghari, Section Organization & Governance

PREFACE AND ACKNOWLEDGEMENT

This documents beholds my master thesis, conducted as fulfilment of the Master program Complex Systems Engineering and Management (CoSEM) at the Delft University of Technology. This master thesis aims to explore what the customers' acceptance is regarding various technological alternatives designed to increase their online purchase successes of apparel items and reduce unnecessary apparel returns for online apparel retailers in the Netherlands. In order to reach this goal, various methods and tools have been used which were taught to me during the CoSEM Master program. These include interviews, a survey and the application of the novel Bayesian Best-Worst Method (BWM). With the support of the excellent graduation committee, these methods and tools could be successfully employed and the results could be critically examined.

I first of all would like to thank my first supervisor Jafar Rezaei. Not only has he showed great availability throughout the master thesis, but he also provided great insight, direction and useful practical feedback regarding the applied BWM method, survey and interview to fit this specific research case. This has allowed me to critically review my decisions and improve my work during the research. My second supervisor, Hadi Asghari, has provided me with useful feedback in determining the set of decision-criteria, allowing me to critically examine the applicability of each criterion to the specific apparel e-commerce empirical case. Furthermore, he has provided useful feedback regarding the formulation of the survey questions, especially the additional demographical questions. As a result, only the most important questions for this research were included in an ethical way. My chairman Bert van Wee has demonstrated his extensive knowledge by providing useful feedback and clear directions, especially in the theoretical domain.

The combination of practical feedback gained from my first supervisor, the theoretical directions of my chairman and the practical guidelines from my second supervisor, have led me to critically review my delivered work and successfully complete my master thesis at the TU Delft. I would sincerely like to thank all three members of my graduation committee for their delivered input, since without their guidance and support, I would not have been able to do it, especially in these times of a COVID-19 pandemic.

Aside from the committee, I want to thank Jan Anne Annema as well, since with his support I have managed to scope my research such that it could be conducted within the comprised time perioded. Furthermore, he also provided clear and helpful directions, which assisted me to seek the right persons at the faculty of TPM during my master thesis. I also want to thank Prof. Langerak from TU Eindhoven for providing me with necessary feedback regarding the applied theory and set of decision-criteria. Furthermore, I would like to thank Kenny Meesters for providing me with useful feedback regarding the set of decision criteria. I also want to thank the six approached online apparel e-commerce experts for their delivered input and time during the interviews. Without the input of all these people, the research could have also not be successfully completed.

Finally, I want to sincerely thank my family and friends, for providing me with the necessary support and motivation during my master thesis. Thank you all!

Ruchika Kalpoe

SUMMARY

Returns management, especially in the apparel e-commerce industry, has gained increased attention due to the ecological and economic implications the returns of apparel items impose. However, up till now, research has mainly focused on addressing the logistical issues rather than examining how customers' apparel returns can be address during the online screen process of apparel items.

More specifically, research which explores 1) factors which drive customers' apparel returns, 2) customer-based instruments designed to better inform customers during the online screening process of apparel and 3) the users' (customers) preference with respect to these technologies, has not yet been empirically examined in literature. Therefore, this research strives to contribute to this gap by exploring these three aspects. Consequently, the main research question is: "*What is the customers' preference regarding technological alternatives online apparel retailers can employ, during the customers' online apparel screening process, in order to increase customers' online apparel purchase successes and reduce unnecessary apparel returns*?" In order to address this research question, a Multi-Criteria Analysis (MCA) approach is used as this approach suits well with the exploratory nature of this research, since it is too early to quantify the actual impact on apparel returns and impact in monetary terms. Based upon the MCA approach, the research is categorized into the following parts.

The first part is devoted to finding technological alternatives to address the goal of increasing customers' online purchase successes and refrain them from returning apparel items. A literature study was conducted through which first various return reasons were established. Once various return reasons were established, the required apparel attribute information to evaluate apparel items online with was extracted, which are: 1) material information, 2) colour information, 3) fit and size information and 4) style information. Furthermore, the literature study has shown that creating the right expectations regarding the apparel attributes, providing accurate product information, and creating a 'feel' for and perception of apparel items displayed online, were also essential (functional) requirements which need to be considered when designing the technological alternatives. Since customer-based instruments are designed to provide customers online with the required apparel attribute information upfront such that they can make well informed pre-purchase decisions, only return reasons were included which could be addressed by the customer-based instruments.

The literature study had also resulted in the identification of various customers-based instruments. Since some of these included instruments and technologies which on their own cannot provide / visualize all the aforementioned required apparel attribute information nor the other required functionalities, it was necessary to combine some instruments such that they can fulfil the requirements and function as comparable alternatives against the technologies which on their own can fulfil all the requirements. As a result, the following four technological alternatives were established

A1: The bare minimum, which contains: alternative product pictures with mix and match function (to see the overall outfit), zoom function and static height/size chart.
A2: The bare minimum with a fit & size recommendation instrument
A3: Avatar (digital computer-based twin)
A4: Virtual Dressing Room (VDR)

The combination was necessary, as online apparel retailers currently only use some of the instruments combined and not separately on their website. Taking the current way of employment of instruments into consideration, practical and employable alternatives were designed. This was necessary as the goal of this research is to identify what the customers preference is with regard to various technological concepts online apparel retailers can actually employ in practice. The combined technological alternatives were composed based upon one another. The first combined alternative constitutes a basic alternative which is close to the current arrangements used by online apparel retailers in the Netherland. The second combined alternative will attach a new functionality to the previous alternative, while

keeping all other aspects alike. As such A2 builds upon A1 by adding a fit & size recommendation function to A1. A3 builds upon A2 by adding a computerized virtual try-on experience to A2. A4 builds upon A3 by providing a more realistic virtual try-on experience compared to A3. Since the technologies are designed to be used by customers, its success relies greatly on the customers usage. Therefore, the research is mainly approached from the users' (customers') perspective.

Following the MCA approach, the second part consisted of establishing a set of criteria to evaluate the technological alternatives. For this, a literature study regarding the Technology Acceptance Model (TAM) was conducted, within the context of apparel e-commerce. Through the literature study, many criteria were established. Compared to other theories which were acknowledged, TAM was still preferred, since it suits the practical goal of this research the most and its simplicity in predicting the users technology acceptance, by solely two determinants:1) Perceived Usefulness (PU) and 2) Perceived Ease of Use (PEU). After evaluating the criteria, based on the practicality of this research, the decision was made to also include the determinant 'Trust' as additional third determinant to the original TAM.

Since many criteria were established, first a categorization of the criteria was required. As a result, the following main-criteria were used in the research: 1) quality of provided information, 2) information gathering and handling and 3) user-friendliness. Based on the literature study regarding TAM, the main-criterion 'quality of provided information' is mostly perceived as significant external predictor of the determinant PU, the main-criterion 'information gathering and handling' is mostly perceived as significant external predictor of Trust and the main-criterion 'user-friendliness' is mostly perceived as significant external predictor of the determinant PEU. Based on the applicability for the online e-commerce case, the decision was made to also include the three main-criteria and their sub-criteria as such in the research. The sub-criteria belonging to the main-criterion 'quality of provided information' are: 1) reliability of material information, 2) reliability of colour information, 3) reliability of fit & size information and 4) reliability of style information. The sub-criteria belonging to the main-criterion 'information gathering and handling' are: 1) the way of data collection through technology and 2) data handling by online clothing retailer. The sub-criteria belonging to the main-criterion 'user-friendliness' are: 1) responsiveness, 2) search time, 3) availability, 4) attractiveness and 5) required preparatory work time.

The third part consists of determining the optimal group weight per criterion. The preference elicitation method used in this research is the novel Bayesian BWM, through which the optimal group weights of each criterion could be established. Using a BWM imposed online survey, the required input data was obtained for the target group which were individuals which have online shopping experience.

The fourth part consisted of obtaining the performance card and determining the customers technological preference. The scores of each alternative with respect to each criterion was established through six apparel e-commerce experts interviews, using the same imposed structure of the BWM. Using the weighted sum equation, the users' technological preference was determined.

The survey results showed that from all 11 sub-indicators, reliable fit & size information is perceived as the most important criterion affecting users' technology acceptance. Based on this, the observation can be made that the Bayesian BWM is indeed a valid method to predict the importance of criteria, since the literature studies and expert interviews have indeed shown that apparel return reasons mostly stem from issues related to fit & size issues. The second most important criterion for users' technology acceptance, closely followed after 'reliability of fit & size information' is 'data handling by online clothing retailer. The third most important criterion is 'reliability of colour information', followed by 'reliability of material information' and 'the way of data collection through technology' as the fourth and fifth most important criteria for technology acceptance.

The results obtained from the apparel e-commerce expert interviews show that when the reliability of information provision increases, the perceived privacy and security concerns increases and the perceived ease of use decreases. Whilst e.g. A4 is perceived as the most useful (PU) when it comes to providing reliable material, colour and style information and very reliable in providing fit &size information, A4 scores the worst with respect to the privacy and security related criteria (information handling by online clothing retailer and the way of data collection through technology), which are perceived as the second and third most important criteria affecting the users' technology preference. This implies that privacy and security concerns do have a great impact on the customers' technology preference. Furthermore, A4 also scores very low with respect to the perceived ease of use criteria such as responsiveness, search time, availability and required preparatory work time. This is rather similar for A3, on which A4 builds upon qua functionality (by providing a more realistic virtual try-on experience).

Based on the obtained criteria-weights through the survey and the scores from online apparel retail experts, it can be observed that currently A1 is the most preferred alternative (has the highest chance of reaching users' technology acceptance) also proving that the Bayesian BWM is indeed a valid method to predict the users' technology acceptance since A1 is the closest alternative which is currently employed by online apparel companies. A2 is perceived as the second best, closely followed by A4. A3 is currently perceived as the least preferred alternative. As a result, existing companies (especially multi-brand stores) are advised to focus on A1 and A2, especially A2 since: 1) A2 scores better with respect to A1 in terms of providing reliable fit & size information and 2) A2 is perceived to not require that more time, money and expertise to implement compared to A1. New commers in the market are advised to focus on the criteria which have the highest weight. This might help them to become successful in the current market.

Furthermore, on a scientific and methodological level, this research contributes to existing literature of TAM in the field of apparel e-commerce, by identifying 11 criteria which play a role in the users' (online shoppers') technology acceptance. It also contributes to the empirical application of the novel Bayesian BWM in the specific field of apparel e-commerce and proves that through a less data extensive, simpler and even so reliable qualitative approach to operationalize TAM, the customer technology preference can be predicted as well.

Although this research indicated that A1 is the most preferred alternative, this cannot be guaranteed with full certainty. It could be that the technologies could co-exist in practice, since in time the current technological superiority of A1 and A2 over A3 and A4 might change. Since A4 and A3 score the best with respect to providing reliable apparel attribute information (material, colour, fit & size and style information), which is required to make better online pre-purchase decisions and refrain customers' from returning items, future research could examine if they can co-exist. Since the survey results have indicated that reliable fit & size information and privacy and security related criteria are of high importance for the users technology acceptance, there is advised to further examine the different approaches that can be applied to obtain customer body-measurement data for A2, A3 and A4 such that reliable recommendations regarding e.g. fit & size information can be provided whilst preserving customers' privacy and security.

Since this research is still in the exploratory phase, it was too early to monetize the impact and actual efficiency gains such as apparel returns reduction and cost reduction. Therefore, online apparel retailers are advised to perform a cost-benefit-analysis (CBA) study for each technological alternative, in order to determine their actual (additional) feasibility. In the CBA, different technological variables from this recommended additional research, and other regulatory (privacy and security) and financial variables need to be included as well.

Table of Contents

PREFACE AND ACKNOWLEDGEMENT 2

SUMMARY 3

List of figures 8

List of tables 8

1. INTRODUCTION 9

1.1 Knowledge gap and added value of research 9

1.1.1 Scientific knowledge gap 11

1.1.2 Main research question 13

1.2 Research approach and sub-questions 14

1.3 Research methods 15

1.4 Link to MSc CoSEM Program 27

1.5 Report structure 28

2. IDENTIFYING RETURN REASONS & DESIGNING TECHNOLOGICAL ALTERNATIVES 29

2.1 Literature review process 29

2.2 Identified reasons for apparel returns 31

2.3 Interpretation of apparel return reasons 34

2.4 Identified instruments and technologies 34

2.4.1 Included technologies and instruments 42

2.5 Designing technological alternatives 42

2.5.1 Combining instruments 43

2.5.2 Technological alternatives 43

3. AQUISITION OF EVALUATION CRITERIA 48

3.1 Literature Study on the Technology Acceptance Model 48

3.1.1 Predictive validity of TAM 48

3.1.2 Results 49

3.1.3 Interpretation of results 56

3.2 Selection of evaluation criteria 57

3.2.1 Categorization of identified criteria 57

3.2.2 Category selection 58

3.2.3 Finalizing the set of evaluation criteria 60

4. OBTAINING CRITERIA WEIGHTS 66

4.1 Survey as data collection tool 66

4.1.1 Target group 66

4.1.2 Data collection procedure 66

4.1.3 Survey design 66

4.2 Data analysis method 67
4.3 Results 68
4.3.1 Survey Respondents 68
4.3.2 Obtained group criteria weights 70
5. ESTABLISHING CUSTOMERS' PREFERENCE 74
5.1 Data collection tool 74
5.1.1 Target group 74
5.1.2 Data collection process 75
5.1.3 Interview design 75
5.2 Interview results 76
5.3 Customers' preference regarding technological alternatives 79
6. MAPPING THE RESULTS 80
6.1 Data collection tool 80
6.2 Results 81
7. DISCUSSION, CONLUSION & RECOMMENDATIONS 84
7.1 Practical level 84
7.2 Scientific and methodological level 87
REFERENCES 89
APPENDIX A: Scientific Paper 95
APPENDIX B: Established categories of criteria 115
APPENDIX C: Survey used to obtain criteria weights 117
APPENDIX D: Survey results (based on BWM users' survey) 122
APPENDIX E: Interview with apparel e-commerce experts 124
APPENDIX F: Input data retrieved from BWM Interview 139
APPENDIX G: Calculated scores per alternative with respect to each sub-criterion 143

List of figures

Figure 1: Research flow diagram *16*
Figure 2: Technology Acceptance Model *48*
Figure 3: Hierarchy of criteria to evaluate the technologies *65*
Figure 4: Credal ranking of main-criteria *70*
Figure 5: Credal ranking of sub-criteria belonging to 'quality of provided information' *72*
Figure 6: Credal ranking of sub-criteria belonging to 'information gathering and handling' *72*
Figure 7: Credal ranking of sub-criteria belonging to 'quality of provided information' *73*
Figure 8: Alternatives' scores with respect to criterion 'implementation possibility in company' *81*
Figure 9: Credal ranking of alternatives regarding 'employment possibility in company' *83*

List of tables

Table 1: Overview of reviewed literature 10
Table 2: Definition of scores from 1- 9. 26
Table 3: Overview of additional reviewed literature for SQ1 30
Table 4: Identified return reasons regarding apparel attributes 31
Table 5: Customer-based instruments and technologies 35
Table 6: Technological alternatives in columns with their relevant components 44
Table 7: Identified additional determinants and external predictors through literature 50
Table 8: Categorization of criteria 58
Table 9: First subset of criteria 61
Table 10: Second subset of criteria 62
Table 11: Third subset of criteria 65
Table 12: Socio-demographic characteristics 69
Table 13: Users' optimal group weights per main-criterion based upon BWM survey. 70
Table 14: Users' optimal groups' weights per sub-criterion based upon BWM survey 71
Table 15: Characteristics of interviewees 74
Table 16: Alternatives' scores with respect to the sub-criteria 77
Table 17: Ranking of the technological alternatives 79

1. INTRODUCTION

The emergence of e-commerce in the apparel retail industry has fundamentally changed consumers' purchasing behavior. Nowadays, people do not necessarily have to travel to the physical shops to purchase apparel and are more often fitting clothes at home than in the fitting room of stores itself. As a result, less physical space is required to sell and purchase consumer goods (De Leeuw, Minguela-Rata, Sabet, Boter, & Sigurðardóttir, 2016; Kennisinstituut voor Mobiliteitsbeleid, 2017).

Social and Scientific relevance

Although e-commerce has its benefits, it also imposes societal implications. With the increase of online purchases, the number of order returns also increases (Minnema, Bijmolt, & Gensler, 2017). According to Minnema et al. (2017), approximately 30% of all online purchases in the Netherlands are returned to the sender, which imposes structural problems for online retailers. Of all returned products, apparel is the biggest part. According to Wiese, Toporowski, & Zielke (2012), returns for apparel items are more common than for most other products, due to the many apparel attributes. Of all the returned products bought online, 40% are apparel items (Edwards, McKinnon, & Cullinane, 2010).

A few of these implications include: extra quality checks, extra administrative work, re-packaging and storing which furthermore results in an increase of logistic costs (Kennisinstituut voor Mobiliteitsbeleid, 2017). Moreover returns processing, especially checking the quality of the returned goods, is very labour intensive (Griffis, Rao, Goldsby, & Niranjan, 2012). Also, when products are no longer reusable, they get destroyed. According to Kennisinstituut voor Mobiliteitsbeleid (2017), nearly 10% of all returned goods are no longer reusable and get destroyed. Furthermore, all this extra work imposed by returned goods, can cause a company between 10 to 15 euros of extra costs per returned order (Kennisinstituut voor Mobiliteitsbeleid, 2017). In the Netherlands, these costs are on average 12.5% of the invoice amount (Kennisinstituut voor Mobiliteitsbeleid, 2017). Due to the increase in order returns, the number of transport van-movements in residential areas has also increased, which imposes consequences for the air quality, traffic safety, the overall living environment of cities and the congestion problem the Netherlands is currently confronted with (Kennisinstituut voor Mobiliteitsbeleid, 2017).

1.1 Knowledge gap and added value of research

Literature study

Due to the aforementioned ecological and economic implications, 'returns management' has gained increased attention, especially in the online apparel retail domain (Difrancesco, Huchzermeier, & Schröder, 2018; NRC, 2019).

In order to find the added value of this research within the field of online apparel returns management, a literature study was conducted. The consulted databases were the scientific databases Scopus and Web of Science. The aim was to explore what has already been done in the field of online apparel returns. The search was set on finding search terms in the title, abstract and keywords of the articles. Only accessible findings in published papers such as scientific journals, conference proceedings and published book chapters were used.

After having inserted multiple different search terms, the following search terms were used:

(1) *online apparel retailer AND return rate OR fashion OR purchase OR behavior OR customer OR decision.*

(2) *product return prevention instruments OR apparel return prevention instruments.*

The first string of search terms, has led to 15 articles on Scopus and four articles on Web of Science. After selecting scientific journals and conference proceedings, 13 articles were left on Scopus of which six articles were consulted as these articles mainly focused on customer product returns in online apparel retail. Of the four articles found on Web of Science, only one article mainly focused on customer product returns in online apparel retail. However, this article is also found on Scopus and therefore perceived as a duplicate.

The second string of search terms, has led to five articles on Scopus. Based on the literature selection criteria, only one article was useful. However, this article was also found by using the other search terms *(1),* thus is perceived as duplicate. This was the same for Web of Science.

Thus, after removing duplicates, the primary research has led to six relevant articles for the literature review process. After the primary search, each eligible article on Scopus was checked for other relevant articles. This secondary search has led to four additional relevant articles.

To sum up, this search has together resulted in 10 relevant articles for the literature review process. Through this literature study, insight was gained regarding online apparel returns management. Various reason behind apparel returns were identified, along with means which have been designed to address the reasons for apparel returns.

Table 1: Overview of reviewed literature

Search engine	Study	Objective	Methods/Theory
Scopus - primary search	Peng & Al-Sayegh (2014)	Examining the application of a **size recommendation service** – Shapemate, for matching customized apparel data with personalized 3D human body data, for a Leading fashion E-commerce and retailers in UK.	• Semi-structured focus group interviews and questionnaires. • Sample size: nine participants.
	Wang, Ramachandran & Sheng (2016)	Examining the causal impact of fit-valance and fit-reference information in **online reviews** on apparel returns at the largest online-only specialty retailer in Western USA.	• Regression analysis. • Quasi-experiments. • Sample size: reviews of 942 products (696 had fit-reference-information, 246 did not).
	Gallino & Moreno (2018)	Examining the impact of a **virtual fitting room** on apparel returns at an online apparel retailer in Latin America.	• Regression analysis. • Randomized field experiments, using online customers' records. • Sample size: over 60.000
	De Leeuw, Minguela-Rata, Sabet, Boter, & Sigurðardóttir (2016)	Examining the effect of a **Virtual Dressing Room (VDR)** on order returns, for Wheelchair-Bound customers in Scandinavia, Copenhagen.	• In-depth interviews • Sample size: six people with different roles in E-commerce apparel retail companies.
	Brooks & Brooks (2014)	Examining the impact of **image interactivity** on order returns, using the website of Guess.com.	• Public response surveys (wheel chair survey respondents) for a VDR simulation system. • Sample size: 31 • Focus group interviews with seven customers and three experts (sales staff).
	Fiore & Jin (2003)	Examining the impact of **image interactivity** on order returns, using the website of Guess.com.	• Statistical tests (paired t-test). • Repeated-measures experimental design. • Sample size: 103 undergraduate university majors in Mid-western USA.

Scopus - secondary search	Saarijärvi, Sutinen, & Harris (2017)	The study examines various **reasons for order returns.**	• Semi-structured interviews • Sample size: 21 random university students.
	Walsh, Möhring, Koot & Schaarschmidt (2014)	Identifying **three types of instruments** to reduce product return rates: monetary, procedural and customer-based instruments.	• Grounded theory • Literature review • In-depth qualitative interviews • Sample size: eight managers from online retailers active in the field of fashion.
	Walsh & Möhring (2017)	Examining the impact of **product reviews** on order returns	• Risk theory. • Field experiment, using the order records dataset of a well-known European online retailer in Germany.
	Ding, Xu & Tan (2015)	Investigate the impact of the amount of product page viewing on product returns.	• Fixed effect logit model • Customers' dataset of an E-commerce website in China. • Sample size: over 58.000 customers.

1.1.1 Scientific knowledge gap

- Gap 1: Empirical research regarding customer-based instruments is lacking

Reasons for apparel returns

Saarijärvi et al. (2017) identify and describe 22 reasons for order returns, categorises these 22 reasons of order returns in 10 categories of online returning behavior, and identifies six different moments in which the return decisions are made. This article also indicates a link between the 10 categories of online returning behavior and the six moments wherein customers' return decisions are made. However, the role and potential of different measures or tools by which online apparel retailers can reduce the different forms of online returning behavior and product returns is not examined. Literature also indicated that product information uncertainty is the main factor which contributes to customers' order returns (Gallino & Moreno, 2018). According to Peng & Al-Sayegh (2014), uncertainty about garment size accounts for 30-40% of the total order returns.

Apparel return prevention instruments

Based on a literature review conducted by Walsh et al. (2014), three categories of technologies and instruments that can be employed were identified in order to prevent and reduce order returns. These categories are: 1) monetary instruments, 2) procedural instruments and 3) customer-based instruments. According to Walsh et al. (2014), "monetary instruments are aimed at financially disincentivizing (or financially incentivizing) customers from returning products, procedural instruments are designed to either reduce transparency (in relation to the return process) for customers to identify return sinners and to increase the efficiency of the order and delivery process and customer-based instruments attempt to increase the ease of the order process from the consumer perspective by reducing consumers' perceived pre-purchase uncertainty" (p.6). According to the author, the distinction between the three instrument categories is necessary to study the performance of each preventive instrument more effectively.

However, literature studies conducted by Walsh & Möhring (2017) and Walsh et al. (2014) indicate that prior research has mainly focused on monetary instruments and that existing research about procedural instruments and mostly customer-based preventive product return instruments is sparse. Furthermore, Walsh et al. (2014) interviewed apparel retail managers and discovered that none of the interviewed retailers have implemented any of the proposed customer-based instruments such as virtual try-ons yet. According to the authors, the reason for this is that although online retailers perceive these instruments as useful, they also perceive them as challenging due to expense and lack of experience.

Wang, Ramachandran & Sheng (2016) and De Leeuw et al. (2016) examined the impact of product information provision on consumer purchase-related-decision-making. De Leeuw et al. (2016) suggest that appropriate and precise product information provision prevents consumers unnecessary product returns. However, research conducted by De Leeuw et al. (2016) was only approached from the retail perspective and not from a customers' perspective. Only experts from the apparel e-commerce industry were interviewed. Therefore, the customers' perception of what appropriate and precise product information is, is lacking. The authors indicate that addressing this study from the point of view of customers may lead to complementary findings.

Wang, Ramachandran & Sheng (2016) suggest that ''a combination of fit-reference information such as body-size and fit-valance expressions such as 'true to size' in online reviews reduce product misfit, purchasing error and thus the product return rate'' (p. 0). Whilst this study provides useful insight, it solely focuses on apparel which people can return due to a lack of fit-reference-information. Walsh & Möhring (2017) examine the impact of one customer-based instrument, namely product reviews on customers' order returns. The experiment shows that product reviews reduces the order returns. However, the impact of and comparison to other customer-based instruments on customers' order returns and the customers preference is not examined.

In order to reduce apparel misfits, Peng & Al-Sayegh (2014) generated the 3D body scanning garment size recommendation application named 'ShapeMate' to determine the right size, which according to the authors can be easily integrated in online apparel retail shops. The effectiveness of this application is only tested on a small focus group (nine participants). Therefore, the authors indicate that there is room for improvement regarding the capabilities of this particular customer-based instrument. Brooks & Brooks (2014) examined the use of Virtual Dressing Rooms (VDRs) especially for the wheelchair-bound community. However, the full capabilities are not yet know, as the VDR is still being tested in the laboratory. Gallino & Moreno (2018) examined the impact of Virtual Fitting-Room technologies on product-fit-uncertainty and order return costs and found that Virtual Fitting-Room technologies reduces misfits and costs and increases conversion rates and order value (Gallino & Moreno, 2018). Fiore & Jin (2003) examined the image interactivity function for Guess.com. The authors indicate that the usage-image-interactivity-function such as 'mix-and match' on websites, by which customers can see how apparel will coordinate, can reduce product mismatch and consumers purchase uncertainty.

Ding, Xu & Tan (2015) examined the link between online product page viewing and order returns using a dataset from the largest e-commerce website in China. They indicate that an increase in product page viewing increases the probability of product returns. However, this relationship is only explored through examining whether people had clicked on / used: 1) product recommendations, 2) extra visualizations (extra photos) and 3) product reviews to buy products. Moreover, the proposed instruments and the product scope were very general (e.g. there was not specifically mentioned what information was examined in the reviews and the research did not specifically focus on apparel). Furthermore, the research did not include causes of customers' return behaviour nor the relationship between the instruments and the return reasons.

In conclusion, whilst research has been conducted about monetary instruments and efficient transport routing and handling of order returns (post-purchasing), not much empirical research is conducted so far about customer-based instruments that can be used during the customers' online screening process of apparel in order to prevent unnecessary apparel returns (pre-purchasing). Consequently, this literature study has showed that no empirical studies have yet been carried out within the online apparel retail domain which examines / compares the perceived effectiveness of various customer-based technological concepts in addressing online purchased apparel returns, especially from the customers perspective.

- Gap 2: Empirical research regarding TAM in the apparel e-commerce sector is lacking

According to Walsh (2014), “customer based instruments attempt to increase the ease of the order process from the consumer perspective by reducing consumers’ perceived pre-purchase uncertainty” (p. 6). Walsh, et al. (2014) indicate that “the purpose of using these instruments is to communicate suitable information about the product to customers, so they can evaluate the personal fit more precisely and refrain from returning it because of a possible misfit” (p. 8). Since customer-based instruments are designed to be used by customers, its success relies greatly on the customers usage.

Based on this, the application of customer-based instruments fits well within the concept of technology acceptance, which can be evaluated through the application of the Technology Acceptance Model (TAM) developed by Davis (1986). The TAM merely predicts the use of a technology through two determinants, namely: 1) Perceived Usefulness (PU) and Perceived Ease of Use (PEU) (Davis, 1986). This aspect of TAM suits well with the research goal, since the acceptance of different technologies and instruments must be assessed from the clients’/users’ perspective (customers) before they can be evaluated from the decision-makers’ perspective (online apparel retailers). In other words, first the customers’ preference regarding the technological concepts designed to prevent unnecessary apparel returns needs to be examined, such that online apparel retailers know if and how they can adapt the current arrangements in order to reduce apparel returns.

In order to examine the application of TAM in this specific research domain, a literature study on TAM was conducted. This literature study, as presented in section 3.1.2, presents an overview of the research fields where TAM has been used. Based on this, the observation was made that in neither of the identified studies, the empirical setting of returns management within the apparel e-commerce sector was examined. While the identified fields wherein TAM has been used might to a certain extent have an overlap with apparel e-commerce returns management, the identified determinants and their external predictors (i.e. external variables) used in the studies might not be as applicable since they are in their particular field of study. This implies that the customers’ preference and acceptance for the customers-based technological concepts have never been evaluated in the scientific context of apparel e-commerce returns management. Another observation that cloud be made was that TAM is mostly operationalized using SEM, which aims to measure the structural relationship between indicators (measured variables) and latent variables (non-observable variables) such as the users attitude towards using a technology and the customers’ technology acceptance (Hox & Bechger, 1999; Nachtigall et al., 2003). However, this research does not aim to determine the relationship between the factors used to predict the user technology acceptance. Instead, this research aims to find a ranking of technological alternatives from the users’ perspective in the field of apparel e-commerce returns management, based on a set of indicators (i.e. criteria) which affect the users’ technology acceptance. Therefore, it was necessary to find other ways to still examine and predict the customers’ technology acceptance and produce reliable results.

1.1.2 Main research question

In sum, although ‘returns management’ within the online apparel industry has received increased attention in literature due to the ecological and economic issues apparel returns impose, prior research has paid limited attention on customer-based technological concepts which can be used to prevent unnecessary online purchased apparel returns. Consequently, reducing online purchased apparel returns for retailers in the Netherlands, wherein 1) the customers’ reasons for apparel returns and 2) customer-based instruments are taken into consideration especially from the point of view of customers is a gap in literature. Furthermore, empirical knowledge of using the Technology Acceptance Model (TAM) to predict the users’ technology acceptance in the context of apparel e-commerce returns management is scarce. Therefore, this research strives to address these gaps by exploring and analysing 1) which reasons drive customers’ apparel returns, 2) which customer-based technological concepts can be used to address these reasons for apparel returns and 3) what the customers' preference/acceptance with regard to using the customer-based technological concepts is. As a result, the main research question is:

"*What is the customers' preference regarding technological alternatives online apparel retailers can employ, during the customers' online apparel screening process, in order to increase customers' online apparel purchase successes and reduce unnecessary apparel returns?*"

The above mentioned research question has the following added scientific and practical values. Firstly, this research contributes to the knowledge of operationalizing the Technology Acceptance Model (TAM) in the empirical setting of apparel e-commerce. This is done by identifying and measuring relevant criteria (i.e. indicators) to evaluate the customers' acceptance of the technological alternatives in the apparel e-commerce sector. Secondly, technological alternatives are designed which assists customers during the online screening process of apparel items, such that they can more accurately evaluate apparel items such that their online apparel purchase successes increase and unnecessary apparel returns are prevented. Thirdly, in order to measure the customers' preference regarding the technological alternatives, the novel Bayesian Best-Worst Method tool is applied. Through this, the research contributes to the empirical application of the novel Group Multi-Criteria Decision-Making method. Fourthly, by comparing the technologies to one another, the users' (customers') preference regarding the technological alternatives is established. Through this, decision-makers of online apparel retail shops know if and how they can adapt current arrangements in order to reduce apparel returns.

The rest of this chapter is structured as follows. In section 1.2, the research approach and sub-questions are described. The research methods and research flow diagram are described in section 1.3. In section 1.4, the link to the MSc. Program is described.

1.2 Research approach and sub-questions

Based on the identified knowledge gap, this research aims to explore what the customers' preference is with regard to technological alternatives online apparel retailers can employ to 1) increase the customers online purchase successes and 2) prevent unnecessary returns of online purchased apparel items. Since there is little practical and scientific knowledge in this specific research domain especially from the customers perspective, this research can be perceived as typical exploratory (Brown, 2006). As a result of this, new insights might lead to changes of the research direction (Saunders, Lewis & Thornhill, 2015).

MCA as approach

In order to address the research question, a Multi-Criteria Analysis (MCA) approach is used as this approach suits well with the exploratory nature of this research. Since this research aims to find a ranking of technological alternatives from the users' perspective based on a set of indicators (i.e. criteria) which affect the users' technology acceptance, an MCA approach is used to evaluate the technological alternatives and predict the users' technology acceptance. The advantage of using MCA compared to other evaluation methods such as the Cost-Benefit Analysis (CBA), is that the customers' preference and perceived effectiveness of different alternatives can be tested against each other on the basis of various (non-monetary) factors (Annema, Mouter, & Rezaei, 2015). Since this research is exploratory, it is still too early to indicate the actual impact of various technologies and instruments on actual apparel returns and convert that into monetary terms. Both of these steps can be considered as difficult as firstly, many impacts of these technologies and instruments on actual apparel returns are still not known and might be difficult to establish. Therefore, MCA is preferred over e.g. CBA. Another reason why an MCA approach is chosen, is because the opinions of different stakeholders can be incorporated, which leads to better informed and fair decisions (Annema, Mouter, & Rezaei, 2015).

Following the MCA approach, the research is categorized into three parts. The first part is devoted to finding technological alternatives to address the goal of increasing customers' online purchase successes and prevent them from returning apparel items. In order to do so, first understanding to why customers' return online purchased apparel items was essential. Consequently, specific reasons behind apparel returns needed to be established, by which the specific empirical research and solution domain could be specified. Then, alternatives could be designed through which the various reasons of apparel returns could be addressed. These designed alternatives had to be detailed enough to understand how each technology or instrument can contribute to addressing the reasons for customers' apparel item returns.

Following the MCA approach, the second part consisted of finding criteria that are important for evaluating the customers' acceptance regarding the designed alternatives. In this research, the established criteria are linked to determinants that can be applied to operationalize TAM in the field of apparel e-commerce returns management. Subsequent to the MCA approach, the third part consisted of identifying the relevance (weight) of each criterion from the point of view of customers, as they are the ones 1) who actually return the online purchased apparel and 2) whom have to use the information technologies to make well-informed pre-purchase decisions such that unnecessary apparel returns can be prevented. The fourth part consisted of comparing the alternatives (technologies) based on the criteria, weighted criteria and performance score, in order to rank the alternatives based on the customers' preference. Based on this, recommendations could be provided to online apparel retail decision-makers on how to adapt current arrangements during the customers' online apparel screening process such that customers online purchase successes can be increased and unnecessary apparel returns can be prevented.

Based on the findings, the next step was to identify if online apparel retailers in the Netherlands have already employed the preferred alternative(s) yet and why (not) and what the managerial implications of each technological alternative are which can inhibit their adoption in current practises.

Sub-questions

Based on the MCA approach, the following five sub-research questions were constructed:
SQ1. What technological alternatives can be used to address customers reasons for apparel returns?
SQ2. What criteria are important for evaluating these technological alternatives?
SQ3. What is the relevance that customers assign to these identified criteria?
SQ4. Based upon these criteria and their obtained weights, how do these technological alternatives compare in terms of preference?
SQ5. To what extent does the preferred technological alternative map the current arrangements used by online apparel retailers in the Netherlands and why?

Research scope

This research solely focusses on online apparel purchases for online apparel retailers in the Netherlands. Since the knowledge gap has indicated that research regarding customers-based instruments is lacking, this research solely focuses on these type of instruments. The aim of these instruments is to provide necessary product information to customers, such that they can examine the product and personal fit more accurately such that the pre-purchase uncertainty is reduced and returns as a result of possible misfit are prevented (Walsh, et al., 2014). As a result, this research is conducted from the customers' perspective, as the goal is to explore how customers can make better pre-purchase decisions during the online screening process of apparel in order to reduce apparel returns.

1.3 Research methods

In this paragraph, the methods and tools which are used to answer each aforementioned sub-question are discussed in a step-by-step approach. As a result, a Research Flow Diagram (RFD) is constructed (see figure 1) which gives an overview of the research design through which the main research question is addressed. Per sub-question, the applied methods and tools along with the inputs and outputs are indicated.

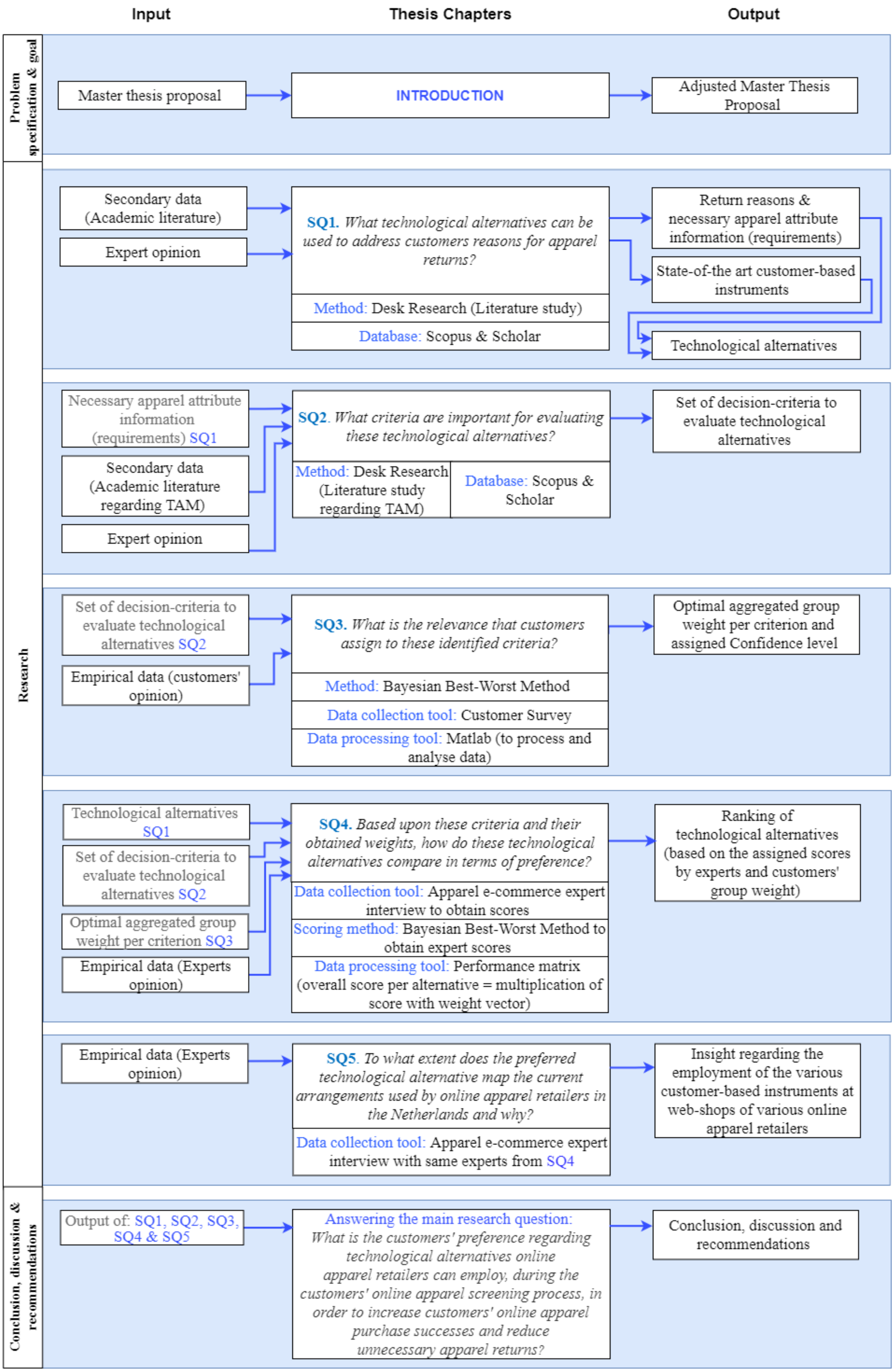

Figure 1: Research flow diagram

SQ1. What technological alternatives can be used to address customers reasons for apparel returns?

Identifying reasons for apparel returns

In order to answer this question, first the reasons that drive apparel returns needed to be established. Based on this, the specific research and solution domain could be specified. Various return reasons for online purchased apparel items were identified through an extensive literature study.

Since this research solely focusses on apparel return prevention information technologies and instruments which can be employed such that customers can evaluate apparel items and the personal match more precisely during the online screening process, only return reasons are included which can be treated by these category of technologies and instruments (Customers-based instruments). The included return reasons for online purchased apparel items are indicated in table 4 of section 2.2. The apparel return reasons were finalized with the opinion of two domain experts.

Of these two experts, one has a background in the e-commerce industry and functions as quality assurance inspector for nearly 5.5 years in the fourth biggest e-commerce retailer in the Netherlands and the second biggest online fashion retailer in the Netherlands. This person makes the technical translation from the styling/design phase to the technical application & visualization of clothing on the web shop. Currently, this expert has the lead of the returns management project for reducing returns within the company. The other expert is the local online marketing manager of an online apparel retailer in the Netherlands, which has over 10 years of experience. The acquired insights of these two experts has led to a better understanding about the reasons which drive customers' apparel returns and the formulation of the necessary apparel attribute information.

Establishing required apparel attribute information

Once the return reasons were identified, the apparel attribute information which is needed such that customers can evaluate the apparel items in detail and make a well-informed online pre-purchase decision, could be extracted. The necessary apparel attribute information was extracted based on the synergy of the identified return reasons.

Identifying technologies and instruments

Once the required apparel attribute information was extracted, the next step was to identify various technologies and instruments which can be used to address the required apparel attribute information during the online screening / evaluation process of apparel. In other words, a Morph chart was created which can be seen in table 6 of section 2.5.2.

Designing technological alternatives

Once these technologies and instruments were identified, instruments were combined to form technological alternatives that are comparable to one another. The combination of instruments was based on 1) the ability to fill in all the required apparel attribute information, which was extracted from the reasons of apparel returns and 2) the combination of instruments which are already used by most of the online apparel retailers. Once the alternatives were established, they were finalized with the opinion of a domain expert. Whilst multiple professors of various universities in the Netherlands and other experts from online apparel stores in the Netherlands were approached, only one had responded at that time. The experts were approached based on their comprehensive knowledge of the topic.

The expert who was available at that time, was the same quality assurance inspector that was approached for the first sub-question (SQ1), which makes the technical translation from the styling/design phase to the technical application & visualization of clothing on the web shop, has nearly 5.5 years of experience and has the lead of the returns management project for reducing returns at the fourth biggest e-commerce retailer in the Netherlands and the second biggest online fashion retailer in the Netherlands. The acquired insight of this expert regarding the identified instruments and technologies has also led to a better understanding of each instrument and technology and the formulation of the designed alternatives.

Method

Through the literature study that was used to analyse the knowledge gap (see section 1.1), various reasons for apparel returns along with technologies and instruments which can be used to address these reasons were already identified. The consulted databases were Scopus and Web of Science. However, since returns management in the online apparel retail industry is getting increased attention, relevant information regarding apparel returns might also be found in (un)published literature. Therefore, an additional Literature Study was conducted to find possible additional reasons for apparel returns and technologies and instruments using Scholar as database as well, aside from Scopus.

The most commonly known main advantages of conducting Desk Research compared to Field Research is that it is less labour intensive and less time consuming, as secondary data is used and the research can be done without the inclusion of respondents. On the other hand, the most commonly known disadvantage of Desk Research is that it can be criticized for its lack of full transparency about the generation of data, results and conclusions. In order to address this issue, only the most legitimate and reliable databases as mentioned before were consulted.

In sum, the first sub-question (SQ1) was addressed by conducting a thorough literature study which has led to the findings of 1) required apparel attribute information extracted from the identified apparel return reasons and 2) technological alternatives which are designed by partly combining instruments such that they can fulfil the required apparel attribute information. Both the list of the required apparel attribute information and the designed alternatives were finalized with the opinion of apparel e-commerce experts.

SQ2. What criteria are important for evaluating these technological alternatives?

Technology Acceptance Model

Ideally, criteria should be defined during interviews or workshops with decision-makers, which provides them the possibility to insert their perspective and ensures transparency in the overall process which increases the level of understanding. However, since this research is mainly approached from the customers' perspective, the perspective was clear. As a result, the designed alternatives needed to be evaluated with respect to criteria which define the customers' acceptance. Since enough literature was found which could be assessed to identify a set of decision-criteria, a literature study served as input instead. Consequently, in order to establish a set of relevant decision-criteria for evaluating and comparing the technological alternatives, a literature study regarding the Technology Acceptance Model (TAM) which is developed by Davis (1986) was conducted. Figure 2 in section 3.1 provides an overview of the original TAM, followed by the literature review results regarding TAM presented in table 7 in section 3.1.2. The identified criteria are external predictors of determinants of the TAM which can be used as input for the MCA, to determine the users' technology acceptance with respect to technological concepts in the field of online apparel e-commerce returns management. Since this research does not aim to estimate the structural relationship between the factors (in the TAM model), TAM is instead used as theoretical foundation with the aim to theoretically underpin the criteria which are necessary to evaluate the customers' technology acceptance and rank the technological alternatives in the context of apparel e-commerce returns management.

Aside from TAM, there a many other innovations frameworks such as the Feitelson & Salomon (2004) framework and the Multi-level perspective on Technology Transitions framework developed by Geels (2002) which can be used to evaluate the acceptance of the technological alternatives in society. However, since the goal of this research is to solely determine the technology acceptance from the users' (customers') perspective, as they are the ones whom have to use it, the TAM was the most suitable model as this particular model only includes two determinants to measure the use of an innovation (technology) namely, 1) the Perceived Usefulness (PU) and 2) the Perceived Ease of Use of the technology (PEU). As other frameworks also consider other factors such as political, institutional and financial determinants for technology acceptance, they were deemed too broad for this research.

To determine the technology acceptance, the theory of reasoned action (TRA) or the theory of planned behaviour (TPB) can also be used. TRA includes the determinants relative importance of attitudes (i.e. the users' feeling towards a particular behaviour) and subjective norms (i.e. the way perception of the users' significant other affects the users' performance and behavior) to predict users behavioural intention (Fishbein & Ajzen, 1975). TPB, which is built upon TRA, also includes perceived behavioural control (i.e. the users perceived control over expressing their own behaviours and attitudes) to measure the users behavioural intention (Ajzen, 1985). However, Davis (1986) observed that these additional determinants did not have a high correlation with the use of technology and excluded these two determinants. As a result, the TAM solely predicts the technology acceptance based on two determinants: 1) PU and PEU. Based on the simplicity of the TAM, this model suited the practical goal of this research better, which is solely determining the customers (users') preference regarding technological alternatives.

Predictive validity of TAM

Since the development of the TAM, nearly 30 years ago, the model has mostly been criticized for being incomplete as it only includes two perspectives of looking at technology acceptance, namely 1) the perceived usefulness (PU) and 2) perceived ease of use (PEU) (Legris, Ingham, & Collerette, 2003; Chutter, 2009). More specifically, the attitude towards using a technology is only determined by the two aforementioned determinants. Therefore, since its development, multiple extensions have been made of the TAM as critics argue that the attitude towards using a technology can also be directly influenced by other determinants. In section 3.1.1, this is elaborated in detail.

As a result, in order to increase the predictive validity of the TAM in this specific research, criteria were not solely selected based on the two original determinants of TAM. Instead, criteria related to determinants which also directly influence the attitude towards using a technology were considered as well. As a result, the original TAM was expanded by adding a third determinant to the model to predict the technology acceptance. Based on the reviewed literature, the determinant 'Trust' was also important to include which meant that the TAM was operationalized with external predictors (i.e. criteria) of the following three determinants: 1) PU, 2) PEU and 3) Trust.

Since this research aims to establish the customers' preference regrading technologies, the MCA was approached from the users' (customers') perspective. In other words, criteria that are important for the customers are only included in the MCA. Organizational and financial criteria for example were not deemed important to determine the customers' preference regarding the technologies and were therefore not included in the MCA. However, in SQ.5, these are reflected upon, such that insight can be gained regarding managerial implications that could inhibit the adoption from the point of view of apparel e-commerce decision-makers.

Establishing main-criteria set

Using many criteria is not convenient, as it can become difficult to handle and compare information (Choo, Schoner, & Wedley, 1999). Through the literature study regarding the TAM, a total of 17 criteria were identified, along with the sub-criteria through which they are measured (see table 7 in section 3.1.2). Because there were more than nine criteria, it was essential to add a level to the hierarchy of the problem by making meaningful clusters or categories (Rezaei, 2015).

Based on the synergy between the (sub)-criteria, five categories were made (see Appendix B). Of these five categories, three were included in the MCA (see table 8, in section 3.2.2). The remaining two categories were not included as main-criteria, since the interest did not lay in evaluating the psychological reasoning behind a customer's choice in depth. On the contrary, the interest laid in finding a set of decision-criteria which could be used to get a general overview of the customers preference, by examining the customers' perceived usefulness, perceived ease of use of a technology or instrument and trust regarding information gathering and handling.

This being said, in this research, the decision was made to not include attitude related, knowledge related or subjective norm related criteria in the MCA analysis. Furthermore, since Davis (1986) observed that subjective norms and relative importance of attitude did not have a high correlation with the use of technologies, criteria related to these determinants were excluded as well. Based on the demographic information such as gender, age and education level, the representativity of the dataset was examined.

Establishing subsets of criteria

In order to acquire usable and manageable sets of sub-criteria, it was necessary to aggregate some sub-criteria since using many criteria is not convenient, as it can become difficult to handle and compare information (Choo et al., 1999). The aggregation of the sub-criteria was based on the synergy between the sub-criteria.

Once the main-criteria set and the sub-criteria sets were established, they were finalized with the opinion of domain experts. The experts were approached based on their comprehensive knowledge of the topic. The first expert has a background in academia and functions as a full Product Development & Management Professor of the Innovation, Technology Entrepreneurship & Marketing (ITEM) group at the Industrial Engineering department of Eindhoven University of Technology (TU/e). The field of expertise this expert is active in is 'Product Developments & Management', where process optimization of bringing innovative products to market and managing products after they have been launched is examined.

The second expert was the same expert that was approached to finalize the designed technological alternatives at SQ1. This expert has a background in the e-commerce industry and functions as quality assurance inspector in the fourth biggest e-commerce retailer active in the Netherlands and the second biggest online fashion retailer active in the Netherlands.

The third expert that was approached, had a background in the Industry and Academia, and functions as a professor at Tilburg School of Economics and Management. This expert has a background in Business Engineering, ICT systems and Management and previously worked as a software developer and IT project manager for commercial firms. In the course 'Management and Information Systems' provided by the expert, the importance of information and information technology (IT) is examined from different perspectives in the decision-making process, in order to optimize business processes and improve the service provided towards customers.

Once the main criteria and sub-criteria were finalized with the opinion of domain experts, the data gathering process could start. Figure 3 in section 3.2.3 provides an overview of the main-criteria and sub-criteria that were used in this research.

Consequently, the second sub-question (SQ2) was answered by conducting a thorough literature study regarding the TAM, through which a set of relevant decision- criteria was established that functioned as input for the evaluation of the technological alternatives. Since there were more than nine criteria that were deemed important for this research, the criteria were first categorized into three main-criteria categories, which were further subcategorized into three subsets of sub-criteria. The categorization was based on the synergy between the identified (sub)-criteria. Afterwards, the main-criteria and sub-criteria were finalized with the opinion of three domain experts after which the data gathering process could start. In total, four sets (one main set and three subsets) of decision-criteria were established for the evaluation of the technological alternatives.

SQ3. What is the relevance that customers assign to these identified criteria?

As already indicated, customers are the ones who have to use the technologies. Therefore it is interesting to know what criteria customers value the most or least when it comes to accepting a technology they can use to better inform themselves online, such that their online purchase successes increase and unnecessary returns are prevented. This being said, this research is approached from the customers perspective.

In order to gather the weights of each criterion based on the preference of the decision-makers, the Best-Worst Multi-Criteria Decision-Making Method, or in short BWM, was applied. The BWM is a relatively novel method, developed by Rezaei (2015), which has been used in various fields of study (bestworstmehod.com), such as quality assessment in airline industry (Rezaei, Kothadiya, Tavasszy, & Kroesen, 2018), airport evaluation and ranking (Shojaei, Seyed Haeri, & Mohammadi, 2018), supplier selection (Gupta & Barua, 2017; Rezaei, Wang, & Tavasszy, 2015), sustainability assessment (Kaswan & Rathi, 2020; Zhao, Guo, & Zhao, 2018), selection of agricultural machines and equipment (Hafezalkotob, Hami-Dindar, Rabie, & Hafezalkotob, 2018), software vulnerability assessment (Anjum, Kapur, Agarwal, & Khatri, 2020), smart metering value sensitive design (van de Kaa, Rezaei, Taebi, van de Poel, & Kizhakenath, 2020) to name a few.

BWM for criteria weights derivation

In literature, various Multi-Criteria Decision-Making methods (MCDMs) can be found. Two very common MCDM tools are the Analytic Hierarchy Process (AHP) and the Analytic Network Process (ANP) which are used to infer the weights of decision-criteria based on the preference of the decision-makers (Saaty, 2004).

Rezaei (2015) presented the Best-Worst Multi-Criteria Decision-Making Method, or in short BWM, which is an appropriate method to address multi-dimensional and perception-based constructs. In the BWM, the number of required comparisons is $2n-3$ comparisons (Rezaei, 2015). The number of comparisons when using AHP, is $n(n-1)/2$ comparisons (Saaty, 2004). This means that when there are seven criteria (n=7), 11 comparisons are required when using BWM and 21 comparisons are needed when using AHP. In this research, the BWM is chosen, because compared to the other MCDM methods it 1) requires less comparisons and is therefore faster to compute and 2) leads to more consistent comparison data, implying that it produces more reliable weights (Rezaei, 2015, 2020). Since the decision-maker chooses a best and worst criterion before conducting the pairwise comparisons when using BWM, a clear understanding regarding the range of evaluation is gained upfront which could lead to more consistent pairwise comparisons, hence more reliable weights (Rezaei, 2020). Incidentally, fewer data does not necessarily mean less reliable results. For clarity purposes and since customers do not necessarily like spending a lot of time filling in surveys, which is the tool that was used to obtain the criteria weights, a low amount of pairwise comparisons was necessary.

The most criticized feature of MCA is the subjective or biased value judgments of decision-makers, which can affect the final outcome (Annema et al., 2015; Choo et al., 1999). As humans are very rarely consistent with their thoughts, behaviours and actions of our return behaviour might differ over time. Therefore, the consistency of the respondents was also checked and the ones which were acceptable were considered (Liang, Brunelli, & Rezaei, 2020).

Steps of BWM

The BWM steps as provided by Rezaei (2015) are described below:

Step 1. Establishing a set of decision- criteria.

The first step of the BWM is to identify a set of *n* decision criteria ($\{c_1, c_2, c_3, \ldots, c_n\}$) which the decision-maker can use to evaluate the designed alternatives.

Through the literature study regarding TAM, see previous sub-question (SQ2), a set of decision criteria was already identified.

Step 2. Defining the Best criterion and the Worst criterion.

In the second step, the decision-maker chooses the best (most important or preferable) criterion and the worst (least important or least preferable) criterion for determining his/her preference of using an alternative.

Step 3. Obtaining the Best-to-Others (BO) comparison vector.

In the third step, the decision-maker determines the preference of the best (most important) criterion against all other criteria by using a scale from 1-9. A value of 1 implies that the two criteria are of equal importance, whereas a 9 suggests that the best criterion is absolutely more important than the other one. As a result, a BO vector is obtained:

$A_B = (a_{B1}, a_{B2}, a_{B3},...., a_{Bn})$, where a_{Bj} is the preference of the best criterion B over the other criterion j. In addition, it is quite straight forward that the preference of the best perceived criterion B against itself is 1, i.e. $a_{BB} = 1$.

Step 4. Obtaining the Others-to-Worst (OW) comparison vector.

In the fourth step, the decision-maker determines the preference of all other criteria against the worst (least important) criterion by using the same scale from 1-9. A value of 1 implies that the two criteria are of equal importance, whereas a 9 suggests that the other criterion is absolutely more important than the worst one. As a result, an OW vector is obtained:

$A_W = (a_{1W}, a_{2W}, a_{3W},...., a_{nW})^T$, where a_{jW} is the preference of the other criterion j over the worst criterion W. In addition, it is quite straight forward that the preference of the worst criterion against itself is 1, i.e. $a_{WW} = 1$.

Step 5. Establishing optimal group weights of criteria.

- Bayesian BWM

Since the goal of this research is to examine what the customers' preference is regarding the technological alternatives, this research involves a group performance evaluation of the effectiveness of the technological alternatives with regard to the identified criteria from the perspective of customers.

Therefore, in this research a novel multi-criteria group decision-making approach of the BWM called the Bayesian BWM is applied, to calculate the optimal weights of each criterion, which provides a probabilistic interpretation of the initial BWM.

By using the Bayesian BWM, the optimization problem of the initial BWM is replaced with a probabilistic model (Mohammadi & Rezaei, 2019). Whilst the primary input data (step 1 till step 4) stays the same, the input data and output data have to be modelled as probabilistic distributions, instead of multinomial distribution. According to Mohammadi & Rezaei (2019), "from a probabilistic perspective, the criteria are seen as the random events, and their weights are thus their occurrence likelihoods" (p.3). This being said, only the application of the last step (step 5) of the initial BWM differs when using the Bayesian BWM. As a result, the Bayesian BWM has the following additional sub-steps which are undertaken in step 5.

Step 5.1. Constructing the probability distribution

Assume that there are k decision-makers ($k = 1, 2, ...K$), there are j evaluation criteria ($c_j = c_1, c_2, ..., c_n$), then A_B^k represents the Best-to-Others (BO) vector of one decision-maker and A_W^k the Others-to-

Worst (OW) vector of once decision-makers. If the optimal weights of one decision-maker is w^k, the optimal group weight after aggregation is w^{agg}. The vector, $A_B^{1:K}$ represents the BO vector of all decision-makers and $A_W^{1:K}$ indicates the OW vector of all decision-makers. Based on this, the equation for the joint probability distribution of the group decision for the Bayesian BWM is formulated as:

$$P\left(w^{agg}, w^{1:K} \middle| A_B^{1:K}, A_W^{1:K}\right)$$

If the probability in the aforementioned equation is calculated, the following probability rule can be used to compute the probability of each individual variable:

$$P(x) = \sum_{y} P(x, y)$$

with x and y representing arbitrary random variables (Mohammadi & Rezaei, 2019).

Step 5.2. Calculating the optimal group weight

The aggerated weight w^{agg} is dependent on the optimal weight of every individual decision-maker w^k, which is calculated by the input BO and OW vectors (A_B and A_W). Each time new input data (pairwise comparison data) is inserted, w^{agg} is updated. As a result of the previous concepts, conditional independence is present between variables. Taking this independence into consideration, the equation for the joint probability of the Bayesian BWM can be presented as:

$$P\left(w^{agg}, w^{1:K} \middle| A_B^{1:K}, A_W^{1:K}\right) \propto P\left(A_B^{1:K}, A_W^{1:K} \middle| w^{agg}, w^{1:K}\right) P\left(w^{agg}, w^{1:K}\right)$$

The above equation, can further be presented as:

$$P\left(A_B^{1:K}, A_W^{1:K} \middle| w^{agg}, w^{1:K}\right) P\left(w^{agg}, w^{1:K}\right) = P(w^{agg}) \prod_{k=1}^{K} P\left(A_W^k \middle| w^k\right) P\left(A_B^k \middle| w^k\right) P\left(w^k \middle| w^{agg}\right)$$

Based on the above equation, the corresponding probability can be found by specifying the distribution of each element. As a result, $A_B^k|w^k$ and $A_W^k|w^k$ can be defined as follows:

$$A_B^k \Big| w^k \sim multinomial\left(\frac{1}{w^k}\right),\ \forall_k = 1,2,\ldots,K;\ A_W^k \Big| w^k \sim multinomial\left(w^k\right),$$
$$\forall_k = 1,2,\ldots,K$$

Furthermore, w^k under w^{agg} conditioned can be composed as underlying Dirichlet distribution:

$$w^k|w^{agg} \sim \mathrm{Dir}(\gamma \times w^{agg}),\ \ \forall_k = 1,2,\ldots,K$$

with w^{agg} being the averaged value of the distribution and γ is a non-negative parameter (Mohammadi & Rezaei, 2019).

Since γ is a non-negative parameter, it needs to obey the underling gamma distribution where a and b represents the shape and the scale parameters of the gamma distribution.

$$\gamma \sim gamma(a, b)$$

Ultimately, the aggregated or group optimal weight w^{agg} abides to the Dirichlet distribution, with the parameter α being set to 1.

$$w^{agg} \sim Dir(\propto)$$

Once the probability distribution of all parameters is finalized, the posterior distribution is calculated by using the Markov-chain Monte Carlo (MCMC) technique (Mohammadi & Rezaei, 2019).

Compared to the initial BWM, the Bayesian BWM has the following benefits. First, it can instantly combine the final criteria weights for a group of decision-makers, when multiple decisions-makers are involved (Mohammadi & Rezaei, 2019). This means that the researcher does not have to obtain the criteria weights of each decision-maker first separately and aggregate the weights afterwards using the arithmetic mean, which according to Mohammadi & Rezaei (2019) is prone to outliers and provides limited information to decision-makers (in this case apparel e-commerce decision-makers).

Step 5.3. Credal ranking and Confidence level

The Bayesian BWM provides a credal ordering of each and every pair of criteria (c_i, c_j) for all $(c_i, c_j \epsilon\ C)$, with C being the set of criteria. In order to understand whether the rankings of the criteria (based on their group weights) are consistent with the evaluation of all experts, the confidence level (CL) is computed in the weight directed graph (Mohammadi & Rezaei, 2019). The CL indicates the probability or confidence (P) that c_i is better that c_j and is computed as follows:

$$P(c_i > c_j) = \int I\left(w_i^{agg} > w_j^{agg}\right) P(w^{agg})$$

In the above equation, I represents a conditional parameter which can only be computed if $\left(w_i^{agg} > w_j^{agg}\right)$ is detained, or else it is 0. Evidently, the CL is obtained by the number of samples Q acquired by the Markov-chain Monte Carlo technique (MCMC).

$$P(c_i > c_j) = \frac{1}{Q} \sum_{q=1}^{Q} I\left(w_i^{agg_q} > w_j^{agg_q}\right); P(c_j > c_i) = \frac{1}{Q} \sum_{q=1}^{Q} I\left(w_j^{agg_q} > w_i^{agg_q}\right)$$

In the above equation, w^{agg_q} represents q $w^{agg's}$ from MCMC samples. If $P(c_i > c_j) > 0.5$, then the criterion *i* is more important than criterion *j* (Mohammadi & Rezaei, 2019). The total probability is equal to 1, $(P(c_i > c_j) + P(c_j > c_i) = 1$.

- Advantages of Bayesian BWM

The initial BWM approach cannot combine the preference of several decision-makers at once, when multiple decisions-makers are involved (Mohammadi & Rezaei, 2019). Instead, it obtains the criteria weights of each decision-maker first and then aggregates the weights by the arithmetic mean. Nonetheless, averages are prone to outliers and give minimal insight about the overall preferences of decision-makers (Mohammadi & Rezaei, 2019).

On the contrary, with the use of probabilistic modelling and interpretation of the data provided by the Bayesian BWM, the combined distribution and each-and-every individual preferences are computed at the same time, resulting in more reliable criteria weights (Mohammadi & Rezaei, 2019). Consequently, by using the probabilistic modelling approach, the final optimal group weights for each criterion can be calculated at once.

The calculated optimal aggregated (group) criterion weight (w^{agg}), indicates the total preference of all decision-makers for a specific criterion.

A second advantage of the Bayesian BWM is the credal ranking and the confidence level it provides in the weight directed graph (step 5.3), wherein each node represents a criterion and each edge indicates the obtained confidence (Mohammadi & Rezaei, 2019). According to Mohammadi & Rezaei (2019), "the confidence level represents the extent to which one can be certain about the superiority of a criterion over one another" (p.2). The assigned confidence levels indicate the groups' perceived importance of one criterion over another, which can provide decision-makers (in this case apparel e-commerce decision-makers) with more information on how to adapt current arrangements (Mohammadi & Rezaei, 2019).

- Computing the criteria weights using Bayesian BWM

In order to acquire the groups' optimal weights and the assigned confidence levels with the Bayesian BWM, the Bayesian BWM solver needs to be used which is operationalized in MATLAB (Mohammadi & Rezaei, 2019). In addition "just another Gibbs sampler" (JAGS) needs to be installed as well. Both the solver and JAGS can be downloaded from the website: http://bestworstmethod.com/home/software/ (Mohammadi & Rezaei, 2019).

After conducting step 1 till 4 of the initial BWM, now both the optimal weight of all decision-makers for each criterion and their confidence levels (visualized in the weight directed graph) can be computed simultaneously by running the model in MATLAB. Afterwards, the obtained criteria weights and confidence levels can be analysed. Since there were four sets of criteria, the model was ran four times.

Data collection tool (BWM customer survey) & target group

To obtain the necessary data for the pairwise comparison analysis (step 2 till 4 of the BWM), an online survey was constructed using the imposed structure of the BWM. In appendix C, an overview of the survey is provided in English. As there were three main-criteria (categories), which each are subcategorised into three sub-criteria sets (see figure 3, section 3.2.3), a total of four BWM comparison analysis needed to be conducted. Only participants were included which had online shopping experience (had purchased apparel items online in the last 6 months).

Data collection procedure

Before sending out the survey, the survey was tested amongst a panel of 15 respondents, with different age groups, gender and education levels. The aim was to test 1) if the survey was easy to understand for everyone and 2) how much time it takes to fill in the survey. Data was collected over a period of 17 days. A total of 216 respondents completely filled in the online survey.

The obtained data from the questionnaire was analysed using the Bayesian BWM. The results of the Bayesian BWM calculations are the optimal group weights (w^{agg}) of the relevant criteria. Before calculating the optimal group weights, the consistency of the respondents was also checked and the ones which were acceptable were considered (Liang et al., 2020). After excluding the pairwise comparisons with an unacceptable consistency ratio, different sample sizes for different levels of the model were acquired and used.

In sum, the third sub-question (SQ3) was answered through an online survey using the imposed structure of the BWM. The optimal criteria weights for customers were inferred using the novel Bayesian group BWM. Before handing out the survey online, the survey was tested with a panel. After 17 days, 216 respondents completely filled in the online BWM survey. However, after excluding the pairwise comparisons with an unacceptable consistency ratio, different sample sizes for different levels of the model were acquired.

SQ4. Based upon these criteria and their obtained weights, how do these technological alternatives compare in terms of preference?

Through this sub-question, the preference for each technological alternative (at SQ1) was determined.

Data collection tool and target group

In order to complete the MCA, first a performance matrix needed to be obtained which shows the performance score of each alternative with respect to each criterion. For this, data was collected through structured interviews with online apparel retail experts in the Netherlands.

In order to set-up interview meetings with experts, the companies and experts were approached via email and also by phone. A total of 28 online apparel retail companies were approached. Of these 28, it was possible to interview experts from four companies. As a result, data was collected through structured interviews with six online apparel retail experts in the Netherlands, namely three online marketing managers, two quality assurance mangers and one online product specialist.

Ideally, the performance score per alternative with regard to the criteria, is derived from the same group of decision-makers from which the optimal criteria weights are acquired to ensure a high validity in the data. However, in this research, experts' opinion is gathered instead to obtain the performance scores, since experts have the knowledge about the technologies and instruments and how effective each composed alternative is in addressing each criteria. As the alternatives are not already fully used in practice, the assumption is made that customers might find it difficult to score the alternatives, especially the ones composed out of state-of-the-art technologies such as VDRs and Avatars. Therefore, experts were approached to obtain the performance scores on behalf of customers, to determine which alternative customers will prefer with respect to the criteria.

Interview design

Since the interviews were conducted via telephone, the interview was send upfront via email to the interviewee, such that the interviewee could walk through the provided upfront information and interview questions together with the interviewer.

Before conducting the interview, first the interview protocol was mentioned. Then the structure of the interview was explained, followed by an introduction of the case and the goal of the interview. Afterwards, the designed alternatives were explained followed by five interview topics which were addressed during the interview (see appendix E). After each interview, the interviews were immediately transcribed.

To obtain the performance score per alternative with regard to the criteria, the Bayesian BWM was again applied. As a result, the interview was constructed using the imposed structure of the BWM.

First, the decision-maker (apparel e-commerce expert) was asked to choose a best alternative with respect to a criterion. Secondly, the expert was asked to choose a worst alternative with respect to the same criterion. Thirdly, the expert was asked to compare the best alternative to the other alternatives with regard to that specific criterion and assign a score from 1-9. The same scores (1-9) proposed by (Rezaei, 2015) were used, with some slight adjustments. The meaning of the numbers 1-9 were defined as follows:

Table 2: Definition of scores from 1- 9, based on Rezaei (2015).

1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

Fourthly, the expert was asked to compare the other alternatives to the worst alternative regarding a specific criterion, and assign a score using the same scale from 1-9.

Once this was done, the criteria scores were computed based on the same way as explained in the previous sub-question (SQ3). The results of these BWM calculations were the optimal group (experts) weights (i.e. scores) of the alternatives with respect to the criteria and the confidence level that are visualized using the provided weight directed graph, based on the evaluation of six experts.

The reason why the BWM was used instead of a regular scale from for example extremely low (1) to extremely high (10) to obtain the performance score of each alternative, is because using the BWM produces more reliable results, even though this way of scoring alternatives is more time consuming.

After obtaining the experts' scores, these scores were multiplied with the obtained customers' optimal group criteria-weights from SQ3, resulting in the performance matrix as indicated in table 17 of section 5.3.

In sum, the fourth sub-question (SQ4) measured the customers preference regarding the technological alternatives which were identified at SQ1. By finalizing the MCA and merging its results with the customers' optimal criteria weights obtained in the previous sub-question (SQ3), the fourth sub-question was answered. The score of each alternative with respect to each criterion was obtained through six structured interviews with apparel e-commerce experts from four online apparel retail companies in the Netherlands, using the imposed structure of the BWM.

SQ5. To what extent does the preferred technological alternative map the current arrangements used by online apparel retailers in the Netherlands and why?

Once the customers' preference for an alternative was determined, the next step was to establish what the employment / perceived implementation possibility of these alternatives are from the point of view of online apparel retailers.

In this 'reflective' sub-question, the findings of the previous sub-questions are thus mapped. Information for this sub-question was also obtained during the structured expert interviews used at the fourth sub-question (SQ4), since six experts within the online apparel retail industry were already approached to answer the fourth sub-question. At the end of the interview (see appendix E, topic 5), the online apparel retail experts were also asked to indicate which of these alternatives are already used by the company and to explain why (not). Then by applying the BWM again, in the same manner as in the previous sub-question, the experts were also asked to indicate which alternative is the best and which one is the worst in terms of the employment / implementation possibility in the company. Consequently, similarly to SQ4, the experts were asked to conduct the BMW pairwise comparison, but now with respect to the criterion 'employment possibility in company'. Based on this, managerial implications were obtained and recommendations were formulated that acquire further attention.

In sum, through the fifth sub-question (SQ5) managerial implications were acquired which can hinder the adoption of the technological alternative for the point of view of apparel e-commerce decision-makers. This data was acquired through the same expert interviews used for SQ4.

1.4 Link to MSc CoSEM Program

Within the Complex System Engineering and Management (CoSEM) MSc program, the main goal is not to simply design innovations such as technologies and systems, but to also find its central embeddedness in society, which itself is a complex socio-technical environment. In order to do so, the perspective of not only decision-makers needs to be considered, but also the perspective of the users which adds to the complexity of achieving adoption of innovations in society. Within this specific research, the customers' preference of technologies is assessed by including the perspective of decision-makers (online apparel retail experts) and users (customers). Since the goal of this research is to evaluate the technological alternatives using both perspectives, the link with the CoSEM MSc program is fulfilled.

Since apparel e-commerce, which has become normal in our daily lives, also inflicts negative externalities on society that are complex by nature (see section 1), addressing ‘returns management’ within the apparel e-commerce domain for the Netherlands entails sufficient socio- technical complexity and thus suits the research objectives of the CoSEM study program. This link with the CoSEM MSc program can be further explained by looking at the methods and tools which are used to address the research goal. The technological complexity first lays in identifying apparel return reasons and designing alternatives, using a step-by-step systems engineering approach as taught in the MSc program. Furthermore, the socio-technical complexity lays in designing the online BWM survey and designing and conducting the expert interviews as well, through which the weights and scores are obtained in order to evaluate the technological alternatives. Each tool needed to be critically constructed and examined before usage to ensure clarity, e.g. the research perspective and goals needed to be clear, such that valid results could be obtained. Furthermore, computing the weights using the novel Bayesian BWM and analysing the data through modelling tools such as MATLAB in a critical manner, was also needed to rightfully present the data.

1.5 Report structure

The remainder of this report follows the structure as presented in the Research Flow Diagram (figure 1). As a result, the upcoming 4 chapters will be devoted to answering the sub-questions. Consequently, in Chapter 2, the first sub-question is addressed. Chapter 3 is dedicated to answering the second sub-question and so on. At the end of each chapter, the results obtained per sub-question will be presented and interpreted. Finally, in chapter 7, the findings are summarised and the main-research question is addressed. Furthermore, discussions and suggestions for further research are presented in chapter 7 as well.

2. IDENTIFYING RETURN REASONS & DESIGNING TECHNOLOGICAL ALTERNATIVES

This chapter is devoted to answering the first sub-question, namely: SQ1. *What technological alternatives can be used to address customers reasons for apparel returns?* First, in section 2.1, the literature review process is described which is used to identify apparel return reasons and return prevention technologies and instruments. In section 2.2, the identified reasons for online purchased apparel returns are explained. The apparel attribute information which is required in order to evaluate apparel items online is described in section 2.3. In section 2.4, the identified technologies and instruments are described, followed by the designed technological alternatives in section 2.5.

2.1 Literature review process

Through the literature study, as presented in the previous chapter, various reasons for customers' apparel returns were identified along with various product return prevention instruments. However, since returns management in the online apparel retail industry is getting increased attention, relevant information regarding apparel returns might also be found in (un)published literature on Scholar. Therefore, an additional literature study was conducted to find possible additional reasons for apparel returns and technologies and instruments using Scholar as database as well, aside from Scopus.

The search on Scopus and Scholar was set on finding search terms in the title, abstract and keywords of the articles. Only accessible findings in published papers such as scientific journals and conference proceedings and published book chapters were used. The search in Scopus resulted in findings published from 2001 up till now (2019). Therefore, the year 2001 up till now is used as literature selection criteria for Scholar. Furthermore, language was also considered as selection criteria. Only literature which is published in English or Dutch is considered.

After having tried the same search terms as *(1)* and *(2)* as indicated in section 1.1, too many records were found on Scholar. After multiple trials, the following search terms were eventually used:

(3) (online OR e-commerce) AND (apparel OR clothing) AND returns

Based on the aforementioned search terms and selection criteria, this search has led to 55 articles on Scopus and 69100 records on Scholar. Since there were still too many records on Scholar, the 'advanced search' function on Scholar was used. After using the 'advanced search' function, wherein many different keywords were inserted, the use of the words ''e-commerce return'' in the title of the article eventually resulted in 77 records, of which 54 were actual literature records. Of these 54 records, 25 were accessible and thoroughly examined. Of these 25 records in Scholar, 4 articles were used in the literature review process as most of these studies focused on logistic / distribution problems.

Of the 55 articles found on Scopus, 25 were accessible of which 16 were relevant. However of these 16 articles, 6 articles (Peng & Al-Sayegh, 2014; Wang, Ramachandran, & Sheng, 2016; Gallino & Moreno, 2018; De Leeuw et al., 2016; Brooks & Brooks, 2014 and Fiore & Jin, 2003) were already found during the first search trail (using search terms *(1)*). This being said, the search in Scopus had resulted in 10 new articles. To sum up, this search in Scholar and Scopus (using the search terms *(3)*) has resulted in 14 additional relevant articles for the literature review process. Underlying table 3, indicates the additional literature that was found using the search terms *(3).*

Table 3: Overview of additional reviewed literature for SQ1

#	Search Engine	Study	Objective
1	Scopus	Li, Zhou, Zhu & Mok (2017)	To visualize **virtual 3D try-on** effects of apparel on customized human models / **Avatars** for the online evaluation of fit and size.
2	Scopus	Gu & Tayi (2015)	Investigating the influence of **fit uncertainty** on consumers' post-purchasing behaviours.
3	Scopus	Algharabat & Shatnawi (2014)	Investigating the influence of perceived usefulness, perceived social presence and perceived enjoyment on **3D quality**, by using My Virtual Model™ technology.
4	Scopus	Kristensen et al. (2013)	Establishing a **Virtual Dressing Room (VDR**) application, using contemporary ICT such as camera sensors to reduce apparel returns in Copenhagen (Denmark).
5	Scopus	Apeagyei (2010)	This paper investigates the application of an **Avatar** as virtual try-on experience, through 3D body scanning technology.
6	Scopus	Misra, Wan & McAuley (2018)	The goal of the research is to predict fitness for associated catalogue sizes by using the customers' size recommendation information provided in e.g**. reviews**
7	Scholar	Lohse, Kemper & Brettel (2017)	Investigates the impact of **customer reviews** on return behavior in online fashion.
8	Scholar	Zhang (2018)	Enhancing the effectiveness of marketing efforts in the e-commerce sector, through **online customer reviews (OCRs).**
9	Scopus	Seewald et al. (2019)	Reducing returns through compensating customer bias (e.**g. misjudgement of correct sizes**) and manufacturer bias (inconsistent/incorrectly communicated product sizes) on the size-matching process.
10	Scopus	Jang & Burns (2004)	Examines the differences among four different types of apparel Web retailers regarding three components which are merchandise, promotion and **customer service.**
11	Scopus	Nasibov, Vahaplar, Demir & Okur (2017)	Assessing the application of a **size recommendation application**, through a fuzzy logic approach, such that the best fitted apparel size for customers' body measurements in each brand can be predicted.
12	Scholar	Shen, Shang & Dai (2019)	Provides a 4PL strategy / concept model to improve reverse logistics management for online apparel retailers.
13	Scopus	Hidellaarachchi, Gunatilake, Perera, Sandaruwan & Weerasinghe (2019)	Examining the impact of a **2D avatar** model on selecting properly fitted clothing using both human body measurements and garment measurements.
14	Scholar	Xiangdong Liu & Ming Lei (2008)	Analyses the impact of exogenous factors (i.e. **reasons for product returns**) on the return price, optimal product demand**, amount of returns** and revenues for three channel strategies.

To sum up, the literature study (presented in section 1.1) and this literature study (presented in table 3) have together resulted in 24 relevant articles for the literature review process, which were accessible and thoroughly examined to identify 1) reasons for customers' apparel returns and 2) product return prevention instruments which can be used during the online screening/ evaluation process of apparel items.

Inclusion of apparel return reasons

Since the knowledge gap has indicated that research regarding customers-based instruments is lacking, this research solely focusses on these type of instruments. According to Walsh, et al. (2014), “the purpose of these instruments is to communicate suitable product information to customers, such that they can evaluate the product and personal fit more precisely such that the pre-purchase uncertainty is reduced and returns as a result of possible misfit are prevented” (p. 8).

Through this literature study, many return reasons for apparel items were identified. However, in this research only return reasons were included which can be addressed by the customers-based technologies and instruments. As a result, only return reasons were included which stem from a mismatch between customers’ expectations based on the provided online apparel attribute information during the online screening process (virtual), and the actual product (realism). Furthermore, only return reasons were included which stem from an actual purchase intention. E.g. ‘just trying out and sending it back’ as return reason, is not included.

2.2 Identified reasons for apparel returns

Underlying table 4 indicates the identified return reasons based upon the examined literature. The categorization along with the reasons of returns are based upon Saarijärvi et al. (2017, p. 20-23).

Table 4: Identified return reasons regarding apparel attributes

Return moment: Online screening process				
#	Description of category	Reason for returns (Return factors)	Scopus	Scholar
1	***Disconfirmation driven:*** The quality of apparel differs or is not what was expected on the basis of the information provided on the website regarding apparel attributes such as material and colour.	A different material quality than what was expected	(Saarijärvi et al., 2017), (Algharabat & Shatnawi, 2014), (Gallino & Moreno, 2018), (Brooks & Brooks, 2014)	(Zhang, 2018)
		The colour hue differed from what was expected	(Saarijärvi et al., 2017), (Algharabat & Shatnawi, 2014)	
		Misleading information (apparel description, apparel images)	(Saarijärvi et al., 2017)	
		An unexpected negative aspect that was not visible in the apparel images (e.g. rips or tears)	(Saarijärvi et al., 2017)	
2	***Size (chart) driven:*** The apparel size is not correct, although the customer exactly chose his or her size (Small, medium, large etc.) **Size** is your actual measurement (think waist, inseam, neck, etc.).	Size variations, inconsistencies or mismatches: The size of apparel items are too big or too small	(Saarijärvi et al., 2017) (Hidellaarachchi et al., 2019), (Seewald et al., 2019), (Misra et al., 2018), (Nasibov et al., 2017), (Li et al., 2017), (Algharabat & Shatnawi, 2014), (Kristensen et al., 2013), (Apeagyei, 2010), (De Leeuw et al., 2016), (Wang, Ramachandran, & Sheng, 2016), (Peng, F., Al-Sayegh, 2014b), (Brooks & Brooks, 2014)	(Shen et al., 2019)
3	***Feeling driven:*** When actually trying on the apparel item, the customer does not feel ‘right’.	The apparel item does not match the customers’ style	(Saarijärvi et al., 2017), (De Leeuw et al., 2016)	
		The feeling of the apparel item is not right	(Saarijärvi et al., 2017), (Brooks & Brooks, 2014)	(Xiangdong Liu & Ming Lei, 2008)
		Customers’ misjudgement /misperception of the right fit.	(Saarijärvi et al., 2017), (Seewald et al., 2019), (Li et al., 2017), (Gu & Tayi, 2015), (Algharabat & Shatnawi,	(Zhang, 2018)

			2014), (Kristensen et al., 2013), (Apeagyei, 2010), (Gallino & Moreno, 2018), (Wang, Ramachandran, & Sheng, 2016) , (Peng, Al-Sayegh, 2014)	
4	***Benefit Maximization driven:*** The customer orders multiple apparel items with the aim to keep only one or few of the items.	Ordering multiple sizes of the same apparel item, in order to keep only one	(Saarijärvi et al., 2017), (Gallino & Moreno, 2018), (De Leeuw et al., 2016), (Brooks & Brooks, 2014)	
		Ordering the same apparel item in multiple colours, in order to keep only one	(Saarijärvi et al., 2017), (De Leeuw et al., 2016), (Brooks & Brooks, 2014)	
		Ordering alternative apparel items (e.g. different styles) for the same need, in order to not keep all of them	(Saarijärvi et al., 2017), (De Leeuw et al., 2016)	

According to Gallino & Moreno (2018), product uncertainly is defined as “the buyer’s difficulty in assessing the product’s characteristics and predicting how the product will perform in the future” (p.5). The inability to try-on apparel items also increases the customers pre-purchase uncertainty (Algharabat & Shatnawi, 2014). According to Liu & Lei (2008) touch and feel plays an important role for customers, in order for them to determine how well apparel suits their tastes and needs. However, since apparel is displayed online, it is not possible to actually touch and feel apparel items.

Saarijärvi et al. (2017) established various reasons for apparel returns, which were divided in different categories. Of these return categories, four are directly linked to the product, in this case apparel. These four categories are: 1) disconfirmation driven, 2) size chart driven, 3) feeling driven and 4) benefit maximization driven. The categories represent the behavioural explanation behind online apparel return reasons. To explain the return reasons, this categorization as provided by Saarijärvi et al. (2017) is also used in this research.

Disconfirmation driven

The returning behaviour ‘disconfirmation driven’ arises due to apparel quality concerns which on its own arise due to the variation in apparel attributes as shown online and in reality (Saarijärvi et al., 2017). This mismatch between perceived benefit and realized benefit or the mismatch between expectations and reality results in cognitive dissonance which further leads to product returns (Zhang, 2018).

According to Algharabat & Shatnawi (2014) and Gallino & Moreno (2018), customer often face risks when shopping online, as visual information such as material (fabric texture) might not be clear. Not being able to get a feel of the material (texture) and material flow also leads to a disconfirmation of expectations (Brooks & Brooks, 2014). Various studies, as indicated in table 4, indicate that material quality variation which refers to e.g. the variation in stretchability, thickness, texture and sewing, positively contributes to customers’ disconfirmation of expectations. Besides this, customers’ disconfirmation of expectations is also caused by misleading product description and pictures and the absence of apparel style related features in pictures such as a rips or tears in the apparel items (Saarijärvi et al., 2017). Furthermore, customers also return apparel items when the hue of the colour is different than was expected based on the provided information on the website (Saarijärvi et al., 2017). According to Algharabat & Shatnawi (2014), customer often face risks when shopping online, as visual information such as colour might not be clear as well.

Size chart driven

According to the reviewed literature, customers also return apparel items as a result of the apparel item not being the correct size, although the customer exactly ordered her or his size (Small, medium, large etc.). Saarijärvi et al. (2017) refer to this returning behaviour as ‘size chart driven’. Based on the reviewed literature, sizing problems were perceived as the most common reasons for returns of online purchased apparel items. The high return rate in the apparel e-commerce industry is mainly caused due to the size mismatches (Hidellaarachchi et al., 2019) or improper size (Shen et al., 2019). Reasons such as the size being too big or too small were the main reasons behind apparel returns regarding size. These are mostly caused by the customers bias (misjudgement of the correct size) and retailers and manufacturers bias (inconsistent or even incorrect reported product sizes / size charts) (Seewald et al., 2019; De Leeuw et al., 2016). Currently, most returns are caused by size selection and fit, as these attributes are still difficult to address (Li et al., 2017; Peng & Al-Sayegh, 2014; Misra et al., 2018). Furthermore, brand size inconsistency, e.g. one brand's size 10 is another's 8 or 12, also is a primary reason for apparel returns (Brooks & Brooks, 2014; Kristensen et al., 2013).

Feeling driven

According to the reviewed literature, customers also return apparel items when the feeling of the product, when worn, is not 'right'. Saarijärvi et al. (2017) refers to this returning behaviour as 'feeling driven'. This returning behavior refers to return reasons regarding both functional and aesthetic features of apparel.

Aesthetic features of apparel refer to apparel attributes such as style, colour and material (Jin & Black, 2012). Style refers to the shape, cut, design and pattern of the garment such as slim, regular or loose for a shirt and for trousers casual, baggy, slim, ripped/ teared design etc. Functional features of apparel refer to apparel attributes such as fit and size (Jin & Black, 2012). Good fitting for apparel items is essential for every individual, as along with appearance it results in physical and psychological comfort (Kim & Damhorst, 2010). However, as the apparel items are virtually displayed, an actual feeling of the apparel cannot be gained, which also leads to unnecessary online apparel returns (Brooks & Brooks, 2014).

Customers also experience a pre-purchase uncertainty since information regarding fit and silhouette of the apparel item might not be clear to them (Algharabat & Shatnawi, 2014). Furthermore, as already indicated, the inability to try-on apparel items also increases the customers pre-purchase uncertainty (Algharabat & Shatnawi, 2014). As fit is difficult to assess prior to purchase, customers often make incorrect purchase decisions which results in increased apparel returns (Wang et al., 2016).

According to Zhang (2018); Seewald et al. (2019); Gu & Tayi (2015) and Gallino & Moreno (2018), lack of fit is the primary issue of order returns. Peng & Al-Sayegh (2014) also indicate that apparel not fitting as expected is the primary issue of order returns. According to Nasibov et al. (2017); Apeagyei (2010) and Kristensen et al. (2013) poor fit & size are also perceived as primary reasons for customers' dissatisfaction and the barrier of online apparel retailing.

Based on the examined literature, the identified return reasons which stem from the 'feeling driven' return behavior are that the products cannot be touched and felt pre-purchasing in order to evaluate the tastes and needs (Liu & Lei, 2008), the product does not match the customers style (De Leeuw et al., 2016) and that the customers' perception/judgment of the fit is incorrect, based on the online proved information or the way the information was provided. According to De Leeuw et al. (2016), upfront information regarding size, colours and styles is important, as it will reduce unnecessary order returns.

Benefit Maximization driven

The returning behavior 'benefit maximization' refers to customers ordering many apparel items, with the purpose to keep only one or some of them (Saarijärvi et al., 2017); Brooks & Brooks, 2014); De Leeuw et al., 2016). The identified return reasons which stem from this returning behaviour are: 1) ordering multiple sizes of the same apparel item with the purpose to keep only one (Gallino & Moreno, 2018); (De Leeuw et al., 2016) 2) ordering the same apparel item in multiple colours with the purpose to keep only one (De Leeuw et al., 2016) and 3) ordering alternative apparel items (e.g. different styles) for the same need with the intention not to keep all of them (De Leeuw et al., 2016).

2.3 Interpretation of apparel return reasons

According to the examined literature, customer experience inherent risk or pre-purchase uncertainty when purchasing apparel items online, since they cannot evaluate the personal fit accurately in reality. When customers purchase apparel online, they cannot touch, feel or actually see the apparel items nor try them on to test the personal match / fit regarding the apparel attributes. There is thus a gap between the 'virtual' and 'reality' when purchasing apparel online.

In the examined literature, the observation that was made is that fit and size are mostly covered together. Consequently, features of 'fit' relate to apparel features of 'size' such as fit being referred to as apparel being either too small/ too short or too long/too large on the body shape (Jin & Black, 2012).

Based on the identified reasons for apparel returns, four apparel attributes were identified which are necessary to inform individuals about upfront (during the online screening process) such that they can evaluate apparel more accurately and make well-informed pre-purchase decisions. These four attributes are: 1) material, 2) fit & size, 3) style and 4) colour. Based on these identified necessary apparel attribute information, technological alternatives need to be designed such that customers can make a well-informed pre-purchase decision.

Another observation that could be made is that apparel items were returned because the attributes were different from what was expected, the provided information was misleading, style related features were not visible and customers were unable to get a 'feel' and perception of the apparel items. Based on this, creating the right expectations regarding the apparel attributes, providing accurate product information, and creating a 'feel' for and perception of apparel items displayed online are essential requirements which need to be considered when designing the technological alternatives. As a result, the technological alternatives also need to be evaluated, based on the ability to provide these features.

In the following section, these technologies and instruments are discussed. Once the necessary apparel attributes and requirements were established, these were validated with experts opinion (see section 1.3, SQ1). The acquired insights of these two apparel retail experts has also led to a better understanding about the reasons which drive customers' apparel returns and the formulation of the aforementioned necessary apparel attribute information and requirements.

2.4 Identified instruments and technologies

As already indicated, Walsh et al. (2014) refer to customer-based instruments as "instruments which aim at influencing the customer before and during the online screening process of apparel items. The purpose of using these instruments is to communicate suitable information about the product to customers, so they can evaluate the personal fit more precisely and refrain from returning it because of a possible misfit" (p. 8).

Since the knowledge gap has indicated that research regarding customers-based instruments is lacking, this research strives to examine the effectiveness of these instruments in providing the necessary apparel attribute information customers need, in order to make a well-informed pre-purchase decision such that their purchase successes increase and unnecessary apparel returns are prevented.

As indicated in section 2.1, a literature study was conducted to identify customers-based instruments. The aim was set on finding: 1) what upfront apparel attribute information can be provided to customers with each technology or instrument during the online screening / evaluation process such that the personal fit can be evaluated more precisely and 2) the perceived advantages and disadvantages for both customers (users) and e-commerce apparel retailers when using the technologies and instruments.

Underlying table 5, gives an overview of the identified customer-based instruments and technologies based on the reviewed literature.

Table 5: Customer-based instruments and technologies

#	Study	Examined instruments/technologies
1	Peng & Al-Sayegh (2014)	Size recommendation service called Shapemate.
2	Wang, Ramachandran & Sheng (2016)	Online customer reviews
3	Gallino & Moreno (2018)	Virtual Fitting Room
4	De Leeuw et al. (2016)	Product advice (by the retailer itself)
5	Brooks & Brooks (2014)	Virtual Dressing Room (VDR)
6	Walsh & Möhring (2017)	Online customer reviews
7	Li, Zhou, Zhu & Mok (2017)	Virtual 3D try-on / Avatars
8	Algharabat & Shatnawi (2014)	Mix-and match and zoom technology using My Virtual Model™ for 3D virtual product presentation and evaluation
9	Kristensen et al. (2013)	Virtual Dressing Room (VDR)
10	Apeagyei (2010)	Avatar as virtual try-on experience,
11	Misra, Wan & McAuley (2018)	Online customer reviews
12	Lohse, Kemper & Brettel (2017)	Online customer reviews
13	Zhang (2018)	Online customer reviews
14	Seewald et al. (2019)	Height/size chart
15	Jang & Burns (2004)	Customer hotline e.g. live chat line, alternative product photo's & zoom technology
16	Nasibov, Vahaplar, Demir & Okur (2017)	Size recommendation application
17	Hidellaarachchi, Gunatilake, Perera, Sandaruwan & Weerasinghe (2019)	Avatar

1. Virtual try-on experience

According to Kristensen et al. (2013) virtual try-on can be seen as a 'personalized online shopping approach'. Virtual try-on is an interactive technology, which allows consumers a more real-world view of what e.g. apparel items might really look like, without actually having to try the apparel on. According to Apeagyei (2010), "the main feature of this technology is the capability of virtually testing fit on the individual or a retail-specified size model"(p. 65).

Product presentation on websites where customers have the possibility to view items together (mix and match function), use zoom technology and 3D viewing can give customers a feel for apparel details (Jang & Burns, 2004). The interactivity of virtual try-on allows customers to zoom in on product attributes and virtually view the product in various colours on the individual itself or on the retail-specified models. Furthermore, through the interactivity function the product can be viewed from different angels, as it can be rotated (Fiore & Jin, 2003). The image interactivity function of the virtual try-on experience provides the ability to create and change images of a product or environment on a website (Fiore & Jin, 2003). Through the interactivity function, customers can mix and match images of apparel items (e.g. a shirt and trouser) either virtually on themselves or on virtual retail-specified size models to see what the final outfit will look like. Whereas some mix and match functions use 3D illustrations of apparel items, others use photographic illustrations (2D graphics) (Fiore & Jin, 2003). According to Fiore & Jin (2003), Virtual Dressing Rooms (VDRs) and avatars are examples of 'interactivity' of a website. Based on the reviewed literature, a virtual try-on can be created by using Avatars or Virtual Dressing Rooms (VDRs).

Prerequisite for Virtual try-on experience (Avatar and VDR):

A prerequisite of the virtual try-on experience with an Avatar or VDR, is that different virtual apparel items customers can try on first need to be digitized through a series of images that scan each apparel item from a 360-degree angle, which takes less than 10 minutes per apparel item and has an estimated cost of 5 British pounds or about Euro 5.53 or US$6.11 (Gallino & Moreno, 2018).

(1) Avatar for virtual try-on

According to the examined literature, a virtual try-on experience can be gained by using avatars. An avatar is a 3D computer-generated / animated body model which retailers can provide to customers in order to virtually try-on apparel items in a computer-generated environment. By using technologies such as scanning techniques, traditional image based technologies and manual insertion of measurement data, customer can create a personalised avatar or a so called ‘digital twin’ (Hidellaarachchi, 2019; Apeagyei, 2010). These body models can be customized based on various characteristics such as personal body measurements, facial features, hair colour and body shape (Hidellaarachchi et al., 2019; Gallino & Moreno, 2018). With so called ‘pressure maps’, the fit and size information are indicated.

According to Li et al. (2017); Gallino & Moreno (2018); Brooks & Brooks (2014); Apeagyei (2010) and Hidellaarachchi et al. (2019), avatars have high resolution details and realistic appearance and are designed for accurate fit (shape) and size evaluation. Therefore they can be used to provide personalized size recommendations based on fit information (Gallino & Moreno, 2018). According to Hidellaarachchi et al. (2019), using avatars for virtual try-on can guarantee 80% pre-purchase information accuracy regarding true fit & size. Through the provided visualization of an avatar, customers can also get a perception of the personal fit regarding style of the apparel items (Brooks & Brooks, 2014).

According to Gallino & Moreno (2018), a virtual fitting room application provides retailers with benefits similar to those provided by physical fitting rooms, and effectively reduces information uncertainty for customers. Through the provided 3D virtual product presentation, mix-and-max function and the possibility to zoom in and out on the virtual model, customers can get a better perception on how apparel fits on their bodies which reduces the customers’ perceived pre-purchase risks and increases their purchase intention (Algharabat & Shatnawi, 2014). By inserting data for apparel pattern modification, a perception of the material flow and fit on the customers body-shape can be created, by using draping technology and simulation (Apeagyei, 2010). Through the 3D virtual product presentation provided by avatars, a better perception and feel of visual information with respect to product attributes such as colour, body shape, material texture and fit can be created for customers (Algharabat & Shatnawi, 2014). Furthermore, compared to the indirect experiences (i.e., static pictures), users gain more insight from 3D virtual experiences (Algharabat & Shatnawi, 2014). Furthermore, providing the possibility for customers to view items together through the mix and match function, to use zoom technology and 3D viewing can give customers a feel for apparel details (Jang & Burns, 2004).

Advantage of Avatar for virtual try-on experience:

Since avatars can providing accurate fit & size information and information regarding style, they can also be used to accumulate accurate anthropometric data for the development of size charts (Apeagyei, 2010). The image interactivity allows customers to see apparel items on Avatar (customers virtual body shape) in 360 degree which along with the zoom function can create a feel for visual information with respect to product attributes such as silhouette, material texture, colour and fit (Algharabat & Shatnawi, 2014). Mix and max functions allows customers to experiment / create outfits, through which an overall perception of the outfit can be created (Gallino & Moreno, 2018). Furthermore, the avatar can be stored and moved from site to site, showing its data-reuse possibility and compatibility (Gallino & Moreno, 2018). In addition, the tool can have benefits for the online retailers as they can collect data and evaluate the customer’s choice set (Gallino & Moreno, 2018). The modelling process is efficient, meeting the requirement on real-time application (Li et al., 2017). User involvement is simple and easy to handle, just by e.g. taking two photographs without restrictive clothing conditions. (Li et al., 2017). Aside from an increased accuracy of targeted fit through using avatars and the fact that consumers do not have to travel for tedious fitting sessions, the amount of fit trial sessions undertaken by live fashion models at apparel retail shops and with that that the money spend on the fit trial sessions can be reduced (Apeagyei, 2010). Furthermore, the use of technologies such as avatars as virtual try-on experience, can increase the pleasure of online shopping (Apeagyei, 2010). According to Algharabat & Shatnawi (2014), 3D virtual models with high quality allow customers' to find the actual apparel attribute information, enhances the enjoyment of shopping and allows customers' to feel the social presence. As a result, a high-quality 3D virtual model can diminish users' perceived psychological risk and increase the customers' intention to purchase items online (Algharabat & Shatnawi, 2014).

Disadvantage of Avatars for virtual try-on experience:

However, aside from these advantages, the technology also has some implications which might inhibit its adoption. First of all, computing a 3D virtual human body model through which a consumer can experience a virtual-try-on, requires huge computational power which makes it difficult to use in real-time applications (Hidellaarachchi et al., 2019). The use of data generated is extensive but requires computer-aided design technologies (Apeagyei, 2010). Furthermore, the model generation process can be time consuming (Hidellaarachchi et al., 2019; Li et al., 2017). In addition, customer might perceive discomfort or privacy and security concerns regarding sharing personal body-measurements data and the way body-measurements data can be obtained e.g. through scanning the body or manually uploading pictures or inserting body dimensions and used by online apparel retailers (Hidellaarachchi et al., 2019; Apeagyei, 2010). Whilst personalized size recommendation can be very advantageous, customers' trust and privacy concerns might inhibit the adoption of the technology (Peng & Al-Sayegh, 2014). Replicating the tactile feel of materials e.g. indicate the difference between silk and cotton, which can be gained when browsing in a physical shop, along with recreating a material's physical dynamic on the avatar is challenging and if done not very accurately might inhibit the adoption as well (Brooks & Brooks, 2014). A lack of availability of the technology might also inhibit the adoption (Hidellaarachchi et al., 2019).

(2) Virtual Dressing Room for virtual try-on

According to the reviewed literature, a virtual try-on experience can also be gained by using Virtual Dressing Rooms (VDRs). As is the case with an avatar, with VDRs individuals are not fitting apparel items in an animated computer-generated world. Instead, images are shown over the real world through an interface, so that individuals perceive both at the same time.

Using contemporary ICT such as e.g. Kinect camera sensors, allows individuals to be scanned such that their face and body shape can be identified (e.g. such as snapchat uses facial recognition technology to capture the facial features and elements). Furthermore, the technology also allows the dynamic movements of the person to interact with the physical dynamic flow of the apparel. In other words, aside from the ability to project static pictures or representation of apparel items over the individuals' mirrored image, it also can project a dynamic representation of apparel items on individuals. With the latter, the apparel will move according to the movement of the individual and a real live fitting experience is recreated (Brooks & Brooks, 2014).

Once the body shape is acquired using camera sensors on a computer/laptop, mobile phone, tablet etc., customers can view their mirror image mirrored in a system interface (Brooks & Brooks, 2014), such as snapchat does. Then they can select apparel items from a menu and 'swipedragg' apparel items as an overlay on their mirrored image (Brooks & Brooks, 2014). Using the interactivity functions, individuals can mix and max apparel items, change colours, 3D view apparel outfits, zoom in and out etc.

As already indicated, a VDR and avatar are both developed for a virtual try-on / fitting room experience. In other words, they are both developed to recreate a the psychical fitting room experience for customers whom shop online.

However, a key distinction between avatars and VDRs is that whilst using VDRs, apparel items are not tested in a computerized environment using computerized digital body twins. Instead, customers are looking into a virtual mirror and are trying on apparel on their mirrored self. VDRs are designed to reduce apparel returns via an improved online purchasing experience (Brooks & Brooks, 2014). According to Brooks & Brooks (2014), “the VDR is designed as a single-user application (in first person) with collaborative potentials to include the sharing of images over smart phone personal networks for advised purchasing – a friend/family input (i.e. social shopping)” (p. 588).

VDRs are designed to provide an overall feel and perception of the fit and size of apparel items, style and feel of the material on the individuals’ body shape (Brooks & Brooks, 2014; Kristensen et al., 2013). Virtual Fitting rooms can reduced misfits and social costs such as energy and reverse logistics costs (Gu & Tayi, 2015).

To provide accurate size and fit information, the same tools such as scanning techniques, traditional image based technologies and manual insertion of measurement data which can be used to obtain the body-measurement data for the avatars can also be used for the VDR. For example, in the research conducted by Brooks & Brooks (2014), scanning technology was used to collect accurate anthropometric data of wheelchair-bound customers.

Advantage of VDR for virtual try-on experience:

Since VDRs provide the ability to test apparel items on the individuals own mirrored self, it can provide an overall feel and perception of the fit and size of apparel items, style and feel of the material on the individuals’ body shape (Brooks & Brooks, 2014; Kristensen et al., 2013). Virtual Fitting rooms can reduced misfits and social costs such as energy and reverse logistics costs (Gu & Tayi, 2015). The enjoyment of the online shopping can be enhanced by using virtual try-on technologies, through the interactivity and customer involvement it provides (Brooks & Brooks, 2014). Although interactivity functions may differ on aspects such as the perceived enjoyment, control and involvement, high levels of control and enjoyment may make up for a lower level of realism or vividness to maintain the image interactivity functions (Fiore & Jin, 2003). In addition, 3D virtual product presentation with high quality allow customers’ to find the actual apparel attribute information, provides them with enjoyment and allows customers’ to feel the social presence (Algharabat & Shatnawi, 2014). Consequently, high-quality 3D virtual presentation can reduce users’ perceived psychological risk and increase the customers’ intention to purchase items online (Algharabat & Shatnawi, 2014). Furthermore, such as is the case with the Avatars, it provides the possibility for online retailers to collect data and evaluate the customer’s choice set more accurately, through which a better understanding of the reasoning behind online purchasing behavior can be obtained (Kristensen et al., 2013; Gallino & Moreno, 2018). In addition, no extra assistance for trying out apparel is required and the problems of fitting rooms not being spatial enough to accommodate customers in a wheelchair is solved (Brooks & Brooks, 2014).

Disadvantage of VDR for virtual try-on experience

However, aside from these advantages, the technology also has some implications which might inhibit its adoption. First of all, replicating the textile feel of materials e.g. indicate the difference between silk and cotton, which can be gained when browsing in a physical shop, along with recreating a material’s physical dynamic that interacts with an individuals’ mirrored image motion is challenging and if done not very accurately might inhibit the adoption (Brooks & Brooks, 2014). Furthermore, a lack of availability of the technology might also inhibit the adoption (Hidellaarachchi et al., 2019). Another disadvantage that might inhibit the adoption of the technology is the customers’ trust and privacy concerns regarding sharing personal body-measurements data and the way body-measurements data can be obtained (Hidellaarachchi et al., 2019; Apeagyei, 2010).

2. Online Customer reviews (OCRs)

Product reviews is an example of 'interactivity' of a website (Fiore & Jin, 2003). According to Walsh et al. (2014), "customers reviews are a form of customer feedback on shopping websites which can be used to prevent apparel returns"(p.8). According to Walsh & Möhring (2017), "reviews are a virtual source of customer opinions about and experiences with products so other customers use this instrument to improve their decision making about the product" (p. 345).

Through the extra information that is provided by these customer-based instruments, customers can evaluate apparel items more accurately and can make better / well-informed purchase decisions which in turn reduces order returns (Walsh & Möhring, 2017).

Lohse et al. (2017) acknowledged the involvement theory developed by Zaichkowsky in 1985 which defines involvement as "a person's perceived relevance of the object based on inherent needs, values, and interests" (Zaichkowsky, 1985, p. 342). Lohse et al. (2017) saw a significant impact of product involvement in fashion e-commerce returns and identified a positive impact of positive OCRs on sales after returns and return behaviour. Therefore, instruments which include customers' involvement, such as online product reviews can potentially reduce product returns.

According to Misra et al. (2018), information regarding true fit and size which is based on objective information ((mis) match between the product true size and customers measurements) and the subjective characteristics such as style and other properties can be made available to customers through OCRs. According to Wang, Ramachandran & Sheng (2016) "this upfront provided product evaluation information regarding 'true-to-size/product size purchased' (fit valance) and 'reviewers' body-size' (fit reference) reduces fit and size pre-purchase uncertainty" (p. 0). The reason for this is that customers whom are in the presence of fit-reference information, can better evaluate the suitability of fit-valence expression found in online reviews (Wang, Ramachandran & Sheng, 2016). According to Zhang (2018), OCRs aid customers with the primary issue of online purchases which is lack of fit. Furthermore, OCRs aid customers in better searching for product information (Zhang, 2018).

Advantage of Online Customer Reviews (OCRs):

The advantages of online customers reviews are that they 1) have a unprecedented volume and reach compared to offline product feedback provision and with this a viral dispersion effect across different platforms and 2) have an on-demand availability (Lohse, Kemper & Brettel, 2017, p. 2635). Aside from this, detailed product information can be provided through online customer reviews (Zhang, 2018).

Disadvantage of Online Customer Reviews (OCRs):

Based on the opinion of a quality assurance inspector and current project leader of the return management project in the fourth biggest e-commerce retailer in the Netherlands and the second biggest online fashion retailer in the Netherlands, product reviews can be used to provide additional personalized advice to customers e.g. regarding personal fit & size. However, as they are based on customers opinions, an amount of uncertainty is present regarding the reliability of the information. According to the expert, customer biases can negatively influence the quality of the provided product review which negatively affects the reliability of the information in online reviews and the return rate. This drawback is also supported by Minnema, Bijmolt, Gensler, & Wiesel (2016), which argue that above-average and favourable reviews positively influence purchases but also enhance the return probability. Furthermore, according to the expert, one individuals' body-shape might be different than the other in the sense that individuals differ in detailed local shape characteristics. As a result, the apparel suitability might not be the same for each individual when based on the opinions regarding true-to-size/product size purchased and the reviewers' body-size. Therefore, according to the expert, online customer reviews can better be used as an additional supportive instrument, to assist customers with their online purchase decisions. Since online customers reviews can be perceived as additional supportive instruments, the decision was made not to include them in this research.

3. Product advice

Walsh et al. (2014) indicate that product advice such as size recommendations can be provided through e.g. avatars as virtual shopping assistant, in order to provide a successful match between the customer and apparel items. According to an interview carried out with an online marketing manager, the study indicated that the return rate can be reduced as customers simply find what they are searching for (Walsh et al., 2014, p. 8). Research conducted by De Leeuw et al. (2016) indicates that upfront information regarding colour, size, and styles is important, as it will reduce unnecessary order returns.

This suggest that through using the other seven instruments, product advice / upfront (personalized) product information can be provided such that consumers can make well-informed purchase decisions which can reduce the mismatch between expectation and reality. As the instrument 'product advice' can be perceived as a result of the other seven instruments, it is inherently present within these eight instruments and therefore not included as separate instrument.

4. Height/size instruments

Based on the examined literature, two types of height/size instruments were established, namely: 1) height/size charts and 2) size recommendations application.

(1) Height/size chart

A hight/size charts can best be described by a static tables which is indicated on apparel websites. People can first measure their own body in order to determine to which size category they belong, based on key body-measurements such as bust, waist and hips size. According to Apeagyei (2010), the objective of apparel sizing is "to divide standardized dimensions for the body and apparel into categories, with the aim to fit the maximum number of people with the minimum number of sizes; however size charts vary" (p. 59). Apeagyei (2010) and De Leeuw et al. (2016) identified inconsistencies within size charts which are used nowadays and actual body-sizes. E.g. a size 'small' (S) for example can represent other key body-measurements regarding bust, waist, hip size per country or region. This inconsistency means that sizes are not standardly defined and result in less size-reliability and therefore in increased misfits (Apeagyei, 2010). In order to reduce apparel returns due to fit & sizing issues, it is important that these standardized measurements are accurate for the body sizes they illustrate. Apeagyei (2010), De Leeuw et al. (2016) and Seewald et al. (2019) emphasized the importance of providing more detailed product-specific and accurate measurements regarding key body measurements charts, to reduce misfits and returns due to sizing issues. Furthermore, they emphasize the importance of keeping up with changes regarding anthropometric demographics over time and its translation into apparel sizes indicated in the charts.

(Dis)advantage of an adequate height/size chart:

An online provided size chart allows customers to check their size immediately (Jang & Burns, 2004). However, asking customers to measure their own body first and to consult "fit charts" can be time-consuming (Brooks & Brooks, 2014).

(2) Size recommendation application

An apparel size recommendation app is an instrument which apparel retailers can provide on their websites. According to the same quality assurance inspector and current project leader of the return management project, with size recommendation tools such as the 'Fit finder' customer can insert their body-size measurements and based on data gathered from other customers, the personal fit & size information can be provided. According to Nasibov et al. (2017), poor fit & size is the primary reason for customers dissatisfaction and the main barrier for online apparel retailers. The authors developed a fuzzy logic method which can be used to find the right size of apparel for each individual per brand. Peng & Al-Sayegh (2014) examined an apparel size 3D scanning recommendation app 'Shape mate', through which body- measurements are compared with relevant apparel sizes, to provide accurate size and fit advice to customers.

(Dis)advantage of a size recommendation application:

An online provided size measurement section allows customers to check their size immediately (Jang & Burns, 2004). Furthermore, accurate personal fit & size information can be produced for customers (Peng & Al-Sayegh, 2014). However, asking customers to type in their body measurements in order to obtain the recommend fit & size information, can be time-consuming (Brooks & Brooks, 2014). In addition, privacy issues can occur with regard to the way body-measurements data is gathered and used (Peng & Al-Sayegh, 2014). However, research conducted by Peng & Al-Sayegh (2014) indicated that using a service which is perceived as appropriate to use and which can assist customers with valid fit & size information is mostly perceived as more important than privacy concerns. Based on a focus group consisting of nine male individuals between the age of 20 and 49 years, they found that the use of cookies in the browser to track the accumulated data in an anonymous manner can enhance the perceived appropriateness and that simplifying the steps taken to use the system will also enhance the user experience (Peng & Al-Sayegh (2014).

5. Product-availability information

According to Walsh & Möhring (2017), 'product-availability information' is perceived as a customer-based instrument. It simply means "the provision of stock information which indicates whether the product is available for purchase or not in the online shops, in order to prevent apparel returns caused by late delivery" (Walsh et al., 2014, p. 8). According to Jang & Burns (2004) the provision of stock information is relevant such that consumers can check whether immediate purchase is possible. However, in this research, return reasons which stem from a late delivery are not included. Only return reasons which are related to apparel attributes which customers need in order to evaluate apparel items accurately and prevent unnecessary order returns are included. Therefore 'product-availability information' is not included as instrument in this research.

6. Customer Hotline

According to Jang & Burns (2004), customers service is associated with webservices that provide necessary information that allows customers to place orders and make shopping fast and easy. According to Fiore & Jin (2003), customers service is an example of 'interactivity' of a website. Telephone-mediated customers service falls under customer hotline (Walsh et al., 2014). Some web shops also offer a live chat line that allows customers to ask questions about apparel items (e.g. via WhatsApp) to a sales assistant of the company and a friend at the same time (Jang & Burns, 2004). Providing a customer hotline can address customers dissatisfaction, as it allows customers to ask questions about the purchased merchandise (Jang & Burns, 2004).

Based on the opinion of the same quality assurance inspector and current project leader of the return management project, instruments such as a 'customers hotline' are additional supportive customer service instruments designed to provide customers the ability to ask questions about the apparel items and address customers dissatisfaction regarding the apparel, pre-and post-purchasing. Therefore, according to the expert, customers service instruments such as a customer hotline can better be used as an additional supportive instrument, to assist customers with their online purchase decisions. As a result, a customer hotline as instrument is not included in this research.

7. Alternative product photos

According to Jang & Burns (2004), the purpose of providing 'alternative product photos' is to create a feel for the apparel details. The authors indicate that the presentation function for product image, such as altering apparel colour by click swatch', 'viewing every colour of the apparel item', or 'see all at once' provide customers to examine many parts of the product image in detail and compare with a various apparel items at once. Through this, a better perception and feel for the apparel details is created

(Jang & Burns (2004). However, according to Algharabat & Shatnawi (2014) , users learn more from 3D virtual product presentation compared to the indirect experiences (i.e. static pictures).

8. Zoom technologies

According to the reviewed literature, presentation functions for product image such as zoom technologies and mix and match functions also fall under the 'interactivity function' as described earlier. According to Jang & Burns (2004), zoom technology is made for feel of the apparel details as 'advanced zoom technology' provide the customer to examine various parts of the product image in detail. According to Algharabat & Shatnawi (2014), 'zoom technology' is also made for 'fit', as it provides the ability to zoom in or out on apparel items or virtual models to see how the apparel fit on their bodies. Through instruments such as zoom functions, customers can get a better perception of the fit, which reduces the customers' perceived pre-purchase risks and increases their purchase intention (Algharabat & Shatnawi, 2014).

2.4.1 Included technologies and instruments

The literature study has resulted in the identification of eight instrument categories. However, based on the opinion of an apparel e-commerce expert, only the following instruments and technologies are considered in this research: 1) virtual try-on experience technologies (avatar and virtual dressing room), 2) height/size instruments (static height/size chart and size recommendation application), 3) alternative product photos and 4) zoom technologies. The approached expert is a quality assurance inspector and current project leader of the return management project within the fourth biggest e-commerce retailer active in the Netherlands and the second biggest online fashion retailer active in the Netherlands.

Based on the opinion of this expert, the decision was made to exclude online customer reviews and customers hotline as according to the expert, these are more additional supportive instruments retailers can employ and can be used in all alternatives. Given the uncertainty factor of online customer reviews, as the provided information is based upon customers opinion, online customer reviews can better not be used as a separate alternative according to the expert.

Furthermore, the instrument 'product advice' can be perceived as a result of the other instruments. Given the reason that it is inherently present within the instruments, the decision is made to not include it as a separate instrument (alternative).

In this research, 'product-availability information' which according to Walsh & Möhring (2017) is an instrument that refers to the provision of stock information is also not included, as return reasons which stem from late delivery are not treated in this research. Only return reasons which are related to apparel attributes which customers need in order to evaluate apparel items accurately and prevent unnecessary order returns are included.

As a result, the following four instruments are not included in the research: 1) online customers reviews, customer hotline, 3) product advice and 4) product-availability information.

2.5 Designing technological alternatives

Now that the necessary apparel attribute information and requirements have been established along with the various instruments and technologies which can be used to communicate the required information, technological alternatives were designed through which customers can evaluate the apparel items during the online screening process of apparel. As indicated in section 2.3, in order for customers to make a well-informed online pre-purchase decision, they should be able to evaluate apparel items on the following attributes: 1) material, 2) colour, 3) fit & size and 4) style. Furthermore, through the literature study there was also established that creating the right expectations regarding the apparel attributes, providing accurate product information, and creating a 'feel' for and perception of apparel items displayed online are also essential requirements which need to be considered when designing the technological alternatives.

However, some of the included instruments and technologies in this research (see section 2.4) on their own cannot provide / visualize all these requirements as can be observed from their description provided in section 2.4. Therefore, is was necessary to combine some instruments such that they can fulfil the requirements, and function as comparable alternatives against the technologies which on their own can fulfil all the requirements.

2.5.1 Combining instruments

According to an approached expert who has a background in Academia and who functions as full professor at the Information and Communication Technology (ICT) section at Delft University of Technology, information technology instruments have many application possibilities and therefore it is important to break them down and approach them from a functional level. E.g. an online system such as an online platform or website can be designed according to the functionalities/ functional requirements the client wants it to fulfil. By taking this and the current practical applications of online retailers into consideration, the combined technological alternatives were designed. Consequently, the combined technological alternatives are based upon the functionalities of each instrument and the combination of instruments already employed by online apparel retailers in the Netherlands. The combination was necessary, since online apparel retailers currently only use some of the instruments combined and not separately on their website. Taking the current way of employment of instruments into consideration, practical and employable alternatives were designed. This was necessary as the goal of this research is to identify what the customers preference is with regard to various technological alternatives that online apparel retailers can actually employ in practice.

The combined technological alternatives were composed based upon one another. The first combined alternative constitutes a basic alternative which is close to the current arrangements used by online apparel retailers in the Netherland. The second combined alternative will attach a new functionality to the previous alternative, while keeping all other aspects alike.

The third technological alternative stands on its own, as it represents a technology which on its own can provide the necessary requirements. It already obtains the functionality of the second combined alternative, and has some new/ additional functionalities as well. The same accounts for the fourth technological alternative. Each alternative thus varies in functionality and every successive alternative has a new function added to it, implying 1) a higher (technical) complexity for both the customer (user) and retailer, and 2) costs that are made as a result of it.

Since all alternatives (A1, A2, A3 and A4) have the same fundamentals and only vary in some functionalities, many features of the subsequent alternative have already been explained in the previous alternative. The main advantage of this is that the required time to explain each alternative is minimized.

The combination of instruments which has led to the combined alternatives A1 and A2 had to be done very carefully and were finalized with the opinion of an expert. The inclusion of A3 and A4 in this research were also finalized based on the opinion of this expert, which is a quality assurance inspector and current project leader of the return management project within the fourth biggest e-commerce retailer active in the Netherlands and the second biggest online fashion retailer active in the Netherlands.

2.5.2 Technological alternatives

Underlying table 6 gives an overview of the technological alternatives which will later be used in the comparison analysis to determine the customers' preference. In the first column, the apparel attribute information is indicated which is needed in order for customers to evaluate apparel items, to make well-informed pre-purchase decisions and through this prevent unnecessary apparel returns.

Table 6: Technological alternatives in columns with their relevant components

Requirements *Apparel attributes:*	A1: The bare minimum *	A2: The bare minimum with a fit & size recommendation instrument *	A3: Avatar (digital computer-based twin)*	A4: Virtual Dressing Room (VDR) *
Material quality information	Alternative product pictures, Zoom function	Alternative product pictures, Zoom function	Alternative product pictures displayed on avatar, Zoom function, draping technology	Alternative product **dynamic images displayed on individuals' real mirrored self** (using e.g. ICT such as augmented reality), Zoom function.
Colour information	Alternative product pictures, Zoom function	Alternative product pictures, Zoom function	Alternative product pictures displayed on avatar, Zoom function	Alternative product **dynamic images displayed on individuals' real mirrored self**, Zoom function
Fit & size information	**Static height/size chart**	**Size recommendation application**	**Virtual try- on experience through avatar (personalized or retail-specified)**, Mix and match function, Size recommendation application	**Virtual try- on experience on the individuals' real mirrored self, using camera sensors and contemporary ICT such as augmented reality,** Mix and match function, Size recommendation application
Style information	Alternative product pictures, Mix and match function, Zoom function	Alternative product pictures, Mix and match function, Zoom function	**Virtual try- on experience through avatar (personalised or retail-specified),** Mix and match function, Zoom function	**Virtual try- on experience on the individuals' real mirrored self, using camera-based sensors and ICT such as augmented reality,** Mix and match function, Zoom function

* A photo of the apparel item, information about the apparel item in text form such as size, material and style are also indicated and the ability for colour selection exists, as this information and functionality is already provided and employed by most online apparel retailers in the Netherlands. This is the case for all four alternatives.

A1: The bare minimum

A1 is a combined alternative, composed of the following instruments: 1) Alternative product pictures, 2) mix and match function (to see overall outfit), 3) zoom function and 4) static height/size chart. In current practises, most online apparel retailers in the Netherlands use a combination of instruments 1, 3 and 4. However, the mix and match function, where photos of different apparel items can be placed together such the a perception of the entire outfit and style can be gained and the entire outfit can be evaluated, is not yet employed. Consequently, the mix and match function in this alternative is an addition compared to instruments used in current practices.

In this alternative, the size chart is provided by a picture of a body-measurements table. Customers have to measure their own body measurements first with a measuring band in order to know to which size category they belong. Alternative pictures of an apparel item, from different angles, exist. Pictures of apparel items can be mixed and matched in order to get a perception of the overall outfit (e.g. to get a perception of a certain coloured and style blouse on a certain coloured and style trouser) and the possibility to zoom in and out on pictures exists to examine the aesthetic apparel attributes (material/colour/style) more closely. By providing 'alternative product photos', a better perception and feel for the apparel details is created (Jang & Burns, 2004). The ability to change product colour and to view all colours simultaneously, allows the customer to examine various parts of the product image in detail and compare with various apparel items at the same time. Furthermore, presentation functions for product image such as zoom technologies also fall under the 'interactivity function' as described earlier. The ability to zoom in and out on apparel items can create a feel of the apparel details as 'advanced zoom technology' allows the customer to evaluate various parts of the product image in detail (Jang & Burns, 2004).

A2: The bare minimum with a fit & size recommendation instrument

A2 is also a combination of instruments. This alternative differs from the previous one, in the sense that it uses an interactive size recommendation application such as e.g. 'Fit finder', instead of a static height/size chart. Customers can type in their body measurements and based on this and on behalf of the collected body-measurement data from other customers, the recommended fit and size is provided.

A3: Avatar (digital computer-based twin)

Based on the information proved from the literature study regarding technologies and instruments (see section 2.4), the avatar is a technology which on its own can fulfil all the requirements. Based on the literature study, an avatar distinguishes itself from the previous alternatives, in the sense that it provides a virtual try-on experience. Consequently, an avatar is a computerized digital twin designed to create a virtual try-on experience for customers. With avatars, apparel items can be virtually tried on in a computer-generated environment. Body-measurement data can be inserted manually by customers or via pictures or body scanning technology. Different apparel items can be mixed and matched in order to determine the overall outfit and the possibility to zoom in and out exists. Is also provides personalized size recommendations, based on fit. Through 'pressure maps' the personal fit and size information is indicated. The apparel material dynamic flow (fit) on the avatar is mimicked using draping technology.

A4: Virtual Dressing Room (VDR)

Based on the information proved from the literature study regarding technologies and instruments (see section 2.4), the VDR is a technology which on its own can fulfil all the requirements. A VDR differs from an Avatar, as people now see themselves (it is not computer based). A VDR, such as an Avatar, is a state-of-the-art technology which can be used to create a virtual try-on experience for customers. It uses e.g. body scanning sensors to determine the body shape and augmented reality (such as Snapchat filters) by which apparel can be virtually 'swipedragged' on the body of customers and mixed and matched in order to determine the overall fit of outfits. The possibility to zoom in and out exists to examine the aesthetic apparel attributes (material/colour/style) more closely. Through the scanning and augmented reality technology, the ability to provide personalized size recommendations exist**.**

Main differences between technological alternatives

Based on the information gained from the literature study as indicated in section 2.4, product presentation on websites where customers have the possibility to view items together (mix and match function) and use zoom technology can give customers a feel for visual apparel details (Jang & Burns, 2004). The ability to adjust apparel colour by click swatch', 'viewing every colour of apparel items', or 'see all at the same time' provide the customers to examine various parts of the apparel image in detail and compare with a lot of apparel items simultaneously. Through this, a better perception and feel for the apparel details is created (Jang & Burns, 2004). According to Algharabat & Shatnawi (2014), 'zoom technology' is also made for 'fit', as it provides the ability to zoom in or out on apparel items or virtual models to see how the apparel fit on their bodies. Through instruments such as zoom functions, customers can get a better perception of the fit, which reduces the customers' perceived pre-purchase risks and increases their purchase intention (Algharabat & Shatnawi, 2014). As all four alternatives (A1, A2, A3 and A4) have these functionalities, they can all create a perception and feel and for apparel details.

The main functional difference between A1 and A2 compared to A3 and A4, is that A1 and A2 do not provide the ability to test the personal apparel fit & size and style on the individuals body. A1 and A2 solely provide the ability to mix and match alternative pictures of apparel items to get a perception and feel of the total outfit. On the other hand, an Avatar (A3) and VDR (A4) provide a 'personalized online shopping approach' through the virtual try-on interactive technology, which allows users to establish a more real live perception of what their products might look like on their own body-shape without actually having to try the apparel on. The image interactivity allows customers to see apparel items on the Avatar (customers virtual body shape) in 360 degree which along with the zoom function can create a feel for visual apparel information such as fit, material and colour (Algharabat & Shatnawi, 2014). The personalized online shopping approach might require more user intervention (Hiavaidellaarachchi et al., 2019).

The main functional difference between the Avatar (A3) and a VDR (A4) lays in the level of realism. With the VDR (A4), apparel items can be virtually tested on a virtual dynamic image of an individuals' mirrored self. Therefore, the VDR (A4) provides a higher level of realism compared to the Avatar (A3) and a higher level of realism compared to A1 and A2.

Based on the literature study regarding the instrument and technologies, as indicated in section 2.4, more user intervention is required through the level of interactivity present at virtual try-on applications (Hiavaidellaarachchi et al., 2019). Furthermore, the perceived enjoyment of online shopping can be enhanced when using virtual try-on technologies, through the interactivity and customer involvement they provide (Brooks & Brooks, 2014). However, since interactivity functions may vary on levels of perceived enjoyment, control and user intervention, high levels of control and enjoyment (e.g. at the Avatar) may make up for less realism or vividness (Fiore & Jin, 2003).

Furthermore, according to the literature study, privacy and security concerns might also hinder the adoption of the alternatives with regard to how the technologies acquire body-size measurements data and how data is used by the online apparel retailers. As established through the literature study, various approaches, such as manual insertion of body measurement, scanning technology, uploading pictures etc., can be used to acquire body measurement data for personal fit & size recommendation applications.

According to the examined literature (see section 2.4), these approaches all vary in the level of accuracy, and with that also vary in the customers' perceived privacy and security concerns. For example, 3D scanning technology where customers have to disrobe in order to get accurate body-measurements data to virtually try-on apparel, might be perceived as uncomfortable and raise ethical implications which can inhibit the adoption of the technology by users (customers) (Apeagyei, (2010); Hidellaarachchi et al. (2019)). However, currently scan-based methods are the most accurate in acquiring actual body dimensions and shapes, but aside from being the most expensive approach, they also impose high privacy concerns (Li et al., 2017). Manual insertion of measurements data by customers, has an inherent accuracy risk as customers might insert incorrect /less accurate body-measurements data (Li et al., 2017). According to the authors, manually customised avatars are very accurate (in terms of sizes) and are comparable to avatars created by using scan measurements. Furthermore, the modelling process is efficient, meeting the requirement on real-time application (Li et al., 2017). However, a study conducted by Peng & Al-Sayegh (2014), indicated that a service which provides accurate fit & size information in an appropriate manner (using e.g. cookies in the browser to track accumulated data in an anonymous manner) and is easy to use, is mostly perceived as more important than privacy concerns. This implies that privacy and security concerns can indeed obstruct the users' acceptance of the technological alternatives, but if taken care of appropriately and securely, trade-offs can be made.

Since this research focusses on gaining insight about the overall customers' perception and preference regarding technological alternatives they can use to make well-informed pre-purchase decisions, in order to increase their online purchases successes and reduce unnecessary apparel returns, the decision is made not to focus on the different technical approaches which can be used to obtain e.g. body measurements data in order to provide personal fit & size information. Furthermore, to preserve clarity for customers and due to restricted time of this research, the technical approaches retailers can use to provide personal fit & size information are excluded in this research. Within this research, the technical complexity is thus approached from the customers (users) perspective. Consequently, the various approaches which online apparel retailers can employ to obtain / extract body shape and size information such as uploading pictures (2D approach to extract 3D body measurements data) and different scanning techniques (3D approach) along with their differences are not treated in this research.

Once insight is first gained regarding the overall customers' perception of these technological alternatives which retailers can employ, further in depth research can be conducted about all the different approaches that can be provided by the online retailers to provide personal fit & size information. In addition, since this research is exploratory, it is also still too early to indicate the actual impact of the technological alternatives on actual apparel returns and convert that into monetary terms. Both of these steps can be considered as difficult as many impacts of the examined technologies and instruments on actual apparel returns are still not defined.

Based on the acquired information, the following can be implied: 1) the technological alternatives vary in functionality, 2) level of complexity for customers (users), 3) realism, 4) privacy and security concerns and 5) required user intervention. Following the MCA approach, in order to evaluate the customers preference regarding the technological alternatives (A1, A2, A3 and A4), a set of decision criteria needed to be established which also takes these factors into consideration. In the following chapter, the acquisition of the set of decision criteria is described.

3. AQUISITION OF EVALUATION CRITERIA

In this chapter the second sub-question is addressed, namely: *SQ2. What criteria are important for evaluating these technological alternatives?* In order to acquire a relevant set of decision criteria, this chapter is structured as follows. First, in section 3.1, the literature review regarding the Technology Acceptance Model (TAM) is indicated, wherein the TAM is described, the applicability of the TAM in this specific research case is indicated and the identified criteria are presented and interpreted. In section 3.2, the set of decision criteria that is used to evaluate the customers preference of each technological alternative is presented.

3.1 Literature Study on the Technology Acceptance Model

In 1986, Davis developed the technology acceptance model (TAM) to measure the use of a technology. The TAM is a way of looking at technologies and deciding whether they will be accepted or not by communities. According to Davis (1986), the actual use of a technology is determined by the attitude towards using a technology, which in turn is a function of two main determinants which are: 1) the perceived usefulness (PU) and 2) the perceived ease of use (PEU) (see figure 2). The Attitude Toward Using a technology is determined via these two determinants (PU and PEU) which in their own are influence by external variables. These external variables or so called external predictors are indicated with X1, X2 and X3 in underlying figure 2).

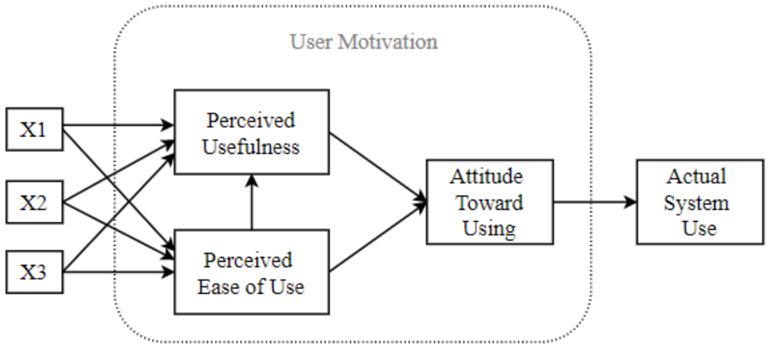

Figure 2: Technology Acceptance Model (retrieved from: Davis, 1986, p. 24)

Perceived usefulness (PU): is defined as “the degree to which an individual believes that using a particular system would enhance his or her job performance” (Davis, 1986, p. 26). Consequently, PU expresses the effectiveness of a technology in addressing a specific function.

Perceived ease of use (PEU): is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis, 1986, p. 26). According to Davis (1986), “PEU plays a crucial role in understanding an individual’s response to information technology” (p. 26). When a technology is not easy to use, there is no positive attitude towards the technology.

3.1.1 Predictive validity of TAM

In literature, TAM is a very prominent model in explaining the technology acceptance and has mostly focused on information technology (IT). Currently, many variants of the TAM model exists in literature as can be seen in the review conducted by Marangunić & Granić (2015). However, since its introduction, nearly 30 years ago, TAM has been mainly criticized on its predictive validity, as it only predicts the acceptance based on two determinants (Legris, Ingham, & Collerette, 2003; Chutter, 2009). Since new technologies have rapidly been introduced, it becomes more difficult now to predict the actual use, simply by using the PU and PEU as determinants (Hong & Tam, 2006).

Therefore, in order to increase the validity of TAM, Marangunić & Granić (2015) conducted an extensive literature review and established three categories of modifications that were made over the years to the original TAM model. According to Marangunić & Granić (2015), the Attitude Toward Using a technology is not only influenced by the determinants PU and PEU which are on their own influence by external predictors or external variables. However, a category of additional variables which the authors refer to as 'factors from other theories' also influence the Attitude Toward Using a technology. According to Marangunić & Granić (2015), this category refers to factors such as subjective norm, expectations, user participation risk theory and trust which aside from the PU and PEU influence the Attitude Toward Using a technology.

A second additional category of factors identified by Marangunić & Granić (2015) which directly influences the Actual System Use is referred to as 'usage measures'. Usage measures includes factors such as attitude toward technology and usage perception.

The third category of additional factors developed by Marangunić & Granić (2015), is named 'contextual factors' which include factors such as gender and cultural diversity, which according to these authors also have a moderating effect on overall technology acceptance.

Therefore, to increase the predictive validity of the TAM in this research, additional external variables which can directly influence 1) the Attitude Toward Using a technology and 2) the Actual System Use are also examined through the literature review presented in paragraph 3.1.2.

3.1.2 Results

In this paragraph the variables, which were identified through the literature study regarding TAM are presented, which can later be used to evaluate the technological alternatives.

In order to acquire relevant literature, the scientific database Scopus was consulted. The aim was to find criteria which are relevant to determine the acceptance of a technology by a community. More specifically, the aim was to identify external variables which influence the Perceived Usefulness (PU) and the Perceived Ease of Use (PEU) of each technology or additional categories consisting of external variables which directly influence the Attitude Toward Using a technology and the Actual System Usage. The search was set on finding search terms in the title, abstract and keywords of the articles. Only accessible findings in published papers such as scientific journals and conference proceedings and published book chapters were used. After having inserted multiple different search terms, the following search terms were used:

(1) TAM AND "success factors" AND "e-commerce"

This string of search terms has resulted in 18 articles on Scopus. Only one article specifically included apparel. Of the 18 articles, 15 were included because three articles could not be included due to access restrictions.

Underlying table 7 gives an overview of the identified external variables and determinants of the TAM.

Table 7: Identified additional determinants and external predictors through literature

#	Main criteria (external predictor)	Category	Study	Method	Field of study	Significant sub-criteria
1	Information quality	As external predictor of PU	Fedorko, Bacik, & Gavurova (2018)	Qualitative exploratory method (using online questionnaires)	E-commerce web site visiting	Completeness of provided information
						Truthfulness of the of provided information
		As external predictor of PU	Landeweerd, Spil, & Klein (2013)	Qualitative interview model-based research method	Google Health	Providing the right information
		As external predictor of PU	Zhou & Zhang (2009)	SEM	E-commerce Website Quality	Information accuracy
		As external predictor of PU and PEU	Cao, Zhang, & Seydel (2005)	Exploratory factor analysis	B2C e-commercial web sites (e.g. apparel websites)	Information accuracy
2	Service quality	As external predictor of PU	Fedorko et al. (2018)	Qualitative exploratory method (using online questionnaires)	E-commerce web site visiting	Speed of feedback provision to users
		Trust as third additional determinant to TAM	Zhou & Zhang (2009)	SEM	E-commerce Website Quality	-
		As external predictor of PU and PEU	Cao et al. (2005)	Exploratory factor analysis	Online shopping	Trust (i.e. keeping site secure and reliable)
3	System quality	As external predictor of PU	Landeweerd et al. (2013)	Qualitative interview model-based research method	Google Health	Reliability of information
		As external predictor of PEU	Zhou & Zhang (2009)	SEM	E-commerce Website Quality	PEU

		As external predictor of PU and PEU	Cao et al. (2005)	Exploratory factor analysis	B2C e-commercial web sites (e.g. apparel websites)	Responsiveness, which referred to reducing loading time
		As external predictor of PU and PEU				Search facility, which refers to searching time and making searching easier.
		As external predictor of PEU	Fedorko et al. (2018)	Qualitative exploratory method (using online questionnaires)	E-commerce web site visiting	Availability of the web shop on the Internet (e.g. on computer or mobile phone)
		As external predictor of PEU				Attractiveness
		As external predictor of PEU				Navigation (suitable navigation and which is easy to use and control)
		As external predictor of PEU	Pires & Lai (2009)	Regression analyses	E-Government	Usability, which refers to how visually appealing, consistent, fun and easy to use a website is
						Navigation, which refers to how easy users can find the required information on the website
4	Social pressure	As external predictor of PU	Landeweerd et al. (2013)	Qualitative interview model-based research method	Google Health	-
5	Perceived compatibili ty	As external predictor of PU	Landeweerd et al. (2013)	Qualitative interview model-based research method	Google Health	The extent to which an innovation is perceived as compatible with prior experiences
						The extent to which an innovation is perceived as compatible with existing work practices
						The extent to which an innovation is perceived as compatible with an individuals' values and norms

6	(Perceived) trust	As external predictor of PU	John (2012)	SEM	E-Commerce	Perceived privacy, has the highest significant impact
						Perceived security
						Service quality
		As external predictor of PU	Zhou & Zhang (2009)	SEM	E-commerce Website Quality	Service quality
		Trust as third additional determinant to TAM	Landeweerd et al. (2013)	Qualitative interview model-based research method	Google Health	Privacy regarding the disclosure of privacy-sensitive information such as personal health information
		Trust as third additional determinant to TAM	Shabana & Arif (2011)	Review	E-commerce	Perceived risk: data privacy which refer to users' data not being used for other purposes without the consent of users and data anonymity,
						Perceived risk: providing users' security through encryption such that data cannot be accesses by unauthorized parties
						Users' web experience
						Perceived site quality
		Trust as third additional determinant to TAM	Shabana & Arif (2011)	SEM	Online shopping	Perceived market orientation.
		Trust as third additional determinant to TAM	Salam, Iyer, Palvia, & Singh (2005)	Exploratory research of a framework regarding the concept of trust	E-commerce	Prior experience of using the internet
						Communication in Internet use
						Online relationships, between vendor and customers
7	User satisfaction	As external predictor of PU	John (2012)	SEM	E-Commerce	Online customer trust positively influences their satisfaction level

		As external predictor of PU	Pires & Lai (2009)	Regression analyses	E-Government	Information quality
		As third additional determinant to TAM	Chiu et al. (2009)	SEM	Online shopping	Trust
		As third additional determinant to TAM	Tella (2012)	Regression analysis (SEM)	E-payment systems	Service quality,
						Security
						Perceived enjoyment
						Perceived speed of technology, which refers to the payment speed of e-payment system
8	User involvement	As external predictor of PU and PEU	Wang & Liu (2009)	Multiple regression analysis	Online Customization for Apparel	-
9	(Performance, Perceived) Risk	As external predictor of PU, as third additional determinant to TAM	Lee (2009)	SEM	Internet banking	Users' security risk
		As third additional determinant to TAM	Shi (2013)	SEM	Small Enterprises E-Commerce	-
10	Attractiveness	As external predictor of PU and PEU	Cao et al. (2005)	Exploratory factor analysis	B2C e-commercial web sites (e.g. apparel websites)	Playfulness i.e. appealing and exciting
11	Experience	As third additional determinant to TAM	Changchit, Cutshall, Lonkani, Pholwan, & Pongwirittho n (2019)	SEM	Online shopping	Past experiences of online shopping
		Contextual factors (additional factor)	Chiu et al. (2009)	SEM	Online shopping	-

12	Perceived security	As third additional determinant to TAM	Changchit et al. (2019)	SEM	Online shopping	Keeping financial and user' private information secure while conducting the Internet transactions (e.g. authentication of user identity and status)
13	Perceived uncertainty	As third additional determinant to TAM	Changchit et al. (2019)	SEM	Online shopping	-
14	Fairness	Trust as third additional determinant to TAM	Chiu et al. (2009)	SEM	Online shopping	Distributed fairness Procedural fairness Interactional fairness
15	User innovativeness	As third additional determinant to TAM	Wang & Liu (2009)	Multiple regression analysis	Online Customization for Apparel	Trust
16	Perceived benefits (e.g. financial and transaction speed)	As third additional determinant to TAM	Lee (2009)	SEM	Internet banking	-
17	User Characteristics & company characteristics	Contextual factors (additional factors)	Olson & Boyer (2003)	Hierarchical cluster analysis	Internet purchasing in small organizations	Education level Tenure in the workforce Annual training received

Cao et al. (2005) examined the influence of aspects of system, information and service quality and attractiveness on the effectiveness of e-commerce web site quality. The results indicated that customers add more value to finding accurate information, searching fast and securely placing orders.

Zhou & Zhang (2009) investigated the same factors as Cao et al. (2005), except for the factor attractiveness, on both the PU and PEU. Furthermore, they also included the factor 'Trust' to determine the PU. The results show that the PEU is highly influenced by system quality, whereas PU is highly affected by information quality. Furthermore, service quality significantly affects the user trust, and has no significant effect on PU and PEU.

Lee (2009) examined the influence of performance risk on PU in the field of online banking, and found that performance risk negatively influences the PU. In addition, the external variable 'perceived benefit', has a significant positive impact on the attitude towards the use of online banking. Perceived benefits refers to benefits such as financial benefits (e.g. lower transaction fees) and faster transaction speed.

Amongst the four types of risks examined by Lee (2009), security risk was deemed the most essential for the success of internet banking (Lee, 2009).

John (2012) highlighted the importance of consumers trust in online systems to examine the success of e-commerce and examined critical variables that affect users' trust in online information systems amongst online shoppers. The study examined the impact of trust and user satisfaction on PU. The results of the study indicated that user satisfaction has more influence on PU than trust. Furthermore, based on research conducted by John (2012), "perceived security, perceived privacy, service quality, system quality and vendor familiarity are the significant variables of online trust in the context of B2C e-commerce" (p. 3). However, amongst these five variables, the first three are deemed the most significant, and from these three variables, 'perceived privacy' has the highest significant impact on trust.

Wang & Liu (2009) investigated the impact of users' involvement and innovativeness towards online apparel customization. The influence of the variable 'involvement' was examined on the PU as on PEU. The results indicated that more users' involvement results in a higher positive relationship between PU and customers' attitude regarding the acceptance of online apparel customizations. This is also the case for the relationship between PEU and the customers' attitude toward acceptance of online customizations. Wang & Liu (2009) used 'trust' to obtain more accurate correlations when investigating the impact of users innovativeness on the Attitude Toward Using a technology. The results show that more user innovativeness results in a higher positive attitude and users' behavioural intention to use technologies for online customization of apparel. According to the authors, this could be explained by the following that customers with a higher innovativeness level are more willing to use new technologies and to buy new apparel items (Wang & Liu, 2009).

Fedorko et al. (2018) analysed experience variables on web site visiting (technology) and found out that information quality has a significant positive impact on PU. Furthermore, the availability of the e-shop on the Internet, which is and aspect of system quality, has a significant positive influence on the PEU. Furthermore, they indicate that the attractiveness of e-commerce websites and suitable navigation which is easy for users to control, are both aspects of system quality which are of high importance to increase the PEU (Fedorko et al., 2018).

Pires & Lai (2009) investigated the acceptance and satisfaction of an e-government portal developed to improve public administration efficiency. The authors indicate that information quality is the most important element of users' satisfaction and that usability and navigation are the most important proportions of system quality.

According to Landeweerd, Spil, & Klein (2013), information quality needs to be aligned adjacent with usefulness, as the usefulness of services is determined by the provision of the right information. Furthermore, the authors indicate that perceived compatibility refers to the possibility of products to integrate with other Google products and indicate that a higher level of compatibility enhances the PU of a service. According the authors, the system quality was perceived reliable as all necessary information was readily available online. Landeweerd et al. (2013) also indicate that social pressure increases the probability of consistent service usage. In addition, privacy regarding the disclosure of personal information is an important variable for the adoption of technologies or products, especially regarding health information.

Shabana & Arif (2011) highlight the importance of trust in e-commerce by providing definitions of trust and various basic models to secure trust in e-commerce. Furthermore, the authors highlight the most common risks customers face and what security measures can be taken to address these risks. Shabana & Arif (2011) indicate that the success of e-commerce depends essentially on consumers trust. According to the authors, customers experience pre-purchase uncertainty as they do not have direct contact with the merchandise. Without appropriate security practises, private customers' information can be stolen by hackers or insiders and or commodified to third parties which imposes serious privacy and data integrity concerns. According to Shabana & Arif (2011), trust is dependent on variables such as perceived risk, participation in e-commerce, users' web experience, perceived site quality and perceived market orientation.

Salam et al. (2005) investigated the impact of prior experience and communication on customers' trust in the context of e-commerce. The authors indicate that prior experience and communication in Internet use and online relationships (between vendor and customers) are important variables in the formation of consumers' usage beliefs.

Chiu et al. (2009) integrated trust and fairness and examined the impact of these two variables on customers' loyalty intentions towards online shopping and found that procedural, interactional and distributive fairness were strong external predictors of the determinant trust, which on its own impacts the satisfaction of customers. Of these three variables, interactional fairness has the strongest influence on trust. According to Chiu et al. (2009), the strong impact of interactional fairness on trust might be caused by a complete and transparent information provision from vendors to customers. Chiu et al. (2009) also examined the impact of the contextual variables such as internet experience and shopping experience on customers' loyalty intention towards online shopping. However, the results have shown that neither shopping nor Internet experience have significant impact on customers' online purchasing loyalty intention.

On the other hand, Olson & Boyer (2003) investigated various contextual variables that affect Internet purchasing adoption in order to identify clusters or users' patterns with regard to how various segments view internet purchasing. The authors indicate that these users' characteristics and companies characteristics might have significant influence on user behavior and therefore affect technology adoption differently. In total, six clusters were identified which each vary in attitude towards the using Internet technology. This research thus indicates that contextual variables can impact the individual level of technology acceptance.

Changchit et al. (2019) expanded the original TAM model with the variables 'past experiences', 'perceived security' and 'perceived uncertainty' and found that these factors along with the PEU are significantly impact the attitude of Thai consumers with respect to online shopping. Shi (2013) expanded the original TAM model as well and found that perceived risk negatively influences the e-commerce adoption.

Tella (2012) investigated the impact of users' satisfaction as dependent variable to predict and explain the successful adoption of an e-payment system. Tella (2012) found that security, service quality, perceived speed and perceived enjoyment have a significant impact on users' satisfaction, and that this variable is important for the success of the e-payment system.

3.1.3 Interpretation of results

Based on the examined literature, no external variables were identified which directly influence the Actual System Use. However, as this research is more exploratory by nature, it is still too early to determine the Actual System Usage and impact on actual apparel returns. Consequently, this research focusses on identifying the customers' preference towards the technological alternatives (identifying the attitude / intention to use a technology) instead of determining the Actual System Use.

Another observation that could be made, based on the examined literature, is that the TAM is mostly operationalized with regression analysis through Structural Equation Modelling (SEM), which corroborates the correlation between the two determinants of TAM (PU and PEU) based on data collected through surveys. Olson & Boyer (2003) used hierarchical clustering analysis (HCA) to gain insight regarding how various segments of respondents perceived internet purchasing and (Cao. et. al) used exploratory factor analysis (EFA) to validate a set of factors on web site quality. In both studies, data was also collected through surveys.

Studies which used a qualitative approach were Fedorko et al. (2018) and Landeweerd et al. (2013). Fedorko et al. (2018) used a qualitative approach, wherein the original TAM was extended with additional variables and surveys were used to determine the impact of these variables on online shoppers e-commerce website visiting. Landeweerd et al. (2013) used a literature study to identify success factors and used qualitative interviews to predict the potential user adoption of the google products. Based on the examined literature, the observation could be made that the qualitative operationalization of TAM in literature is still scarce. The qualitative nature of this research contributes to this, by applying a qualitative application of TAM using an MCA approach wherein the novel Bayesian BWM is applied. Furthermore, by handling the identified variables in a qualitative manner, by using interviews for example, more explanation can be provided regarding the customer preference for the technological alternatives compared to evaluating the preference by simply using a quantitative modelling approach.

3.2 Selection of evaluation criteria

In the previous section, variables which determine the acceptance of technologies were identified through a literature review conducted about the TAM. In this section, variables which are relevant for this specific research are selected which are later used as criteria to evaluate the customers preference for each technological alternative. The selection of the criteria are established by the literature review regarding TAM (presented in section 3.1.2) and were finalized with the opinion of domain experts.

From the literature review results, a total of 17 variables (hereafter criteria) were identified. Based on the importance of each criterion as indicated in literature, their significant impact on technology acceptance and the applicability of the criteria to this specific online apparel case, the inclusion of each criterion is determined. Since in this research a survey is used to infer the criteria-weights using the BWM, it was very important to formulate the criteria and their corresponding sub-criteria as clearly as possible such that customers understand the survey questions and fill it in accordingly to preserve the research validity.

3.2.1 Categorization of identified criteria

From the literature review results presented in section 3.1.2, a total of 17 main-criteria were identified, which are on their own influenced by other characteristics or significant sub-criteria. However, using many criteria is not convenient, as it can become difficult to handle and compare information (Choo et al., 1999). Since the literature study has resulted in the identification of more than nine criteria, first a categorization of the criteria was required (Rezaei, 2015).

Appendix B, gives an overview of the categorization of the criteria. The categories were constructed, based on the synergy between the (sub)-criteria. To illustrate, e.g. completeness, truthfulness, rightness and accuracy of information have been identified through the literature study as significant sub-criteria for the main-criterion 'information quality' (see table 7 of section 3.1.2). However, 'reliability of information' has been used as sub-criterion to measure the main-criterion 'system quality' (see table 7). As these five sub-criteria all have in common that they describe the quality of provided information, they are assigned to the category called 'Characteristics of the quality of provided information'.

Main-criteria & sub-criteria

As a results, the main-criteria used within this research were formulated based upon the categories e.g. the main-criterion 'quality of provided information' was retrieved from the category 'Characteristics of the quality of provided information'. The sub-criteria (i.e. completeness, truthfulness, rightness, accuracy and reliability of information) were the characteristics of this main-criterion.

When a criterion was used as main-and sub-criterion, the significant sub-criteria used to measure the main-criterion were included. For example, the criterion 'service quality' is used by Fedorko et al. (2018) and Cao et al. (2005) as main-criterion, but Zhou & Zhang (2009) and John (2012) used 'service quality' as sub-criterion (see table 7). Since Fedorko et al. (2018) and Cao et al. (2005) used the significant sub-criteria 'speed of feedback provision to users ' and 'trust, keeping site secure and reliable' to measure the main-criterion 'service quality', these second level criteria (sub-criteria) were included instead of 'service quality'.

Based on the literature review results as indicated in table 7 of section 3.1.2, the following five categories of criteria were established: 1) characteristics of the quality of provided information, 2) characteristics of trust regarding information gathering and handling, 3) ease of use technological characteristics, 4) characteristics of attitude /behavioural component and 5) characteristics of social interaction. In appendix B, an overview of the five categories along with the corresponding sub-criteria and descriptions is presented. However, given the practical goal of this research, not all of these categories nor their characteristics (sub-criteria) are important to include. In the following section, this will be further described.

3.2.2 Category selection

This research aims to determine what the customers' preference is regarding technological alternatives which provides them with the necessary apparel attribute information upfront such that they can 1) more accurately evaluate apparel items online and determine the personal apparel match, 2) increase their online purchase success and 3) prevent unnecessary order returns. As a results of this, a comparison analysis is required, wherein the alternatives are compared based on a set of decision-criteria that define the effectiveness of each technological alternative in addressing the necessary apparel attribute information, the functional requirements as identified in section 2.3 and the users' (customers') technology acceptance.

This being said, within this research, the interest does not lay in evaluating the psychological reasoning behind a customers' choice /preference in depth, hence SEM would be applied. On the contrary, the interest lays on establishing the customers' preference of each technological alternative, by including the most essential set of decision-criteria based on TAM literature that define the effectiveness of each alternative.

As a result, the fourth category 'Characteristics of Attitude/behavioural components' and the fifth category 'Characteristics of Social interaction' are not included in this research, as they do not represent criteria through which the effectiveness of the technologies can be evaluated and compared to one another. Furthermore, Davis (1986) observed that perceived behavioural control and subjective norms did not have a high correlation with the use of technologies. As a result, TAM excludes these determinants and predicts the actual use (i.e. behaviour) simply by using PU and PEU as determinants. Underlying table 8, gives an overview of the three categories along with the corresponding sub-criteria and descriptions which were included in the research.

Since this research is exploratory, it is still too early to indicate the actual impact of various technologies and instruments on actual apparel returns and convert that into monetary terms. Therefore, financial factors were also not included in this research.

Table 8: Categorization of criteria

#	Category	(Sub) criteria	Description	Sub-criteria retrieved from underlying study
1	Characteristics of the quality of provided information	Completeness	The completeness of the provided information	Fedorko et al. (2018)
		Truthfulness	The truthfulness of the provided information	Fedorko et al. (2018)
		Rightness	Providing the right information	

		Reliability	The reliability of the provided information	Landeweerd et al. (2013)
		Accuracy	The accuracy of the provided information	Zhou & Zhang (2009)
				Cao et al. (2005)
2	Characteristics of Trust, regarding information gathering and handling	Data Privacy	The disclosure of privacy-sensitive information such as personal health information.	Landeweerd et al. (2013)
			The ability of customers to control what personal information is shared online and used in an unauthorised way, while conducting online transactions.	John (2012)
			Data privacy which refer to users' data not being used for other purposes without the consent of users and data anonymity,	Shabana & Arif (2011)
		Security	Providing users' security through encryption such that data cannot be accesses by unauthorized parties.	Shabana & Arif (2011)
			Keeping site secure and reliable	Cao et al. (2005)
			The perception of customers regarding the level of protection against privacy and security threats	John (2012)
			The possibility of the portals to provide access to all service an functionalities in a secure manner	Tella (2012)
			Users' security risk	Lee (2009)
			The possibility to secure financial and personal information secure, when transmitted online	Changchit et al. (2019)
3	Ease of use Technological Characteristics	Responsiveness	Reducing loading time	Cao et al. (2005)
		Speed	The speed of feedback provision to users	Fedorko et al. (2018)
			Perceived speed of technology, which refers to the payment speed of e-payment system	Tella (2012)
			Benefits such as faster transaction speed	Lee (2009)
		Search facility	Searching time and making searching easier	Cao et al. (2005)
		Availability	Availability (whether on a laptop, computer, mobile phone etc.)	Fedorko et al. (2018)
		Attractiveness	Playfulness	Fedorko et al. (2018)
		Navigation	Refers to how easy users can find the required information on the website.	Pires & Lai (2009)
			Navigation, that suits the users and can be easily be controlled.	Fedorko et al. (2018)
		Perceived enjoyment	Perceived enjoyment when using the technology	Tella (2012)
		Usability	Refers to how visually appealing, consistent, fun and easy to use a website is.	Pires & Lai (2009)

		Playfulness	The extent to which the technology is appealing and exciting	Cao et al. (2005)

In conclusion, from the five identified categories as indicated in appendix B, only the first three categories are included (see table 8) in this research to obtain a final set of decision-criteria. These categories are: 1) characteristics of the quality of provided information, 2) characteristics of trust regarding information gathering and handling and 3) ease of use technological characteristics. In the following section, the inclusion of the categories along with their characteristics are described.

3.2.3 Finalizing the set of evaluation criteria

In this section, the final set of main-criteria and sub-criteria is presented. This is done by first describing the importance of the each main-criterion in this specific research. Secondly, the application of the main-criteria with respect to the TAM is explained. Thirdly, the sub-criteria which are included per main-criterion are explained. In order to acquire a usable and manageable set of sub-criteria, it was necessary to aggregate some sub-criteria since using many criteria is not convenient, as it can become difficult to handle and compare information (Choo et al., 1999). The aggregation of the sub-criteria was based on the synergy between the sub-criteria.

Main-criteria 1: Quality of provided information

Importance of criterion: Since the main goal of this research is to examine which information technology is best accepted by customers in informing them upfront such that unnecessary apparel returns can be prevented, the main-criteria 'information quality' as identified through the literature study is inevitable. Therefore, the composed alternatives need to be compared based on characteristics which explain 'information quality' per alternative. As a result, all significant sub-criteria which were identified through the literature study (presented in table 7) which were used to measure the main-criterion 'information quality', were assigned to the category 'characteristics of the quality of provided information'. From this category, the main-criteria 'quality of provided information' was retrieved, which is used in the research.

Determinant of TAM: Based on the examined literature (see table 7 of section 3.1.2), the identified main-criterion 'information quality' along with its significant sub-criteria are mostly used as external predictors of the determinant PU. Therefore, the decision is made to also use the criterion 'quality of provided information' as external predictor (i.e. main-criterion) of the determinant PU.

Aggregated sub-criteria 'reliability': based on the literature study, the identified characteristics or sub-criteria of 'quality of provided information' are: 1) completeness, 2) truthfulness, 3) rightness 4) reliability and 5) accuracy (see table 8). The assumption is made that if information is complete, accurate, truthful and right, it is reliable. Therefore, reliability of information is used as characteristic of 'quality of provided information'. Since this research aims to determine which technological alternative is best accepted by customers in informing them upfront such that unnecessary apparel returns can be prevented, the reliability of information needs to concern the required apparel attribute information to evaluate apparel items during the online screening process. Consequently, the reliability of information refers to the reliability of: 1) material information, 2) colour information, 3) fit & size information and 4) style information. As a result, the following sub-criteria were composed as characteristics of the main-criterion 'quality of provided information': 1) reliability of material information, 2) reliability of colour information, 3) reliability of fit & size information and 4) reliability of style information. In underlying table 9 this is presented.

Table 9: First subset of criteria

Main-criteria	Sub-criteria	Description
Quality of provided information	Reliability of material information	Refers to the accuracy, completeness and truthfulness of the **information provided on apparel material** such as: material thickness, stretch-ability, texture and stitching (sewing).
	Reliability of colour information	Refers to the accuracy, completeness and truthfulness of the **information provided about the colour** of apparel.
	Reliability of fit & size information	Refers to the accuracy, completeness and truthfulness of the **information provided about the fit & size** of apparel .
	Reliability of style information	Refers to the accuracy, completeness and truthfulness of the **information provided about the style** of apparel.

Main-criteria 2: Information gathering and handling

Importance of criterion: In order to produce the recommended fit & size information, such that customers can evaluate apparel items more accurately and refrain from returning it, technologies such as VDRs (A4), Avatars (A3) and size recommendation applications (A2) require customers' body-measurements input data.

According to the examined literature and based upon the opinion of two experts, customers' might not be willing to use the technologies as a result of how the technology collects body-measurement data (e.g. through scanning or manually typing in the information). From these two experts, one was the same quality assurance inspector (with 5.5. year experience) approached for SQ1 and the second expert had a background in Business Engineering and ICT systems and Management and currently functions as a professor at Tilburg School of Economics and Management.

Furthermore, according to the examined literature and the aforementioned experts, the way data is handled by online apparel retailers will explain the level of trust customers assign to the use of each technology and thus will further explain the preference to use a technology. Based on this, there can be concluded that personal body-measurement information might be perceived as privacy sensitive information and that the way of data gathering and handling might negatively affect the customers' acceptance of information technologies.

Therefore, it was important that the composed alternatives were compared based on characteristics of trust regarding information gathering and handling as well. As a result, the main-criteria 'information gathering and handling' was constructed.

Trust as additional determinant to TAM: According to Davis (1986), the actual use of a technology is determined by the attitude towards using a technology, which in turn is a function of two main determinants which are the perceived usefulness (PU) and perceived ease of use (PEU). However, as already indicated in paragraph 3.1.1, to increase the inherent validity of TAM, the option to modify the original model was considered for this research. Based on the literature study conducted about the TAM, this was also necessary.

Studies which used the main-criterion 'perceived trust' as external predictor of the determinant PU in the TAM are John (2012) and Zhou & Zhang (2009) (see table 7). However the majority of the studies (Landeweerd et al. (2013); Shabana & Arif (2011); Chiu et al. (2009) and Salam et al. (2005)), used 'perceived trust' as an additional determinant to the TAM (see table 7) which directly influences the intention to use a technology or system. Given the fact that the majority of the studies used 'trust' as an additional determinant to the TAM, in order to evaluate the technology acceptance, the decision was made to also use 'trust' as an additional determinant in this research.

As a result, the TAM used in this research consists of three determinants which are: 1) PU, 2) PEU and 3) Trust.

Based on the literature study regarding TAM, the sub-criteria 'privacy' and 'security' were identified as characteristics of the main category 'characteristics of trust regarding information gathering and handling' (see table 8). As indicated by Chiu et al. (2009), trust is a significant predictor of customer satisfaction. Furthermore, trust significantly affects the attitude towards using a technology (Salam et al., 2005). John (2012) indicates that perceived privacy has the highest significant impact on trust According to Landeweerd et al. (2013), trust refers to the privacy regarding the disclosure of privacy-sensitive information, as most people are very reticent to share personal information. However, the users' willingness to share privacy-sensitive information with online apparel retailers is an important aspect for the successful adoption of e-commerce (Landeweerd et al., 2013). According to Shabana & Arif (2011), the perceived technical trustworthiness also depends on variables such as data security and privacy. According to the authors, trust highly affects customers' online activities and is inevitable for e-commerce success. Shabana & Arif (2011) indicate that most consumers experience a lack of trust regarding the use of their data afterward by web shops, such as sharing and commodifying the data to third parties without users' consent.

Based on the external predictors (i.e. privacy and security) of the determinant 'trust' there can be concluded that data privacy and security concerns negatively affect the adoption of information technologies. However, in this research the different approaches (e.g. uploading pictures (2D approach), various scanning approaches (3D approach) etc.) which can be used to acquire body-measurement data are not treated in detail, as these are decisions which online apparel retailers can make later on, when the general customers' preference regarding the technological alternatives are obtained. Consequently, in this research, the different approaches to acquire body-measurement data are not compared to one another. As of this, the privacy and security concerns regarding each separate approach are not examined in this research.

However, in order to still acquire general insight regarding the value customers assign to the aspects privacy and security and how these aspects influence their technology preference, based on expert input the decision was made to cover these aspects in a more general manner. In order to know how this influences the customers preference, the generally formulated criteria 'the way of data collection through technology' and 'data handling by online apparel retailer' was included in the research.

In sum, within this research, the original TAM is extended by adding 'trust' as additional determinant. As of this, the TAM consists of three determinants which are: 1) PU, 2) PEU and 3) Trust. The main-criterion 'information gathering & handling' was used as external predictor for the additional determinant 'trust'. Furthermore, the following two sub-criteria were composed as characterises of the main-criterion 'information gathering & handling': 1) the way of data collection through technology and 2) data handling by online apparel retailer. In underlying table 10 this is presented.

Table 10: Second subset of criteria

Main-criteria	Sub-criteria	Description
Information gathering & handling	The way of data collection through technology	The way in which the technology acquires your information (for example through scanning, facial recognition or manually inserting body-measurement information) and how this affects your preference for using a technology.
	Data handling by online apparel retailer	The way the online apparel retailer uses and stores the collected information for its services.

Main criteria 3: User-friendliness

Determinant of TAM: Based on the examined literature, criteria assigned to the category 'Ease of use technological characteristics' are mostly used as external predictors for the determinant PEU (see table 7 of section 3.1.2). Therefore, the decision is made to also use these criteria as external predictors for PEU in this research. The PEU is perceived as a relevant determinant for technology adoption which refers to the "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis, 1986, p. 26). According to Davis (1986), when a technology is not easy to use, there is no positive attitude toward using the technology. Therefore it is important to also include technological characteristics through which the perceived ease of use of each technological alternative can be measured in the comparison analysis.

Through the literature review regarding TAM, various 'ease of use technological characteristics' or sub-criteria were identified (see table 8). After defining the various 'ease of use technological characteristics' via reviewing the literature, there was concluded that the criterion 'user-friendliness' captures the category 'ease of use technological characteristics' better. Therefore, the decision is made to use 'user-friendliness' as main-criterion instead of 'ease of use technological characteristics'.

Aggregation of sub-criteria: Based on the literature review regarding TAM, all sub-criteria were deemed important and were therefore used for the comparison analysis in this research. However, aside from the sub-criteria 'responsiveness' and 'availability', an aggregation of the other criteria was made, based on their synergy. The following aggregated sub-criteria were composed: 1) search time and 2) attractiveness. Based on the literature review regarding the technologies and instruments (see section 2.4), the sub-criteria 'required preparatory work time' was established.

Sub-criterion 'responsiveness': according to Cao et al. (2005), responsiveness refers to the loading time of the technology, and a high loading time negatively affects the ease of use of a technology. This criterion was relevant to include as through the literature review regarding the information technologies there was also established that these technologies differ in the required computational power to be able to display the needed apparel related attribute information. According to Hidellaarachchi et al. (2019), Li et al. (2017) and Apeagyei (2010), virtual try-on technologies such as VDRs and Avatars have a high level of interactivity and computing Avatars requires high computational power. Therefore, the assumption is made that a required high computational power might affect the loading time of virtual try-on technologies negatively. Based on these reasons, it was deemed relevant to include the sub-criterion 'responsiveness' as well to compare the alternatives.

Sub-criterion 'availability': Fedorko et al. (2018) indicated that the availability of the web shop on the Internet, whether on a computer or a mobile phone has a significant positive influence on PEU. Therefore the inclusion of the ability to use the different alternatives on different devices was also deemed necessary. As a result, the decision was made to also use the sub-criterion 'availability' to identify how much value customers assign to the ability of using the alternatives on multiple devices.

Aggregated sub-criterion 'search time': through the literature review regarding TAM, amongst other sub-criteria the following sub-criteria were identified (see table 8): 1) speed, 2) search facility and 3) navigation. As these sub-criteria all have in common that they refer to the ease of use with regard to the time spend searching and navigating in order to find needed information, these criteria were aggregated into the sub-criteria 'search time'. As information technologies especially those whom have a high interactivity function might vary on the level of control Fiore & Jin (2003), it is interesting to know how much value customers assign to the time spend on navigating, searching and controlling the technologies in order to evaluate apparel items and acquire the needed apparel attribute information. Based on this insight, retailers can decide how they can present information as clearly as possible in order to increase the perceived ease of use in terms of, navigation and control.

Aggregated sub-criterion 'attractiveness': through the literature review regarding TAM, amongst other sub-criteria the following sub-criteria were identified (see table 8): 1) usability, 2) attractiveness, 3) perceived enjoyment and 4) playfulness. All of these sub-criteria have in common that they refer to the ease of using the technologies and to the aesthetic aspects of the technologies which determines the perceived attractiveness of users when using the technologies.

Cao et al. (2005) and Fedorko et al. (2018), examined the influence of aspects such as attractiveness on the effectiveness of e-commerce web site quality and found that attractiveness which refers to the playfulness i.e. how appealing and exciting the website is, are important aspects to determine the effectiveness of e-commerce websites. Tella (2012) found that perceived enjoyment is a significant criterion to determine users satisfaction. According to the reviewed literature regarding information technologies (see section 2.4), the perceived enjoyment of using technologies can enhance the online shopping experience (Brooks & Brooks, 2014). Therefore, the decision was made to also include the aggregated sub-criterion 'attractiveness' in this research, to establish the customers preference regarding the technological alternatives.

According to the reviewed literature regarding information technologies (see section 2.4), the interactivity function of virtual try-on technologies such as VDRs and Avatars increase the perceived enjoyment of online shopping (Apeagyei (2010) & Brooks & Brooks (2014)). According to Fiore & Jin (2003) interactivity functions may differ on the enjoyment level, and high levels of enjoyment may compensate for less realism or vividness. As the composed alternatives from left to right al increase in the level of interactivity, it is interesting to see what value customer assign to the sub-criteria 'attractiveness' and what trade-offs are made between different sub-criteria such as attractiveness and search time and required user intervention.

Sub-criteria 'required preparatory work time (user intervention): Based on the literature study regarding the technologies and instruments, there was established that information technologies, especially those with high interactivity functions, may vary in the level of required user intervention. Hidellaarachchi et al. (2019) indicate that more user intervention is needed when 3D approaches (i.e. scanning technology) are used to create virtual try-on experiences compared to 2D approaches such as uploading photos to acquire body-size measurements data in order to provide e.g. fit & size information recommendations, through which customers can accurately evaluate apparel items online.

However, aside from privacy and comfort related issues, 3D approaches which require more user intervention, are more time consuming. However, in order to be able to use the technologies designed to accurately evaluate the apparel items online with respect to e.g. fit & size, user intervention is inevitable. This implies that customers have to conduct preparatory work in advance, in order to be able to use the technologies designed to recommend e.g. accurate fit & size information to customers during the online evaluation process of apparel. Therefore, within this research, 'user intervention' refers to the customers' 'required preparatory work time'. Based on the opinion of two experts, is was necessary to also include this sub-criterion, since it might inhibit the acceptance of such technologies as customers might perceive the required preparatory work as highly inconvenient. From these two experts, one was the same quality assurance inspector (with 5.5. year experience) approached for SQ1 and the second expert had a background in Business Engineering and ICT systems and Management and currently functions as a professor at Tilburg School of Economics and Management.

In sum, the main-criterion 'user-friendliness' was used as external predictor for the determinant PEU. Furthermore, the following five sub-criteria were used as characteristics of the main-criterion 'user-friendliness': 1) responsiveness, 2) search time, 3) availability, 4) attractiveness and 5) required preparatory work time. In underlying table 11 this is presented.

Table 11: Third subset of criteria

Main-criteria	Sub-criteria	Description
User-friendliness	Responsiveness	Loading time of technology.
	Search time	Navigation time (number of clicks) when using a technology during the online evaluation process. The evaluation process is when you have selected an item of clothing and are trying to determine your personal match / fit with regard to the apparel attributes such as material, colour, size and style.
	Availability	The ability to use the technology on any device.
	Attractiveness	The extent to which the technology is visually appealing and playful when used.
	Required preparatory work time	The amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit (such as measuring your own body, typing in or uploading body-measurements data or pictures or scanning the body).

In underlying figure 3, an overall overview of the main-criteria and sub-criteria is presented which are used to evaluate the customers' preference regarding the technological alternatives.

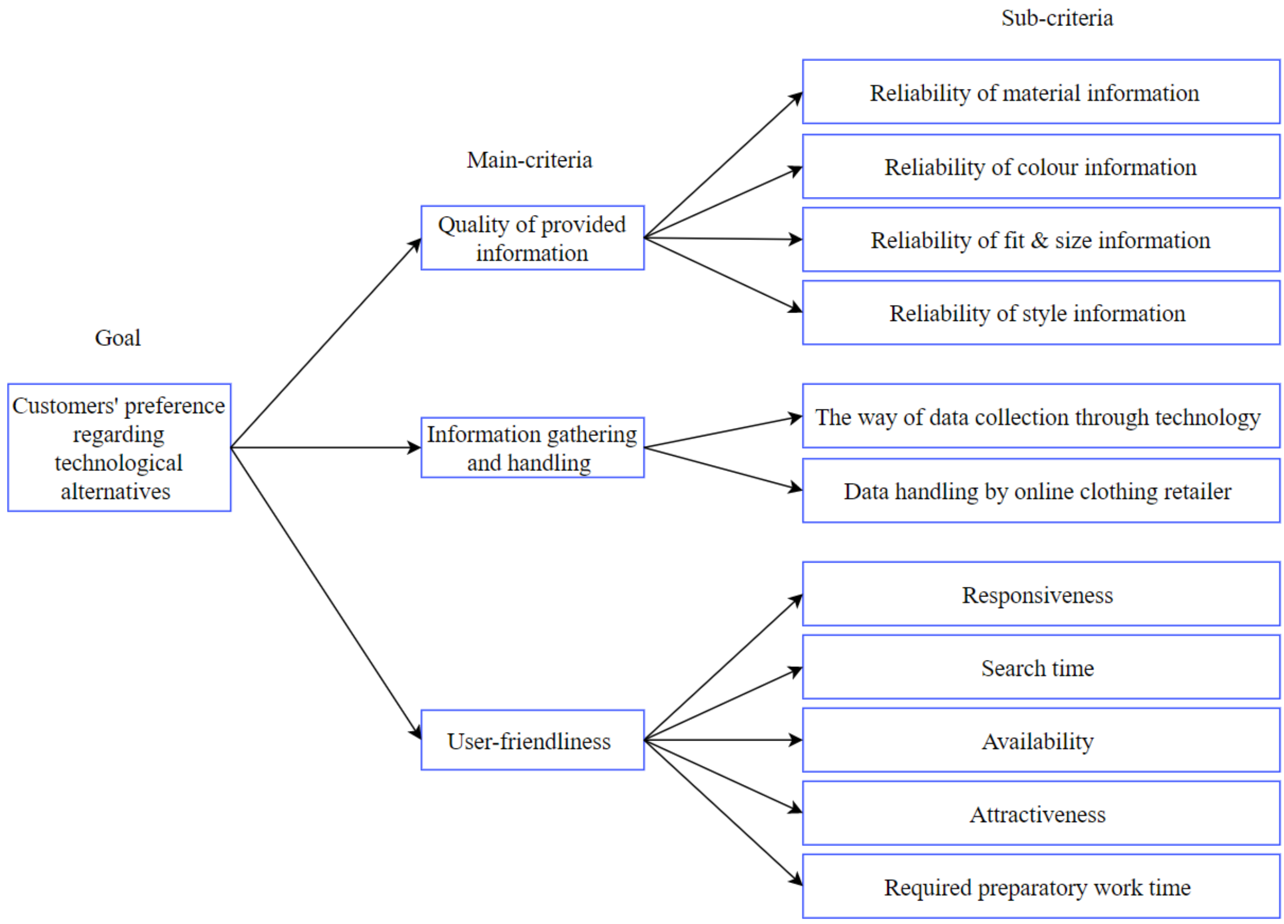

Figure 3: Hierarchy of criteria to evaluate the technologies

4. OBTAINING CRITERIA WEIGHTS

This chapter is devoted to addressing the third sub-question, namely: SQ3. "*What is the relevance that customers assign to these identified criteria?*" In this chapter, the main-criteria and sub-criteria that were identified through SQ2 are weighted. As a results, in section 4.1 the goal of the survey, the target group and the data collection procedure are described. In section 4.2, the applied data analysis method is described, followed by the obtained optimal group weights based upon the users' survey data in section 4.3.

4.1 Survey as data collection tool

In order to acquire the criteria-weights of customers, an online survey was used. In appendix C, an overview of the survey is provided. The survey was constructed using the imposed structure of the BWM. In the previous chapter, three main-criteria which are further subcategorized into sub-criteria were identified (see figure 3). Therefore, four BWM comparison analysis were required (Rezaei, 2015). The goal of the survey was to obtain the optimal group weights of the criteria.

4.1.1 Target group

As this research is conducted in the field of apparel e-commerce, only respondents were included which have online shopping experience. Only respondents which have purchased apparel items online in the last six months were included.

4.1.2 Data collection procedure

Before finalizing the survey, the survey was tested amongst a panel containing of 15 respondents of different age groups, gender and education level.

Based on the test results, no drastic changes were required. The results showed that some criteria might be formulated differently, such that it would be more easier to read and comprehend. As the survey is meant for every individual who purchases apparel items online, the survey needed to be understandable for everyone, no matter the age group, education level etc. Furthermore, a more understandable survey is faster to complete. Based on the acquired feedback, some criteria were reformulated using more generally known words. E.g. since the word 'clothing' is more generally known than the word 'apparel', in the survey the word 'clothing' is used instead.

The survey took about 10 minutes to complete. The final survey was deployed over a period of 17 days. Respondents were reached via an online survey that was shared through the survey link with as many individuals as possible. Respondents were recruited by making use of the researchers' social network and direct contacts. In total 216 complete responses were obtained.

4.1.3 Survey design

As indicated in section 1.3 (SQ3), the survey was constructed using the imposed structure of the BWM. The survey was structured as follows. First the survey protocol was described, wherein 1) the purpose of the survey was explained and 2) was indicated how the collected data will be processed. Furthermore, the number of parts (5 parts), questions (18) and time approximation (10 minutes) for filling in the survey was indicated.

Secondly, the case study was introduced and the goal of the survey was indicated. By doing so, respondents were informed about the research topic, researched problem, and the purpose of conducting the survey. Furthermore, a brief explanation of the four alternatives was provided, such that respondents knew to what alternatives the criteria were referring, for which comparison analysis of alternatives the criteria will be used.

Thirdly, the respondents were brought to the third page, were de data collection part actually started. In this first part of the data collection process called ‘Part 1. Background information’ background information of respondents was acquired. Socio-demographical information such as gender, age and education level was acquired to obtain examine the representativity of the obtained sample.

Fourthly, the respondents were brought to the fourth page, where the data collection process through the BWM started. As already indicated, the survey was constructed using the imposed structure of the BWM. This being said, following the steps proposed by (Rezaei, 2015), the respondents first had to choose the best (most important) criterion and the worst (least important) criterion. Then the respondents had to compare the best criterion to the others and the others to the worst criterion, using a scale from 1 to 9 where 1 represents equal importance and 9 absolute more importance. This was repeated for four sets of criteria. The survey was constructed as such that when conducting the Others-to-Worst pairwise comparison, the best criterion was not displayed again as in the Best-to-Others pairwise comparison the best criterion is already compared to the worst criterion.

During the testing phase of the survey, it became clear that some respondents had difficulties with conducting the comparison analysis in the right manner (comparing the Best-to-Others and Others-to-Worst). Therefore, before proceeding with the second data collection part, first instructions were provided regarding how to conduct the pairwise comparison analysis with the BWM. As a result, in the second part of the survey called ‘Part 2. Main-criteria’, the BWM was applied for the main-criteria, by which the weights of the main-criteria were acquired.

In every part where respondents were asked to conduct the pairwise comparison, a table was provided wherein the criteria were indicated and described. The aim was to achieve valid results and reduce inconsistencies as much as possible, by providing thorough understanding regarding the content of the criteria upfront.

Fifthly, in the third part of the data collection process called ‘Part 3. Quality of provided information’, respondents were asked to conduct the BWM comparison analysis for the first sub-set of sub-criteria. This was also the case for the fourth part called ‘part 4. Information gathering & handling’ and fifth part of the online survey called ‘Part 5. User-friendliness’ of the survey.

At the end of the online survey, each respondent was thanked for filling in the survey and the link of the survey was provided, such that the survey could be shared through others as well.

4.2 Data analysis method

To obtain the weights of each criterion based on the preference of decision-makers, the BWM was applied, as it requires the least pairwise comparisons between the criteria and produces more reliable results compared to other preference elicitation methods such as AHP (Rezaei 2015). Furthermore, it is easier to combine with other MCDM methods compared to for example AHP (Rezaei 2015).

Since the goal of this research is to examine what the customers’ preference is regarding the technological alternatives, this research involves a group performance evaluation of the effectiveness of the alternatives with regard to the identified criteria from the perspective of customers.

However, the original BWM approach developed by Rezaei (2015) cannot amalgamate the preference of the multiple decision-makers. Therefore, an average operator such as the geometric mean is mostly used instead, to combine the preferences when multiple decisions-makers are involved (Mohammadi & Rezaei, 2019). Since averages are prone to outliers and give limited information about the overall preferences of the decision-makers, Mohammadi & Rezaei (2019) developed the Bayesian BWM through which the optimal aggregated (group) criterion-weight (w^{agg}) can be calculated at once, using probabilistic modelling and interpretation of the inputs and outputs.

As a result, in this research the collected data through the survey was analysed using the Bayesian BWM. By applying this novel group decision-making method, the optimal aggregated group criterion-weights (w^{agg}) were calculated at once. Furthermore, the Bayesian BWM provides a credal ranking of the criteria and visualizes it in a weight directed graph. Each node represents a criterion (e.g. A and B) and each edge indicates the obtained confidence, which indicates that e.g. a criterion A is more important than B, with the confidence *d* (Mohammadi & Rezaei, 2019). According to Mohammadi & Rezaei (2019), "the confidence level represents the extent to which one can be certain about the superiority of a criterion over one another" (p.2).

4.3 Results

In section 4.3.1, the representativity of the acquired sample size is discussed. Then, in section 4.3.2, the obtained weights are presented followed by an interpretation of the results in section 4.3.3.

4.3.1 Survey Respondents

In total, 216 respondents filled in the survey completely. Before calculating the optimal group weights, the consistency of the respondents was also checked and the ones with an acceptable consistency ratio were considered (Liang et al., 2020). As a result, per set of criteria, a different sample size was used. A sample size of 113 respondents (n=113) was used for the main-criteria set, a sample size of 77 respondents was used for sub-criteria-set 1 (n=77) and a sample size of 73 was used for sub-criteria-set 3 (n=73). For sub-criteria-set 2, a sample size of 113 instead of 216 was used (n=113), to obtain more reliable results.

Since there are only two criteria in sub-criteria-set 2, the mistake of not assigning the highest value to the best-worst vector cannot occur. As a result, all 216 respondents are only included in this subset, since technically this mistake cannot occur if there are only two criteria (best and worst). If subset 2 would for example consist of more than two criteria, the Likelihood of conducting the pairwise comparison wrongfully could also be present, resulting in the elimination of respondents in this set (as is the case for the main set, subset 1 and subset 3). Since they are only included based on a technicality, implies that the final result will be less reliable. Therefore, the decision was made to consider the second highest sample size (closest to n=216) to determine the criteria weights for subset 2, containing of respondents which have correctly operationalized the pairwise comparison. As a result, the respondents of the main set were used (n=113).

The target audience is a subset of the Dutch population, more specifically online apparel shoppers. However statistics showing the socio-demographic factors of simply apparel purchases were not obtainable, as such the Dutch population is used for comparison instead. Comparisons are made concerning the gender, age and education level of the survey respondents. The statistics of the Dutch population are retrieved from CBS (2019a, 2019b).

In table 12, the socio-demographic characteristics of the survey respondents are presented.

Table 12: Socio-demographic characteristics

	Dutch Population 2019 (N=17282163)		Samples							
			Main-criteria set (n=113)		Sub-criteria set 1 (n= 77)		Sub-criteria set 2 (n=113)		Sub-criteria set 3 (n=73)	
Variables	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Gender										
Male	8581086	50%	51	45%	38	49%	51	45%	36	49%
Female	8701077	50%	62	55%	39	51%	62	55%	37	51%
Age										
<18	3791838	22%	6	5%	3	4%	6	5%	2	3%
18-25	1079925	6%	10	9%	4	5%	10	9%	9	12%
25-34	2198769	13%	24	21%	22	29%	24	21%	19	26%
35-44	2056681	12%	21	19%	13	17%	21	19%	11	15%
45-54	2512575	15%	27	24%	17	22%	27	24%	16	22%
55-64	2328371	13%	10	9%	7	9%	10	9%	5	7%
>64	3314004	19%	15	13%	11	14%	15	13%	11	15%
Highest level of education										
None/basic education	1407000	10%	3	3%	1	1%	3	3%	1	1%
LBO/VMBO;VMBO/MAVO	2950000	20%	8	7%	6	8%	8	7%	2	2%
HAVO;VWO	1353000	9%	13	11%	8	11%	13	11%	7	10%
MBO	3999000	28%	28	25%	23	30%	28	25%	21	29%
HBO;WO	4478000	31%	61	54%	39	50%	61	54%	42	58%
Unknown	215	1%	-	-	-	-	-	-	-	-

The result show that the characteristics (age, gender and education level) of the group of survey respondents differ from when a random sample had been taken of the Dutch population. Looking at the variable gender, there can be concluded that the sample of mostly subset 1 and 3 represent the Dutch population very well. In the sample used for the main set and thereby also subset 2, there is a slight overrepresentation of females (+ 5%) compared to the entire Dutch population, and a slight underrepresentation of male respondents (- 5%) compared to the entire Dutch population. Whilst a rather diverse group of respondents has been observed with respect to age, the group does not entirely represents the Dutch population. In each sample, there is mainly an overrepresentation of the age group 25-34 and 45-54 and an underrepresentation of the age group younger than 18 years old. Looking at the variable education, within each set there is an underrepresentation of the education levels 'none/basic education' and 'LBO/VMBO;VMBO/MAVO' and an overrepresentation of the education levels 'HBO/WO'.

As a result, it cannot be concluded that the samples represent the entire Dutch population with high accuracy, since the samples used for the main set and subset 2 contain biases regarding gender. Furthermore, the samples of each set contains biases regarding age and education. Self-selection might be a possible explanation for this biases. Since the respondent have filled in the survey online, biases could have occurred due to misunderstandings resulting from the perceived complexity of the pairwise comparison method. However, this was addressed as much as possible during the testing phase of the survey (with a panel consisting of respondents with various levels of education) by constructing the questions as clearly as possible and providing a description of how the method should be applied. This perceived complexity of pairwise comparison methods such as BWM might be explained by the education level or willingness to spend time filling in a survey.

4.3.2 Obtained group criteria weights

- Obtained main-criteria weights

In underlying table 13, the optimal group weights of the main-criteria are presented, which were obtained from the online BWM survey.

Table 13: Users' optimal group weights per main-criterion based upon BWM survey.

Main-criteria	Weight
C1. Quality of provided information	0.441
C2. Information gathering and handling	0.235
C3. User-friendliness	0.324

Based on the obtained weights in table 13, 'quality of provided information' is the most important main-criteria (category) for determining the customers' technology preference, with a weight of 0.441 ($w^{agg} = 0.441$). This implies that customers assign more value to obtaining reliable apparel attribute information when purchasing apparel items online, rather than the ease of use of the technology and the perceived privacy and security concerns when using the technology and sharing body-measurements data with online apparel retailers. It is not surprising that 'quality of provided information' is perceived as the most important main-criterion for technology acceptance, since 72 of the 113 respondents (64%) of the respondents selected the criterion as best criterion, whilst only 16 (14%) out of 113 respondents chose it as worst criterion.

In figure 4, the assigned confidence levels are indicated. When looking at the figure, there can be observed that the criterion 'quality of provided information' has a high confidence level of 1 compared to both other two criteria 'data gathering and handling' and 'user-friendliness', implying that the degree of certainty about the criterion is also evident. In other words, we can be very sure about the superiority of C1 over C3 & C2, that 'quality of provided information' is certainly more important than 'user-friendliness' of the technology and 'information gathering and handling'.

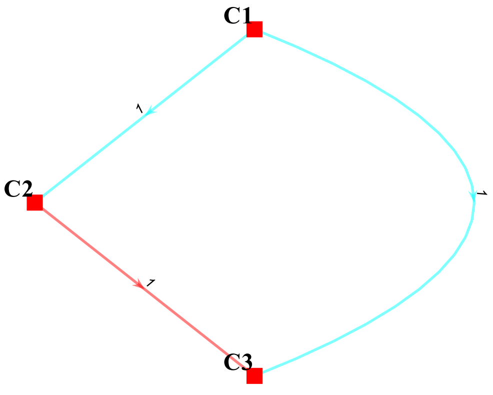

Figure 4: Credal ranking of main-criteria

- Obtained sub-criteria weights

In underlying table 14, the users' (online apparel shoppers') optimal group weights are indicated. In column 1, the main-criteria (categories) are indicated, followed by the sub-criteria in column 2. In column 3, the obtained optimal groups' local weights per sub-criterion are presented. These local weights can be used to only compare the importance of the sub-criteria that belong to the same main-criterion (see ranking in column 4). On the other hand, per sub-criterion a global weight can be obtained by multiplying each local weight of the sub-criterion by the weight of its corresponding category (i.e. main-criteria). These weights are referred to as 'global weights', since these weights can be compared to one another in terms of importance, regardless of the category (main-criteria) they belong (see column 5 and the ranking in column 6).

Table 14: Users' optimal groups' weights per sub-criterion based upon BWM survey

Main-criteria	Sub-criteria	Local Weight per sub-criterion	*Ranking within category*	Global Weight[1] per sub-criterion	*Overall ranking of sub-criteria*
C1. Quality of provided information	c1.1. Reliability of material information	0.242	*3*	0.107	*4*
	c1.2. Reliability of colour information	0.248	*2*	0.110	*3*
	c1.3. Reliability of fit & size information	0.318	*1*	0.140	*1*
	c1.4. Reliability of style information	0.192	*4*	0.084	*6*
C2. Information gathering & handling	c2.1.The way of data collection through technology	0.432	*2*	0.101	*5*
	c2.2. Data handling by online clothing retailer	0.568	*1*	0.133	*2*
C3. User-friendliness	c3.1. Responsiveness	0.207	*2*	0.067	*8*
	c3.2. Search time	0.223	*1*	0.072	*7*
	c3.3. Availability	0.190	*4*	0.062	*10*
	c3.4. Attractiveness	0.179	*5*	0.058	*11*
	c3.5. Required preparatory work time	0.201	*3*	0.065	*9*

Table 14 shows that, 'reliability of fit & size information' is perceived as the most important criterion for the users' technology acceptance (w^{agg} = 0.140). This implies that out of all 11 sub-criterion, users assign the most value to obtaining accurate, complete and truthful fit & size information. According to Mohammadi & Rezaei (2019), the closer the Confidence Level (CL) is to 1, the more evident the degree of certainty about the relation is, suggesting that a criterion is certainly perceived as more important compared to the other. Looking at the category 'quality of provided information' and the confidence level in figure 4, 'reliability of fit & size information' is certainly more important (CL = 1) than the criteria reliability of colour, material and style information. However, this is not surprising and can be explained by the following. In the survey, the respondents were first asked to establish the best and worst sub-criteria. The results are presented in appendix D. The results show that amongst all four criteria, 'reliability of fit & size information' was chosen the most amount of times as best criterion and the least amount of times as worst criterion. The credal ranking in figure 5 also shows that within this category, 'reliability of colour information' is perceived as the second most important apparel attribute, followed by material and style. This implies that users assign the least amount of value to obtaining style information, that their preference of using a technology determines the least on acquiring style information compared to fit & size, colour and material information.

[1] The global weights are obtained by multiplying each local weight by the weight of its corresponding category (i.e. main criteria) . For instance, the global weight of reliability of material information is obtained as follows: Global weight of reliability of material information = local weight of reliability of material information * weight of quality of provided information or: 0.107=0.242*0.441.

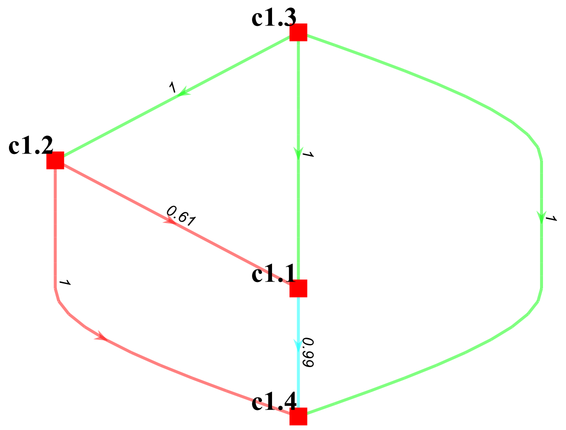

Figure 5: Credal ranking of sub-criteria belonging to 'quality of provided information'

Table 14 also shows that out of all 11 criteria, 'data handling by online clothing retailer' is the second most important criterion affecting users' technology acceptance ($w^{agg} = 0.133$). This implies that after 'fit & size information', user' technology acceptance is mostly influenced by the way the online apparel retailer uses and stores customers' collected information. Looking at the credal ranking and confidence levels indicated in underlying figure 6, 'data handling by online clothing retailer' is certainly perceived as more important than 'the way of data collection through technology' with a CL of 0.99 out of 1. Consequently, users' technology acceptance certainly depends more on how retailers handle their information instead of how the technology acquires data such as body-measurement data. This however is not surprising since the majority of respondents 63 (56%) out of 113 respondents chose data handling by online clothing retailer' as best criterion and 50 (44%) out of 113 respondents chose this criterion as worst (see Appendix D).

Figure 6: Credal ranking of sub-criteria belonging to 'information gathering and handling'

Table 14 also shows that out of all 11 sub-criteria, 'reliability of colour information' is perceived as the third most important criterion affecting the users' technology acceptance ($w^{agg} = 0.110$), subsequent to 'reliable fit & size information' and 'data handling by inline clothing retailer'.

The table also shows that 'reliability of material information' is perceived as the fourth most important criterion affecting the users' technology acceptance ($w^{agg} = 0.107$). This implies that the provision of accurate, complete and truthful material information such as material thickness, stretch-ability, texture and stitching (sewing) is perceived as fourth most important aspect influencing the customers technology preference and as the third most important apparel attribute for evaluating apparel items accurately online (subsequent to fit & size information and colour information).

Table 14 also shows that out of all 11 sub-criteria, 'the way of data collection though technology' is perceived as fifth most important criterion affecting the users' technology acceptance ($w^{agg} = 0.101$). This suggest that the way in which the technology acquires users' information (for example through scanning, facial recognition or manually inserting body-measurement information) also significantly affects the customers' preference for using a technology.

Looking at the sub-criteria that define 'user-friendliness' and the underlying weight directed graph, the criterion 'search time' is certainly perceived as the most important (CL is higher than 0.80) for determining the users' perceived ease of use, compared to the criteria 'responsiveness', 'required preparatory work time', 'availability' and 'attractiveness'. When looking at the figure 7, the results suggests that we can be very sure about the superiority of c3.2 search time over c3.4 attractiveness, but slightly less sure about the superiority of c3.2 search time over c3.1 responsiveness of the technology.

Appendix D, shows the distribution of the best and worst selected criteria. The results show that even though more respondents chose ‘availability’ as best criterion, ‘search time’ is still perceived as the most important criterion affecting the perceived ease of use of the technologies. This can be explained by the following. The comparison input data showed that even when ‘search time’ was not chosen the most times as best criterion, most respondents have assigend a low score (using a 1/9 to 9 scale) in the BO comparison, when the best criterion was compared to ‘search time’. This implies that ‘search time’ is mostly perceived as evenly important, or between evenly or moderately important, or moderately important compared to the selected ‘best’ criterion. As a result, ‘search time’ is assigend the highest optimal group weight within this category.

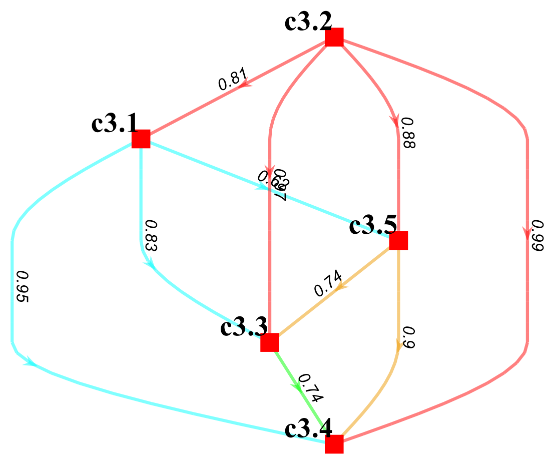

Figure 7: Credal ranking of sub-criteria belonging to ‘user-friendliness’

5. ESTABLISHING CUSTOMERS’ PREFERENCE

In this chapter, the customers’ preference regarding the technological alternatives from SQ1 is calculated, using the identified criteria from SQ2 and the obtained optimal groups’ criteria weights from the users’ survey data from SQ3. By combining the outputs of these three sub-questions, this chapter is devoted to answering the following fourth sub-question: SQ4. “*Based upon these criteria and their obtained weights, how do these technological alternatives compare in terms of preference?*” In this chapter first in section 5.1 the data collection method is described through which the scores of each alternative with respect to each sub-criterion are acquired. In section 5.2, the obtained scores stemming from BMW interviews conducted with six apparel e-commerce experts of four online apparel retail shops in the Netherlands are presented. Ultimately, using the weighted sum equation, the customers’ preference regarding the technological alternatives is calculated and presented in section 5.3

5.1 Data collection tool

In order to acquire the performance scores of each technological alternative with respect to the criteria, expert interviews were held. Since some of the examined technologies and instruments are not yet employed in practice, customers users' experience regarding these technologies and instruments is lacking. Therefore, experts were approached to provide the scores, since they have more in depth knowledge and experience regarding these technologies and instruments. However, since this research focusses on examining the customers' preference regarding the technological alternatives, it was essential to make very explicit that the situation needed to be addressed from the perspective of the customers.

In order to obtain the scores, the BWM was again applied. Usually, the BWM is used to acquire the weights of criteria. However, compared to using a scale from e.g. 1 to 10, applying the BWM results in more valid scores and is therefore used instead. On the contrary, it takes more time to acquire the scores. Within this research, the process of obtaining scores per expert interview with the BWM application, took approximately 60 minutes per expert (based on an average of six expert interviews). The duration of the interview can also be explained by the acquisition of a rational per assigned score.

5.1.1 Target group

Since this research involves technologies and instruments which can be used to better inform customers online with regard to apparel attributes, the search was set on finding experts active in the field of online apparel e-commerce, which have knowledge about the technologies and instruments. As a results, the following six experts from four companies were interviewed.

Table 15: Characteristics of interviewees

Expert	Company (anonymized)	Name (anonymized)	Function	Expertise	Years of experience
1	WAK	AHO	Quality Assurance Inspector	Technical translation from styling/design to the technical application & visualization of clothing on the web shop, lead of the returns management project	5.5 years
2	WAK	EFE	Quality Assurance Inspector	Responsible for the fit & size of apparel & material quality for woman’s department	7 years
3	VOT	ASC	Local marketing manager	Online marketing and retour analysis.	10 years

4	BOK	PKL	Local marketing manager	Omnichannel marketing (physical and digital marketing)	3 year
5	BOK	SHE	Online product specialist	Retour analysis of apparel items, product information optimization.	3 years
6	KLE	KMA	Country online marketing manager	Product recommendations for online apparel items, online marketing campaigns, making the technical translation from styling/design of brands to the technical application & visualization of apparel on the web shop.	2 years

5.1.2 Data collection process

In order to set-up interview meetings with experts, online apparel retailers in the Netherlands were approached. The companies and experts were approached via email and also by phone. In total, six experts from four companies were interviewed. All interviews were conducted via telephone, after which each interview was transcribed.

5.1.3 Interview design

In appendix E, an overview of the interview is presented. Since the BWM was also applied to obtain the scores, an interview structure was imposed by the method. As a result, the interview was constructed using the imposed structure of the BWM. The same definition assigend to the scores (1-9) implied by the BWM was used, with some slight adjustments (see section 1.3 SQ4).

The applied interview structure does not differ that much from the online survey which was used earlier to obtain the users' optimal group criteria weights, as the imposed survey structure also stems from the BWM. However, in the interviews, the alternatives are also explained and discussed in detail with the interviewee.

Since the interview was conducted via phone, the interview document was send upfront via email to the interviewee, such that the interviewee could walk through the provided upfront information and interview questions whilst being interviewed.

The interview was structured as follows. First, the interviewee was informed about the interview protocols. Secondly, the case was introduced and the aim of the survey was described, which was to obtain performance scores of each designed alternatives with regard to the selected criteria. Thirdly, the three subsequent questions which were asked to the interviewee, as a result of the BWM method, were described. These three questions, per sub-set of criteria, were:

1) Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to addressing a particular criterion and which one is the worst?

	Best Alternative	Worst Alternative
Criterion (X)		

2) How much better is the best alternative compared to all other alternatives with regard to this particular criterion, on a scale from 1 to 9?

Best compared to Others		**A1:** Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **static** height/size chart	**A2:** Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing rooms (augmented reality such as snapchat filters)
Best Alt.:					

3) How much better are all other alternatives compared to the worst alternative with regard to this particular criterion, on a scale from 1 to 9?

Others to Worst	Worst Alt.:	
A1		
A2		
A3		
A4		

Fourthly, the four alternatives (A1, A2, A3 and A4) were described. Fifthly, the data collection part regarding the collection of personal background information of the interviewee was described. The purpose of obtaining this information was to get a better indication of the level of expertise of each interviewee. Finally, the four sets of decision criteria were described, along with the scores (1 – 9) which can be assigned. After this, the interview was conducted, using the three subsequent questions as previously indicated.

5.2 Interview results

In appendix E, an overview of the interview is presented. In appendix F, the interview results based upon six apparel e-commerce experts are presented per sub-criterion. In underlying table 4, an overview of the obtained scores (i.e. weights) of each alternative with respect to each sub-criterion is presented.

In the first column, the sub-criteria are presented. In column 2 till 5, the scores of each alternative with respect to the sub-criteria are presented.

The alternatives are:

A1: The bare minimum
A2: The bare minimum with a fit & size recommendation application
A3: Avatar (digital computer-based twin)
A4: Virtual Dressing Room (VDR)

Table 16: Alternatives' scores with respect to the sub-criteria (based on 6 experts' interviews)

Sub-criteria	Scores			
	A1	A2	A3	A4
Reliability of material information	0.140	0.204	0.299	0.358
Reliability of colour information	0.217	0.217	0.228	0.338
Reliability of fit & size information	0.093	0.186	0.362	0.359
Reliability of style information	0.106	0.144	0.359	0.392
The way of data collection through technology	0.424	0.325	0.143	0.108
Data handling by online clothing retailer	0.503	0.278	0.123	0.096
Responsiveness	0.438	0.339	0.101	0.122
Search time	0.274	0.332	0.171	0.223
Availability	0.475	0.307	0.096	0.122
Attractiveness	0.104	0.144	0.316	0.436
Required preparatory work time	0.304	0.361	0.151	0.185

Table 16 shows that based on the experts opinion, A4 is perceived to be the best in providing reliable material, colour and style information. In appendix G, the weight directed graphs of each alternative with respect to all sub-criteria are presented. The nodes in the graph are the alternatives and each edge indicates the preference of the alternatives compared to one another with respect to each specific sub-criterion. The results presented in appendix G show that when it comes to providing reliable material, colour and style information, A4 is certainly perceived as more reliable compared to A1 and A2 (the CLs of A4 compared to A1 and A2 with respect to these three apparel attributes are always higher than 0.89 implying that the relation is evident). However, A4 is perceived as more reliable when it comes to providing reliable style information compared to A3, with a CL of 0.65, suggesting that we can be sure about the superiority of A4 over A3 with respect to providing reliable style information. Furthermore, A4 is also perceived as more reliable when it comes to providing material information compared to A3 (CL = 0.74) and certainly more reliable than A3 (CL=1) in providing colour information.

The main reason why A4 is perceived as the best with respect to these three sub-criteria (material, colour and style) stems from its ability to try-on apparel items on the virtual appearance of the individuals' own body, which gives a better perception and feel of the apparel style and colour. Furthermore, the dynamic movement which can be created gives an better feel and perception of the material quality, which makes it a superior alternative, is very affective to evaluate the personal match of apparel items with online. Furthermore, A4 is perceived as the most attractive to use for customers, since its ability to try-on apparel items on one's own mirrored image along with the dynamic movement where apparel moves with the individuals' body movements, makes it more exciting, playful and visually appealing for customers to use.

The results also show that A3 scores the best with respect to 'reliability of fit & size information', closely followed by A4. Four out of six experts (67%) (expert 1, 3, 5 and 6) chose A4 as the alternative which is the best when it comes to providing reliable fit & size information. However the other two remaining experts chose A3 as the best alternative. The experts who chose A4, indicated that the ability to see how the apparel items fit on the individuals own body shape provided a better perception and feel of fit. According to expert 5, everyone has a different physique, and through the ability to try-on apparel items on their mirrored selves can form a better perception regarding the personal fit & size. Furthermore, per brand the size can be different (apparel can fall larger or smaller), so the visual aspect (the virtual fitting of apparel) has a lot of added value compared to the more static alternatives A1 and A2.

According to expert 2 and 4, A3 an A4 are perceived almost as equally best, but A3 scores better because according to these two experts A3 can still better represent the actual size and body shape more precisely compared to a VDR. The obtained local weights as indicated in table 16 show that A3 is perceived as the most reliable alternative when it comes to providing reliable fit & size information. This can be explained by looking at the assigned values in the comparison analysis (Appendix F). Three out of the four experts whom have chosen A4 as best alternative, have indicated that A4 scores between equally and moderately better, or modality better compared to A3 with respect to the criterion 'reliability of fit & size information'. According to expert 3, A4 scores between evenly and reasonably better than A3, as A3 also offers a virtual try-on experience, but a less realistic try-on experience compared to A4.

One expert (expert 6) even indicated that A4 scores equally good compared to A3. According to this expert, A3 is equally good at providing fit & size information, as one can also make the avatar very realistic on the basis of their own body dimensions and thus receive the correct fitting information. Consequently, all four experts have indicated that the difference between A4 and A3 with respect to providing reliable fit & size information is small. As a result, A3 has the highest assigend score, which compared to A4 is not even that high of a difference.

This can also be seen in the weight directed graph as indicated in appendix G. Based upon the scores and the credal ranking, the alternatives can be ranked as follows with respect to the criterion 'reliability of fit & size information': A3>A4>A2>A1. Based upon the obtained confidence level (see appendix G), A3 is certainly more reliable than A1 and A2 in providing reliable fit & size information (CL =1) However, compared to A4, the degree of certainty is less evident, since the confidence level is only 0.52.

The results have also show that although A4 and A3 score the best with respect to providing reliable apparel attribute information online shoppers need to evaluate apparel items online and refrain them from returning the items, these state-of-the-art technological alternatives score the lowest with respect to the criteria related to privacy and security. Table 16 shows that A1 scores the best with respect to the criteria 'the way of data collection through technology' and 'data handling by online clothing retailer'. This implies that according to the experts, users' are perceived to have the least to no privacy and security concerns whilst using A1, since no data is collected in order to be able to use the instruments to evaluate apparel items with hence no data can be stored and used by online clothing retailers. Based on the weight directed graph as indicated in appendix G, A1 certainly involves the least amount of privacy and security concerns compared to all other alternatives. Based upon the obtained confidence level (see appendix G), A1 is certainly perceived as the alternative with the least privacy and security concerns with respect to 'the way of data collection through technology' compared to A3 and A4 (CL= 1) and A2 (CL= 0.88) suggesting that degree of certainty is evident. Based upon the obtained confidence level (see appendix G) the relation of A1 over the other three alternatives with respect to 'data handling by online clothing retailer' is also evident (CL =1).

Furthermore, based on all six experts, A1 scores the best with respect to responsiveness and availability, since is it is perceived as the least technically complex and requires the least amount of storage capacity as only pictures and a size table need to be uploaded along with the mix and match function. As such, the loading time will be the least negatively affected using A1. Since A1 requires the least amount of computational power, it can easily be made available on any device compared to the other alternatives. Based upon the obtained confidence levels as presented in appendix G, A1 is certainly perceived as the alternative which can be made the most available on any device. In addition, A1 is also certainly perceived as the most responsive alternative compared to all other three alternatives, based on the experts interviews.

Table 16 also shows that A2 scores the best with respect to search time and required preparatory work time. A2 is perceived as the best alternative with respect to search time, since the number of clicks/ effort needed to acquire the necessary apparel attribute information to evaluate apparel items with is considered to be the lowest for A2. Users only have to fill in a format wherein they indicate their body-measurements information, which according to the experts is relatively easy and quick to do. In terms of 'required preparatory work time' A2 scores the best, since most of the experts perceived A2 to require the least amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit. Based upon the obtained confidence levels as presented in Appendix G, A2 certainly requires the least amount of preparatory work time compared to A3 (CL=0.95) and A4 (CL=0.99). Compared to A1, the assigend confidence level of A2 is 0.74 implying that the level of certainty to which A2 is perceived to require less preparatory work time over A1 is 0.74. In addition, A2 requires the least amount of search time with a assigend confidence level of over 0.72 compared to the other alternatives.

5.3 Customers' preference regarding technological alternatives

In the final step of the MCA, the alternatives are ranked based on the criteria weights and the alternatives' assigned scores. Underlying table 17, shows the evaluation of the four technological alternatives with respect to the various sub-criteria, and the final column shows the weight of the sub-criteria. Using the underlying weighted value function, the overall score for each technological alternative can be obtained (final row of table 17) which can be used to rank the customers preference to use an alternative.

$$V_i = \sum_{j=1}^{n} w_j p_{ij}$$

w_j : represents the assigned weight to criterion j
p_{ij} : represents the score of each alternative i with respect to each criterion j
(V_i) : represents the overall value of alternative i and is simply determined by multiplying the score p_{ij} with the respective weight w_j of criterion j ($w_j \geq 0, \sum w_j = 1$) (Rezaei, 2015).

Table 17: Ranking of the technological alternatives

Sub-criteria	Scores of Technological Alternatives				Global weights
	A1	A2	A3	A4	
Reliability of material information	0.140	0.204	0.299	0.358	0.107
Reliability of colour information	0.217	0.217	0.228	0.338	0.110
Reliability of fit & size information	0.093	0.186	0.362	0.359	0.140
Reliability of style information	0.106	0.144	0.359	0.392	0.084
The way of data collection through technology	0.424	0.325	0.143	0.108	0.101
Data handling by online apparel retailer	0.503	0.278	0.123	0.096	0.133
Responsiveness	0.438	0.339	0.101	0.122	0.067
Search time	0.274	0.332	0.171	0.223	0.072
Availability	0.475	0.307	0.096	0.122	0.062
Attractiveness	0.104	0.144	0.316	0.436	0.058
Required preparatory work time	0.304	0.361	0.151	0.185	0.065
Total Score	0.2748	0.2517	0.2221	0.2516	
Ranking	1	2	4	3	

Based on the obtained criteria-weights through the online survey and the scores from online apparel retail experts, it can be observed that A1 has the highest chance of reaching users' technology acceptance. A2 is perceived as the second best, closely followed by A4. A3 is perceived to have the lowest chance of reaching users' technology acceptance.

6. MAPPING THE RESULTS

In this chapter, the results of SQ4 were mapped on the current arrangements of online apparel retailers in the Netherlands. By doing so, the perceived employment possibility of the customers preferred technological alternative in companies was examined from the perspective of online apparel retailers. Consequently, this chapter is dedicated to answering the fifth sub-question: SQ5. "*To what extent does the preferred technological alternative map the current arrangements used by online apparel retailers in the Netherlands and why?*" In this chapter, first the data collection method is described in section 6.1, followed by the results in section 6.2.

6.1 Data collection tool

For time related reasons, the decision was made to combine the data gathering process of SQ4 and SQ5. As a result, the data for this sub-question (SQ5) was also collected through the six experts' interviews from SQ4. At the end of the interview, the experts were first asked the following two questions:

1) What is the return rate (in %) of clothing purchased online by customers with respect to clothing attributes such as material quality, fit & size, colour and style?
2) Which alternative(s) does the company already use?

Based on this information, insight was gained regarding technologies and instruments which are currently employed on the web shop of the company.

At the end of the interview, experts were asked to score all four technological alternatives with respect to the criterion 'employment possibility in company', by using the BWM approach. The three questions which were asked, as a result of the imposed structure of the BWM, are:

3) Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**employment/implementation in the company**" and which one is the worst? Please explain your answer.

	Best Alternative	Worst Alternative
Implementation possibility in company		

4) How much better is the best alternative compared to all other alternatives with regard to "**implementation in the company**", on a scale from 1 to 9? Please explain your answer.

.Best to Others		**A1:** Alternative product pictures with mix and match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix and match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:					

5) How much better are all other alternatives compared to the worst alternative with regard to "**implementation in the company**", on a scale from 1-9? Please explain your answer.

Others to worst	Worst Alt.:	
A1		
A2		

A3	
A4	

Based on this information, the employment possibility of each technological alternative within the companies was evaluated. Through this, insight was gained about the managerial implications regarding each alternative. Consequently, insight was gained about factors which can encourage or inhibit the adoption of each technological alternative, from the point of view of online apparel retailers in the Netherlands.

6.2 Results

Based upon the literature review results as provided in section 2.2, most apparel returns stem from fit & size related issues. The experts interview results have also resulted in the same findings as indicated in literature, since expert 2 has mentioned that the majority (40%) of all apparel returns in the company are indeed a cause of fit & size issues (e.g. the size chart that is not accurate enough so that apparel does not fit) and also 40% of the customers have indicated that the apparel items are not as expected (disconfirmation driven). According to expert 1, the amount of apparel returns stemming from fit & size issues are even higher, approximately 52% whilst for material, colour and style it is 6% for each attribute. According to expert 4, in total 37% of all apparel is returned as a result of fit & size issues (18 % too small and 19% too big), whilst style is 31% (e.g. the style does not look as good on me as expected) and for material and colour the amount of returns is a combined 2% (e.g. other hue, or unclear pictures of apparel items). Based on this, the observation can be made that the identified apparel return reasons from literature used in this research are indeed valid.

Underlying figure 8 provides an overview of the scores (i.e. weights) of each alternative with respect to the criterion ‘implementation possibility in company’. In appendix G, the results obtained from the experts, are described per company.

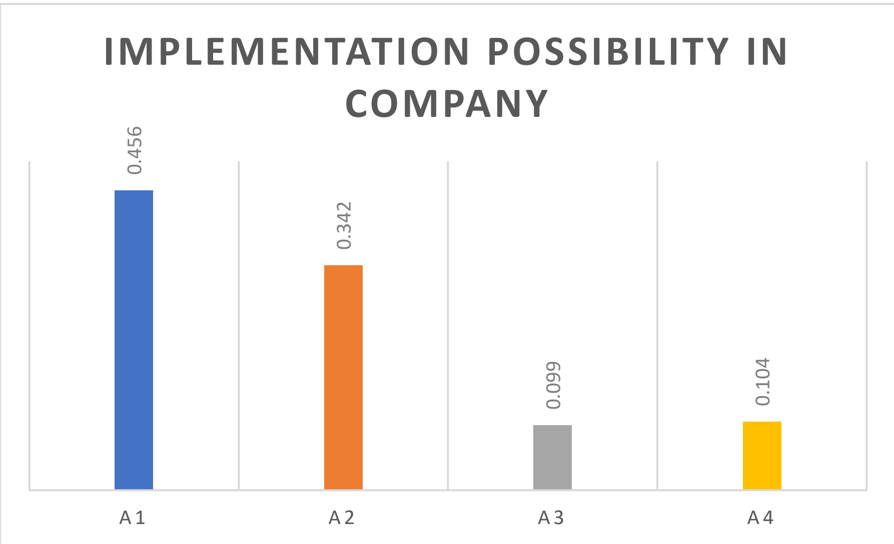

Figure 8: Alternatives' scores with respect to criterion ‘implementation possibility in company’

Based upon the scores, obtained from six apparel retailers, as indicated in figure 8 and the credal ranking presented in underlying figure 9, the alternatives can be ranked as follows with respect to the criterion ‘implementation possibility in company’: A1>A2>A4>A3. The observation can be made that the obtained ranking from solely the experts interview is similar to the customer preference with regard to the technological alternative as presented in table 17 of section 5.3.

Based upon the obtained confidence levels as presented in underlying figure 9, A1 is certainly perceived as the most practically implementable alternative compared to all other three alternatives. Three out of six experts (experts 3, 4 and 6) (50%) chose A1 as most employable in the company, whilst the other three indicated that A2 is the most employable. However, from the experts which have chosen A2 as the most employable alternative, one expert (expert 2) has indicated that A2 scores equally good as A1 and the other two experts (expert 1 and 5) indicated that A2 is between equally and moderately better than A1. As a result, A1 has obtained the highest weight, is perceived as alternative which is currently the most preferred to employ in practise. When it came to choosing the least employable alternative, 50% of the experts (expert 2, 4 and 5) chose A4 and the others experts chose A3. However all three experts which have chosen A4 as worst alternative have indicated that A4 scores somewhat between equally and moderate with respect to the implementation possibility in the company. As a result, A3 is still the least preferred alternative in terms of employment in the company.

Based upon the apparel e-commerce experts interviews, A1 is perceived to have the least amount of managerial implications, since out of all four alternatives A1 is for the most part already employed, aside from the mix and match function to evaluate the entire outfit with. Since all companies already use a lot of product photography, it now comes to the matter of mixing and matching the photos, which according to the experts is perceived as a lot easier than making it very interactive with e.g. A3 or A4.

A2 is perceived as the second best, since it is perceived as the most technically and financially feasible for the companies, after A1. Company WAK recently employed a fit & size recommendation application called a 'fit finder' in order to reduce apparel returns. Since most of the functionalities are already employed in the company to operationalize A2, the mix-and-max function can be easily added to A2 or A1. According to expert 4 of company BOK, A2 is the easiest to implement, if the users' data (e.g. body size measurements) is already stored in the company and the strategic goals are also focused on reducing returns. According to expert 5, the focus within BOK is still insufficiently on the returns department, despite the fact that it is very important. However, currently the capacity in terms of personnel and finance are not available for it yet an the strategic goal is currently more focused on increasing sales and conversion instead of reducing returns.

Looking at the state-of-the-art-technologies, A4 is perceived as the third best alternative with respect to the employment possibility in companies, closely followed by A3 which is perceived to have the most amount of managerial implications. A2 is much better than A3 and A4 when it comes to implementation in the company, as based on the experts results for both A3 and A4 experts have to be hired as the current developers do not have the knowledge to implement A3 and A4. This requires more time and money. Furthermore, testing the technologies and gathering customer opinions also takes much more time, effort and money compared to A1 and A2, according to expert 5.

According to expert 1 and 3, A3 and A4 are both technically and financially equally less feasible for WAK, since experts need to be hired to operationalize A3 and A4 and the whole fashion chain needs to be adapted to the imposed digital way of working, since WAK is a multi-brand online fashion retailer. When a lot of external parties are involved, which is mostly the case with a multi-brand store, this is not deemed possible. Expert 3 indicated that it is more practical to use the state-of-the-art technologies for a company which only trades in its own brands, and that it is even more practical for companies which already work in an advanced manner right from the start with their suppliers, since they already have a lot of required information in the right 3D format (in the right degree of delivery). Such companies already have the necessary software and compared to WAK, they do not have to deal with external parties whom have to provide extra information and data.

According to expert 1, A3 and A4 are more feasible for start-ups, since the digital way of working can then be imposed as the standard way of working. According to expert 1, A4 would be the most difficult to implement, due to the inherent privacy and security concerns which can occur, even more so than with A3. Furthermore, companies have to deal with getting apparel items digitally, which is perceived as very difficult according to expert 1. Since avatars can be made a little more anonymous compared to A4, privacy wise A3 would be less complex to implement into the company. According to expert 3, retailers have to acquire the necessary software for that, which does exist, but they also have to store all the customer data which requires huge storage capacity. On the other hand, the biggest problems which multi-brand stores like WAK face is the requirement of all data such as dimensions and pattern data of all apparel items. The whole chain needs to be aligned on the imposed digital way of working.

Expert 4 also indicated that state of the art technologies such A3 and A4 could have added value for BOK. However, it all depends on how well they work, and how easy they are to implement. According to this expert these kind of applications are often used for home furnishing applications (to find out how e.g. a couch stands in the living room). But for apparel items there is so much more data needed to realize this. The difference between reality and virtual reality is not the same according to the expert. Only if it is real and realistic then it really has added value for BOK according to the expert. For apparel, it is still perceived as very difficult, also considering the number of products and brands BOK has.

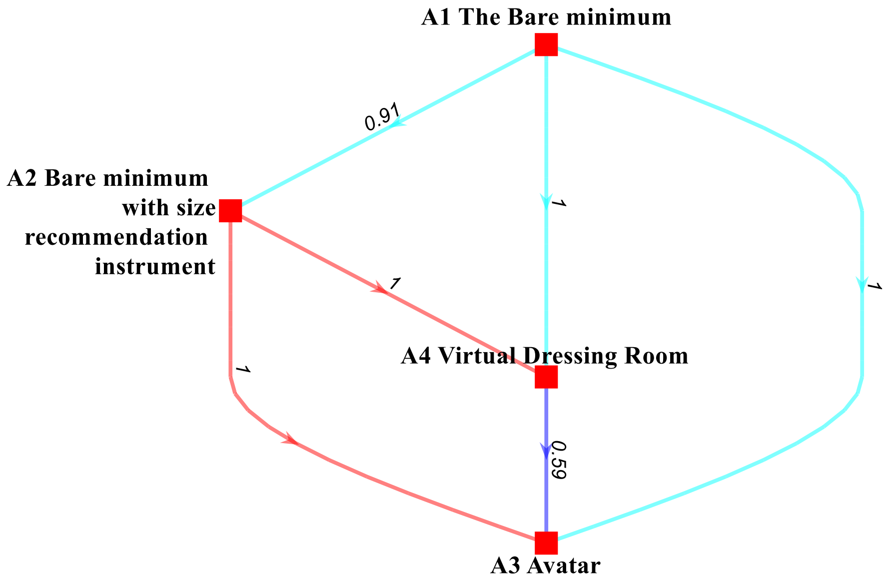

Figure 9: Credal ranking of alternatives regarding 'employment possibility in company'

7. DISCUSSION, CONLUSION & RECOMMENDATIONS

This research paper aimed to explore "*what the customers' preference is regarding technological alternatives online apparel retailers can employ, during the customers' online apparel screening process, in order to increase customers' apparel purchase successes and reduce unnecessary apparel returns*". In order to address this research question, five sub-questions were formulated on the basis of which discussions, conclusions and recommendations were defined at: 1) the practical level and 2) the scientific and methodological level. In the following sections, the discussions, conclusion and recommendations will be addressed per level.

7.1 Practical level

This research aimed to examine what the customers preference is with regard to different technological alternatives that can be used by customers, in order to increase their online purchase successes and reduce unnecessary apparel returns for both users (customers) and online apparel retailers in the Netherlands. Based on the survey results of online apparel shoppers, reliable fit & size information is perceived as the most important criterion affecting users' technology acceptance. Based on this, the observation can be made that the novel Bayesian BWM is indeed a valid method to predict the importance of criteria, since the literature studies and expert interviews have indeed shown that apparel return reasons mostly stem from issues related to fit & size of apparel items.

Based on the obtained criteria-weights through the online survey and the scores from online apparel retail experts, it can be observed that currently A1 is the most preferred alternative, has the highest chance of reaching users' technology acceptance. A2 is perceived as the second best, closely followed by A4. A3 is currently perceived as the least preferred alternative, has the lowest chance of reaching users' technology acceptance. Once the customers' (users') technology preference was determined, the feasibility for online apparel retailers in the Netherlands was also examined by mapping the findings on the current arrangements of online apparel retailers. The findings in chapter 6 show that for the online apparel retailers, efficiency gains and reduction of costs are important, whilst at least being able to offer the same service quality. As a result, A1 currently also has the highest employment possibility in companies, followed by A2, A4 and then A3.

Since the technological alternatives build upon each other in terms of functionality, the level of perceived technological complexity and data required increases as well. As a result of this, the results show that with each subsequent alternative, the reliability of information provision increases, the privacy and security concerns increases and the ease of use decreases. This can also be seen in table 17. For example, whilst A4 is perceived as the most useful (PU) when it comes to providing reliable material, colour and style information and very reliable in providing fit & size information, A4 scores the worst with respect to the privacy and security related criteria (information handling by online clothing retailer and the way of data collection through technology), which are perceived as the second and third most important criteria affecting the users' technology preference (see table 17). This implies that privacy and security concerns do have a great impact on the customers' technology preference. A possible explanation could be that nowadays in the more technologically advanced and data driven society we live in, individuals have become more privacy conscious and assign more value to data privacy and security. Furthermore, A4 also scores very low with respect to the ease of use criteria such as responsiveness, search time, availability and required preparatory work time. This is rather similar for A3, on which A4 builds upon qua functionality (by providing a more realistic virtual try-on experience).

The reason why A3 and A4 are the least preferred, despite that these alternatives score the best with regard to the provision of the necessary apparel attribute information customers need to better evaluate apparel items online, might be due to the fact that they are relatively state-of-the-art and therefore an amount of uncertainty is present regarding the exact amount of knowledge about these technologies (avatars and VDRs). Furthermore, while it was true that online apparel retail experts were asked to approach the scoring of the alternatives with respect to the criteria representing the determinants Trust and PEU from the customers perspective, it might still be possible that the acquired scores are (somewhat) bias. Variables such as experience with functionalities of A1 and A2, low trust in new technologies, time of adoption /low level of innovativeness and conservatism could all be underlying reasons explaining the experts' assigend scores.

Furthermore, since only six experts stemming from four companies could be approached, the individual influence of the assigned scores are higher, which also impacts the end results. However, when looking at the experts data, most experts shared the same arguments and opinions implying that data saturation was reached. However, to find experts with sufficient expertise especially about A3 and A4 (state-of the art technologies) to participate in the BWM interview, was rather difficult. Although the interviewed experts mostly shared the same opinions, there is still advised to continue this research and interview more experts whom are more active in the field of product IT development, to explore the two newer technologies (A3 and A4) better. Since these two technologies scored the best when it comes to addressing the necessary apparel attributes which customers need to evaluate apparel items accurately and as a results are deemed the most effective in addressing apparel returns (PU determinant), it is deemed necessary to explore how the trust related issues (privacy and security concerns) and PEU issues (responsiveness, search time, availability and required preparatory work time issues) can be addressed such that A4 and A3 become more dominant alternatives compared to the A1 and A2 which are less new and more known alternatives.

Since the nature of technology development regarding A4 and A3 is the main reason for this discrepancy, future research could attempt to explain how the ranking will be, which of these two new technologies will have the highest chance of achieving success once they are sufficiently mature. As a result, there is recommended to explore different approaches that can be used for A3 and A4, such as different scanning technology along with their 2D and 3D applications, in further depth before deciding to employ the alternatives in the company. Since this research examined the customers /users' trust with regard to privacy and security concerns in a rather broad manner, more detailed attention should be paid to the different approaches and their impact on privacy and security concerns, such that reliable recommendations regarding e.g. fit & size information can be provided with A2, A3 and A4 whilst preserving customers' data privacy and security.

Based on the results the following is also advised. Existing companies (especially multi-brand stores) could first focus on A1 and A2, especially A2, since 1) A2 scores better with respect to A1 in terms of providing reliable fit & size information (see table 17) and 2) A2 is perceived to not require that more time, money and expertise to implement compared to A1 (see section 6.2). As a result, A2 should be considered first. Moreover, since the survey results have shown that out of the four apparel attribute information, style is perceived as the least important, there is suggested to first focus on the other apparel attributes especially fit & size information since apparel returns are most likely more reduced by addressing these apparel attributes. For new commers in the market, there is advised to focus on the sub-criteria which have the highest weight. This might help them to become successful in the current market.

Although this research indicated that A1 is the most preferred alternative, this cannot be guaranteed with full certainty. It could be that the technologies could co-exist in practice, since in time the current technological superiority of A1 and A2 over A3 and A4 might change. Table 17 shows that A4 scores very similar to A2 (the second best alternative). Furthermore, A3 scores the best with respect to the sub-criterion 'reliability of fit & size information' (the main reason for apparel returns). However, A4 is not far behind. Since the obtained expert results might be (somewhat) bias, the effectiveness of A3 over A4 with respect to fit & size cannot be fully guaranteed. Since A4 and A3 score the best with respect to providing reliable apparel attribute information (material, colour, fit & size and style information), which is required to make better online purchase decisions and refrain customers' from returning items, future research could examine if they can co-exist. In addition, future research could also examine what the customers' technology preference will be amongst different customer segments, by first identifying different clusters (groups) based on characteristics such as age, gender, shopping experience etc., and then computing the different optimal group criteria weights. Through this, more in depth insight might be gained regarding the estimated use of the different technological alternatives and the possibility of co-existence.

Furthermore, since this research is still in the exploratory phase, it was too early to monetize the impact and actual efficiency gains such as apparel returns reduction and cost reduction. Therefore, in order to determine the actual feasibility of each technological alternative and the possibility of co-existence, online apparel retailers are recommended to conduct a cost-benefit-analysis study for each technological alternative. Consequently, since the alternatives (A1 till A4) build upon each other qua functionalities, online apparel retailers are advised to further conduct (partial) CBAs to examine the additional cost and benefits of adding functionalities to A1 (currently the most preferred alternative), under different customer segments, which cannot only lead to a reduction of apparel returns, but can also attract new customers implying an increase in revenue as well.

If the addition of a fit & size recommendation application to A1 (to acquire A2) has deemed feasible, the feasibility of adding an Avatar amongst different customer segments needs to be examined. This can then also be done for A4, by gradually adding a virtual dressing room application to A3. If the additional functionalities deem feasible, the implication could be that multiple options can be provided on the website, such that customers can choose the alternative they prefer the most e.g. feel the most comfortable to use due to inherent privacy and security related issues. By further examining the technology preference amongst different customer segments, more detailed insight can be obtained regarding the estimated use per technological concept per segment.

Since the relatively state-of-the-art technologies (A3 and A4) were perceived to be the most useful in addressing the necessary apparel attribute information and were perceived as most attractive alternatives, it is worth for online apparel retailers to further examine if and how the addition/provision of virtual try-on technologies such as an Avatar or a Virtual Dressing Room can co-exist in practice, taking heterogeneity amongst online apparel shoppers into consideration as well.

In the CBAs, different technological variables stemming from the various measurement approaches which can be used to obtain customers body-measurements along with different applications of recommendations tools need to be included as well. Furthermore, since the privacy and security related criteria (i.e. information handling by online clothing retailer and the way of data collection through technology) were perceived as the second and third most important criteria affecting the users' technology preference, different approaches regarding how these values can be preserved whilst still benefiting from successful purchases and a reduction of apparel returns also need to be included in the CBAs.

If the CBAs produce feasible results (for different customers clusters), a gradual implementation strategy is advised, using a stepwise implementation approach. Consequently, if the addition of a fit & size recommendation application to A1 (to acquire A2) has operated successfully in increasing the customers' online purchase successes and reducing unnecessary apparel returns during a particular period of time, an Avatar can be gradually implemented to again examine the use of A3 amongst different customer segments and its effectiveness in increasing online purchase success and reducing apparel returns. This can then also be done for A4, by gradually adding a virtual dressing room application to A3.

Future steps should also take into consideration that this research was established in the Netherlands, involving shoppers and online apparel retailers in the Netherlands, implying that the obtained weights which have led to this ranking of the customers technology preference might be different based on other contextual variables and empirical setting.

Given the uncertainty factor of online customer reviews (see section 2.4.1), since the provided information is based upon customers opinion, online customer reviews along with customers hotline instruments were not included in this research. Therefore, the way in which these instruments can contribute to the reduction of online apparel returns can also be further explored.

7.2 Scientific and methodological level

Since the aim of this research is to determine what the customers' preference is regarding various technological alternatives, TAM suited the practical goal of this research the most since it solely predicts the technology acceptance based upon two predictors, PU and PEU. Most often, TAM is operationalized using SEM. However, given the exploratory nature of this research and the research goal, it was more ideal to use a more qualitative approach. As a result, the scientific gap this research has aimed to address is the operationalization of TAM, using an MCA approach instead, whereby data is collected through a survey and expert interviews. Moreover, the empirical setting (apparel e-commerce sector), is one where hardly any earlier research could be found wherein TAM was applied. As such, the following added values on the scientific and methodological levels are proposed which also can be used for future research.

Following the MCA approach, one main set and three subsets of decision-criteria (i.e. indicators) were identified and used which play an essential role in the customers' acceptance of the technological alternatives. These criteria sets were identified through an extensive literature study regarding TAM in the apparel e-commerce sector and with the inclusion of experts opinion. Afterwards, the novel Bayesian BWM was applied to operationalize TAM and explore the customers' technology acceptance, using a combination of qualitative methods (expert interviews and an online survey) and quantitative methods (Bayesian BWM and Weighted Sum Equation). The suggestion is made that the higher the obtained weights of a criterion, the more significant influence the criterion has on customers' technology acceptance. Furthermore, by applying the Bayesian BWM, the decision-makers' preference of a criterion could explicitly be confirmed with a certain confidence level.

The results show that the MCA approach and the application of the novel Bayesian BWM indeed produce useful results in a rather exploratory stage, since A1 which currently is mostly employed by online apparel retailers also has the highest chance of reaching technology acceptance (is the most preferred alternative). This proves that the Bayesian BWM is indeed a useful method through which users' technology acceptance can be predicted. To examine the robustness of the applied MCA approach and the novel Bayesian BWM in operationalizing TAM, more research should be conducted whereby the method is used in the exploratory phase of different research problems. Since technology assessment considering the perception of clients/users with the application of TAM, MCDM and the (Bayesian) BWM is very novel but has proven to produce reliable results, there is highly recommended to conduct further research to examine the robustness of the methodology in different empirical settings.

Furthermore, since it is unclear how these compare with conventional TAM studies, which conventionally use SEM, future research could focus om the application of both approaches in the same empirical context. Furthermore, since the criteria were used as external predictors of the determinant PU, PEU and Trust, future research could also focus on applying the criteria as direct determinants of TAM to see how they contribute to the predictive power of TAM. Afterwards, the eligibility of the set of decision-criteria can be established for various technologies in different empirical settings (e.g. in the identified field of studies as indicated in table 7 of section 3.1.2) and also amongst different customer clusters.

The results have also show a difference between obtaining data through an online survey and an interview. Whilst a larger sample size might be obtained through using an online survey instead of conducting interviews, the likelihood of mistakes proved to be higher when using online surveys where respondents e.g. just have to rely on the provided explanation in the survey on how to rightfully conduct the pairwise comparisons. To exemplify, to obtain the customers weights, a survey was used, using the implied structure of the BWM. The testing phase of the survey had shown that an explanation regarding the pairwise comparison was essential, since the pairwise comparison of the BWM was perceived as rather difficult to understand and perform correctly. However, the obtained survey data has still shown that many respondents still faced difficulties in conducting the pairwise comparisons rightfully. As a result, different sample sizes for different levels of the model were used, after excluding the pairwise comparisons with an unacceptable consistency ratio. On the contrary the data from all experts, which was acquired through a BWM interview, was acceptable and used, since the direct communication made it easier to keep the interviewees on the topic and perform the pairwise comparisons in a rightful manner. This shows that the quality of the data does depend on instruments used to collect the data. Since pairwise comparison methods often use surveys to obtain the required input data, different ways of minimizing this impact of the perceived complexity especially when multiple levels of criteria are present, need to be explored.

REFERENCES

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl, J. and Beckmann (Ed.), *Action Control* (pp. 11–39). https://doi.org/10.1007/978-3-642-69746-3_2

Algharabat, R. S., & Shatnawi, T. (2014). The effect of 3D product quality (3D-Q) on perceived risk and purchase intentions: The case of apparel online retailers. *International Journal of Electronic Business*, *11*(3), 256–273. https://doi.org/10.1504/IJEB.2014.063031

Anjum, M., Kapur, P. K., Agarwal, V., & Khatri, S. K. (2020). Assessment of Software Vulnerabilities using Best-Worst Method and Two-Way Analysis. *International Journal of Mathematical, Engineering and Management Sciences*, *5*(2), 328–342. https://doi.org/10.33889/IJMEMS.2020.5.2.027

Annema, J. A., Mouter, N., & Rezaei, J. (2015). Cost-benefit Analysis (CBA), or Multi-criteria Decision-making (MCDM) or Both: Politicians' Perspective in Transport Policy Appraisal. *Transportation Research Procedia*, *10*, 788–797. https://doi.org/10.1016/j.trpro.2015.09.032

Apeagyei, P. R. (2010). Application of 3D body scanning technology to human measurement for clothing fit. *International Journal of Digital Content Technology and Its Applications*, *4*(7). https://doi.org/10.4156/jdcta.vol4.issue7.6

Brooks, A. L., & Brooks, E. (2014). Towards an inclusive virtual dressing room for wheelchair-bound customers. *2014 International Conference on Collaboration Technologies and Systems (CTS)*, 582–589. https://doi.org/10.1109/CTS.2014.6867629

Brown, R. B. (2006). The Problem With the Research Process. In *Doing Your Dissertation in Business and Management* (pp. 52–62). https://doi.org/10.4135/9781849209069.n5

Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: an empirical examination. *Industrial Management & Data Systems*, *105*(5), 645–661. https://doi.org/10.1108/02635570510600000

CBS. (2019a). Bevolking; hoogstbehaald onderwijsniveau en onderwijsrichting. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82816NED/table?ts=1589121974431

CBS. (2019b). Bevolking; kerncijfers. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37296ned/table?fromstatweb

Changchit, C., Cutshall, R., Lonkani, R., Pholwan, K., & Pongwiritthon, R. (2019). Determinants of Online Shopping Influencing Thai Consumer's Buying Choices. *Journal of Internet Commerce*, *18*(1), 1–23. https://doi.org/10.1080/15332861.2018.1496391

Chiu, C.-M., Lin, H.-Y., Sun, S.-Y., & Hsu, M.-H. (2009). Understanding customers' loyalty intentions towards online shopping: an integration of technology acceptance model and fairness theory. *Behaviour & Information Technology*, *28*(4), 347–360. https://doi.org/10.1080/01449290801892492

Choo, E. U., Schoner, B., & Wedley, W. C. (1999). Interpretation of criteria weights in multicriteria decision making. *Computers & Industrial Engineering*, *37*(3), 527–541. https://doi.org/10.1016/S0360-8352(00)00019-X

Chutter, M. Y. (2009). *Overview of the Technology Acceptance Model: Origins, Developments and Future Directions*. https://doi.org/http://sprouts.aisnet.org/9-37

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: theory and results* (MIT). Retrieved from http://hdl.handle.net/1721.1/15192

De Leeuw, S., Minguela-Rata, B., Sabet, E., Boter, J., & Sigurðardóttir, R. (2016). Trade-offs in managing commercial consumer returns for online apparel retail. *International Journal of Operations & Production Management*, *36*(6), 710–731. https://doi.org/10.1108/IJOPM-01-2015-0010

Difrancesco, R. M., Huchzermeier, A., & Schröder, D. (2018). Optimizing the return window for online fashion retailers with closed-loop refurbishment. *Omega*, *78*, 205–221. https://doi.org/10.1016/j.omega.2017.07.001

Ding, Y., Xu, H., Tan, B. C. Y. (2015). A conflicting relationship between online product page viewing and product returns. *2015 International Conference on Information Systems: Exploring the Information Frontier*. Retrieved from https://www2.scopus.com/record/display.uri?eid=2-s2.0-84964655314&origin=resultslist

Edwards, J. B., McKinnon, A. C., & Cullinane, S. L. (2010). Comparative analysis of the carbon footprints of conventional and online retailing: A "last mile" perspective. *International Journal of Physical Distribution and Logistics Management*, *40*(1–2), 103–123. https://doi.org/10.1108/09600031011018055

Fedorko, I., Bacik, R., & Gavurova, B. (2018). Technology acceptance model in e-commerce segment. *Management & Marketing*, *13*(4), 1242–1256. https://doi.org/10.2478/mmcks-2018-0034

Feitelson, E., & Salomon, I. (2004). *The Political Economy of Transport Innovations*. https://doi.org/10.1007/978-3-540-24827-9_2

Fiore, A. M., & Jin, H. (2003). Influence of image interactivity on approach responses towards an online retailer. *Internet Research*, *13*(1), 38–48. https://doi.org/10.1108/10662240310458369

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Gallino, S., & Moreno, A. (2018). The Value of Fit Information in Online Retail: Evidence from a Randomized Field Experiment. *Manufacturing & Service Operations Management*, *20*(4), 767–787. https://doi.org/10.1287/msom.2017.0686

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, *31*(8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Griffis, S.E., Rao, S., Goldsby, T.J., Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. *Operations Management*, *30*(4), 282–294. https://doi.org/10.1016/j.jom.2012.02.002

Gu, Z. J., & Tayi, G. K. (2015). Consumer mending and online retailer fit-uncertainty mitigating strategies. *Quantitative Marketing and Economics*, *13*(3), 251–282. https://doi.org/10.1007/s11129-015-9161-2

Gupta, H., & Barua, M. K. (2017). Supplier selection among SMEs on the basis of their green innovation ability using BWM and fuzzy TOPSIS. *Journal of Cleaner Production*, *152*, 242–258. https://doi.org/10.1016/j.jclepro.2017.03.125

Hafezalkotob, A., Hami-Dindar, A., Rabie, N., & Hafezalkotob, A. (2018). A decision support system for agricultural machines and equipment selection: A case study on olive harvester machines. *Computers and Electronics in Agriculture*, *148*, 207–216. https://doi.org/10.1016/j.compag.2018.03.012

Hidellaarachchi, D., Gunatilake, H., Perera, S., Sandaruwan, D., & Weerasinghe, M. (2019). Towards a virtual garment fitting model for male upper body for online marketplace. *18th International Conference on Advances in ICT for Emerging Regions, ICTer 2018 - Proceedings*, 322–331. https://doi.org/10.1109/ICTER.8615576

Hong, S.-J., & Tam, K. Y. (2006). Understanding the Adoption of Multipurpose Information Appliances: The Case of Mobile Data Services. *Information Systems Research*, *17*(2), 162–179. https://doi.org/10.1287/isre.1060.0088

Hox, J., & Bechger, T. (1999). An Introduction to Structural Equation Modeling. *Family Science Review*, *11*, 354–373. Retrieved from https://www.researchgate.net/publication/27706391%0D

Jang, E., & Burns, L. D. (2004). Components of apparel retailing Web sites. *Journal of Fashion Marketing and Management*, *8*(4), 375–388. https://doi.org/10.1108/13612020410559975

Jin, H., & Black, C. (2012). Assessing functional and aesthetics clothing needs of young male tennis players. *International Journal of Fashion Design, Technology and Education*, *5*(2), 145–150. https://doi.org/10.1080/17543266.2012.658088

John, S. P. (2012). *How online trust influence B2C e-Commerce adoption? An empirical study among Asian online shoppers*. 3699–3710. Retrieved from http://aisel.aisnet.org/amcis2012/proceedings/AdoptionDiffusionIT/10

Kaswan, M. S., & Rathi, R. (2020). Investigating the enablers associated with implementation of Green Lean Six Sigma in manufacturing sector using Best Worst Method. *Clean Technologies and Environmental Policy*, *22*(4), 865–876. https://doi.org/10.1007/s10098-020-01827-w

Kennisinstituut voor Mobiliteitsbeleid. (2017). *Stedelijke distributie en gedrag*. Retrieved from https://www.kimnet.nl/binaries/kimnet/documenten/rapporten/2017/06/06/gedrag-en-stedelijke-distributie/Stedelijke+distributie+en+gedrag.pdf.

Kim, H., & Damhorst, M. L. (2010). The Relationship of Body-Related Self-Discrepancy to Body Dissatisfaction, Apparel Involvement, Concerns With Fit and Size of Garments, and Purchase Intentions in Online Apparel Shopping. *Clothing and Textiles Research Journal*, *28*(4), 239–254. https://doi.org/10.1177/0887302X10379266

Kristensen, K., Borum, N., Christensen, L. G., Jepsen, H. W., Lam, J., Brooks, A. L., & Brooks, E. P. (2013). Towards a next generation universally accessible "online shopping-for-apparel" system. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8006 LNCS*(PART 3), 418–427. https://doi.org/10.1007/978-3-642-39265-8_47

Landeweerd, M., Spil, T., & Klein, R. (2013). *The Success of Google Search, the Failure of Google Health and the Future of Google Plus*. https://doi.org/10.1007/978-3-642-38862-0_14

Lee, M.-C. (2009). Factors influencing the adoption of internet banking: An integration of TAM and TPB with perceived risk and perceived benefit. *Electronic Commerce Research and Applications*, *8*(3), 130–141. https://doi.org/10.1016/j.elerap.2008.11.006

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management*, *40*(3), 191–204. https://doi.org/10.1016/S0378-7206(01)00143-4

Li, R., Zhou, Y., Zhu, S., & Mok, P. Y. (2017). Intelligent clothing size and fit recommendations based on human model customisation technology. *Computer Science Research Notes*, *2702*(May), 25–32. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072782683&partnerID=40&md5=1e114146c16a0dfd8b35ab16e23cfeee

Liang, F., Brunelli, M., & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, *96*, 102175. https://doi.org/10.1016/j.omega.2019.102175

Lohse, T; Kemper, J; Brettel, M. (2017). How online customer reviews affect sales and return behavior - An empirical analysis in fashion E-commerce. *25th European Conference on Information Systems (ECIS)*, 2635–2644. Retrieved from https://aisel.aisnet.org/ecis2017_rip/16

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Minnema, A, Bijmolt, T.H., Gensler, S. (2017). 3. Oorzaken en gevolgen van het terugsturen van online aankopen. In *Ontwikkelingen in het martktonderzoek: Jaarboek Markt Onderzoek Associatie* (42nd ed.). Retrieved from http://moa04.artoo.nl/clou-moaweb-images/images/bestanden/pdf/Jaarboeken_MOA/MOA_JAARBOEK_2017_HFST3.pdf

Minnema, A., Bijmolt, T. H. A., Gensler, S., & Wiesel, T. (2016). To Keep or Not to Keep: Effects of Online Customer Reviews on Product Returns. *Journal of Retailing*, *92*(3), 253–267. https://doi.org/10.1016/j.jretai.2016.03.001

Misra, R., Wan, M., & McAuley, J. (2018). Decomposing fit semantics for product size recommendation in metric spaces. *RecSys 2018 - 12th ACM Conference on Recommender Systems*, 422–426. https://doi.org/10.1145/3240323.3240398

Mohammadi, M., & Rezaei, J. (2019). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, 102075. https://doi.org/10.1016/j.omega.2019.06.001

Nachtigall, C.; Kroehne, U.; Funke, F.;Steyer, R. (2003). (Why) Should We Use SEM? Pros and Cons of Structural Equation Modeling. *Methods of Psychological Research Online, 8*(2), 1–22.

Nasibov, E., Vahaplar, A., Demir, M., & Okur, B. (2017). A fuzzy logic Approach to predict the best fitted apparel size in online marketing. *Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings*. https://doi.org/10.1109/ICAICT.2016.7991773

NRC. (2019). Retourzendingen zijn een structureel probleem voor webwinkels. Retrieved from https://www.nrc.nl/nieuws/2019/02/27/te-coulant-bij-terugsturen-a3679975

Olson, J. R., & Boyer, K. K. (2003). Factors influencing the utilization of Internet purchasing in small organizations. *Journal of Operations Management*, *21*(2), 225–245. https://doi.org/10.1016/S0272-6963(02)00089-X

Peng, F., Al-Sayegh, M. (2014a). Personalised virtual fitting for fashion. *International Journal of Industrial Engineering and Management*, *5*(4), 233–240. Retrieved from https://www2.scopus.com/record/display.uri?eid=2-s2.0-84921647586&origin=inward&txGid=53a79fc31123e6dcc450bf3b9a54aad8

Peng, F., Al-Sayegh, M. (2014b). Personalised virtual fitting for fashion. *International Journal of Industrial Engineering and Management*, *5*(4), 233–240. Retrieved from https://www2.scopus.com/record/display.uri?eid=2-s2.0-84921647586&origin=inward&txGid=ee1be944a6bbaf4f929048d3976cf150

Pires, Guilherme; Lai, K. (2009). *Testing of a model evaluating e-government portal acceptance and satisfaction*. Retrieved from https://www.researchgate.net/publication/228429021_Testing_of_a_Model_Evaluating_e-Government_Portal_Acceptance_and_Satisfaction

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

Rezaei, J. (2020). A Concentration Ratio for Nonlinear Best Worst Method. *International Journal of Information Technology & Decision Making*, 1–17. https://doi.org/10.1142/S0219622020500170

Rezaei, J., Kothadiya, O., Tavasszy, L., & Kroesen, M. (2018). Quality assessment of airline baggage handling systems using SERVQUAL and BWM. *Tourism Management*, *66*, 85–93. https://doi.org/10.1016/j.tourman.2017.11.009

Rezaei, J., Wang, J., & Tavasszy, L. (2015). Linking supplier development to supplier segmentation using Best Worst Method. *Expert Systems with Applications*, *42*(23), 9152–9164. https://doi.org/10.1016/j.eswa.2015.07.073

Saarijärvi, H., Sutinen, U.-M., & Harris, L. C. (2017). Uncovering consumers' returning behaviour: a study of fashion e-commerce. *The International Review of Retail, Distribution and Consumer Research*, *27*(3), 284–299. https://doi.org/10.1080/09593969.2017.1314863

Saaty, T. L. (2004). Decision making — the Analytic Hierarchy and Network Processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, *13*(1), 1–35. https://doi.org/10.1007/s11518-006-0151-5

Salam, A. F., Iyer, L., Palvia, P., & Singh, R. (2005). Trust in e-commerce. *Communications of the ACM*, *48*(2), 72–77. https://doi.org/10.1145/1042091.1042093

Saunders, M., Lewis, P. & Thornhill, A. (2015). *Research Methods for Business Students 7th Edition*. Pearson Education Limited.

Seewald, A. K., Wernbacher, T., Pfeiffer, A., Denk, N., Platzer, M., Berger, M., & Winter, T. (2019). Towards minimizing e-commerce returns for clothing. *ICAART 2019 - Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, *2*, 801–808. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064806285&partnerID=40&md5=d6a7695f178a65e4eaafef6489cfb4db

Shabana, & Arif, M. (2011). *Trust as a Standard for E-Commerce Infrastructure*. https://doi.org/10.1007/978-3-642-17878-8_47

Shen, J., Shang, X., & Dai, J. (2019). Garment E-commerce Return and Exchange 4PL Innovation. *DEStech Transactions on Social Science, Education and Human Science*, (isehs). https://doi.org/10.12783/dtssehs/isehs2019/31614

Shi, B. (2013). An Empirical Study on Small Enterprises E-Commerce Adoption Decision Based on TAM Model & TOE Framework. *Advanced Materials Research*, *712–715*, 2521–2524. https://doi.org/10.4028/www.scientific.net/AMR.712-715.2521

Shojaei, P., Seyed Haeri, S. A., & Mohammadi, S. (2018). Airports evaluation and ranking model using Taguchi loss function, best-worst method and VIKOR technique. *Journal of Air Transport Management*, *68*, 4–13. https://doi.org/10.1016/j.jairtraman.2017.05.006

Tella, A. (2012). Determinants of E-Payment Systems Success. *International Journal of E-Adoption*, *4*(3), 15–38. https://doi.org/10.4018/jea.2012070102

van de Kaa, G., Rezaei, J., Taebi, B., van de Poel, I., & Kizhakenath, A. (2020). How to Weigh Values in Value Sensitive Design: A Best Worst Method Approach for the Case of Smart Metering. *Science and Engineering Ethics*, *26*(1), 475–494. https://doi.org/10.1007/s11948-019-00105-3

Walsh, G., Möhring, M., Koot, C., Schaarschmidt, M. (2014). Preventive product returns management systems - A review and model. *22nd European Conference on Information Systems*. Retrieved from https://www2.scopus.com/record/display.uri?eid=2-s2.0-84905845167&origin=inward&txGid=22c32ed1368d3f2aa3bbcc85b31c98cc

Walsh, G., & Möhring, M. (2017). Effectiveness of product return-prevention instruments: Empirical evidence. *Electronic Markets*, *27*(4), 341–350. https://doi.org/10.1007/s12525-017-0259-0

Wang, Y, Ramachandran, V., & Sheng, O. (2016). The causal impact of fit valence and fit reference on online product returns. *2016 International Conference on Information Systems, ICIS 2016*. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019398893&partnerID=40&md5=ea608397a24d1e2020d7e2ed9219e128

Wang, Yun, & Liu, C.-Y. J. (2009). Online customization for apparel: The roles of involvement and innovativeness. *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, 2236–2240. https://doi.org/10.1109/IEEM.2009.5373076

Wiese, A., Toporowski, W., & Zielke, S. (2012). Transport-related CO2 effects of online and brick-and-mortar shopping: A comparison and sensitivity analysis of clothing retailing. *Transportation Research Part D: Transport and Environment*, *17*(6), 473–477. https://doi.org/10.1016/j.trd.2012.05.007

Xiangdong Liu, & Ming Lei. (2008). Optimal uniform pricing and return policies for channel strategies in E-commerce age. *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, 1487–1492. https://doi.org/10.1109/SOLI.2008.4686636

Zaichkowsky, J. L. (1985). Measuring the Involvement Construct. *Journal of Consumer Research*, *12*(3), 341. https://doi.org/10.1086/208520

Zhang, Y. (2018). *Enhancing e-Commerce Performance: Product Return and Online Customer Review Perspectives* (Michigan State University). Retrieved from https://search.proquest.com/openview/2a2a29c9b3429e93ab271ff0dc9de55c/1?pq-origsite=gscholar&cbl=18750&diss=y

Zhao, H., Guo, S., & Zhao, H. (2018). Comprehensive benefit evaluation of eco-industrial parks by employing the best-worst method based on circular economy and sustainability. *Environment, Development and Sustainability*, *20*(3), 1229–1253. https://doi.org/10.1007/s10668-017-9936-6

Zhou, T., & Zhang, S. (2009). Examining the Effect of E-commerce Website Quality on User Satisfaction. *2009 Second International Symposium on Electronic Commerce and Security*, 418–421. https://doi.org/10.1109/ISECS.2009.24

APPENDICES

APPENDIX A: Scientific Paper

A novel group multi-criteria decision-making approach for establishing users' technology acceptance in the context of apparel e-commerce

Ruchika Kalpoe
R.A.S.KALPOE@student.tudelft.nl
Delft University of Technology – Faculty Technology, Policy and Management – Transport and Logistics

ABSTRACT

Returns management, especially in apparel e-commerce, has gained increased attention due to the ecological and economic implications it imposes. However, research which explores the relationship between 1) reasons which drive customers' apparel returns and 2) customer-based instruments designed to reduce online apparel returns, has not yet been empirically examined in literature, especially from the point of view of customers. As a result, this research aims to examine the customers' technology acceptance of four technological alternatives designed to prevent unnecessary apparel returns. To determine the customers' technology acceptance, the Technology Acceptance Model (TAM) is used, which is mostly operationalised using Structural Equation Modelling (SEM). However, this research uses a new, less data extensive, simpler and even so reliable qualitative approach to operationalize TAM. As such, a Multi-Criteria Decision-Analysis (MCDA) approach is used, wherein the novel Bayesian Group Best-Worst Method (BWM) is applied to infer the optimal group weights of the indicators (i.e. criteria) that influence customers'(users') technology acceptance (TA). This is done within the context of apparel e-commerce and with the application of qualitative tools such as an online BWM survey and expert interviews. This research contributes to the empirical application of the novel Bayesian BWM, in the specific field of apparel e-commerce and proves that users' technology acceptance can be predicted by applying the aforementioned MCDA approach as well. This paper recommends to apply the approach in other empirical cases using the same or other methods as well to further examine its robustness and predictive validity.

Keywords: Returns Management, Apparel e-commerce, Customer-based information technologies, Multi-Criteria Decision-Making, Bayesian Best-Worst Method, Technology Acceptance Model, User acceptance
Word count: 8370

1. INTRODUCTION

Nowadays, more people are purchasing apparel online instead of in physical shops. As a result of this growing apparel e-commerce business, the number of apparel returns is also increasing (Minnema, Bijmolt, & Gensler, 2017). Approximately 30 percent of online purchased products in the Netherlands are returned (Minnema et al., 2017). Of this, 40 percent are apparel items (Julia B. Edwards, McKinnon, & Cullinane, 2010).

Aside from unsuccessful purchases and a reduced amount of revenues for apparel e-commerce retailers, as indicated by studies conducted by for example the Kennisinstituut voor Mobiliteitsbeleid (2017) and Griffis, Rao, Goldsby, & Niranjan (2012), other negative societal implications are inflicted as well. E.g. more retours results in more van-movements, hence more CO_2 emission, less traffic safety, reduced air quality and overall living environment of cities and more traffic congestion (Kennisinstituut voor Mobiliteitsbeleid, 2017). Due to these ecological and economic implications, returns management has gained increased attention, especially in the online apparel retail domain (Difrancesco, Huchzermeier, & Schröder, 2018; NRC, 2019).

In order to reduce the amount of returns and the negative externalities, Walsh, Möhring, Koot & Schaarschmidt (2014) proposed three categories of preventive instruments to reduce return rates, namely 1) monetary instruments, 2) procedural instruments and 3) customer-based preventive instruments. According to Walsh et al. (2014), the distinction between the three instrument categories is necessary to study the performance of each preventive instrument more effectively. According to Walsh et al. (2014), monetary instruments "are aimed at financially disincentivizing (or financially incentivizing) customers from returning (retaining) products" (p.6). Furthermore, "procedural instruments are designed to either reduce transparency (in relation to the return process) for customers, to identify 'return sinners' and to increase the efficiency of the order and delivery process" (Walsh et al., 2014, p. 6). According to the authors, "customer-based instruments attempt to increase the ease of the order process from the consumer perspective by reducing consumers' perceived pre-purchase uncertainty" (Walsh et al., 2014, p. 6).

However, literature studies conducted by Walsh & Möhring (2017) and Walsh et al. (2014) indicate that prior research has mainly focused on monetary instruments and that existing research about procedural instruments and mostly customer-based preventive product return instruments is sparse. Based on a literature study regarding apparel returns and preventive instruments, the observation could also be made that so far many research has mostly focused on addressing the logistic problems post purchasing. The result have shown that not much empirical research has been conducted so far on how to prevent apparel returns pre-purchasing or during the online screening / evaluation process of apparel items. Consequently, research regarding technologies and instruments which can be used to influence the customers' online pre-purchase decision in order to prevent unnecessary apparel returns is lacking. There is thus a need for online apparel retailers to create an overview of reasons which drive customers' apparel returns and instruments through which apparel returns can be reduced. Therefore, managing apparel returns for online apparel retailers thus concerns an exploratory research for customers' apparel return behavior and technology preference. Since customers have to use the technologies and instruments, this gap needs to be addressed from the customers (users) perspective, since the customers perceived effectiveness of these technologies will determine the impact on unnecessary apparel returns. In other words, the customers' acceptance towards these technologies will determine the impact on unnecessary apparel returns.

Therefore, this paper aims to identify what the customers' acceptance is regarding various technological alternatives designed to prevent unnecessary apparel returns within the context of apparel e-commerce. This is done by applying a more qualitative approach and operationalization of the Technology Acceptance Model (TAM), whereby less data is required to produce reliable results. Since this research is conducted in the empirical setting of apparel e-commerce, the applicability and reliability of the applied approach is presented in the apparel e-commerce sector. In addition, since the technologies are designed to be used by customers, its success relies greatly on the customers usage. Therefore, the research is mainly approached from the users (customers) perspective.

The remainder of this paper is structured as follows. In section 2, the theoretical model (TAM) is explained in virtue of which the customers' acceptance regarding the various technologies is determined. The applied qualitative approach used to operationalize TAM is first described in section 3. In the fourth section, the application of the research approach and the results are presented. At last, in section 5, the conclusions, discussions and recommendations are presented.

2. THEORY (TAM)

In order to understand the users acceptance or rejection of technologies, various theories exist in literature which can be used, such as the theory of reasoned action (TRA) developed by Fishbein and Ajzen (Fishbein & Ajzen, 1975), the theory of planned behaviour (TPB) developed by Ajzen (Ajzen, 1985) and the Technology Acceptance Model (TAM) developed by Davis (Davis, 1986). To predict users behavioural intention, which measures the likelihood of a behaviour occurring, TRA uses the determinants relative importance of attitudes (i.e. the users' feeling towards a particular behaviour) and subjective norms (i.e. the way perception of the users' significant other affects the users' performance and behavior) (Fishbein & Ajzen, 1975). To improve the predictive power of TRA, the TPB was developed by Ajzen (Ajzen, 1985), wherein the additional determinant perceived behavioural control (i.e. the users perceived control over expressing their own behaviours and attitudes) was included as well to predict the users behavioural intention.

However, Davis (1986) observed that this additional determinant did not have a high correlation with the use of technologies. As a result of this, TAM was developed, to predict the technology acceptance, wherein subjective norm and perceived behavioural control were excluded and the actual use (i.e. behaviour) was simply predicted by solely using two determinants which are: 1) the Perceived Usefulness (PU) and 2) the Perceived Ease of Use (PEU) Davis (1986). Since this model solely predicts the acceptance of technologies via two determinants PU and PEU, this model suits the practical research goal better compared to the aforementioned theories. As a result, the TAM developed by Davis (1986), which is a very prominent model in explaining the technology acceptance (especially for information technologies) is used. Underlying figure 1 provides an overview of the TAM, and X1, X2, and X3 are the external predictors used to measure the two determinants (PU and PEU).

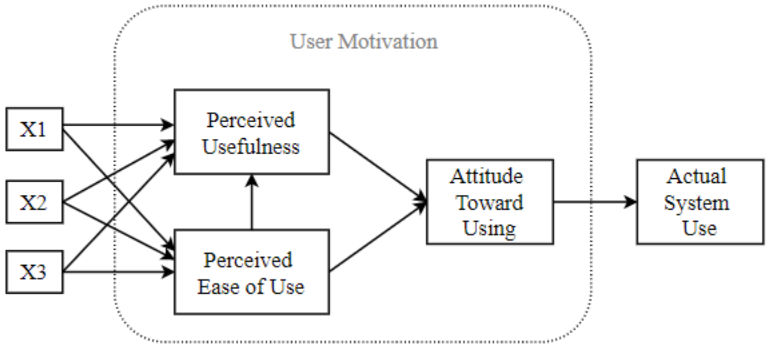

Figure 1: Technology Acceptance Model (retrieved from: Davis, 1986, p. 24)

Perceived usefulness (PU): is defined as "the degree to which an individual believes that using a particular system would enhance his or her job performance" (Davis, 1986, p. 26). Consequently, PU expresses the effectiveness of a technology in addressing a specific function.

Perceived ease of use (PEU): is defined as "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davis, 1986, p. 26). According to Davis (1986), "PEU plays a crucial role in understanding an individual's response to information technology" (p. 26).

TAM finds it popularity in its simplicity, as it solely uses two determinants to predict technology acceptance, PU and PEU, which makes the model highly versatile and easy to apply (Vogelsang, Steinhüser, & Hoppe, 2013). In literature, other various frameworks exits which can be used to evaluate the customers' acceptance regarding technologies, e.g. the Feitelson & Salomon (2004) framework and the Multi-level perspective on Technology Transitions framework developed by Geels (2002). However, since these frameworks also include determinants such as political, institutional and financial determinants to predict the technology success, they are deemed too broad for this research goal, which is to solely determine the customers preference regarding various technologies.

2.1 Predictive validity

In the past 30 years, multiple extensions have been made of TAM, since the model has mainly been criticized on its predictive validity, as it is perceived as incomplete since it only predicts the acceptance based on solely two determinants (Legris, Ingham, & Collerette, 2003; Chutter, 2009). Based on an extensive literature study conducted by Marangunić & Granić (2015) regarding the application of TAM, many changes and extension of the TAM were identified to increase the predictive validity of the model.

Therefore, to increase the predictive validity of TAM in this research, the decision was also made to examine additional determinants to predict the customers' acceptance regarding the technological alternatives. This was done through a literature study regarding TAM and the inclusion of experts opinion with a background in academia and the apparel e-commerce industry. In section 4, this is further elaborated.

3. METHODOLOGY

Based on the extensive literature study conducted by Marangunić & Granić (2015) and the own literature study conducted regarding TAM, the observation could be made that almost all publications used SEM to operationalize TAM, implying that the fraction of qualitative approaches is still very small. SEM is a statistical approach, mostly used in the field of psychology (Nachtigall, Kroehne, Funke, & Steyer, 2003). It can be considered as a combination of factor analysis, multiple regression analysis and path modelling, and it is applied to evaluate the structural relationship/ correlation between indicators (measured variables) and latent variables (non-observable variables) without measurement error (Hox & Bechger, 1999; Nachtigall et al., 2003). Since non-observable variables such as attitude toward using a technology and the users' acceptance of technologies cannot be measured directly, indicators (observable variables) through which they can be measured are required.

However, since the aim of this paper is to explore the customers' acceptance regarding various technological alternatives based on the ranking of a set of indicators (i.e. criteria), and not to determine the correlation between the indicators used to predict the users technology acceptance, another more qualitative approach was used to operationalize TAM.

Consequently, within this research an MCA approach is applied whereby TAM is used as theoretical foundation to identify and theoretically underpin the indicators (i.e. criteria) which are necessary to evaluate the customers' technology acceptance and rank the technological alternatives in the context of apparel e-commerce returns management. As a result, an MCA approach was applied, in order to quantify the importance/influence of indicators (i.e. criteria) and determine which indicators are perceived as the most important for achieving users' technology acceptance. In the following section, this approach is described.

3.1 MCA approach

A MCDM problem is normally presented in a matrix form, as indicated in underlying matrix. Underlying matrix indicates the general form of an MCA approach for the evaluation of a set of alternatives $\{a_1, a_2, \ldots a_m\}$based on a set of decision-criteria $\{c_1, c_2, \ldots, c_n\}$ and p_{ij} is the score of each alternative *i* with respect to each criterion *j* (Rezaei, 2015). Its overall aim is to rank the alternatives and select the best one.

$$
A = \begin{array}{c c}
 & \begin{array}{cccc} c_1 & c_2 & \cdots & c_n \end{array} \\
\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_n \end{array} &
\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix}
\end{array}
$$

By using for example the weighted value function, as presented in underlying equation, the overall value of alternative *i* presented as (V_i) can be calculated. When for example the weight w_j is assigned to criterion *j*, then Vi is simply determined by multiplying the score p_{ij} with the respective weight w_j of criterion *j* ($wj \geq 0, \sum wj = 1$) (Rezaei, 2015).

$$V_i = \sum_{j=1}^{n} w_j p_{ij}$$

Following this MCA approach, first a set of alternatives is needed followed by a set of decision-criteria by which the alternatives can be evaluated. Then, by using a preference elicitation method, the criteria weights should be established. In literature, a variety of methods exist which can be applied to infer the criteria weights. In the next section the applied novel method is described.

3.2 Bayesian BWM

In this research, the BWM is applied as preference elicitation method, since it requires the least pairwise comparisons between the criteria and produces more reliable results compared to other MCDA approaches such as AHP (Rezaei, 2015, 2020). When using the BWM, a number of *2n-3* comparisons are required (Rezaei, 2015). When using AHP, the number of comparisons needed is *n(n-1)/2* comparisons (Saaty, 2004). Furthermore, the BWM is easier to combine with other MCDM methods compared to for example AHP (Rezaei, 2015).

Since this research examines what the technology acceptance is from the perspective of users (customers), a group decision-making method is used to operationalize TAM. Group decision analysis methods can be used to aggregate individual preferences to present the best alternative (Mohammadi & Rezaei, 2019). Since the initial BWM cannot combine the preferences of multiple decision-makers at once, the Bayesian BWM is used instead, which is developed by Mohammadi & Rezaei (2019). The novel Bayesian BWM is preferred over other group MCDM methods such as e.g. group AHP, due to the aforementioned benefits of BWM.

Compared to the initial BWM, the Bayesian BWM solely differs in the last step (step 5) which consists of computing the criteria weights. This being said, it uses the same input data as the initial BWM. With the use of probabilistic modelling and interpretation of the data provided by the Bayesian BWM, the combined distribution and each-and-every individual preferences are computed at the same time, resulting in more reliable criteria weights (Mohammadi & Rezaei, 2019). The Bayesian BWM has the following additional sub-steps which are undertaken in step 5.

Step 5.1. Constructing the probability distribution

Assume that there are *k* decision-makers ($k = 1, 2, \ldots K$), there are *j* evaluation criteria ($c_j = c_1, c_2, \ldots, c_n$), then A_B^k represents the Best-to-Others (BO) vector of one decision-maker and A_W^k the Others-to-Worst (OW) vector of once decision-makers. If the optimal weights of one decision-maker is w^k, the optimal group weight after aggregation is w^{agg}. The vector, $A_B^{1:K}$ represents the BO vector of all decision-makers and $A_W^{1:K}$ indicates the OW vector of all decision-makers. Based on this, the equation for the joint probability distribution of the group decision for the Bayesian BWM is formulated as:

$$P\left(w^{agg}, w^{1:K} \middle| A_B^{1:K}, A_W^{1:K}\right)$$

If the probability in the aforementioned equation is calculated, the following probability rule can be used to compute the probability of each individual variable.

$$P(x) = \sum_{y} P(x,y)$$

with x and y representing arbitrary random variables (Mohammadi & Rezaei, 2019).

Step 5.2. Calculating the optimal group weight

The aggerated weight w^{agg} is dependent on the optimal weight of every individual decision-maker w^k, which is calculated by the input BO and OW vectors (A_B and A_W). Each time new input data (pairwise comparison data) is inserted, w^{agg} is updated. As a result of the previous concepts, conditional independence is present between variables. Taking this independence into consideration, the equation for the joint probability of the Bayesian BWM can be presented as:

$$\mathrm{P}\left(w^{agg}, w^{1:K} \middle| \mathrm{A}_B^{1:K}, \mathrm{A}_W^{1:K}\right) \propto \mathrm{P}\left(\mathrm{A}_\mathrm{B}^{1:K}, \mathrm{A}_W^{1:K} \middle| w^{agg}, w^{1:K}\right) \mathrm{P}\left(w^{agg}, w^{1:K}\right)$$

The above equation, can further be presented as:

$$\mathrm{P}\left(A_B^{1:K}, A_W^{1:K} \middle| w^{agg}, w^{1:K}\right) \mathrm{P}\left(w^{agg}, \mathrm{w}^{1:K}\right) = \mathrm{P}(w^{agg}) \prod_{k=1}^{K} \mathrm{P}\left(A_W^k \middle| w^k\right) \mathrm{P}\left(A_B^k \middle| w^k\right) \mathrm{P}\left(w^k \middle| w^{agg}\right)$$

Based on the above equation, the corresponding probability can be found by specifying the distribution of each element. As a result, $A_B^k|w^k$ and $A_W^k|w^k$ can be defined as follows:

$$A_B^k \left| w^k \sim multinomial \left(\frac{1}{w^k}\right), \forall_k = 1,2,\ldots,K; \; A_W^k \right| w^k \sim multinomial \left(w^k\right), \\ \forall_k = 1,2,\ldots,K$$

Furthermore, w^k under w^{agg} conditioned can be composed as underlying Dirichlet distribution:

$$w^k|w^{agg} \sim \mathrm{Dir}(\gamma \times w^{agg}), \;\; \forall_k = 1,2,\ldots,K$$

with w^{agg} being the averaged value of the distribution and γ is a non-negative parameter (Mohammadi & Rezaei, 2019).

Since γ is a non-negative parameter, it needs to obey the underling gamma distribution where a and b represents the shape and the scale parameters of the gamma distribution.

$$\gamma \sim gamma(a, b)$$

Ultimately, the aggregated or group optimal weight w^{agg} abides to the Dirichlet distribution, with the parameter α being set to 1.

$$w^{agg} \sim Dir(\propto)$$

Once the probability distribution of all parameters is finalized, the posterior distribution is calculated by using the Markov-chain Monte Carlo (MCMC) technique (Mohammadi & Rezaei, 2019).

Compared to the initial BWM, the Bayesian BWM has the following benefits. First, it can instantly combine the final criteria weights for a group of decision-makers, when multiple decisions-makers are involved (Mohammadi & Rezaei, 2019). This means that the researcher does not have to obtain the criteria weights of each decision-maker first separately and aggregate the weights afterwards using the arithmetic mean, which according to Mohammadi & Rezaei (2019) is prone to outliers and provides limited information to decision-makers (in this case apparel e-commerce decision-makers).

Step 5.3. Credal ranking and Confidence level

The Bayesian BWM provides a credal ordering of each and every pair of criteria $\left(c_i, c_j\right)$ for all $\left(c_i, c_j \epsilon\ C\right)$, with C being the set of criteria. In order to understand whether the rankings of the criteria (based on their group weights) are consistent with the evaluation of all experts, the confidence level

(CL) is computed in the weight directed graph (Mohammadi & Rezaei, 2019). The CL indicates the probability or confidence (P) that c_i is better that c_j and is computed as follows:

$$P\left(c_i > c_j\right) = \int I\left(w_i^{agg} > w_j^{agg}\right) P(w^{agg})$$

In the above equation, I represents a conditional parameter which can only be computed if $\left(w_i^{agg} > w_j^{agg}\right)$ is detained, or else it is 0. Evidently, the CL is obtained by the number of samples Q acquired by the Markov-chain Monte Carlo technique (MCMC).

$$P\left(c_i > c_j\right) = \frac{1}{Q} \sum_{q=1}^{Q} I\left(w_i^{agg_q} > w_j^{agg_q}\right); P\left(c_j > c_i\right) = \frac{1}{Q} \sum_{q=1}^{Q} I\left(w_j^{agg_q} > w_i^{agg_q}\right)$$

In the above equation, w^{agg_q} represents q $w^{agg's}$ from MCMC samples. If $P\left(c_i > c_j\right) > 0.5$, then the criterion *i* is more important than criterion *j* (Mohammadi & Rezaei, 2019). The total probability is equal to 1, ($P\left(c_i > c_j\right) + P\left(c_j > c_i\right) = 1$.

Through the provided credal ranking and the assigned confidence levels (CL) in the weight directed graph, the groups' perceived importance of one criterion over one another is visualized, which can provide decision-makers (in this case apparel e-commerce decision-makers) with more information on how to adapt current arrangements (Mohammadi & Rezaei, 2019).

Compared to SEM, which determines the technology acceptance based upon the relationship between the indicators, this research attempts to determine the customers' technology acceptance through the assigned importance/ preference to each indicator (i.e. criterion). Consequently, the influence on technology acceptance is quantified through the computed weights of each indicator (i.e. criteria). Criteria with high aggregated weights are considered to have a significant impact on technology acceptance, suggesting that a high degree of users' (customers') technology acceptance could be realized once scoring well on each and every criterion. In the following chapter, the theory and methodology are applied.

Following the MCA approach, within this research the following steps were initialized to determine the customers technology preference using the Bayesian BWM as method to operationalize TAM.

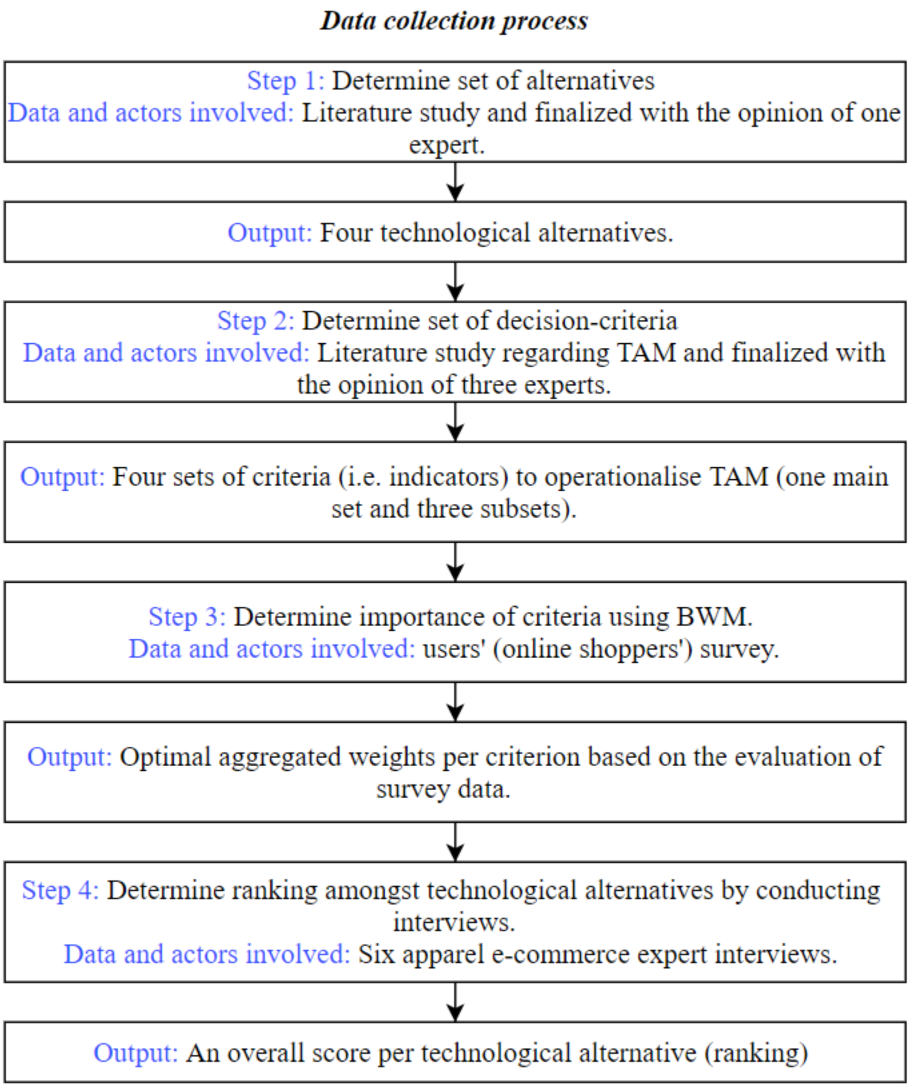

Figure 2: Stepwise data collection process

In the following chapter, the applicability and reliability of this MCA approach in operationalizing users' technology acceptance is analysed within the context of apparel e-commerce returns management.

4. RESULTS

4.1 Set of Alternatives

In order to obtain a set of alternatives, a literature study was conducted as described in step 1 of figure 2. The results showed that research so far has mainly focused on monetary and procedural instruments, and not so much on customer-based instruments, which according to Walsh et al. (2014) "attempt to increase the ease of the order process from the consumer perspective by reducing consumers' perceived pre-purchase uncertainty" (p. 6). As a results, these instruments are treated in this research.

Walsh, et al. (2014) indicate that "the purpose of using these instruments is to communicate suitable information about the product to customers, such that they can evaluate the personal fit more precisely and refrain from returning it because of a possible misfit" (p. 8). As a results, in this research, return reasons were included which can be addressed by these instruments.

The identified drivers of customers online purchase apparel returns were: 1) disconfirmation driven (Saarijärvi et al., 2017), 2) size-chart driven (Saarijärvi et al., 2017), 3) feeling driven (Saarijärvi et al.,

2017) and 4) benefit maximization driven (Saarijärvi et al., 2017; Brooks & Brooks, 2014; De Leeuw, Minguela-Rata, Sabet, Boter, & Sigurðardóttir, 2016). In table 3 of the research conducted by Kalpoe (2020), an overview of the identified reasons linked to these drivers is presented Based on the identified reasons for apparel returns, the following four apparel attributes were extracted, which are necessary for customers to evaluate apparel accurately online: 1) material information, 2) colour information, 3) fit & size information and 4) style information.

Furthermore, creating the right expectations regarding the apparel attributes, providing accurate product information, and creating a 'feel' for and perception of apparel items displayed online were also essential requirements that were identified, which the technologies should fulfil in order to prevent unnecessary apparel items. These features were based on the observation that apparel items were returned because the attributes were different from what was expected, the provided information was misleading, style related features were not visible and customers were unable to get a 'feel' and perception of the apparel items.

This literature study has also lead to the identification of various customer-based instruments and technologies such as: height/size chart, fit & size recommendation application, alternative product photo's, mix-and-match function, zoom technologies, avatars and virtual dressing rooms. However, since some of these identified instruments on their own cannot provide all the aforementioned requirements, it was necessary to combine some instruments such that they can fulfil the requirements and function as comparable alternatives against the technologies which on their own can fulfil all the requirements. The combined alternatives were based upon the current practices of apparel e-commerce retailers. This was done to provide practical solutions to apparel retailers. The set of alternatives was finalized with the opinion of an apparel quality assurance inspector (expert 1 indicated in table 1). The alternatives are composed in such a way that they build upon each other qua functionalities. As such A2 builds upon A1 by adding a fit & size recommendation function to A1. A3 builds upon A2 by adding a computerized virtual try-on experience to A2. A4 builds upon A3 by providing a more realistic virtual try-on experience compared to A3. The composed alternatives are:

A1: The bare minimum
A2: The bare minimum with a fit & size recommendation instrument
A3: Avatar (digital computer-based twin)
A4: Virtual Dressing Room (VDR)

In table 6 of the research conducted by Kalpoe (2020), a thorough description of each alternative is presented.

4.2 Set of criteria

Following the second step of the MCA approach (as indicated in figure 2), a set of decision criteria (i.e. indicators) to operationalize TAM needs to be established._For the MCA, criteria can be established through literature research when sufficient literature is available. Otherwise, the criteria can be established through interviews. Within this research, a literature study regarding TAM functioned as input, since sufficient literature was available in the e-commerce field. The set particularity used for the apparel e-commerce case was finalized with three experts opinions, of which two have a background in academia and one is an apparel quality assurance inspector at the fourth biggest e-commerce retailer in the Netherlands and the second biggest online fashion retailer in the Netherlands.

Underlying figure 3 gives an overview of the criteria sets.

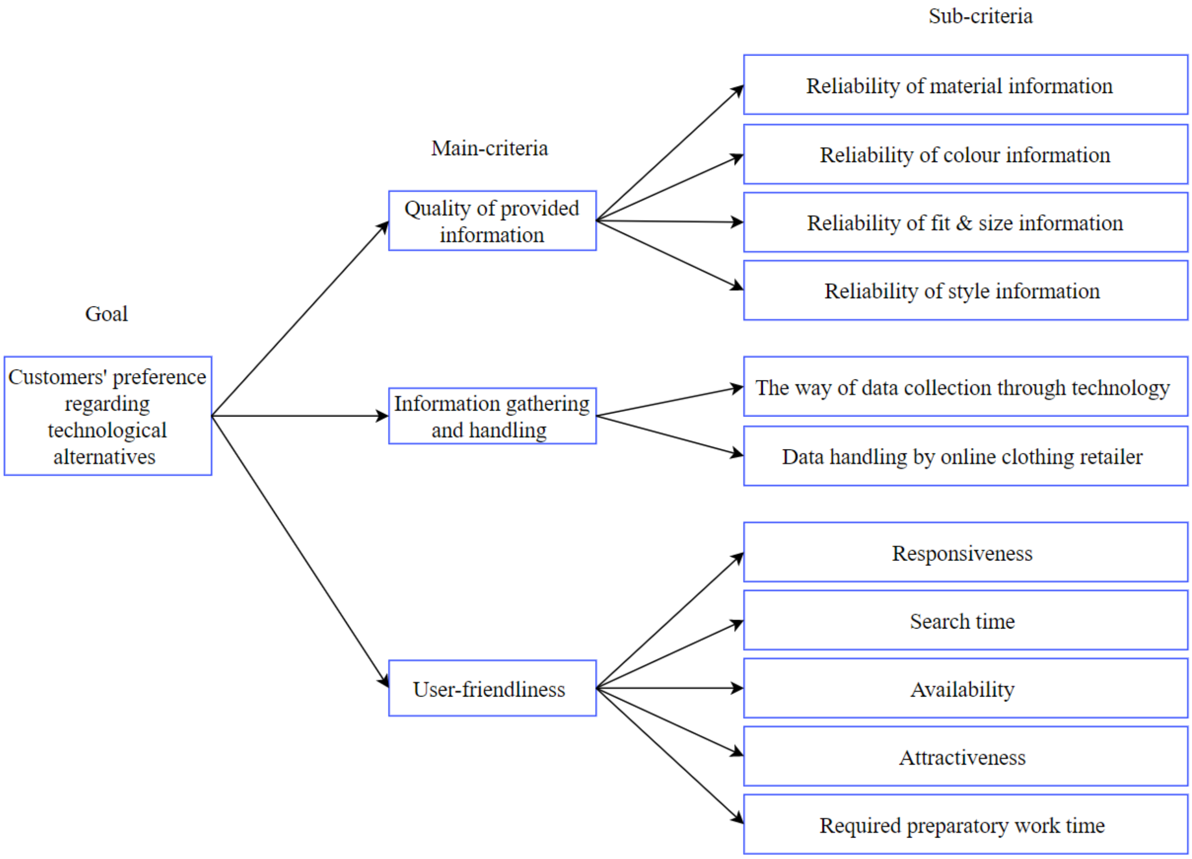

Figure 3: Hierarchy of criteria to evaluate the technologies

As indicated in figure 3, two hierarchy levels exist, namely main-criteria (i.e. main-indicators) and sub-criteria (sub-indicators). The main-criteria were established, based on the synergy between the identified significant sub-criteria through the literature study regarding TAM.

Based on the description of TAM, provided in section 2, the original TAM only has two determinants, namely: 1) the PU and 2) PEU. However, based on literature study results regarding TAM, trust was also an important determinant and is therefore also included as determinant of technology acceptance. As a result, this research provides an extension of the original TAM. Hence, in this research, the following three determinants of technology acceptance are used: 1) PU, 2) Trust and 3) PEU.

Based on the literature study regarding TAM, the main-criterion ‘quality of provided information’ is mostly perceived as significant external predictor of the determinant PU, the main-criterion ‘information gathering and handling’ is mostly perceived as significant external predictor of Trust and the main-criterion ‘user-friendliness’ is mostly perceived as significant external predictor of the determinant PEU. Based on the applicability for the online e-commerce case, the decision was made to also include the three main-criteria and their sub-criteria as such in the research.

4.3 Users (customers) preference

Groups’ optimal weights

The optimal aggregated weight per criterion was established by applying the Bayesian BWM. The input data for the BWM was obtained through an online survey targeted at online apparel shoppers. In total, 216 respondents whom purchased apparel items online were reached. Before calculating the optimal group weights, the consistency of the respondents was also checked and the ones which were acceptable were considered (Liang et al., 2020). After excluding the pairwise comparisons with an unacceptable consistency ratio, different sample sizes for different levels of the model were acquired and used. As a result, a sample size of 113 was used for the main set. For subset 1, a sample size of 77 was used. A sample size of 113 was used for subset 2 and a sample size of 73 was used for subset 3.

Scores

In order to establish the scores, structured expert interviews were conducted, since experts have the knowledge regarding the effectiveness of each technology and instrument in addressing the specific apparel attribute information. A total of six industry experts, stemming from four online apparel retail companies in the Netherlands were interviewed.

Underlying table 1 provides an overview of the interviewed experts.

Table 1: Characteristics of interviewees

Expert	Company (anonymized)	Function	Expertise	Years of experience
1	A	Quality Assurance Inspector	Technical translation from styling/design to the technical application & visualization of clothing on the web shop, lead of the returns management project	5.5 years
2	B	Quality Assurance Inspector	Responsible for the fit & size of apparel & material quality for woman's department	7 years
3	A	Local marketing manager	Online marketing and retour analysis.	10 years
4	C	Local marketing manager	Omnichannel marketing (physical and digital marketing)	3 year
5	C	Online product specialist	Retour analysis of apparel items, product information optimization.	3 years
6	D	Country online marketing manager	Product recommendations for online apparel items, online marketing campaigns, making the technical translation from styling/design of brands to the technical application & visualization of apparel on the web shop.	2 years

In appendix E of the research conducted by Kalpoe (2020), an overview of the BWM interview is presented. The BWM was applied to obtain the scores. As a result, the BWM structure was imposed on the interview, resulting in a structured interview. Since a concrete structure was present, the interviews could be prepared upfront as it was already know what should be asked. Using the BWM as scoring method resulted in more reliable results, compared to e.g. using a scale from 1 to 10. However, it was more time consuming to obtain and analyse the data. Since the Bayesian BWM was applied, the obtained scores are weights as well.

Once the scores were gained, the performance matrix could be constructed, wherein the alternatives are scored with respect to each criterion.

4.3.1 Interpreting optimal group weights

To quantify the importance /influence of the indicators (i.e. criteria) and determine which indicators have the largest influence on technology acceptance, the Bayesian BWM was applied.

Underlying table 2 provides an overview of the obtained weights, based on survey respondents:

Table 2: Customers' group weights of main-criteria and sub-criteria

Main-criteria	Weight	Sub-criteria	Local Weight	Global Weight
C1. Quality of provided information	0.441	c1.1. Reliability of material information	0.242	0.107
		c1.2. Reliability of colour information	0.248	0.110
		c1.3. Reliability of fit & size information	0.318	0.140
		c1.4. Reliability of style information	0.192	0.084
C2. Information gathering and handling	0.235	c2.1.The way of data collection through technology	0.432	0.101
		c2.2. Data handling by online clothing retailer	0.568	0.133
C3. User-friendliness	0.324	c3.1. Responsiveness	0.207	0.067
		c3.2. Search time	0.223	0.072
		c3.3. Availability	0.190	0.062
		c3.4. Attractiveness	0.179	0.058
		c3.5. Required preparatory work time	0.201	0.065

- Interpreting main-criteria weights

Looking at the main-indicators, 'quality of provided information' is the most important main-indicator for technology acceptance (w^{agg} = 0.441). This implies that individuals certainly prefer to obtain reliable apparel attribute information, compared to the perceived ease of use of the technology and the main-indicator information gathering and handling. When looking at the underlying figure, there can be observed that the criterion 'quality of provided information' has a high confidence level of 1 compared to both other two criteria 'data gathering and handling' and 'user-friendliness', implying that the degree of certainty about the criterion is also evident. In other words, we can be very sure about the superiority of C1 over C3 and C2, that 'quality of provided information' is certainly more important than 'user-friendliness' of the technology and 'information gathering and handling'.

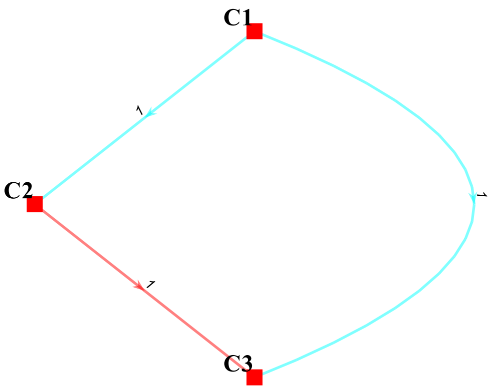

Figure 4: Credal ranking of main-criteria

- Interpreting global weights of sub-criteria

The results show that from all 11 sub-indicators, fit & size information is perceived as the most important for technology acceptance (w^{agg} = 0.140). This implies that individuals assign high value to obtaining reliable fit & size information. Slightly behind it is 'data handling by online clothing retailer' (w^{agg} = 0.133). This implies that the way the online apparel retailer uses and stores the collected information for its services significantly impacts the customers preference and technology acceptance. The third most important sub-indicator is 'reliability of colour information' (w^{agg} = 0.110), implying that the accuracy, completeness and truthfulness of the provided information regarding the colour of apparel items is the third most important sub-indicator affecting technology acceptance. The results also show that ‘reliability of material information’ and ‘the way of data collection through technology’ are the fourth and fifth most important sub-indicators for technology acceptance (w^{agg} = 0.107 and w^{agg} = 0.101). This implies that the provision of accurate, complete and truthful information regarding the material of apparel items which refers to material thickness, stretch-ability, texture and stitching (sewing) also significantly influences the customers’ technology acceptance, followed by the way in which the technology acquires customers information (for example through scanning, facial recognition or manually inserting body-measurement information).

Looking at the main-indicator 'quality of provided information', the sub-indicator 'reliability of fit & size information' is perceived as the most important. Based on the assigend confidence level in underlying figure 5, the relationship is also evident, suggesting that 'reliability of fit & size information' is certainly more important (CL = 1) than 'reliability of material information', 'reliability of colour information' and 'reliability of style information'. On the other hand, table 2 shows that 'reliability of style information' is perceived as the least important. This implies that individuals who purchase apparel items online are the least interested in obtaining style information to evaluate apparel items online, compared to the other three apparel attributes: 1) material, 2) colour and 3) fit & size information.

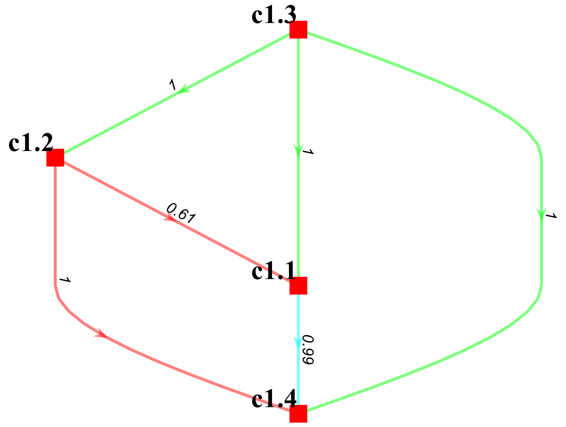

Figure 5: Credal ranking of sub-criteria related to ‘quality of provided information’

Furthermore, the observation can be made that all two sub-indicators related to the main-indicator 'information gathering and handling' are also perceived as highly important indicators for technology acceptance (aggregated weight higher than 0.1). Furthermore, of all the 11 sub-indicators, 'the way of data collection through technology’ is perceived as the second most important indicator for technology acceptance. This implies that customers assign high value to privacy and security concerns, which has a high significant influence on the determinant Trust and through that on the technology acceptance. Looking at the assigend confidence levels in underling figure 6, 'data handling by online clothing retailer' is certainly more important than 'the way of data gathering through technology' in determining the customers’ technology acceptance.

Figure 6: Credal ranking of sub-criteria related to ‘information gathering and handling’

Looking at the sub-indicators belonging to the main-indicator 'user-friendliness', the results show that 'search time' is perceived as the most important for technology acceptance. This implies that the perceived ease of use mostly relies on the number of clicks / effort an individual needs to perform when using the technology to evaluate apparel items online. Looking at the assigend confidence levels in underlying figure 7, 'search time' is certainly more important than all sub-indicators related to user-friendliness, with a confidence level that is higher or equal to 0.81. This implies that 'search time’ is certainly perceived as more important in determining the perceived ease of use of the technologies, compared to the other sub-indicators belonging to the main-indicator ‘user-friendliness’.

The second most important sub-indicator related to the main-indicator ‘user-friendliness’, is 'responsiveness', implying that the loading time of the technology is the second most important aspect influencing the perceived ease of use of the technology. The results show that 'required preparatory work time' is the third most important sub-indicator, implying that the amount of work customers need to do upfront before they can start using the technology to evaluate apparel items with is the third most important indicator influencing the perceived user-friendliness of the technology. The fourth and fifth most important sub-indicators related to the main-indicator 'user-friendliness' are 'availability' and 'attractiveness'. This implies that the ability to use the technology on any device is perceived as the fourth important aspect defining user friendliness and that customers technology acceptance is the least influenced by the aesthetic features of the technologies.

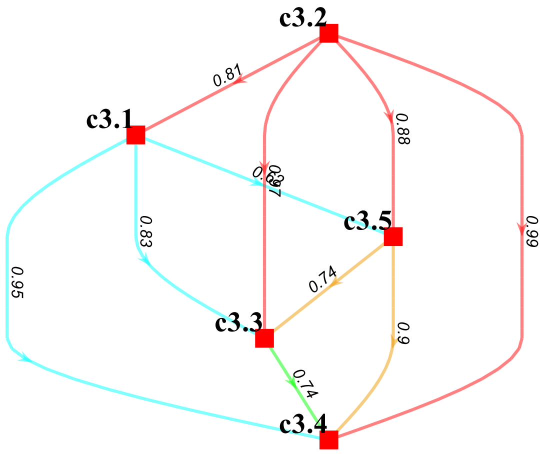

Figure 7: Credal ranking of sub-criteria related to ‘user-friendliness’

4.3.2 Comparing the alternatives

In underlying table 3, an overview of the obtained scores (i.e. weights) of each alternative with respect to each sub-criterion are presented. The scores were obtained from six apparel e-commerce expert interviews.

Table 3: Experts’ scores (i.e. weights) of main-criteria and sub-criteria

Sub-criteria	Local Weight			
	A1	A2	A3	A4
Reliability of material information	0.140	0.204	0.299	0.358
Reliability of colour information	0.217	0.217	0.228	0.338
Reliability of fit & size information	0.093	0.186	0.362	0.359
Reliability of style information	0.106	0.144	0.359	0.392
The way of data collection through technology	0.424	0.325	0.143	0.108
Data handling by online apparel retailer	0.503	0.278	0.123	0.096
Responsiveness	0.438	0.339	0.101	0.122
Search time	0.274	0.332	0.171	0.223
Availability	0.475	0.307	0.096	0.122
Attractiveness	0.104	0.144	0.316	0.436
Required preparatory work time	0.304	0.361	0.151	0.185

Looking at table 3, the result obtained from the expert interviews show that A3 scores the best with respect to the sub-indicator ‘reliability of fit & size information’, however closely followed by A4. According to expert 2, 40% of all apparel returns in the company are indeed a cause of fit & size issues (e.g. the size chart that is not accurate enough so that apparel does not fit) and 40% of all returns also stem from apparel items not being as expected (disconfirmation driven). According to expert 1, the amount of apparel returns stemming from fit & size issues are even higher, nearly 52% whilst for material, colour and style it is 6% for each attribute. According to expert 4, in total 37% of all apparel is returned as a result of fit & size issues (18 % too small and 19% too big), whilst style is 31% (e.g. the style, when worn, does not look as good as expected) and for material and colour the amount of returns is a combined 2% (e.g. other hue, or unclear pictures of apparel items). Based on this, the observation can be made that the identified apparel return reasons from literature used in this research are indeed valid, since the literature study has shown that most returns stem from fit & size issues.

The results presented in table 3 show that A4 scores the best with respect to the indicators reliability of material, colour and style information. The main reason why A4 is still perceived as the best with respect to these three sub-indicators stems from its ability to try-on apparel items on the virtual appearance of the individuals' own body image, which gives a better perception and feel of the apparel style and colour according to the interviewed experts. Furthermore, the dynamic movement which can be created gives a better feel and perception of the material quality, which makes it a superior alternative, is very effective to evaluate the personal match of apparel items with online. Furthermore, A4 also scores the best with respect to the sub-indicator 'attractiveness'. A4 was perceived as the most attractive alternative, due to its ability to try-on apparel items on one's own mirrored image. In addition, the dynamic movement where apparel moves with the individuals' body movements, makes it more exiting, playful and visually appealing for customers to use.

Looking at the same table, A1 scores the best with respect to all two sub-indicators belonging to the main-indicator 'information gathering and handling'. This implies that customers are perceived to have the least to no privacy and security concerns when using A1, since no data is collected in order to be able to use the instruments to evaluate apparel items with. Furthermore, based on all six experts, A1 scores the best with respect to responsiveness and availability, since it is perceived as the least technically complex and requires the least amount of storage capacity as only pictures and a size table need to be uploaded along with the mix-and-match function. As such, the loading time will be the least negatively affected using A1. Since A1 requires the least amount of computational power, it can easily be made available on any device compared to the other alternatives.

A2 scores the best with respect to the sub-indicators search time and required preparatory work time. A2 is perceived as the best alternative with respect to search time, since the number of clicks/ effort needed to acquire the necessary apparel attribute information to evaluate apparel items with is considered to be the lowest for A2. Customers only have to fill in a format wherein they indicate their body-measurements information, which according to the experts is relatively easy and quick to do. In terms of 'required preparatory work time' A2 scores the best, since most of the experts perceived A2 to require the least amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit.

By combining table 2 and 3, underlying table 4 is composed. Table 4, provides an overview of the obtained scores from experts along with the weights obtained through the survey. In the first column, the sub-indicators (i.e. sub-criteria) are indicated. The subsequent four columns indicate the assigned scores of each alternative with respect to each criterion, obtained from 6 expert interviews. In the last column, the global weights are indicated. Using the weighted sum equation (2), the final scores were obtained and the alternatives were ranked based on preference.

Table 4: Ranking of the technological alternatives

Sub-criteria	Scores of Technological Alternatives				Global weights
	A1	A2	A3	A4	
Reliability of material information	0.140	0.204	0.299	0.358	0.107
Reliability of colour information	0.217	0.217	0.228	0.338	0.110
Reliability of fit & size information	0.093	0.186	0.362	0.359	0.140
Reliability of style information	0.106	0.144	0.359	0.392	0.084
The way of data collection through technology	0.424	0.325	0.143	0.108	0.101
Data handling by online apparel retailer	0.503	0.278	0.123	0.096	0.133
Responsiveness	0.438	0.339	0.101	0.122	0.067
Search time	0.274	0.332	0.171	0.223	0.072
Availability	0.475	0.307	0.096	0.122	0.062
Attractiveness	0.104	0.144	0.316	0.436	0.058
Required preparatory work time	0.304	0.361	0.151	0.185	0.065
Total Score	0.2748	0.2517	0.2221	0.2516	
Ranking	1	2	4	3	

Based on the obtained criteria-weights through the survey and the scores from online apparel retail experts, it can be observed that A1 has the highest chance of reaching users' technology acceptance. A2 is perceived as the second best, closely followed by A4. A3 is perceived to have the lowest chance of reaching technology acceptance.

4.4 Managerial implications

In order to establish what the perceived employment possibility of the technological alternatives is in companies from the point of view of online apparel retailers, at the end of the interview, experts were asked to score all four technological alternatives with respect to the criterion 'employment possibility in company', again using BWM.

Through this, insight was gained about factors which can encourage or inhibit the acceptance of each technological alternative, from the point of view of online apparel retailers in the Netherlands.

Table 5: Ranking alternatives based on practical implementation possibility

Alternatives	Implementation possibility in company	Ranking
A1	0.456	1
A2	0.342	2
A3	0.099	4
A4	0.104	3

As can be seen in table 4, A1 is the most accepted alternative, closely followed by A2. Amongst the two newer technologies (A3 and A4), A4 has a higher employment chance for online apparel retailers than A3. The results obtained from the expert interviews, as indicated in table 5, show that when it comes to the practical implementation of the alternatives, the same ranking is obtained as the ranking regarding the customers' acceptance (indicated in table 4). Consequently, A1 is perceived to have the least amount of managerial implications, since out of all four alternatives A1 is for the most part already employed, aside from the mix-and-match function to evaluate the entire outfit with. A2 is perceived as the second best, since it is perceived as the most technically and financially feasible for the companies, after A1. Looking at the state-of-the-art-technologies, A4 is perceived as the third best alternative, closely followed by A3 which is perceived to have the most amount of managerial implications. A2 is much better than A3 and A4 when it comes to implementation in the company, as based on the experts results for both A3 and A4 experts have to be hired as the current developers do not have the knowledge to operationalize the technologies, which costs more money and time. When using A3 and A4, the whole chain needs to be aligned to the digital way of working which is required to operationalize A3 and A4, which according to the expert interviews does not seem feasible for mature multi-brand stores. In addition, testing the technologies and gathering customer opinions also takes much more time, effort and money compared to A1 and A2.

5. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

The purpose of this paper was to predict the customers' acceptance regarding various technological alternatives designed to increase customers online purchase successes and reduce unnecessary apparel returns, by applying a less data extensive, simpler and reliable approach. As a result, the novel Bayesian BWM is applied to operationalize TAM, which involves identifying various indicators (i.e. criteria), quantifying the importance of each indicator through the assigned preference and determining which indicator has the highest impact on technology acceptance through the assigned weight. Furthermore, by applying the Bayesian BWM, the decision-makers' preference of a criterion could explicitly be confirmed with a certain confidence level.

Within this research, 11 sub-indicators for the customers' technology acceptance and four technological alternatives have been analysed. The analysis shows that reliable fit & size information is the most important sub-indicator influencing the customers' technology acceptance, which according to the examined literature and online apparel retailers is indeed perceived as the apparel attribute with the most apparel returns. This proves that the Bayesian BWM is indeed a valid method to predict the importance of criteria. Furthermore, it seems that currently, A1 has the highest chance of reaching technology acceptance.

The results have indicated that predicting the technology acceptance by operationalizing TAM can be done using the aforementioned MCA approach as well. The novel Bayesian BWM developed by Mohammadi & Rezaei (2019), is applied to a real-life problem (apparel e-commerce) to check its robustness. The result show that the technology which has the highest probability of customers' acceptance is also the one which is currently the most employed by online apparel retailers in the Netherlands. This shows that the novel Bayesian BWM method is indeed a successful method which can predict technology acceptance and preference. Since technology assessment considering the perception of clients/users with the application of TAM, MCDM and the (Bayesian) BWM is very novel, but has proven to produce reliable results, there is highly recommended to conduct further research to examine the robustness of the methodology in different empirical settings.

Based on the outcome of the research, more mature companies, especially multi-brand stores, are advised to focus on A2, since compared to A1, A2 requires the least amount of effort (time, money, expertise) to implement. Since 1) the survey results have shown that reliable fit & size information is perceived as the most important indicators for technology acceptance and 2) based on the experts interviews the fit & size recommendation function of A2 can provide more reliable information compared to the static height/size chart of A1, there is suggested to gradually move from A1 to A2. Since out of the four apparel attributes, style information is perceived as the least important, there is suggested to first focus on the other apparel attributes especially fit & size information (the most important indicator). In order to prevent apparel returns, new companies entering the market are advised to focus on the sub-indicators (i.e. sub-criteria) with the highest weight and the alternatives which score the best with respect to the criteria with the highest weights, as this might help them to increase the number of successful sales and prevent unnecessary apparel returns.

Limitations and future research areas

The reason to why A3 and A4 are the least preferred, might be due to the fact that they are relatively state-of-the-art. To find experts with sufficient expertise especially about A3 and A4 to participate in the BWM was rather difficult. Since only six experts stemming from four companies were approached, the individual influence of the assigned scores are higher, which also impacts the end results. However, when looking at the experts data, most experts shared the same arguments and opinions implying that data saturation was reached.

Although the interviewed experts mostly shared the same opinions, there is still advised to continue this research by interviewing more experts whom are more active in the field of product IT development, to explore the two newer technologies (A3 and A4) better, since the nature of technology development of A4 and A3 is the main reason for this discrepancy (A1 and A2 are inferior to A3 and A4 when it comes to technology superiority). This can mostly be seen by looking at the alternatives' weights with respect to the second and third most important sub-indicators affecting the users' technology preference which are 'the way of data collection through technology' and 'information handling by online clothing retailers' (see table 4), implying that the most amount of privacy and security concerns can occur when using A4 and A3. This is also the case for A4 and A3 with nearly all sub-indicators related to the perceived user-friendliness which are responsiveness, search time, availability and required preparatory work time, implying that these two state-of-the-art alternatives are perceived as the least user-friendly. Since the technological alternatives build upon each other in terms of functionality, the level of perceived technological complexity and data required increases. As a result of this, table 4 shows that whilst the reliability of information provision regarding material, colour, fit & size and style increases per subsequent alternative, the privacy and security concerns increase and the perceived user-friendliness (ease of use) of the technology decreases. Therefore, future research could examine how the ranking will be, once the A3 and A4 are sufficiently mature with respect to these sub-indicators.

Furthermore, while it was true that online apparel retailers were asked to approach the scoring of the alternatives with respect to the main-criteria 'information gathering and handling' and 'user-friendliness' from the customers (users) perspective, it is still possible that the scores obtained from the expert interviews are (slightly) bias. Variables such as experience with functionalities of A1 and A2, low trust in new technologies, time of adoption /low level of innovativeness could all be underlying reasons explaining their assigend scores.

Although the results have indicated that A1 is the most preferred alternative, this cannot be guaranteed with full certainty. It could be that the four technologies could co-exist in practice, since in time the current technological superiority of A1 and A2 over A3 and A4 might change. Table 4 shows that A4 scores very similar to A2 (the second best alternative). Furthermore, A3 scores the best with respect to the sub-indicator 'reliability of fit & size information' (the main reason for apparel returns). However, A4 is not far behind. Since the obtained expert results might be (somewhat) bias, the effectiveness of A3 over A4 with respect to fit & size cannot be fully guaranteed. Variables such as experience with functionalities of A3 and A4 could be underlying reasons explaining their assigend scores. For these reasons, further research is required regarding the technical applications of these technologies. Since A4 and A3 score the best with respect to providing reliable apparel attribute information (material, colour, fit & size and style information), which is required to make better online purchase decisions and refrain customers' from returning items, future research could examine if they can co-exist. In addition, future research could also examine what the customers' technology preference will be amongst different customer segments, by first identifying different clusters (groups) based on characteristics such as age, gender, shopping experience etc., and then computing the different optimal group criteria weights. Through this, more in depth insight might be gained regarding the estimated use of the different technological alternatives and the possibility of co-existence.

The results also show a difference between obtaining data through a survey or an interview, since the obtained survey data has still shown that many respondents still faced difficulties in conducting the pairwise comparisons rightfully. As a result, different sample sizes for different levels of the model were used, after excluding the pairwise comparisons with an unacceptable consistency ratio. On the contrary the data from all experts, which was acquired through a BWM interview, was acceptable and used, since the direct communication made it easier to keep the interviewees on the topic and perform the pairwise comparisons in a rightful manner. This shows that the quality of the data does depend on instruments used to collect the data. Since pairwise comparison methods often use surveys to obtain the required input data, different ways of minimizing this impact of the perceived complexity especially when multiple levels of criteria are present, need to be explored.

Future steps should also take into consideration that this research is limited to the apparel e-commerce sector in the Netherlands, implying that the obtained weights which have led to this ranking of the customers technology preference might be different based on other contextual variables and empirical setting. Furthermore, this research did not examine various measurement approaches which can be used to obtain customers body-measurements along with different applications of the included fit & size recommendation tools. As a result, further research can explore the various approaches and different applications in further detail along with their imposed benefits and costs. Given the uncertainty factor of online customer reviews, as the provided information is based upon customers opinion, online customer reviews along with customers hotline instruments were not included in this research. Therefore, the way in which these instruments can contribute to the reduction of online apparel returns can be further explored.

REFERENCES

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In J. Kuhl, J. and Beckmann (Ed.), *Action Control* (pp. 11–39). https://doi.org/10.1007/978-3-642-69746-3_2

Brooks, A. L., & Brooks, E. (2014). Towards an inclusive virtual dressing room for wheelchair-bound customers. *2014 International Conference on Collaboration Technologies and Systems (CTS)*, 582–589. https://doi.org/10.1109/CTS.2014.6867629

Chutter, M. Y. (2009). *Overview of the Technology Acceptance Model: Origins, Developments and Future Directions*. https://doi.org/http://sprouts.aisnet.org/9-37

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: theory and results* (MIT). Retrieved from http://hdl.handle.net/1721.1/15192

De Leeuw, S., Minguela-Rata, B., Sabet, E., Boter, J., & Sigurðardóttir, R. (2016). Trade-offs in managing commercial consumer returns for online apparel retail. *International Journal of Operations & Production Management*, *36*(6), 710–731. https://doi.org/10.1108/IJOPM-01-2015-0010

Difrancesco, R. M., Huchzermeier, A., & Schröder, D. (2018). Optimizing the return window for online fashion retailers with closed-loop refurbishment. *Omega*, *78*, 205–221. https://doi.org/10.1016/j.omega.2017.07.001

Edwards, J. B., McKinnon, A. C., & Cullinane, S. L. (2010). Comparative analysis of the carbon footprints of conventional and online retailing. *International Journal of Physical Distribution & Logistics Management*, *40*(1/2), 103–123. https://doi.org/10.1108/09600031011018055

Feitelson, E., & Salomon, I. (2004). *The Political Economy of Transport Innovations*. https://doi.org/10.1007/978-3-540-24827-9_2

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, *31*(8–9), 1257–1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Griffis, S.E., Rao, S., Goldsby, T.J., Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. *Operations Management*, *30*(4), 282–294.

https://doi.org/10.1016/j.jom.2012.02.002

Hox, J., & Bechger, T. (1999). An Introduction to Structural Equation Modeling. *Family Science Review*, *11*, 354–373. Retrieved from https://www.researchgate.net/publication/27706391%0D

Kalpoe, R. (2020). *A multi-criteria assessment to determine the customers' technology preference in the context of apparel e-commerce*. Delft University of Technology.

Kennisinstituut voor Mobiliteitsbeleid. (2017). *Stedelijke distributie en gedrag*. Retrieved from https://www.kimnet.nl/binaries/kimnet/documenten/rapporten/2017/06/06/gedrag-en-stedelijke-distributie/Stedelijke+distributie+en+gedrag.pdf.

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management*, *40*(3), 191–204. https://doi.org/10.1016/S0378-7206(01)00143-4

Liang, F., Brunelli, M., & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, *96*, 102175. https://doi.org/10.1016/j.omega.2019.102175

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Minnema, A, Bijmolt, T.H., Gensler, S. (2017). 3. Oorzaken en gevolgen van het terugsturen van online aankopen. In *Ontwikkelingen in het marktonderzoek: Jaarboek Markt Onderzoek Associatie* (42nd ed.). Retrieved from http://moa04.artoo.nl/clou-moaweb-images/images/bestanden/pdf/Jaarboeken_MOA/MOA_JAARBOEK_2017_HFST3.pdf

Mohammadi, M., & Rezaei, J. (2019). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, 102075. https://doi.org/10.1016/j.omega.2019.06.001

Nachtigall, C.; Kroehne, U.; Funke, F.;Steyer, R. (2003). (Why) Should We Use SEM? Pros and Cons of Structural Equation Modeling. *Methods of Psychological Research Online, 8*(2), 1–22.

NRC. (2019). Retourzendingen zijn een structureel probleem voor webwinkels. Retrieved from https://www.nrc.nl/nieuws/2019/02/27/te-coulant-bij-terugsturen-a3679975

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

Rezaei, J. (2020). A Concentration Ratio for Nonlinear Best Worst Method. *International Journal of Information Technology & Decision Making*, 1–17. https://doi.org/10.1142/S0219622020500170

Saarijärvi, H., Sutinen, U.-M., & Harris, L. C. (2017). Uncovering consumers' returning behaviour: a study of fashion e-commerce. *The International Review of Retail, Distribution and Consumer Research*, *27*(3), 284–299. https://doi.org/10.1080/09593969.2017.1314863

Saaty, T. L. (2004). Decision making — the Analytic Hierarchy and Network Processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, *13*(1), 1–35. https://doi.org/10.1007/s11518-006-0151-5

Vogelsang, K., Steinhüser, M., & Hoppe, U. (2013). A Qualitative Approach to Examine Technology Acceptance. *Thirty Fourth International Conference on Information Systems*. Milan.

Walsh, G., Möhring, M., Koot, C., Schaarschmidt, M. (2014). Preventive product returns management systems - A review and model. *22nd European Conference on Information Systems*. Retrieved from https://www2.scopus.com/record/display.uri?eid=2-s2.0-84905845167&origin=inward&txGid=22c32ed1368d3f2aa3bbcc85b31c98cc

Walsh, G., & Möhring, M. (2017). Effectiveness of product return-prevention instruments: Empirical evidence. *Electronic Markets*, *27*(4), 341–350. https://doi.org/10.1007/s12525-017-0259-0

APPENDIX B: Established categories of criteria

The categorisation of criteria was based on the synergy between the criteria retrieved from the literature study regarding TAM. Underlying gives an overview of the five categories along with the corresponding sub-criteria and descriptions. The categories along with their sub-criteria which are indicated in the grey rows are not included in the research.

#	Category	(Sub) criteria	Description	Sub criteria retrieved from underlying study
1	Characteristics of the quality of provided information	Completeness	The completeness of the provided information	Fedorko et al. (2018)
		Truthfulness	The truthfulness of the provided information	Fedorko et al. (2018)
		Rightness	Providing the right information	Landeweerd et al. (2013)
		Reliability	The reliability of the provided information	
		Accuracy	The accuracy of the provided information	Zhou & Zhang (2009)
				Cao et al. (2005)
2	Characteristics of Trust, regarding information gathering and handling	Data Privacy	The disclosure of privacy-sensitive information such as personal health information.	Landeweerd et al. (2013)
			The ability of customers to control what personal information is shared online and used in an unauthorised way, while conducting online transactions.	John (2012)
			Data privacy which refer to users' data not being used for other purposes without the consent of users and data anonymity	Shabana & Arif (2011)
		Security	Providing users' security through encryption such that data cannot be accesses by unauthorized parties.	Shabana & Arif (2011)
			Keeping site secure and reliable	Cao et al. (2005)
			The perception of customers regarding the level of protection against privacy and security threats	John (2012)
			The possibility of the portals to provide access to all services and functionalities in a secure manner.	Tella (2012)
			Users' security risk	Lee (2009)
			The possibility to secure financial and personal information secure, when transmitted online	Changchit et al. (2019)
3	Ease of use Technological Characteristics	Responsiveness	Reducing loading time	Cao et al. (2005)
		Speed	The speed of feedback provision to users	Fedorko et al. (2018)
			Perceived speed of technology, which refers to the payment speed of e-payment system	Tella (2012)
			Benefits such as faster transaction speed	Lee (2009)

		Search facility	Searching time and making searching easier	Cao et al. (2005)
		Availability	Availability (whether on a laptop, computer, mobile phone etc.)	Fedorko et al. (2018)
		Attractiveness	Playfulness	Fedorko et al. (2018)
		Navigation	Refers to how easy users can find the required information on the website	Pires & Lai (2009)
			Navigation, that suits the users and can be easily be controlled	Fedorko et al. (2018)
		Perceived enjoyment	Perceived enjoyment when using the technology	Tella (2012)
		Usability	Refers to how visually appealing, consistent, fun and easy to use a website is	Pires & Lai (2009)
		Playfulness	The extent to which the technology is appealing and exciting	Cao et al. (2005)
		Financial benefits	Financial benefit such as lower transaction fees	Lee (2009)
4	Characteristics of Attitude/behavioural component	Social pressure	Refers to the influences of others on an individuals' attitude/intention to use a technology.	Landeweerd et al. (2013)
		Perceived compatibility	Compatibility with prior experience	Landeweerd et al. (2013)
			Compatibility with existing work practice	Landeweerd et al. (2013)
			Compatibility with values	Landeweerd et al. (2013)
		Experience/Knowledge	Prior experience of using internet	Salam et al. (2005)
			User' web experience	Shabana & Arif (2011)
			Prior experience of using the internet	Salam et al. (2005)
			Prior experience of online shopping	Changchit et al. (2019)
			Education level	Olson & Boyer (2003)
			Tenure in the workforce	Olson & Boyer (2003)
			Annual training received	Olson & Boyer (2003)
		Fairness	Fairness is related to customers' trust with regard to e-commerce vender adoption	Chiu et al. (2009)
		User innovativeness	Innovativeness in a technological context refers to willingness of an individual to try new technologies and can be influenced through the factors such as trust.	Wang & Liu (2009)
		Perceived market orientation		Shabana & Arif (2011)
5	Characteristics of Social interaction	Communication	Communication in Internet use	Salam et al. (2005)
		Online relationships	Online relationships between vendor and customers	Salam et al. (2005)
		User involvement	Involvement of consumers in the development of new a technology	Wang & Liu (2009)

APPENDIX C: Survey used to obtain criteria weights

Dear respondent,

This survey is part of my master thesis at TU Delft. The results of this survey will be used to formulate recommendations about the adoption of information technologies based on the customers' preference. The survey has 5 parts, a total of 18 questions and will take about 10 minutes. The information obtained through this survey will only be used for scientific purposes and is strictly anonymous.

It is important that you use the "back" and "next" buttons at the bottom of each page of this survey to move to a different page and not press previous in your browser. The survey will then close.

If there is an star (*) next to the question, this means that the question needs to be answered in order to continue the survey. Once you start the survey, you cannot save it in between. It is important to finish the survey completely once you have started.

Thank you for your participation,

Ruchika Kalpoe

INTRODUCTION

The case:

Imagine that you want to purchase clothing online with the intention to keep it. In order to do so, you need to be able to evaluate the clothing items accurately. However, as clothing items are displayed virtually (on a website), you cannot 1) feel the items to determine the material quality, 2) you cannot see the colour in real life, 3) you cannot try on the clothing to determine the right size and fit and 4) you cannot test your personal match with regard to style of clothing in real life. Due to these reasons, many online customers experience a pre-purchase uncertainty. Most of the return reasons for online purchased clothing items stem from customers not being able to determine 1) the material quality, 2) actual colour, 3) fit & size and 4) the personal fit with regard to style.

As a result of this, means have been developed to assist customers during the online purchasing process, such that they can evaluate these clothing attributes more accurately in order to increase customers' purchase successes and prevent unnecessary order returns.

The provided technological alternatives to address these return reasons are:

(1) Alternative product pictures with mix-and-match function (to see the total outfit), zoom function and **static** height/size chart: In this alternative, the size chart is provided by a picture of a body-measurement table. Customers have to measure their own body measurements first with a measuring band in order to know to which size category they belong.

(2) Alternative product pictures with mix-and-match function (to see the total outfit), but instead with an **interactive** height/size chart: with an interactive height/size chart, you can insert your body-measurements information such as height, chest, waist, hip size etc. and based on that the personal size, fit and style information can be recommended.

(3) Virtual Dressing Rooms (augmented reality such as Snapchat filters): with this technology you can scan your body (such as Snapchat does) through an application on your computer or mobile phone. Once the technology has captured your body-dimensions, you can "swipedrag" filters of clothing items to determine the personal fit.

(4) Avatars (digital computer-based twin): you can use an existing or make your own animated body-double via an application and use this to try on clothing on the website of clothing retailers.

The purpose of this survey is to determine which criteria you perceive as relevant when it comes to choosing a solution (technology) that you can use to better inform yourself about clothing attributes, in order to increase your online purchase success and prevent unnecessary clothing returns.

You are kindly requested to read the questions carefully and to answer them truthfully.

PART 1. BACKGROUND INFORMATION

1. What is your gender?
   - o Male
   - o Female
2. What is your age?
   - o <18
   - o 18-24
   - o 25-34
   - o 35-44
   - o 45-54
   - o 55-64
   - o >64
3. What is your highest level of education?
   - o Geen/Basisschool
   - o LBO/VMBO (kader-of beroepsgericht)
   - o VMBO (theoretisch gemend/MAVO)
   - o HAVO
   - o VWO/Gymnasium
   - o MBO
   - o HBO
   - o WO

INSTRUCTIONS

For the most part, this survey will contain a series of the following 4 questions:

1. Out of these criteria, which one do you perceive as **most important (best)** for determining your preference of using a technology?

2. Out of these criteria, which one do you perceive as **least important (worst)** for determining your preference of using a technology?

3. How important is the **most important (best) criterion compared to the other criteria** on a scale from 1 to 9?

4. How important are **all other criteria compared to the least important (worst) criterion** on a scale from 1 to 9?

Now an example is provided on how to fill in each of these questions.

Example: Lets say we have four criteria A, B, C and D.

1. If you think that "**A**" is the **most important (best) criterion**, choose criterion "A" in the first question.

2. If you think that "**D**" is the **least important (worst) criterion**, choose criterion "D", in the second question.

3. Then compare the most important criterion (best) to all other criteria.
In order to do this, the following table will appear:

	B	C	D
A	-- Please Select -- ▾	-- Please Select -- ▾	-- Please Select -- ▾

Answer the third question as such:

- first ask yourself how important the best criterion "**A**" is compared to B, and assign a score from 1-9.
- then ask yourself how important the best criterion "**A**" is compared to C, and assign a score from 1-9.
- at last, ask yourself how important the best criterion "**A**" is compared to D, and assign a score from 1-9.

4. After this, compare the other criteria to the least important criterion (worst).
In order to do this, the following table will appear:

	D
B	-- Please Select -- ▾
C	-- Please Select -- ▾

Answer the fourth question as such:

- first ask yourself how important criterion B is compared to the worst criterion "**D**", and assign a score from 1-9.
- then ask yourself how important criterion C is compared to the worst criterion "**D**", and assign a score from 1-9.

PART 2. MAIN-CRITERIA

In the underlying table, three main criteria are described:

#	Main-criteria	Description
1	Quality of provided information	How complete, truthful and accurate each technology can provided necessary clothing attribute information such as material, colour, fit & size and style, to determine the personal match/fit.
2	Information gathering and handling	The way your personal information is collected by the technology and processed by the online clothing retailer.
3	User-friendliness	The extent to which the technology is easy to use.

1. Out of these criteria, which one do you perceive as most important (best) for determining your preference of using a technology?
2. Out of these criteria, which one do you perceive as least important (worst) for determining your preference of using a technology?
3. How important is the most important criterion compared to the other criteria on a scale from 1 to 9?
4. How important are the other criteria compared to the least important criterion on a scale from 1 to 9?

PART 3. QUALITY OF PROVIDED INFORMATION

In the underlying table, four clothing attribute related sub-criteria are described, which are relevant for a technology to provide such that customers can make a well-informed online purchase decision and realize a successful online purchase:

	Sub-criteria	Description
Quality of provided information	Reliability of material information	Refers to the accuracy, completeness and truthfulness of the **information provided on clothing material** such as: material thickness, stretch-ability, texture and stitching (sewing).
	Reliability of colour information	Refers to the accuracy, completeness and truthfulness of the **information provided about the colour** of clothing.
	Reliability of fit & size information	Refers to the accuracy, completeness and truthfulness of the **information provided about the fit & size** of clothing.
	Reliability of style information	Refers to the accuracy, completeness and truthfulness of the **information provided about the style** of clothing.

1. Out of these criteria, which one do you perceive as most important (best) for determining your preference of using a technology?
2. Out of these criteria, which one do you perceive as least important (worst) for determining your preference of using a technology?
3. How important is the most important criterion compared to the other criteria on a scale from 1 to 9?
4. How important are the other criteria compared to the least important criterion on a scale from 1 to 9?

PART 4. INFORMATION GATHERING & HANDLING

In the table below, two sub-criteria are described which are characteristics of information gathering and handling:

	Sub-criteria	Description
Information gathering & handling	The way of data collection through technology	The way in which the technology acquires your information (for example through scanning, facial recognition or manually inserting body-measurement information) and how this affects your preference for using a technology.
	Data handling by online clothing retailer	The way the online clothing retailer uses and stores the collected information for its services.

1. Out of these criteria, which one do you perceive as most important (best) for determining your preference of using a technology?
2. Out of these criteria, which one do you perceive as least important (worst) for determining your preference of using a technology?
3. How important is the most important criterion compared to the other criterion on a scale from 1 to 9?

PART 5. UDER-FRIENDLINESS

in the underlying table, five sub-criteria are described which are characteristics of the perceived user-friendliness of a technology:

	Sub-criteria	Description
User-friendliness	Responsiveness	Loading time of technology.
	Search time	Navigation time (number of clicks) when using a technology during the online evaluation process. The evaluation process is when you have selected an item of clothing and are trying to determine your personal match / fit with regard to the apparel attributes such as material, colour, size and style.
	Availability	The ability to use the technology on any device.
	Attractiveness	The extent to which the technology is visually appealing and playful when used.
	Required preparatory work time	The amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit (such as measuring your own body, typing in or uploading body-measurements data or pictures or scanning the body).

1. Out of these criteria, which one do you perceive as most important (best) for determining your preference of using a technology?
2. Out of these criteria, which one do you perceive as least important (worst) for determining your preference of using a technology?
3. How important is the most important criterion compared to the other criteria on a scale from 1 to 9?
4. How important are the other criteria compared to the least important criterion on a scale from 1 to 9?

You have gone through all 18 questions. Clicking on "Submit" will close the survey.

THANK YOU

Thank you for filling out the survey. Your input is greatly appreciated!

APPENDIX D: Survey results (based on BWM users' survey)

1. Main set

The three main-criteria (categories) are:

Main-criteria	Weight
C1. Quality of provided information	0.441
C2. User-friendliness	0.324
C3. Information gathering & handling	0.235

In the survey, the respondents were first asked to determine the best and worst criteria. The main-criterion that was chosen the most amount of times as best criterion was 'quality of provided information'. 72 of the 113 respondents (64%) considered this criterion as the best, while for 'user-friendliness' and 'information gathering and handling' it was respectively 26 (23%) and 15 (13%) out of 113 respondents.

The main-criterion which was chosen the most amount of times as worst criterion was 'information gathering and handling'. 70 of the 113 respondents (62%) considered this criterion the as worst, while for 'user-friendliness' and 'information gathering and handling' it was respectively 27 (24%) and 16 (14%) out of 113 respondents.

2. Subset 1

The four sub-criteria categorized under 'quality of provided information' and which are used to evaluate the customers preference regarding the various technological alternatives are:

Sub-criteria	Local Weight	Ranking within category
c1.1. Reliability of material information	0.242	3
c1.2. Reliability of colour information	0.248	2
c1.3. Reliability of fit & size information	0.318	1
c1.4. Reliability of style information	0.192	4

In the survey, the respondents were first asked to determine the best and worst sub-criteria belonging to the main-criterion 'quality of provided information'. The sub-criterion that was chosen the most amount of times as best criterion based upon the respondents, was 'reliability of fit & size information'. 52 of the 77 respondents (68%) considered this criterion the a best, while for 'reliability of material information', 'reliability of colour information' and 'reliability of style information' it was respectively 16 (21%), 8 (10%) and 1 (1%) out of 77 respondents.

The sub-criteria which was chosen the most amount of times as worst criterion was 'reliability of style information'. 29 of the 77 respondents (38%) considered this criterion a worst, while for 'reliability of material information', 'reliability of colour information' and 'reliability of fit & size information' it was respectively 22 (29%), 14 (18%) and 12 (16%) out of 77 respondents.

3. Subset 2

The two sub-criteria categorized under 'information gathering and handling, and which are used to evaluate the customers preference regarding the various technological alternatives are:

Sub-criteria	Local Weight	Ranking within category
c2.1.The way of data collection through technology	0.432	2
c2.2. Data handling by online apparel retailer	0.568	1

In the survey, the respondents were first asked to determine the best and worst sub-criterion. From the two sub-criteria, the sub-criterion ‘information handling by online clothing retailer’ was perceived best criterion. 63 (56%) of the total amount of respondents (n=113) chose this criterion as best criterion for determining their preference for a technology. On the contrary, 50 (44%) of the respondents considered this criterion s worst criterion.

4. Subset 3

The five sub-criteria categorized under ‘user-friendliness’, and which are used to evaluate the customers preference regarding the various technological alternatives are:

Sub-criteria	Local Weight	Ranking within category
c3.1. Responsiveness	0.207	2
c3.2. Search time	0.223	1
c3.3. Availability	0.190	4
c3.4. Attractiveness	0.179	5
c3.5. Required preparatory work time	0.201	3

In the survey, the respondents were first asked to determine the best and worst sub-criteria belonging to the main-criterion ‘user-friendliness’. The sub-criterion that was chosen the most amount of times as best criterion was ‘availability’. 19 of the 73 respondents (26%) considered this criterion the most important, followed by ‘search time’, ‘required preparatory work time’, ‘attractiveness’ and ‘responsiveness’ which was respectively 16 (22%), 13 (18%), 13 (18%) and 12 (16%) out of 73 respondents

The sub-criteria which was chosen the most amount of times as worst criterion was ‘attractiveness’. 26 of the 73 respondents (36%) considered this criterion the least important, followed by ‘required preparatory work time’, ‘availability’, ‘search time’ and ‘responsiveness’ which was respectively 16 (22%), 15 (21%), 10 (14%) and 6 (8%) out of 73 respondents.

APPENDIX E: Interview with apparel e-commerce experts

Master thesis: Returns Management online clothing purchases
Study: Master Complex Systems Engineering & Management, TU Delft

Interview protocol for scoring alternatives

This interview is conducted om behalf of my master thesis at TU Delft. The results of this semi-structured interview will be used to formulate recommendations about the adoption of information technologies based on the customers preference. The obtained information through this interview will only be used by the researcher. Furthermore, the obtained information will solely be used for this specific research. In the report, sensitive data obtained through the interviews will be decontextualized and anonymized. A transcribed version of the interviews can be shared if requested by the apparel retail experts.

Thank you in advance for your participation.

Structure of the interview

First the case will be introduced along with the interview goal. Secondly, the solutions (alternatives) will be explained. Thirdly, the criteria will be explained which will be used to evaluate the alternatives.

The case

If customers want to by clothing online with the intention to keep it, they need to be able to evaluate the product accurately.

However, as the product is displayed virtually (on a website), the product can 1) not be felt to determine the material quality, 2) the colour cannot be seen in real life, 3) the clothing cannot be tried on to determine the right size and fit and 4) the personal match with regard to style of the product cannot be tested in real life. Due to these reasons, many online customers experience a pre-purchase uncertainty. Most of the return reasons for online purchased clothing items stem from customers not being able to determine 1) the material quality, 2) actual colour, 3) the fit and size and 4) the personal fit with regard to style.

As a result of this, means have been developed to assist customers during the online purchasing process, such that they can evaluate these clothing attributes more accurately in order to increase customers' clothing purchase successes and reduce unnecessary order returns.

Interview goal

The goal of this interview is to obtain performance scores of each designed alternatives with regard to the selected criteria, by asking the following questions:

1. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to addressing a particular criterion and which one is the worst?

	Best Alternative	Worst Alternative
Criterion (X)		

2. How much better is the best alternative compared to all other alternatives with regard to this particular criterion, on a scale from 1 to 9?

Best compared to Others	**A1:** Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **static** height/size chart	**A2:** Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing rooms (augmented reality such as snapchat filters)
Best Alt.:				

3. How much better are all other alternatives compared to the worst alternative with regard to this particular criterion, on a scale from 1 to 9?

Others to Worst	Worst Alt.:
A1	
A2	
A3	
A4	

The provided alternatives are:

	Alternatives	Description
A1	Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **static** height/size chart	In this alternative, the size chart is provided by a picture of a body-measurements table. Customers have to measure their own body measurements first with a measuring band in order to know in which size category they belong to. Different pictures can be mixed-and matched in order to determine the overall outfit and the possibility to zoom in and out on pictures exists to examine the aesthetic clothing attributes (material/colour/style) more closely.
A2	Alternative Product pictures with mix and match function (to see overall outfit), zoom function and **interactive** height/size chart	This solution differs from the previous one, in the sense that it uses an interactive size chart where customers need to type in their body measurements. Based on the typed in body-measurements and on behalf of the collected data from other clients, the recommended fit and size is provided.
A3	Avatars (digital computer-based twin)	An avatar is a computerized digital twin, to create a virtual try-on experience for customers. With avatars, clothing items can be virtually tried on in a computer-generated environment. Body-measurement data can be inserted manually by customers or via pictures or body scanning technology. Different clothing items can be mixed-and matched in order to determine the overall outfit and the possibility to zoom in and out exists. Is also provides personalized size recommendations based on fit. Through 'pressure maps' the personal fit and size information is indicated. The clothing material dynamic flow (fit) on the avatar is mimicked using draping technology.
A4	Virtual Dressing rooms (augmented reality such as snapchat filters)	A VDR differs from an Avatar, as people now see themselves (it is not computer based). A VDR is also

		a state of the art technology which can be used to create a virtual try-on experience for customers. It uses e.g. body scanning to determine the body shape and augmented reality (such as Snapchat filters) by which the clothing can be virtually 'swipedragged' on the body of customers and mixed-and matched in order to determine the overall fit of outfits. The possibility to zoom in and out exists to examine the aesthetic clothing attributes (material/colour/style) more closely. It also has the ability to provide personalized size recommendations.

Interview topics

Topic 1: Background information

Date:
Name:
Function:
Expertise:
Years of experience:

Topic 2: Scoring the alternatives based on the following four sub-criteria

	Sub-criteria	Description
Quality of provided information	Reliability of material information	Refers to the accuracy, completeness and truthfulness of the **information provided on clothing material** such as: material thickness, stretch-ability, texture and stitching (sewing).
	Reliability of colour information	Refers to the accuracy, completeness and truthfulness of the **information provided about the colour** of clothing.
	Reliability of fit & size information	Refers to the accuracy, completeness and truthfulness of the **information provided about the fit & size** of clothing.
	Reliability of style information	Refers to the accuracy, completeness and truthfulness of the **information provided about the style** of clothing.

Topic 3: Scoring the alternatives based on the following two sub-criteria

	Sub-criteria	Description
Information gathering & handling	The way of data collection through technology	The way in which the technology acquires your information (for example through scanning, facial recognition or manually inserting body-measurement information) and how this affects your preference for using a technology.
	Data handling by online clothing retailer	The way the online clothing retailer uses and stores the collected information for its services.

Topic 4: Scoring the alternatives based on the following five sub-criteria

	Sub-criteria	Description
	Responsiveness	Loading time of technology.

User-friendliness	Search time	Navigation time (number of clicks) when using a technology during the online evaluation process. The evaluation process is when you have selected an item of clothing and are trying to determine your personal match / fit with regard to the clothing attributes such as material, colour, size and style.
	Availability	The ability to use the technology on any device.
	Attractiveness	The extent to which the technology is visually appealing and playful when used.
	Required preparatory work time	The amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit (such as measuring your own body, typing in or uploading body-measurements data or pictures or scanning the body).

Scorecard:

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

Topic 5: Technology adoption in company

Scoring the alternatives based on the **employment/implementation possibility** in the company, using the same scorecard.

Topic 1

Date:
Name:
Function:
Expertise:
Years of experience:

Topic 2: Quality of provided information

In the table below, four sub-criteria with corresponding description are indicated, which are relevant such that customers can make a well-informed online purchase decision:

	Sub-criteria	Description
Quality of provided information	Reliability of material information	Refers to the accuracy, completeness and truthfulness of the **information provided on clothing material** such as: material thickness, stretch-ability, texture and stitching (sewing).
	Reliability of colour information	Refers to the accuracy, completeness and truthfulness of the **information provided about the colour** of clothing.
	Reliability of fit & size information	Refers to the accuracy, completeness and truthfulness of the **information provided about the fit & size** of clothing.
	Reliability of style information	Refers to the accuracy, completeness and truthfulness of the **information provided about the style** of clothing.

1. Out of all alternatives (A1, A2, A3, A4), which one is the best in providing **reliable material information** such that customer can evaluate clothing accurately and which one is the worst?

	Best Alternative	Worst Alternative
Material information reliability		

2. How much better is the best alternative compared to all other alternatives in providing **reliable material information** to customers, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

3. How much better are all other alternatives compared to the worst alternative, in providing **reliable material information** to customers, on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

4. Out of all alternatives (A1, A2, A3, A4), which one is the best in providing **reliable colour information** such that customer can evaluate clothing accurately and which one is the worst?

	Best Alternative	Worst Alternative
Colour information reliability		

5. How much better is the best alternative compared to all other alternatives in providing **reliable colour information** to customers, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

6. How much better are all other alternatives compared to the worst alternative, in providing **reliable colour information** to customers, on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

7. Out of all alternatives (A1, A2, A3, A4), which one is the best in providing **reliable fit & size information** such that customer can evaluate clothing accurately and which one is the worst?

	Best Alternative	Worst Alternative
Fit & size information reliability		

8. How much better is the best alternative compared to all other alternatives in providing **reliable fit & size information** to customers, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

9. How much better are all other alternatives compared to the worst alternative, in providing **reliable fit & size information** to customers, on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

10. Out of all alternatives (A1, A2, A3, A4), which one is the best in providing **reliable style information** such that customer can evaluate clothing accurately and which one is the worst?

	Best Alternative	Worst Alternative

Style information reliability		

11. How much better is the best alternative compared to all other alternatives in providing **reliable style information** to customers, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

12. How much better are all other alternatives compared to the worst alternative, in providing **reliable style information** to customers, using a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

Topic 3: Information gathering & handling

In the table below, two sub-criteria with corresponding description are indicated, which are characteristics of information gathering and handling:

	Sub-criteria	Description
Information gathering & handling	The way of data collection through technology	The way in which the technology acquires your information (for example through scanning, facial recognition or manually inserting body-measurement information) and how this affects your preference for using a technology.
	Data handling by online clothing retailer	The way the online clothing retailer uses and stores the collected information for its services.

1. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**The way of data collection through technology**" and which one is the worst?

	Best Alternative	Worst Alternative
The way of data collection through technology		

2. How much better is the best alternative compared to all other alternatives with regard to **"The way of data collection through technology**", on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

3. How much better are all other alternatives compared to the worst alternative, with regard to **"The way of data collection through technology**", on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

4. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**Data handling and use by retailers**" and which one is the worst?

	Best Alternative	Worst Alternative
Data handling and use by retailers		

5. How much better is the best alternative compared to all other alternatives with regard to **"Data handling and use by retailers**", on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9: 1: **Equally** good 2: Somewhat between Equally and Moderately 3: **Moderately** better than 4: Somewhat between Moderate and Strong 5: **Strongly** better than 6: Somewhat between Strong and Very strong 7: **Very strongly** better than 8: Somewhat between Very strong and Absolute 9: **Absolutely** better than

6. How much better are all other alternatives compared to the worst alternative with regard to **"Data handling and use by retailers"**, on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

Topic 4: User friendliness characteristics of technology

In the table below, five sub-criteria with corresponding description are indicated, which are characteristics of the perceived user friendliness of a technology:

	Sub-criteria	Description
User-friendliness	Responsiveness	Loading time of technology.
	Search time	Navigation time (number of clicks) when using a technology during the online evaluation process. The evaluation process is when you have selected an item of clothing and are trying to determine your personal match / fit with regard to the clothing attributes such as material, colour, size and style.
	Availability	The ability to use the technology on any device.
	Attractiveness	The extent to which the technology is visually appealing and playful when used.
	Required preparatory work time	The amount of work customers need to do upfront before they can start using a technology to evaluate the overall personal clothing match / fit (such as measuring your own body, typing in or uploading body-measurements data or pictures or scanning the body).

1. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**Responsiveness**" and which one is the worst?

	Best Alternative	Worst Alternative
Responsiveness		

2. How much better is the best alternative compared to all other alternatives with regard to **"Responsiveness"**, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

3. How much better are all other alternatives compared to the worst alternative with regard to **"Responsiveness**", on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

4. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**Search time**" and which one is the worst?

	Best Alternative	Worst Alternative
Search time		

5. How much better is the best alternative compared to all other alternatives with regard to **"Search time**", on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

6. How much better are all other alternatives compared to the worst alternative with regard to **"Search time**", on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

7. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**Availability"** and which one is the worst?

	Best Alternative	Worst Alternative
Availability		

8. How much better is the best alternative compared to all other alternatives with regard to **"Availability**", on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

9. How much better are all other alternatives compared to the worst alternative with regard to "**Availability**", on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

10. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**attractiveness**" and which one is the worst?

	Best Alternative	Worst Alternative
Perceived attractiveness		

11. How much better is the best alternative compared to all other alternatives with regard to " **attractiveness**", on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

12. How worst are all other alternatives compared to the worst alternative with regard to "**attractiveness**", on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

13. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to "**Required preparatory work time**" and which one is the worst?

	Best Alternative	Worst Alternative
Required preparatory work time		

14. How much better is the best alternative compared to all other alternatives with regard to **“Required preparatory work time**”, on a scale from 1 to 9?

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

15. How much better are all other alternatives compared to the worst alternative with regard to **“Required preparatory work time**”, on a scale from 1-9?

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

Topic 5: Technology adoption in company

Scoring the alternatives based on the **embedment/implementation in the company**, using the same scorecard.

1. What is the return rate (in %) of clothing purchased online by customers with respect to clothing attributes such as material quality, fit & size, colour and style?

2. Which alternative(s) does the company already use?

3. Out of all alternatives (A1, A2, A3, A4), which one is the best when it comes to “**employment/implementation in the company**” and which one is the worst? Please explain your answer.

	Best Alternative	Worst Alternative
Implementation in company		

4. How much better is the best alternative compared to all other alternatives with regard to "**implementation in the company**", on a scale from 1 to 9? Please explain your answer.

Best to Others	**A1:** Alternative product pictures with mix-and-match function, zoom function and **static** height/size chart	**A2:** Alternative product pictures with mix-and-match function, zoom function and **interactive** height/size chart	**A3:** Avatars (digital computer-based twin)	**A4:** Virtual Dressing Rooms (augmented reality such as snapchat filters)
Best Alt.:				

The meaning of the numbers 1-9:
1: **Equally** good
2: Somewhat between Equally and Moderately
3: **Moderately** better than
4: Somewhat between Moderate and Strong
5: **Strongly** better than
6: Somewhat between Strong and Very strong
7: **Very strongly** better than
8: Somewhat between Very strong and Absolute
9: **Absolutely** better than

5. How much better are all other alternatives compared to the worst alternative with regard to **"implementation in the company**", on a scale from 1-9? Please explain your answer

Others to worst	Worst Alt.:
A1	
A2	
A3	
A4	

APPENDIX F: Input data retrieved from BWM Interview

The pairwise comparison data is retrieved from interviews with 6 apparel e-commerce experts.

Expert ID	Best & worst **Alternative** regarding ***Reliability of material information***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A2	B_O	3	1	6	5
	Worst	A3	O_W	3	6	1	2
Expert 2	Best	A4	B_O	5	3	1	1
	Worst	A1	O_W	1	3	5	5
Expert 3	Best	A4	B_O	3	3	2	1
	Worst	A1	O_W	1	1	3	3
Expert 4	Best	A4	B_O	7	5	3	1
	Worst	A1	O_W	1	3	7	7
Expert 5	Best	A4	B_O	6	6	2	1
	Worst	A1	O_W	1	1	6	6
Expert 6	Best	A4	B_O	4	3	2	1
	Worst	A1	O_W	1	3	4	4

Expert ID	Best & worst **Alternative** regarding ***Reliability of colour information***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A2	B_O	1	1	4	3
	Worst	A3	O_W	5	4	1	5
Expert 2	Best	A4	B_O	5	5	1	1
	Worst	A1	O_W	1	2	5	5
Expert 3	Best	A4	B_O	1	1	1	1
	Worst	A1	O_W	1	1	1	1
Expert 4	Best	A1	B_O	1	1	7	5
	Worst	A3	O_W	7	7	1	3
Expert 5	Best	A4	B_O	5	5	4	1
	Worst	A1	O_W	1	1	5	5
Expert 6	Best	A4	B_O	7	7	3	1
	Worst	A1	O_W	1	1	5	7

Expert ID	Best & worst **Alternative** regarding ***Reliability of fit & size information***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A4	B_O	7	3	3	1
	Worst	A1	O_W	1	7	5	7
Expert 2	Best	A3	B_O	9	9	1	4
	Worst	A1	O_W	1	3	9	7
Expert 3	Best	A4	B_O	5	3	2	1
	Worst	A1	O_W	1	4	5	5
Expert 4	Best	A3	B_O	9	7	1	3
	Worst	A1	O_W	1	3	9	7

Expert 5	Best	A4	B_O	8	4	3	1
	Worst	A1	O_W	1	6	7	8
Expert 6	Best	A4	B_O	7	5	1	1
	Worst	A1	O_W	1	5	7	7

Expert ID	Best & worst **Alternative** regarding ***Reliability of style information***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A4	B_O	6	5	1	1
	Worst	A1	O_W	1	3	7	6
Expert 2	Best	A3	B_O	5	4	1	3
	Worst	A1	O_W	1	3	5	5
Expert 3	Best	A4	B_O	5	5	3	1
	Worst	A1	O_W	1	1	5	5
Expert 4	Best	A4	B_O	5	4	2	1
	Worst	A1	O_W	1	3	5	5
Expert 5	Best	A4	B_O	5	5	3	1
	Worst	A1	O_W	1	1	6	5
Expert 6	Best	A4	B_O	7	5	2	1
	Worst	A1	O_W	1	3	5	7

Expert ID	Best & worst **Alternative** regarding ***The way of data collection through technology***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A2	B_O	4	1	4	5
	Worst	A4	O_W	5	5	2	1
Expert 2	Best	A2	B_O	3	1	5	4
	Worst	A3	O_W	3	5	1	2
Expert 3	Best	A1	B_O	1	3	7	9
	Worst	A4	O_W	9	5	3	1
Expert 4	Best	A1	B_O	1	3	7	9
	Worst	A4	O_W	9	5	3	1
Expert 5	Best	A1	B_O	1	4	5	8
	Worst	A4	O_W	8	6	3	1
Expert 6	Best	A1	B_O	1	3	6	7
	Worst	A4	O_W	7	5	2	1

Expert ID	Best & worst **Alternative** regarding ***Data handling by online clothing retailer***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A1	B_O	1	5	9	9
	Worst	A4	O_W	9	5	2	1
Expert 2	Best	A1	B_O	1	2	9	9
	Worst	A3	O_W	9	7	1	2
Expert 3	Best	A1	B_O	1	5	7	9
	Worst	A4	O_W	9	7	3	1
Expert 4	Best	A1	B_O	1	5	7	9

	Worst	A4	O_W	9	5	2	1
Expert 5	Best	A1	B_O	1	3	5	8
	Worst	A4	O_W	8	6	4	1
Expert 6	Best	A1	B_O	1	3	6	7
	Worst	A4	O_W	7	5	2	1

Expert ID	Best & worst **Alternative** regarding ***Responsiveness***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A1	B_O	1	5	9	8
	Worst	A3	O_W	9	7	1	2
Expert 2	Best	A1	B_O	1	1	7	5
	Worst	A3	O_W	7	7	1	3
Expert 3	Best	A2	B_O	1	1	7	5
	Worst	A3	O_W	7	7	1	5
Expert 4	Best	A1	B_O	1	3	7	7
	Worst	A4	O_W	7	5	2	1
Expert 5	Best	A1	B_O	1	3	7	6
	Worst	A3	O_W	7	7	1	3
Expert 6	Best	A1	B_O	1	2	5	7
	Worst	A4	O_W	7	6	3	1

Expert ID	Best & worst **Alternative** regarding ***Search time***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A1	B_O	1	5	9	8
	Worst	A3	O_W	9	5	1	3
Expert 2	Best	A1	B_O	1	1	7	5
	Worst	A3	O_W	7	7	1	3
Expert 3	Best	A2	B_O	3	1	1	1
	Worst	A1	O_W	1	3	3	3
Expert 4	Best	A4	B_O	7	3	2	1
	Worst	A1	O_W	1	4	9	7
Expert 5	Best	A1	B_O	1	2	7	4
	Worst	A3	O_W	7	5	1	5
Expert 6	Best	A2	B_O	3	1	6	6
	Worst	A3	O_W	4	6	1	2

Expert ID	Best & worst **Alternative** regarding ***Availability***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A1	B_O	1	5	9	8
	Worst	A3	O_W	9	6	1	1
Expert 2	Best	A1	B_O	1	1	9	5
	Worst	A3	O_W	9	9	1	4
Expert 3	Best	A1	B_O	1	3	8	7
	Worst	A3	O_W	8	5	1	3
Expert 4	Best	A1	B_O	1	2	7	7
	Worst	A4	O_W	7	5	3	1

Expert 5	Best	A1	B_O	1	3	8	5
	Worst	A3	O_W	8	5	1	3
Expert 6	Best	A1	B_O	1	3	6	6
	Worst	A4	O_W	6	5	2	1

Expert ID	Best & worst **Alternative** regarding ***Attractiveness***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A4	B_O	7	5	5	1
	Worst	A1	O_W	1	2	3	7
Expert 2	Best	A4	B_O	7	7	2	1
	Worst	A1	O_W	1	1	7	7
Expert 3	Best	A3	B_O	7	7	1	3
	Worst	A1	O_W	1	5	7	6
Expert 4	Best	A4	B_O	7	5	2	1
	Worst	A1	O_W	1	3	7	7
Expert 5	Best	A4	B_O	7	5	4	1
	Worst	A1	O_W	1	4	7	7
Expert 6	Best	A4	B_O	7	7	4	1
	Worst	A1	O_W	1	2	4	7

Expert ID	Best & worst **Alternative** regarding ***Required preparatory work time***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A2	B_O	5	1	8	7
	Worst	**A3**	O_W	7	8	1	3
Expert 2	Best	A1	B_O	1	3	9	7
	Worst	**A3**	O_W	9	8	1	4
Expert 3	Best	A2	B_O	5	1	3	3
	Worst	A1	O_W	1	5	3	3
Expert 4	Best	A1	B_O	1	5	9	9
	Worst	**A3**	O_W	9	5	1	3
Expert 5	Best	A1	B_O	1	4	7	5
	Worst	**A3**	O_W	7	5	1	3
Expert 6	Best	A2	B_O	5	1	3	4
	Worst	A1	O_W	1	5	5	4

Expert ID	Best & worst **Alternative** regarding ***Implementation in company***			**A1**	**A2**	**A3**	**A4**
Expert 1	Best	A2	B_O	2	1	7	8
	Worst	A4	O_W	9	8	2	1
Expert 2	Best	A2	B_O	1	1	9	5
	Worst	A3	O_W	9	9	1	3
Expert 3	Best	A1	B_O	1	3	6	7

	Worst	A4	O_W	7	6	2	1
Expert 4	Best	A1	B_O	1	5	8	7
	Worst	A3	O_W	8	6	1	3
Expert 5	Best	A2	B_O	2	1	5	5
	Worst	A3	O_W	8	5	1	1
Expert 6	Best	A1	B_O	1	3	6	7
	Worst	A4	O_W	7	5	2	1

APPENDIX G: Calculated scores per alternative with respect to each sub-criterion

The alternatives' scores are obtained through an interview with 6 apparel e-commerce experts.

Table G.1: experts' scores (i.e. weights) per sub-criterion

	Sub-criteria	Scores			
		A1	A2	A3	A4
Subset 1	Reliability of material information	0.140	0.204	0.299	0.358
	Reliability of colour information	0.217	0.217	0.228	0.338
	Reliability of fit & size information	0.093	0.186	0.362	0.359
	Reliability of style information	0.106	0.144	0.359	0.392
Subset 2	The way of data collection through technology	0.424	0.325	0.143	0.108
	Data handling by online apparel retailer	0.503	0.278	0.123	0.096
Subset 3	Responsiveness	0.438	0.339	0.101	0.122
	Search time	0.274	0.332	0.171	0.223
	Availability	0.475	0.307	0.096	0.122
	Attractiveness	0.104	0.144	0.316	0.436
	Required preparatory work time	0.304	0.361	0.151	0.185

Subset 1

Underlying figure provides an overview of the scores of each alternative with respect to each sub-criterion belonging to the main-criterion 'quality of provided information', which are reliability of: 1) material information, 2) colour information, 3) fit & size information and 4) style information. Note, since the BWM is also used to obtain the scores, the scores are also presented as weights.

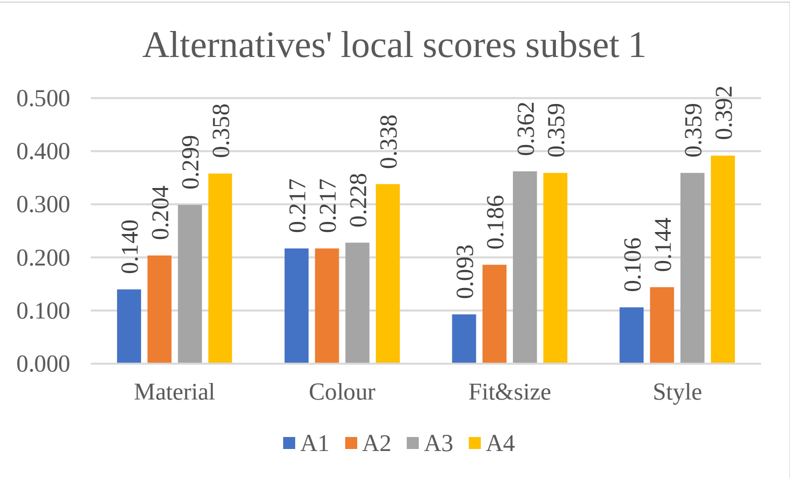

Figure G.1: alternatives' scores with respect to each sub-criterion of subset 1

- Sub-criterion 1.1: reliability of material information

Based upon the scores as indicated in figure G.1 and the credal ranking of underlying figure G.2, the alternatives can be ranked as follows with respect to the criterion ‘reliability of material information’: A4>A3>A2>A1. The nodes in the graph are the alternatives and each edge indicates the preference of the alternatives compared to one another with respect to each specific sub-criterion.

Based upon the obtained confidence levels as indicated in figure G.2, A4 is certainly more reliable than A1 and A2. Compared to A3, A4 is more reliable when it comes to providing material information, with an assigend confidence level of 0.74.

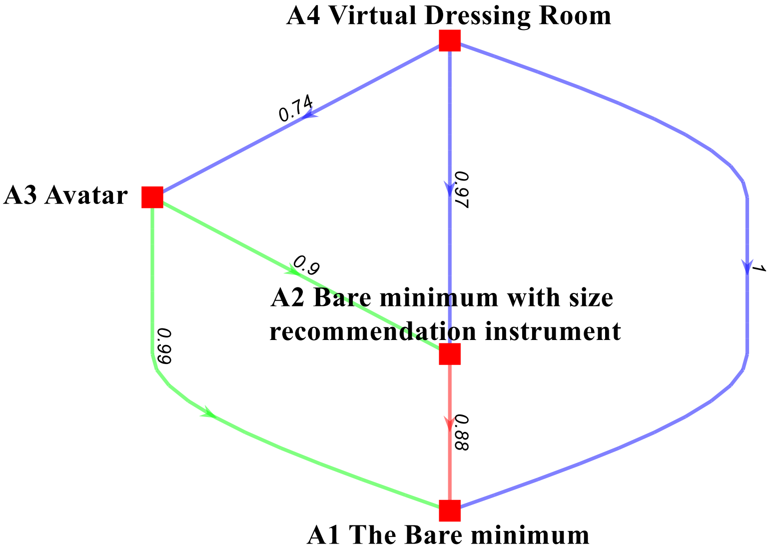

Figure G.2: credal ranking of alternatives regarding reliability of material information

The results show that A4 scores the best with respect to reliability of material, colour, fit & size and style information, followed by A2 as second best, A3 as third best and A1 as least reliable alternative with regard to providing material information. According to five out of six interviewees (every expert except for expert 1) (83%), A4 is perceived as the most reliable when it comes to providing material information, since this alternative can create a better feel of material quality and a better perception of the material fit on the body shape, due to its ability to provide dynamic movement of apparel aligned with the movement of the individuals own body.

Most of these five experts assigend similar values to A3, in the sense that they indicated that A3 is equally good in providing material information, or between equally or moderately, or moderately better than A4 in providing reliable material information (as can be seen in Appendix F). According to these experts, compared to A4, A3 can provide a less real life perception and feel of material since it is computer-based. Furthermore, the material draping technology which replicates the material physical dynamic for avatars is still perceived as difficult according to expert 1. As a result, less reliable information can be provided compared to A4 with respect to material information.

According to expert 2 till 6 (83%), A1 is the least preferred alternative, since people are less able to judge the material quality from a static photo.

These five experts assigned similar values to A2, in the sense that they indicated that A2 is equally good in providing material information or moderately better than A1 in providing reliable material information (as can be seen in Appendix F). This can be justified, since A2 differs from A1 only with regard to the fit & size recommendation application.

- Sub-criterion 1.2: reliability of colour information

Based upon the scores as indicated in figure G.1 and the credal ranking of underlying figure G.3, the alternatives can be ranked as follows with respect to the criterion 'reliability of colour information': A4>A3>A2>A1.

Based upon the obtained confidence levels as indicated in underlying figure G.3, A4 is certainly more reliable in providing colour information compared to all the other three alternatives.

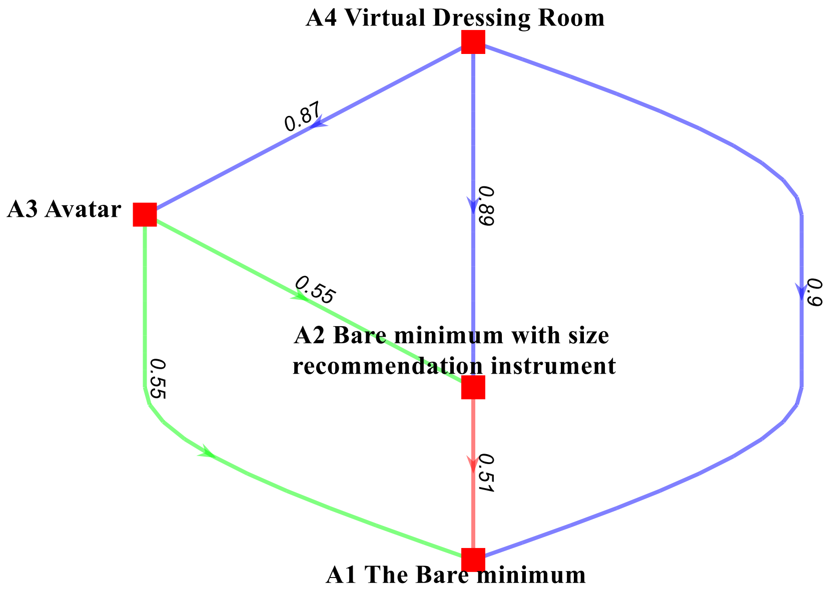

Figure G.3: credal ranking of alternatives regarding reliability of colour information

The results also show that A4 is perceived as the most reliable alternative when it comes to providing colour information and A1 as least reliable. According to four out of six experts (expert 2, 3, 5 and 6) (67%) A4 is chosen as the most reliable alternative, since users can get a better perception of the personal colour match once apparel is tried on their own body than when seen worn on a retail provided fashion model. This allows them to better evaluate the personal colour match before purchasing the apparel items.

According to four out of six experts (67%) (expert 2, 3, 5 and 6), A1 is the least preferred alternative, since the colour must be derived purely from the photographs compared to e.g. A3 and A4. According to expert 5, A1 is the worst, because there are still some apparel returns as a result of colour, so it seems that creating the right colour expectation based on solely the pictures is not really easy for customers. A2 and A1 score the same, as they do not differ based on the provision of colour information.

According to expert 1, A2 is the best and according to expert 4, A1 is the best. Since A1 and A2 only differ in terms of a fit & size recommendation application, they are the same when it comes to providing colour information (by the provided text and through pictures of apparel items). The reason why A2 was chosen as best by expert 1 is because photos can be fine-tuned very well. A1 and A2 are the same when it comes to providing colour information, because they do not differ in functionality. But expert 1 always choses A2 over the others, due to the fit & size recommendation function which the company has just recently employed.

The reason why A1 was chosen as best by expert 4 is, because nothing has happened to the images yet (the images are not yet edited) compared to A3 and A4. With A3 and A4, you can see how apparel items fall and fit on you, but if you look closely at a picture it can be more reliable than if something happens to it (gets edited). Since A1 and A2 do not differ with regard to providing colour information, both are equally good at providing colour information.

Both these experts (1 and 4) chose A3 as worst when it comes to providing reliable colour information as according to expert 4 the colour can be visualised and perceived as different since it is computer-based (animated). According to expert 4, A1 is very much better than A3 when it comes to providing colour information, since the strong/focussed viewing of a picture can provide more reliable information than when the pictures are edited to be viewed on a computer based animated Avatar.

According to expert 2, based on the current technological development, A3 is the least able to convey the colour as realistically as possible. A4 is strongly better than A3, because people see themselves and can virtually try-on apparel on themselves, and therefore can better evaluate the personal colour-match.

Sub-criterion 1.3: reliability of fit & size information

Based upon the local scores as indicated in figure I.1 and the credal ranking presented in underlying figure I.4, the alternatives can be ranked as follows with respect to the criterion 'reliability of fit & size information': A3>A4>A2>A1.

Based upon the obtained confidence levels presented in underlying figure G.4, A3 is certainly more reliable than A1 and A2 in providing fit & size information. However, compared to A4, the degree of certainty is less evident, since the confidence level is 0.52.

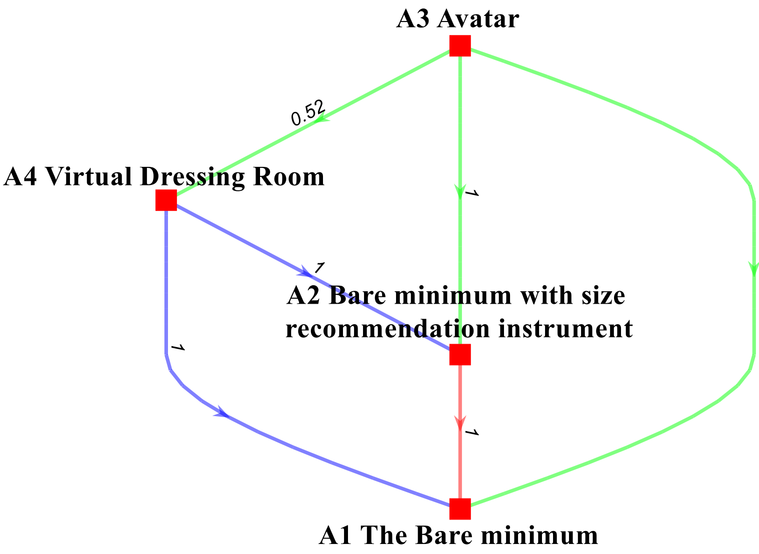

Figure G.4: credal ranking of alternatives regarding reliability of fit & size information

All six experts (100%) have indicated that A1 is the worst, since according to all experts customers most often make mistakes when measuring their own body sizes, thus fit and size wise A1 is the least reliable. The main reason mentioned why A1 is perceived as the least reliable, is because in practice it turns out that people cannot measure themselves well, which makes A1 very debatable. A2 is all about statistics, which is not always reliable either. But since we are talking about reducing returns, A1 is the least reliable.

Four out of six experts (67%) (expert 1, 3, 5 and 6) chose A4 as the alternative which as the best when it comes to providing reliable fit & size information. However the other two remaining experts chose A3 as the best alternative.

The experts who chose A4 as best alternative with respect to providing reliable of fit & size information indicated that the ability to see how the apparel items fit on the individuals own body shape provided a better perception and feel of fit. According to expert 5, everyone has a different physique, and through the ability to try-on apparel items on their mirrored selves can form a better perception regarding the personal fit & size. Furthermore, per brand, the size can be different (apparel can fall larger or smaller), so the visual aspect (the virtual fitting of apparel) has a lot of added value compared to the more static alternatives A1 and A2.

According to expert 2 and 4, A3 an A4 are perceived almost as equally best, but A3 scores better because according to these two experts A3 can still better represent the actual size and body shape more precisely compared to a VDR (A4).

The obtained scores as indicated in figure I1 show that A3 is perceived as the most reliable alternative when it comes to providing reliable fit & size information. This can be explained by looking at the assigned values in the comparison analysis (see Appendix F). Three out of the four experts whom have chosen A4 as best alternative, have indicated that A4 scores between equally and moderately better, or modality better compared to A3 with respect to the criterion 'reliability of fit & size information'. According to expert 3, A4 scores between evenly and reasonably better than A3, as A3 also offers a virtual try-on experience, but a less realistic try-on experience compared to A4.

One expert (expert 6) even indicated that A4 scores equally good compared to A3. According to this expert, A3 is equally good at providing fit & size information, as one can also make the avatar very realistic on the basis of their own body dimensions and thus receive the correct fitting information.

Consequently, all four experts have indicated that the difference between A4 and A3 with respect to providing reliable fit & size information is small. As a result, A3 has the highest assigend local weight (w^{agg} = 0.362) which compared to A4 (w^{agg} = 0.359) is not even that high of a difference.

- Sub-criterion 1.4: reliability of style information

Based upon the scores as indicated in figure G.1 and the credal ranking presented in underlying figure G.5, the alternatives can be ranked as follows with respect to the criterion 'reliability of style information': A4>A3>A2>A1.

Based upon the obtained confidence levels presented in underlying figure G.5, A4 is the certainly more reliable than A1 and A2 in providing style information. However compared to A3, the level of confidence is 0.65.

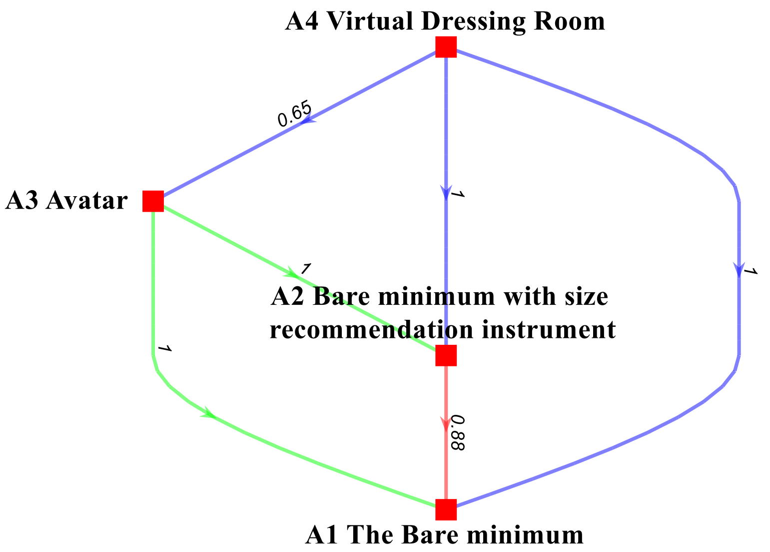

Figure G.5: credal ranking of alternatives regarding reliability of style information

Five out of six experts (except for expert 2) (83%) have indicated that A4 can provide the most reliable style information, since it provides the ability to test the outfit (use the mix-and-match function) on an individuals' mirrored self, which provides a better feel and perception regarding style and the personal match of the entire outfit.

Most of these five experts assigend similar values to A3, implying that A3 is equally good, between equally and moderately, or moderately better than A4 in providing reliable material information (as can be seen in Appendix F). The main argument for this was that with both alternatives (A3 and A4) users can evaluate their total outfit in terms of style by means of a personal try-on (virtual) experience. Based on this, people can make a better pre-purchase decision.

However A4 scores better, since A3 also offers a virtual try-on experience, but a less realistic try-on experience compared to A4 (expert 3 and 4, 5).

All experts (100%) perceived A1 as the least reliable option for providing style information, since in a photo the fit with regard to style is hard to see. According to expert 1, based on their return reasons, people do not understand a 'baggy fit', for example. According to expert 3, the quality of the photography also influences the reliability of the provided information. According to this expert, style information is less easily derived from photos then for example colour information, also since they are photos of fashion models.

All experts assigned similar values to A2, implying that A2 is equally good, between equally and moderately, or moderately better than A1 in providing reliable style information (as can be seen in Appendix F). According to expert 6, style information can also be extracted from the fit recommendation function of A2, as fit and style terminologies are often used not that distinctively as they should. As such, compared to A1, A2 is more reliable with regard to providing style information. According to the experts, compared to A2, A3 is between reasonably and much better in providing style information, because the personal match in style can be tested on an individual's own body double. Compared to A1, A3 scores much better, because from photos (of retail fashion models) customers are less able to judge the personal match in terms of style. From photos, style information can be traced well, but less well than when it is indicated on an own 'body double'.

Subset 2

Underlying figure G.6 provides an overview of the scores of each alternative with respect to each sub-criterion belonging to the main-criterion 'information gathering and handling, which are: 1) the way of data collection through technology and 2) data handling by online clothing retailers.

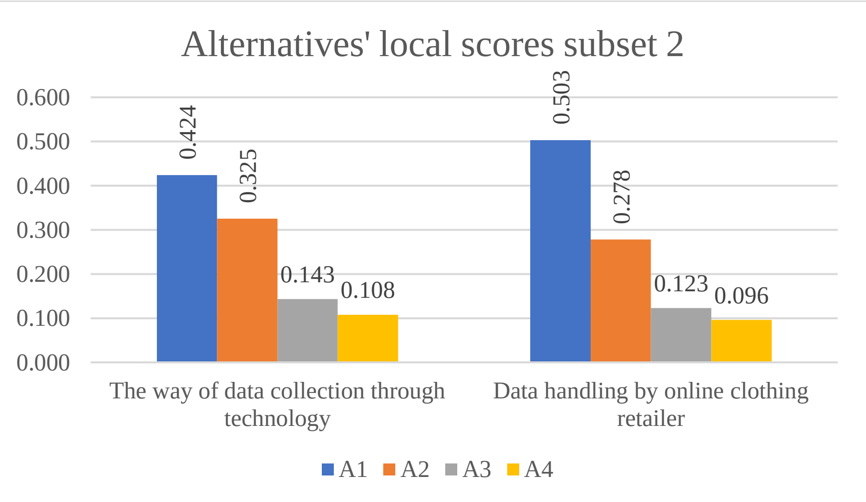

Figure G.6: alternatives' scores with respect to sub-criteria of subset 2

- Sub-criterion 2.1: the way of data collection through technology

Based upon the scores as indicated in figure G.6 and the credal ranking presented in underlying figure G.7, the alternatives can be ranked as follows with respect to the criterion ‘the way of data collection through technology’: A1>A2>A3>A4.

Based upon the obtained confidence levels as presented in underlying figure G.7, A1 certainly involves the least amount of privacy and security concerns compared to all other alternatives.

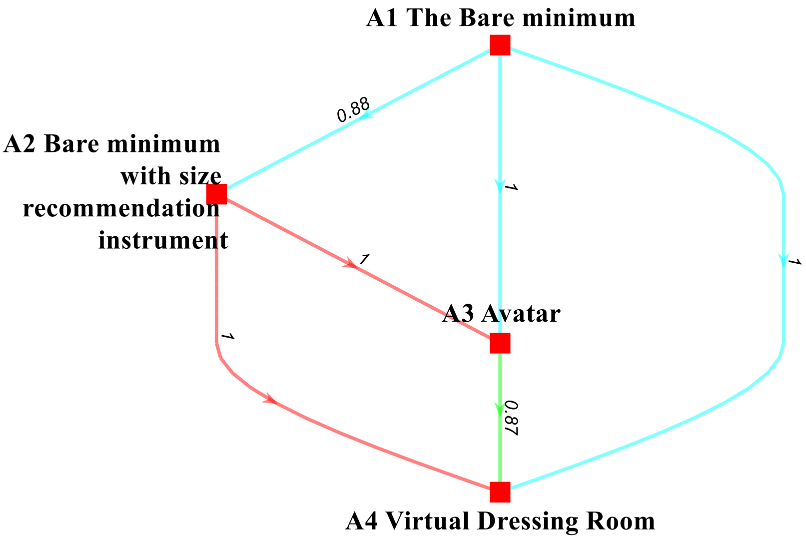

Figure G.7: credal ranking of alternatives regarding ‘the way of data collection through technology’

The results show that A1 scores the best with respect to the way of data collection through the technology and how this affects the customers' preference to use the technology. On the contrary, A4 scores the worst with respect to this criterion.

According to four out of six experts (67%) (expert 3 till 6), A1 scores the best, since customers are the least exposed as they do not have to share personal information such as body-measurements in order to use the instruments to evaluate the personal match with respect to the apparel attributes. On the contrary, five out of six experts (83%) (expert 1, 3, 4, 5 and 6) indicated that the most privacy concerns can arise when using A4, since customers are the most exposed as it uses facial and body scanning and recognition

Furthermore, most of these experts indicated that A3 and A4 also score quite evenly with respect to the criterion ‘the way of data collection through technology’, but that A3 is still reasonably better than A4. Since A3 is computer-based (provided a less realistic appearance of an individual) and more anonymous, people are less exposed and therefore less able to perceive privacy and security concerns.

According to expert 4, A2 is much better than A4, because with A2 data (body measurements data) is provided without an image of the person, whereas at A4 an image along with personal data is provided which is more privacy sensitive.

One expert (expert 2) mentioned that A3 is the worst, since people might be less familiar with A3. Since the Avatar is quite new and requires the creation of an avatar, people will be less inclined to switch to it compared to a VDR that mimics reality much better (by evaluating apparel in a 'digital' mirror) and instantly reproduces ones’ mirrored self-image. The expert did indicate that it was difficult to estimate the specific difference between A3 and A4 with respect to the perceived privacy and security concerns.

Experts 1 and 2 have indicated that A2 is the best and provide the same reason for this. According to the experts, A1 does not collect information, but customers want to gain advice / see results. As such, A2 will still be the most preferred alternative by customers, since compared to A3 and A4 it requires less information and still can guarantee promising results regarding apparel attribute information.

- Sub-criterion 2.2: data handling by online clothing retailer

Based upon the scores as indicated in figure G.6 and the credal ranking presented in underlying figure G.8, the alternatives can be ranked as follows with respect to the criterion 'data handling by online clothing retailer': A1>A2>A3>A4.

Based upon the obtained confidence levels presented in underling figure G.8, A1 is also certainly preferred over all other three alternatives, with respect to data handling by the online apparel retailers.

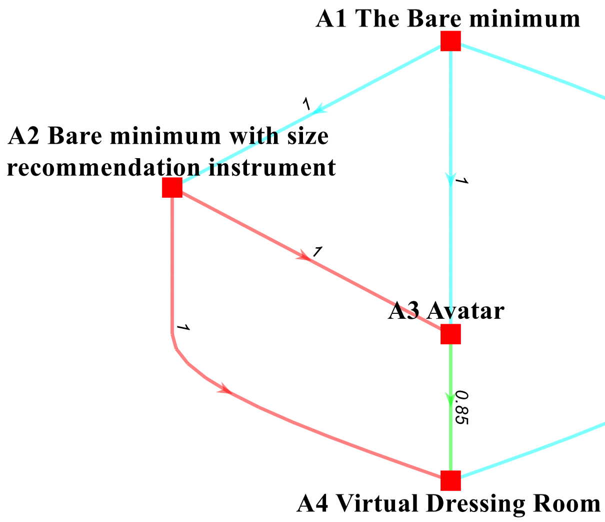

Figure G.8: credal ranking of alternatives regarding 'data handling by online clothing retailer'

The results show that A1 scores the best with respect to data handling by online clothing retailers and how it affects the customers preference to use the technology, implying that a higher weight positively influences the customer preference to use a technology.

On the contrary, A4 scores the worst with respect to this criterion. All six experts have indicated that A1 is the best alternative, since no data is collected, hence no data can be stored and used thus no privacy and security concerns can arise. On the contrary, according to five out of six experts (experts 1,3,4,5 and 6), customers can perceive the most privacy and security concerns when using A4, since e.g. other data than simply body measurements data can be collected as well (such as pictures of individuals virtually trying on apparel on the mirrored image of themselves) which can be used for other purposes by online apparel retailers. As a result, customers might be the least comfortable to use A4, a bit more comfortable to use A3 (since it is computer-based), a bit more comfortable than A3 to use A2, since the avatar is a body-twin and thus requires more data compared to A2 to evaluate apparel items with, and the most comfortable to use A1 when it comes to data handling by the online apparel retailer.

Expert 2 indicated that A3 is the worst alterative, since according to this expert with A3 (an avatar) more information can be given away by adjusting the avatar to the customers' liking, compared to A4 where the technology sees you as you are. Other customers' preferences regarding e.g. the desired physical appearance of the customer can also be gained, stored and used for other practices (e.g. advertisements), since the avatar provides features to design your own avatar with respect to hair colour, facial features etc.

Subset 3

Underlying figure G.9 provides an overview of the scores of each alternative with respect to each sub-criterion belonging to the main-criterion 'user-friendliness', which are: 1) responsiveness, 2) search time, 3) availability, 4) attractiveness and 5) required preparatory work time.

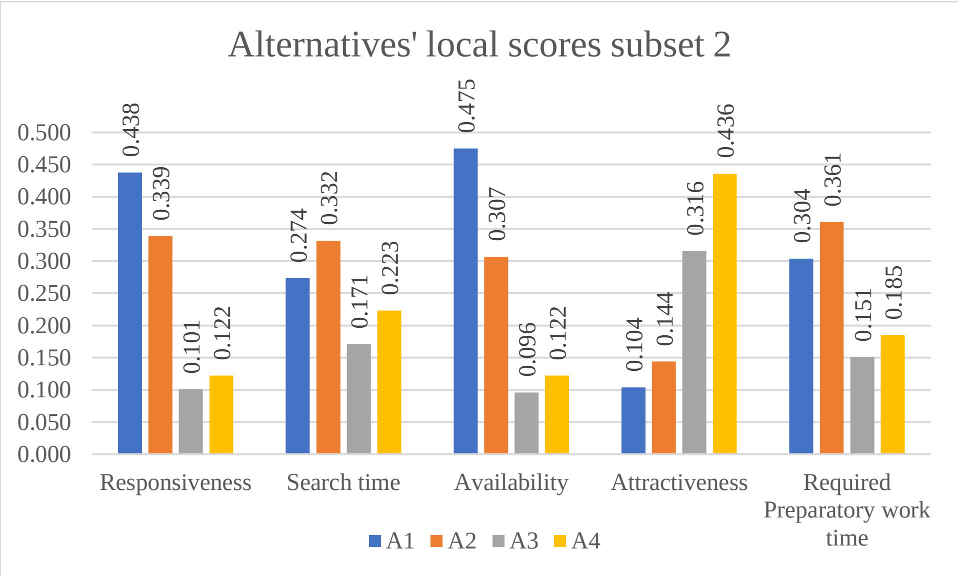

Figure G.9: alternatives' scores with respect to sub-criteria of subset 3

- Sub-criterion 3.1: Responsiveness

Based upon the scores as indicated in figure G.9 and the credal ranking presented in underlying figure G.10, the alternatives can be ranked as follows with respect to the criterion 'responsiveness': A1>A2>A4>A3.

Based upon the obtained confidence levels as presented in underlying figure G.10, A1 is certainly the most responsive alternative compared to all other three alternatives.

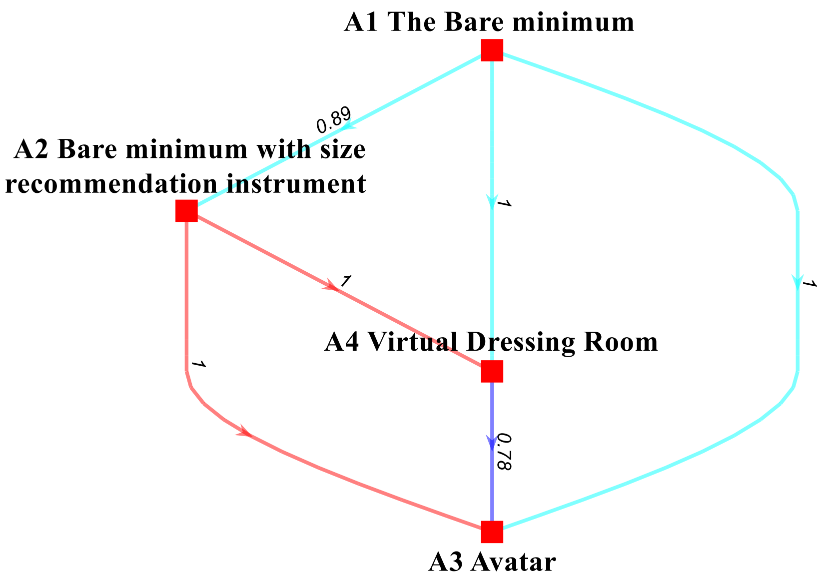

Figure G.10: credal ranking of alternatives regarding 'responsiveness'

Five out of six experts (experts 1, 2, 4, 5 and 6) (83%) have indicated that A1 scores the best with respect to responsiveness (loading time) of the technology. According to 4 out of 6 experts (expert 1, 2, 3 and 5) A3 scores the worst.

The main reason why A1 requires the least amount of loading time, is that only a size table needs to be uploaded according to expert 1. According to expert 2, A1 is the perceived as the best, because it is the least heavy program to run. According to expert 5, A1 is perceived as the best with respect to loading time, since this alternative has the most static data which already exists and one only has to click on it to see the size table.

The main reason why A3 was perceived as the least responsiveness by expert 1, is that a computer-based program can be quite heavy in combination with the data attached to the apparel items. A4 is between even and reasonably better compared to A3, because it is very computer based. However, the expect could not guarantee this with full certainty, since it depends of the software and technical applications of the technologies.

According to expert 2 and 3, A3 is processor and software wise the heaviest to run on devices, which can negatively affect the charging time. A2 compared to A1 is just as good when it comes to responsiveness, since not much computational power is needed to run A2. A4 compared to A3 scores reasonably better when it comes to responsiveness, since A4 is a slightly less heavy program to run than A3.

According to expert 6, A4 was perceived as least responsive, since customers provide a lot more information which requires much more data storage capacity, which can negatively influence the loading time of the technology on the website. A3 scores reasonably better than A4, since according to expert 6, a VDR requires more software as it involves the environment and dynamic movement , which can negatively impact the responsiveness. However, since this is a relative state-of-the art technology, the experts could not fully guarantee this, since it fully depends on the technological availability and applicability qua functionality and the provided software.

According to expert 2, the combination of the apparel items along with the body-measurements data, the required calculations within the images and the possibility to display the apparel items on the individuals' own mirrored self, can be the most difficult for A4 to do well, just a bit more difficult than in comparison to e.g. A3. A1 is reasonably better than A2, as A2 does require some actions to implement the fit & size recommendation application function, but that is neither perceived as much more technically complex than A1 nor requires a lot of storage capacity according to all six experts.

- Sub-criterion 3.2: Search time

Based upon the scores as indicated in figure G.9 and the credal ranking presented in underlying figure G.11, the alternatives can be ranked as follows with respect to the criterion 'search time': A2>A1>A4>A3.

Based upon the obtained confidence levels as presented in underlying figure G.11, A2 requires the least amount of search time with a assigend confidence level of over 0.73 compared to the other alternatives.

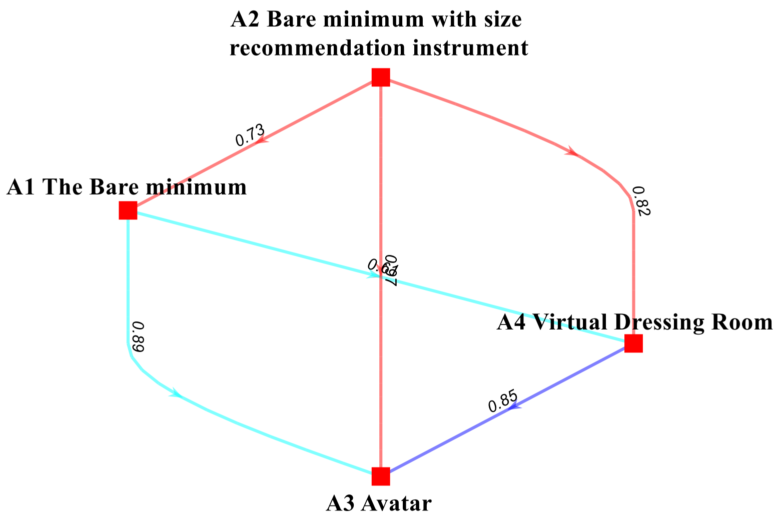

Figure G.11: credal ranking of alternatives regarding 'search time'

The results show that A2 is the best alternative with respect to search time, A1 is the second best, followed by A4. A3 is perceived as the least best with respect to search time which implies that users have to spend the most amount of navigation time (number of clicks) when using the technology during the online evaluation process of apparel.

Two out of six experts (experts 3 and 6) (33%) have indicated that A2 is the best alternative. The main reason for this is that with A2, customers just need to fill in the required information in a format to obtain the personal advice provided. With A1, customers always have to measure themselves and remember e.g. which size category they belong to, then go to the size chart table to establish which size category they belong to, which requires more searching time in order to evaluate apparel items more accurately. Furthermore, A2 is reasonably better than A1 in terms of search time and just as good compared to A3 and A4, since with A3 and A4 the number of clicks will not be that great of a difference. However, both experts have indicated that they cannot say this with very high certainty, since it all depends on how the technologies will be technically operationalized by retailers in practice.

When looking at Appendix F, three experts (expert 1, 2 and 5) have indicated that A1 is the best alternative with research to 'search time'. The reason why A2 is still assigned the highest optimal group weight can be explained by the following. Expert 2 has assigend similar values (1) to A2 implying that A2 scores the same as A1 with respect to 'search time'. Furthermore, expert 5 has indicated that A2 scores between equally and moderately better than A1 with respect to search time. As a results, A2 has the heighest weight assigend.

Four of the six experts (expert 1,2,5 and 6) (67%) have indicated that A3 requires the most amount of search time, since customers have to create an avatar first, which requires more clicks. A3, is perceived as the worst, because customers need to perform more actions to get the Avatar right, in the sense that they are going to 'tweak' more on the Avatar to get a 'perfect' Avatar.

Most of these experts assigned similar values to A4, implying that A4 is equally good, between equally and moderately or moderately better than A3 when it comes to search time, implying that the search time spend on evaluating apparel items with A4 does not differ that much from A3, but is still a bit higher with A3 since it is necessary to develop/adjust a personal avatar to accurately evaluate the appeal items.

According to expert 2, A4 scores reasonably better than A3 when it comes to search time, because with A3 you still have to conduct more actions, because it is computer-based, while with A4 it is a mirrored self that follows your movement. According to this expert, with a VDR (A4), customers do not have to use an extra click to see the fit from all sides (as the technology follows the movement of the individual) while with an avatar you first have to click on the avatar itself to be able to rotate it.

According to expert 6, when using A4, the technology should compute the individuals' mirrored self correctly right away based on the scanning technology and the inserted body-measurements data. However, with an Avatar, customers need to 'tweak' more to get the right appearance of themselves in order to evaluate apparel items more accurately. Consequently, a representative animation body double requires more actions). A2 is pretty much better than A1 when it comes to 'search time', because all customers have to do is to fill in their information and the system allows to remember/save their data.

A2 is between much and very much better than A3 and A4, since according to expert 6 the VDR has less searching time compared to the Avatar, but the difference compared to A2 is pretty much the same.

- Sub-criterion 3.3: Availability

Based upon the scores as indicated in figure G.9 and the credal ranking presented in underlying figure G.12, the alternatives can be ranked as follows with respect to the criterion 'availability': A1>A2>A4>A3.

Based upon the obtained confidence levels as presented in underlying figure G.12, A1 is certainly perceived as the alternative which can be made the most available on any device.

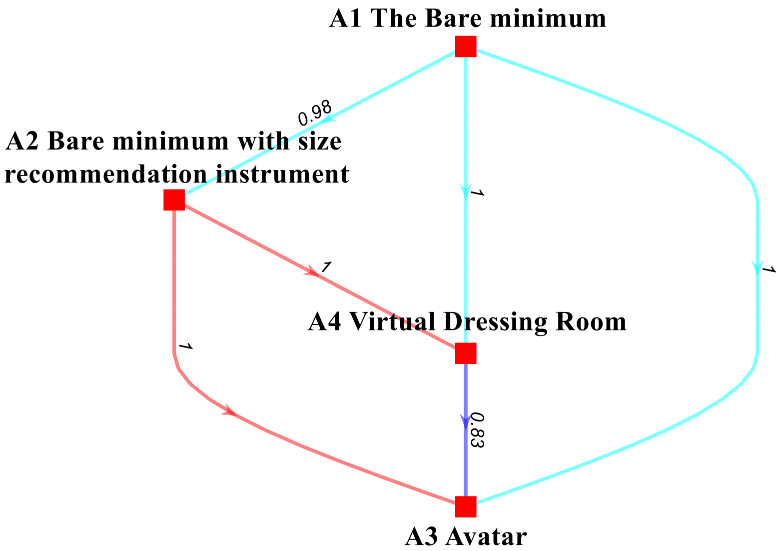

Figure G.12: credal ranking of alternatives regarding 'availability'

When looking at the sores (i.e. weights) of each alternative with respect to the criterion availability, A1 and A2 score relatively the same and A3 and A4 score relatively the same. This can be explained by the following, all six experts have indicated that A1 scores the best with respect to availability on devices, and most of them also indicated that A2 is equally good, or between equally and moderately better, or moderately better that A1 with respect to the criterion availability on various devices. This is also the case between A3 and A4.

The main reason mentioned to why A1 has the highest score with respect to availability, is that A1 is easy to install and therefore easy to make available on any device. According to expert 3, A1 is also the best since as it has already proven itself the most. Also, A1 is the least technically complicated, making it easier to provide it for any device. According to expert 5, A1 requires very low to no computational power, as it is just a matter of loading a size chart and pictures.

Since A2 builds upon A1 in functionality, with the addition of a fit & size recommendation function, the complexity to make A2 available on every device increases. However, most experts have indicated that this requires minimal changes regarding e.g. scalability.

Four out of six experts (expert 1, 2, 3, and 5) (67%) perceived A3 as the most challenging when it comes to making the avatar available on multiple devices, since it requires the most computational power. According to expert 2, A3 is the hardest to install. Compared to A3, A4 is reasonably better to make available on any device, as it is requires less computational software power and resembles snapchat that has already proven itself in part.

According to expert five, A3 is perceived as the worst, because adjusting the program on any device seems to be a bigger challenge and can require a lot of storage capacity, which makes it less easy to run it on any device. A1 scores reasonably better than A2, because creating a data format does not require a lot of work nor high computational power. Accord to expert 5, A4 scores relatively better than A3, since A3 requires more storage capacitance, which cannot be pleasant to use on e.g. Phones. However, the expert has indicated that this cannot be guaranteed with high certainty upfront, since it all depends on how the technologies will be technically operationalized by retailers in practice. Because of this, all four experts (expert 1, 2, 3 and 5) could not guarantee upfront whether A4 or A3 is/will be the most challenging when it comes to availability.

Expert 4 and 6 have indicated that A4 is the most challenging when it comes to making the technology available on any device. According to expert 4, A4 is perceived as the worst, because the scalability of moving/ dynamic images that have to adapt to the body-shape or body-movements is different and a bit more technically complex from static photo's. For this reason, A4 is a bit more difficult to make available on any device compared to A3. According to expert 6, customers also need to have or acquire a camera in order to be able to use A4. According to this expert, A3 scores between evenly and reasonably better than A4, because the difference between a person and the 'background' environment and the dynamic movement of an individual do not have to be taken into consideration when using an animated version (A3) of a virtual try-on experience.

According to all the experts, A1 scores above strongly better than A3 (as can be seen in appendix F)as 'virtual reality' is difficult to make responsive. Also more is asked of e.g. the phones in terms of software, and there are also a lot of people who have an old phone where the software cannot handle the technologies. This also applies for A1 compared to A4.

- Sub-criterion 3.4: Attractiveness

Based upon the scores as indicated in figure G.9 and the credal ranking presented in underlying figure G.13, the alternatives can be ranked as follows with respect to the criterion 'attractiveness': A4>A3>A2>A1.

Based upon the obtained confidence levels as presented in underlying figure G.13 A4 is also certainly preferred over all other three alternatives, with respect to data handling by the online apparel retailers.

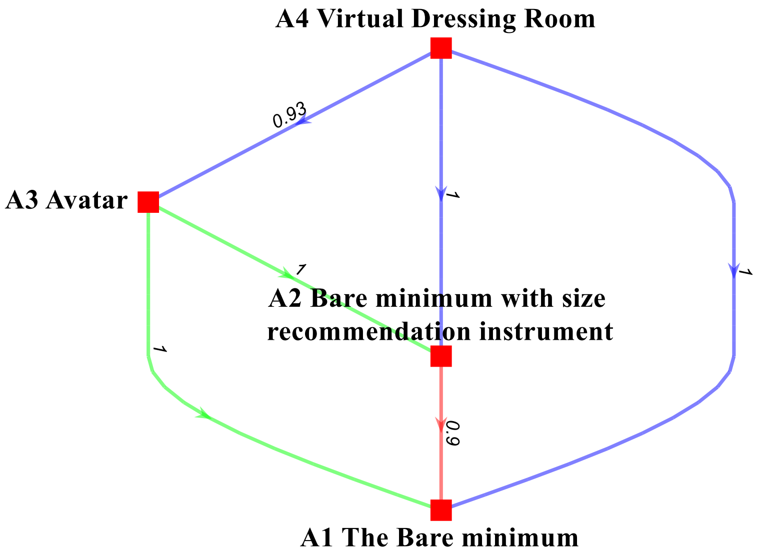

Figure G.13: credal ranking of alternatives regarding 'attractiveness'

The results show that A4 scores the best with respect to attractiveness of the technology, and that A1 scores the worst. Five out of six experts (expert 1, 2, 4, 5 and 6) (83%) have indicated that A4 is the best alternative and all experts indicated that A1 is the worst alternative with respect to being visually appealing and playful.

The reasons mentioned to why A4 was perceived as the most attractive alternative, was due to its ability to see oneself and try on apparel items on one's own mirrored image. The dynamic movement, where apparel moves with the individuals' body movements, makes it more exciting, playful and visually appealing for customers to use.

Only one expert (expert 3) has indicated that A3 is the most attractive, since users do not have to deal with the real appearance of themselves, since the avatar can be altered according to the customer likings (e.g. hair colour etc.). According to this expert, A3 is pretty much more attractive than A4, since people can still tinker with their self-image.

The provided reason explaining why A1 is perceived as the least attractive alternative to use, is because it is the most static alternative / the least interactive which negatively influence the perceived attractiveness.

A3 was perceived as less attractive compared to A4, since A4 provides a more realistic virtual try-on experience compared to A3 which provides a computer-based virtual try on experience.

Since A2 builds upon A1, by adding a fit & size recommendation function, A2 was perceived as more attractive, since the possibility to insert one's own measurements data (the user interactivity) to obtain personal recommendations might be perceived as more attractive compared to the static alternative A1.

- Sub-criterion 3.5: Required preparatory work time

Based upon the scores as indicated in figure G.9 and the credal ranking presented in underlying figure G.14, the alternatives can be ranked as follows with respect to the criterion 'required preparatory work time': A2>A1>A4>A3.

Based upon the obtained confidence levels as presented in underlying figure G.14, A2 certainly requires the least amount of preparatory work time compared to A3 and A4. Compared to A1, the assigend confidence level is 0.74 implying that the level of certainty to which A2 is perceived to require less preparatory work time is 0.74.

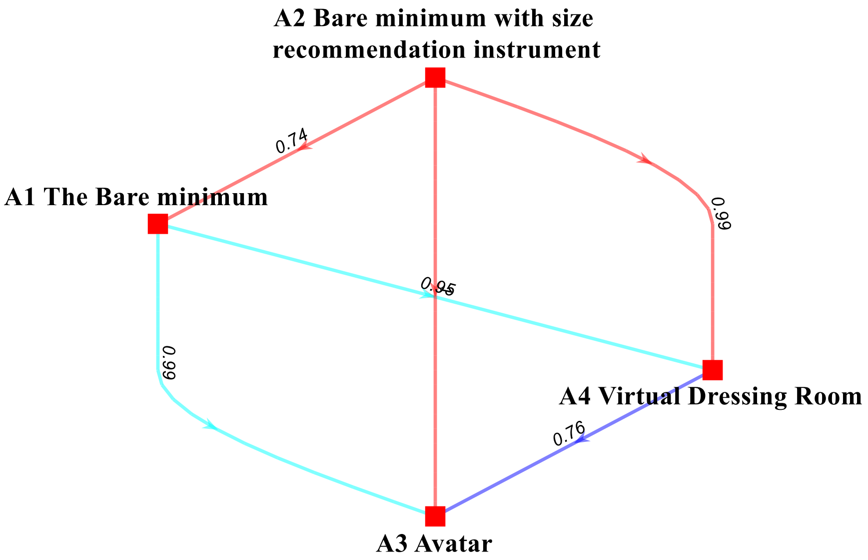

Figure G.14: credal ranking of alternatives regarding 'required preparatory work time'

The results show that A2 scores the best with respect to the required preparatory work time, and that A3 scores the worst. This implies that customers need to do the least amount of preparatory work upfront in order to be able to use A2 to evaluate the personal apparel match with, and for A3 the most amount of preparatory work upfront.

Three of the six experts (expert 1, 3, and 6) (50%) have indicated that A2 requires the least amount of preparatory work time upfront the evaluation process, since customers simply have to insert their body measurements online which according to the experts is quite easy, simple and quick to do. According to these three experts, A2 is perceived to require less preparatory work compared to A1, since A1 requires users to first measure their own body with e.g. a body measuring band, in order for them to know what size category they belong to. Consequently, as indicated by expert 1, inserting ones' body measurement information is more convenient than having to measure one's body on its own. According to expert 1, inserting one's body measurements can be done on the couch, hence is more comfortable and less inconvenient, since people do not have to stand up to measure their body first. Furthermore, due to the possibility to store customers' data, the customer does not have to insert his/her body measurements every time he/she wants to purchase apparel online. Customers do not have to remember their body measurements, since it is stored, which is not the case with A1.

The other three experts (expert 2, 4 and 5) (50%) have indicated that A1 required the least amount of preparatory work time, since people there only have to measure themselves which is the least complex to do.

Although 50% of the experts (expert 2, 4 and 5) chose A1 as best and the other 50% of experts (expert 1, 3 and 6) chose A2 as best alternative (alternative which required the least amount of preparatory work), two experts (expert 3 and 6) chose A1 as worst alternative whilst no experts chose A2 as worst alternative. As a results, A2 is assigned a higher optimal group weight compared to A1, making A2 the best / most preferred alternative (the alternative with the least amount of required preparatory work time).

Four out of six experts (expert 1, 2, 4 and 5) have indicated that A3 requires the most amount of preparatory work time upfront, and is therefore perceived to be the most inconvenient to use. According to these experts, A3 requires the most amount of preparatory work time, since customers first need to create / modify an avatar which takes the most time compare to the other three alternatives. According to expert 1, customers might have to scan their body in order to obtain a highly accurate avatar, since scanning is more accurate than e.g. adjusting the avatar by manually inserting body-measurement data. For this, customers might need to travel to a certain facility or a scan-kit needs to be send to the customers. For the reason that is can be highly inconvenient and perceived as highly complex for customers, the required preparatory time for A3 is perceived to be the highest. In addition, when manually adjusting an online provided avatar, the number of actions required is also perceived the highest for A3, since customers are building virtually which requires more actions.

Compared to A3, A4 requires less preparatory work, since for A3 the number of required actions are still perceived to be more compared to A4, since customers have to create an avatar in order to use the virtual try-on experience. Since there are different approaches to acquire the right body measurements for A3 and A4, the difference between A3 and A4 with respect to 'required preparatory work time' also relies on these approaches. Compared to A3 and A4, A2 requires less preparator work time, because customers do not have to build an avatar or use technologies such as scan technology to be able to use the technology (A2).

Alternatives' employment possibility in company (based upon 6 apparel e-commerce experts)

Company 1: WAK

According to expert 1 and 3 whom are from the same company (WAK), a combination of A2 (interactive size chart) and A1 (static size table) have already been employed without the mix-and-match-function of apparel items to see the total outfit.

According to expert 1, A2 would be the easiest to implement, because most of the functionalities of A2 are already employed in the company and the mix-and-match function can be easily added to it.

Currently, one general size chart (A1) is employed. However, according to these two experts, it would be better to make the size table brand specific, but this requires to a lot of additions to the site. Compared to A1, A2 is easier to implement, because the data can simply be inserted into the system which is already employed in WAK.

According to expert 3, the required software nor the necessary data to employ A3 and A4 are currently available in the company. In operationalizing A1, the company is already good at. WAK has its own photo studio and the suggestions e.g. fit & size recommendations are also linked to A2. If the fit-finder (fit & size recommendation application) of A2 is not linked to a specific brand, a standard size chart (A1) is given as size advice. According to expert 1, A3 and A4 are both technically and financially equally less feasible for WAK, since Information Technology experts need to be hired to operationalize A3 and A4 and the whole fashion chain needs to be adapted to the imposed digital way of working by A3 and A4, since WAK is a Multi-brand online fashion retailer. According to expert 1, A3 and A4 are more feasible for start-ups, since the digital way of working can then be imposed as the nature of working.

According to expert 1, A4 would be the most difficult to implement, due to the inherent privacy and security concerns which can occur, even more so than with A3. Plus companies have to deal with getting apparel items digitally, which is very difficult. Since avatars can be made a little more anonymous compared to A4, privacy wise A3 would be less complex to implement into the company according to expert 3.

According to expert 3, A4 is perceived as the least employable for WAK, because on the one hand retailers have to deal with loading the actual body measurements. Retailer have to acquire the necessary software for that, which does exist, but they also have to store all the customer data which requires huge storage capacity. On the other hand, the biggest problems which multi-brand stores like WAK face is the requirement of all data such as dimensions and pattern data of all apparel items. The whole chain needs to be aligned on the imposed digital way of working.

First of all, WAK needs to get all this data from the brands the apparel items are bought from. However, according to expert 3, those brands do not have all the data in the required format. Furthermore, according to excerpt 3, brands that do have everything in a digital format, will not be willing to give away their entire 3D fit application, because it is their trademark. In other words, WAK has to do it itself per apparel item, which is not financially feasible. A3 and A4 are both technically and financially less feasible. According to expert 3, the whole chain has to be geared to it. The costs and effort are so enormous for A3 and A4, that it is better to invest in improved photography for A1 and A2. External parties make it difficult to employ. WAK has a lot of external parties, is a multi-brand store, so it is not possible. E.g. for a company which only trades in its own brands and for companies which already work in an advanced manner with their suppliers such that they already have a lot of apparel information in 3D format, it can work. Since these companies already have the necessary software, they do not have to deal with external parties whom have to provide extra information and data. These companies, have solved the problem of the required data (in the right degree of delivery) with regard to apparel by working in an advanced way right from the start.

But on the other hand there is also the problem of privacy and security when it comes to customer data. According to expert 3, things like where the data is stored and how and where the Avatars are stored is practically not feasible in terms of e.g. storage capacity. Expert 3 indicated that in an internal interview held with one of the biggest fashion retailer in Europe, the fashion retailer indicated that it is practically not feasible, that they do not want to start with A3, because the storage of all Avatars requires a lot of storage capacity. In other words, for mature multi-brand stores it is not feasible to employ A3.

Company 2: VOT

According to expert 2, who is an online marketing manager, A1 is already used without the mix-and-match function. A2 is perceived as the best, because it is the most technically and financially feasible for the company.

A1 and A2 score equally well when it comes to the implementation possibility in the company. However, A3 is the least employable, because the investments are quite high and technically it will be a bottleneck for VOT. According to expert 2, A4 is moderately better than A3, the VDR module is just a bit easier to implement compared to the Avatar. And an avatar sounds a bit more exciting and complicated than the VDR, because a VDR is a ‘digital mirror’ which provides a virtual-try on experience which is closer to the experience of trying on apparel items in an actual fitting room.

Company 3: BOK

Expert 4 and 5 are from the same online apparel retailers in the Netherlands. Within this company (BOK), also A1 is employed however also without the mix-and-match-function.

According to expert 4, A1 is the most implementable alternative, because it is currently already partly used within BOK and it is the easiest to implement. Currently, a lot of product photography is already employed and now it is just a matter of mixing and matching the photos, which is a lot easier than making it very interactive e.g. with A3 or A4.

According to expert 4 of BOK, the preference to implement an alternative also depends on whether customers want the interactive element. According to this expert, it is a fact that customers cannot get every detail related to apparel attributes out of a photo. However, BOK does invest a great deal in photography. Furthermore, BOK wants to retain a lot of physical stores, as there are a lot of repeat purchases for cosmetics, for example. Based on an internal customers’ research conducted by BOK, customers want to try products such as cosmetics first in the shop, and then conduct their repeat purchases online. This might also be the case for other fashion items such as e.g. apparel items. According to expert 4, A1 is much better to implement than A2, as it is easier to employ and technically and financially more feasible.

A1 is very much better to implement than A4 and between very much better and absolutely better compared to A3, as it is technically even more complex (in terms of software and storage capacity) and also financially not feasible for BOK.

According to expert 4, state of the art technologies such A3 and A4 could have added value for BOK, but this all depends on how well they work, and how easy they are to implement. According to this expert these kind of applications are currently often used for home furnishing applications (to find out how e.g. a couch fits in the living room). But for apparel items there is so much more data needed to realize this. The difference between reality and virtual reality is not the same. Only if it is real and realistic then it really has added value for BOK according to the expert. For apparel, it is still perceived as very difficult, also considering the number of products and brands BOK has.

According to expert 5, A2 was also already tested in the form of the 'fit-finder', but the results were not really positive, since the number of returns increased. It turned out that at that time BOK had not matched the fit & size recommendation applications with customers data and apparel attribute information well after all. However, since it was not a project of this expert, the expert could not provide exact details which had led to the ‘failure’ of the ‘fit-finder’ application.

A2 is the easiest to implement, if the data is still stored within the company. Approximately 1.5 years ago, BOK started their entire e-commerce department, but at that time the website didn't look so good and professional. Now, the focus of BOK is on online shopping, since it is growing and thus is becoming necessary. According to expert 5, the focus within BOK is still insufficient on the returns department, despite the fact that it is very important. There reason why a limited amount of attention is given to order returns management is that currently the capacity in terms of personnel and finance are not available for it yet. According to expert 5, BOK now prefers to focus on increasing sales and conversion instead of reducing returns. A2 is much better than A3 and A4 when it comes to implementation in the company, as for both A3 and A4 experts have to be hired as the current developers do not have the knowledge to do so. This costs more time and money. According to expert 5, testing the technologies and gathering customer opinions also takes much more time, effort and money compared to A1 and A2. A3 is still perceived as worst alternative, because BOK has to create a program for it which can be very have to run, it has to be available on all devices which is more difficult if it is a heavy program to run. The loading time will therefore also be negatively affected.

Company 4: KLE

According to expert 6, A1 is already in use (pictures and a static size chart), without the mix-and-match function. Since A1 is already mostly employed, the best alternative is A1, since it is the easiest to implement and also the easiest to adjust. According to expert 6, A2 is also pretty simple to implement, but it is technically a bit trickier compared to A1. A3 as well as A4 require a lot of technical development for the company, which makes it technically more complex, more difficult and costly to implement. According to this expert, this also requires a lot from the website, since they are heavy programs, which will have a huge impact on the loading time of the website, and will further result in worse SEO (poorer performance of the technology and website). A2 is perceived as a reasonable option, but according to the expert it all relies on how much extra it can actually add in practice compared to A1 which is already mostly employed.

The expert has indicated that A4 is the worst to employ for KLE, because the loading time for the website with A4 will be much worse and this has more negative impact on the customers' willingness to purchase apparel items online, compared to A3.According to the expert, A1 is perceived as reasonably better than A2, because with A2 different sizes per brand have to be taken into consideration as well in the fit & size recommendation application, which is more complex to achieve in a proper manner. Since sizes differ per brand, it is more difficult to employ a more interactive alternative by adding e.g. a fit & size recommendation function to A1 to obtain A2.