

# Shopping cart design evaluation

A quantitative research case study prepared by Dave Hone

# Purpose

Compare two shopping cart designs using an outside-in view of customer preferences



## Quantitative research

Small scale survey (n=300) to quantify customer preferences for shopping cart design

Remote unmoderated quantitative usability test using an interactive mobile app prototype



## Insights gathered

Measure the extent the new shopping cart design meets customer and business needs

Confirm feedback on current design and understand reasons for preferences

# Executive summary

Research was undertaken to inform the decision to proceed with the development of a new shopping cart

Analysis of research results indicate that should development proceed, Design B would likely improve usability and be customers' preferred solution



The purpose of user research is to identify opportunities to improve customer experience.

Qualitative user research, undertaken in 2021 to test customer response to the latest live iteration of the mobile app, identified observable issues with the current shopping cart visual and functional design.

The mobile app team designed two potential solutions.



Further research was needed to identify customer preferences between the designs and the current 'live' design, and which of the two Designs A/B was preferred by customers.

A quantitative online study was undertaken with two interactive prototypes of Design A/B.

The Customer Support team sent email invites to 150,000 recent and engaged customers.



Participation was incentivised with a \$10 bonus token.

A 2% response rate was anticipated for this type of study and incentive.

300 customers participated in two groups. 150 were asked questions about Design A, and 150 about Design B.

Customer data was used to balance the two groups A/B for behaviour and demographics.



Participants used the new design to complete a simulated transaction on their own mobile device and answer questions about their experience.

Participants provided Likert-score and free-text responses.

A statistical analysis of quantitative results was used to infer the likelihood that customer experience would be improved by a new shopping cart.



Analysis indicates both Design A and Design B were preferred to the current live app.

Design B was preferred over the current live app to a greater degree than Design A.

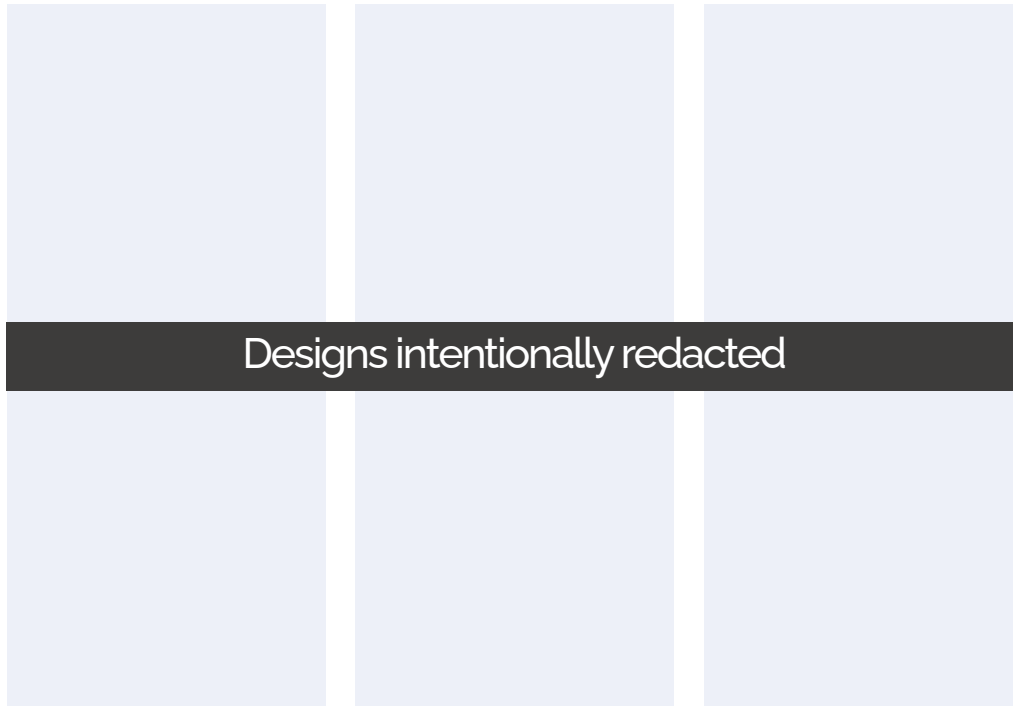
Participants' preference for Design B was statistically significant at 95% confidence ( $p < 0.001$ ).

The following report details findings from analysis.

# Proposed designs

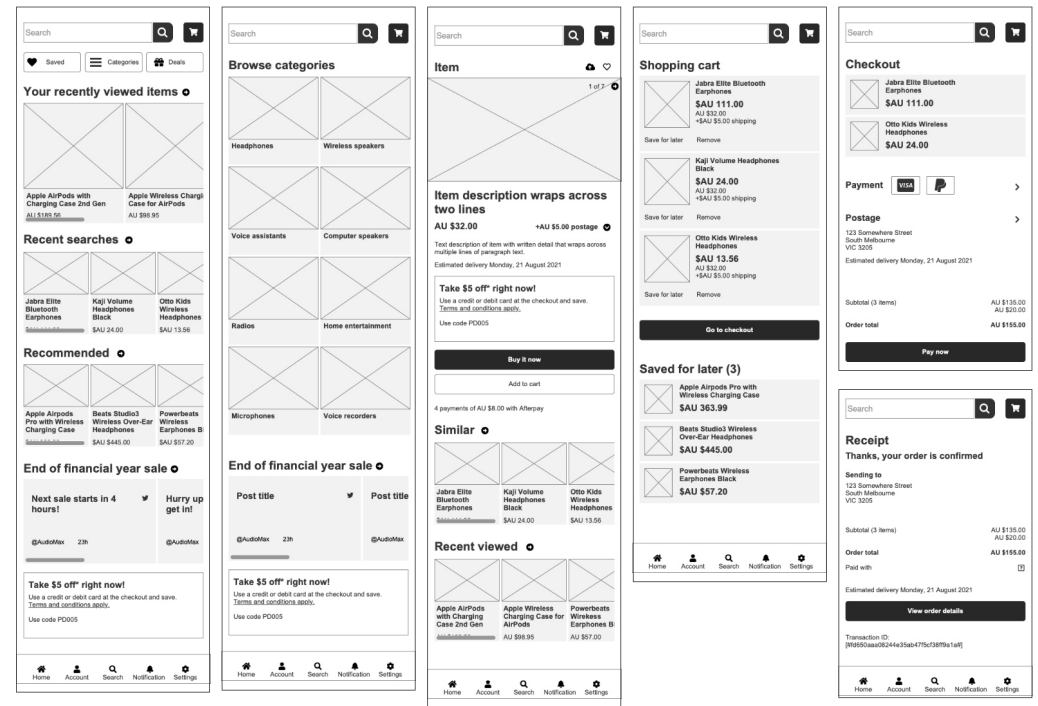
## Design A 'minimal'

Design A reduces the quantity of information displayed on screen and increases the number of steps to be navigated compared to the current 'live' design. The design intent is to address usability issues by reducing complexity.



## Design B 'utility'

Design B condenses information onto fewer pages compared to the current 'live' design, adding filter functions to manage information. The design intent is to address usability issues by enhancing control and comparison functions.



Both Designs A/B were preferred over the current 'live' design

Design B was preferred to a greater degree than Design A

# Method

# Objectives

Inform urgent decisions on mobile app redesign with customer preferences

# Approach

## Unmoderated quantitative study using UserZoom

This study analyses quantitative data from an interactive prototype and attitudinal data from survey questions to evaluate two design options for a shopping cart



**Having participants use the design was important so that we could study customer behaviour (what they did) and attitudes (what they say)**

The UserZoom platform was chosen because it provides the opportunity for participants to use the new shopping cart design to complete a transaction.

An interactive prototype was constructed in Axure to mimic a transaction on the mobile app, from immediately after product selection through to purchase receipt.

Quantitative data gathered from the study included:

- Behavioural data from prototype interactions (actions taken, 'clicks' and scrolls), time taken, journeys taken, and success rates.
- Attitudinal data from survey responses (questions were multi-choice, Likert-scale, and open-ended verbatim responses).



**Participant screening was important so that we could prioritise feedback from the highest-value customers**

Internal stakeholders agreed that introducing a deliberative bias was appropriate to prioritise feedback from high-value and medium-value customer cohorts.

Customer cohort lists were provided by the Customer Support team using data of recency, frequency, and monetary value (RFM).

300 participants were split into two groups (cells A/B n=150) balancing each cell for demographics and behaviour (RFM, product category, tenure, mobile device size, age, gender and engagement).

Each cell was shown one of two designs so that response bias was managed in the (otherwise identical) survey.

Analysis was undertaken to determine if there were any statistically significant differences across the two cells A/B.

# Analysis

## Statistical tests for quantitative attitudinal data

This report draws conclusions by analysing closed-question responses (what responses participants chose) with sentiment analysis of free-text responses (why they chose that), and a statistical test on how customers scored design A/B

- All tests for statistical significance were undertaken at 95% confidence (common convention for this type of study)
- Non-parametric tests have been used to report on statistical significance for any preference of design A or design B (Mann-Whitney and Wilcoxon, as appropriate)
- Non-parametric test are a common convention for comparing two groups two groups of Likert-score data without normal distribution and equal variance assumptions
- Non-parametric statistical analysis is included in this documents as p values. Any p value  $<0.05$  is considered significant. See interpretation method (page 20)
- Affinity mapping analysis has been used to categorise (codify) free-text responses by topic
- Free-text feedback is represented as "quotes", shown grouped by the Likert-scale rating of the associated question.

Key takeaways are indicated in this report in colour

Green (good)

Yellow (caution)

Red (pain points)



# Recruit- ment

# Customer cohorts

Recruitment was screened using behavioural and demographic data of 300 customers, split into two similar cells of n=150

## A mix of customer value

Both cells A/B comprised a similar mix of high-value and mid-value customers, screened using Recency, Frequency, Monetary (RFM) customer data.

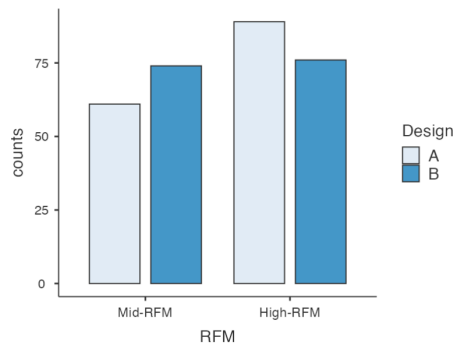


Fig i. Graph showing cells A/B had a similar mix of medium-value and high-value customers.

## A mix of customer engagement

Both cells A/B comprised a similar mix of customer engagement, screened using Net Promoter Score (NPS) customer data from Customer Support surveys.

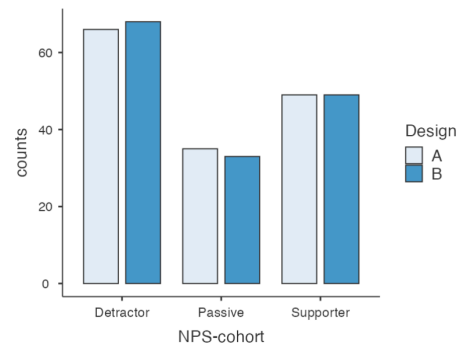


Fig ii. Graph showing cells A/B had a similar mix of NPS 'Detractor', 'Supporter' and 'Passive' scores.

## A mix of customer tenure

Both cells A/B comprised a similar mix of customer tenure (amount of time since registration), screened using customer data registration date.

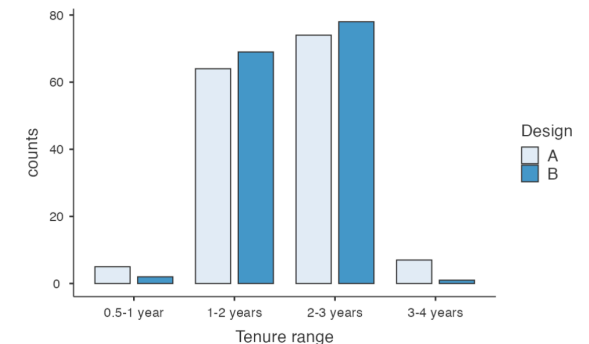


Fig iii. Graph showing cells A/B had a similar mix of new and longer-term customers.

**Conclusion: Cells A/B had a similar mix of customer value, engagement and tenure**

# Demographics

Internal stakeholders agreed that small differences between the two cells A/B were acceptable

## A mix of gender

Both cells A/B comprised a similar mix of gender, screened using registration form customer data.

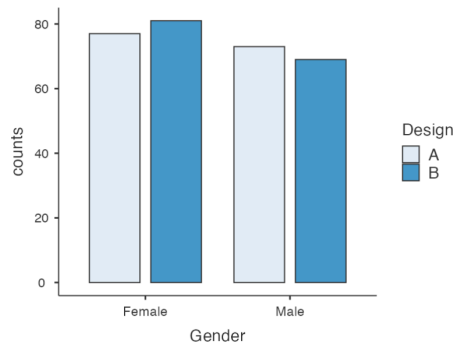


Fig iii. Graph showing cells A/B had a similar mix of gender.

## A mix of age ranges

Both cells A/B comprised a similar mix of age ranges, achieved using registration form customer data.

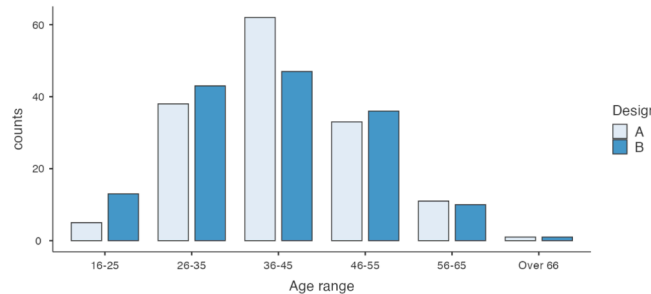


Fig iv. Graph showing cells A/B had a broadly similar mix of ages with small differences.

## A mix of device usage

Both cells A/B comprised a similar mix of device usage, achieved using traffic analysis customer data.

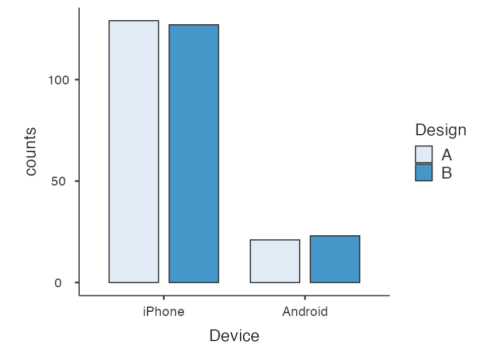


Fig v. Graph showing cells A/B had a similar mix of device usage.

**Conclusion: Cells A/B had a similar mix of gender, ages and device usage**

# Key findings

# Current design

All n=300 recent customers were asked to rate their experience of using the current mobile app

## Confirming the current app experience

Most participants rated their experience using the current app between neutral and 'Very poor', consistent with previous qualitative research.

Figure vi shows cells A/B comprised similar ratings.

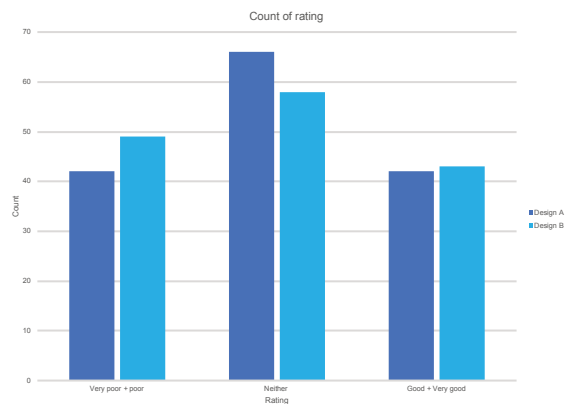


Fig vi. Graph showing mixed feedback on the current design was similar across cells A/B.

## Reasons given

Participants were asked why they gave a 'Poor' or 'Very poor' rating. Verbatim responses included:

*"It's not got the size of Amazon and it's not as cheap as ebay, but it's specialist stuff, the specific stuff I need."*

*"It's a bit slow."*

*"I wouldn't say I particularly like or dislike it, but it gets the job done."*

*"Sometimes it frustrating to have to go back and make changes to an order or compare."*

*"It's a bit clunky."*

*"Too many buttons."*

## Top-of-mind feedback

Figure vii shows sentiment analysis of all verbatim responses indicating that the top-of-mind feedback categories are 'Features and functions', 'Usability and accessibility', 'Journeys and navigation'.

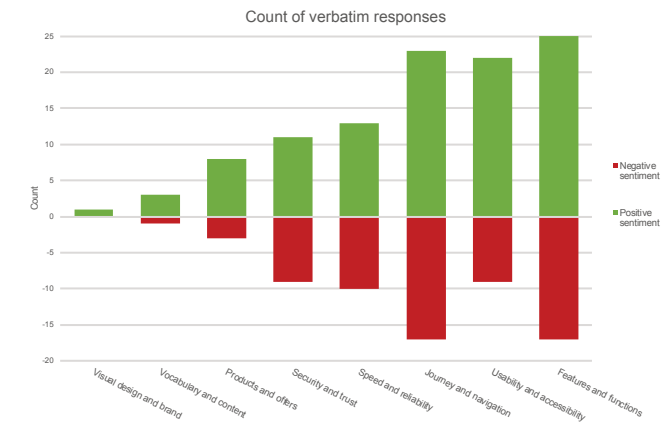


Fig vii. Graph showing top 3 feedback categories on the current design are 'Features and functions', 'Usability and accessibility', 'Journeys and navigation'.

Q1: Rate your experience using the mobile app (Likert-scale 'Poor' to 'Excellent')

Q2: Is there anything you particularly like or dislike about the app? (Optional free text)

**Conclusion: Features, usability and navigation are top-of-mind reasons given for negative feedback on the current app**

# Measuring A/B

After using the new shopping cart, participants were asked to score the new shopping cart design

## Both designs A/B scored well

As illustrated in Figure viii, respondents scored both designs A/B well for ease-of-use:

- n=93 agreed or strongly agreed that Design A was easy to use (scored >2)
- N=106 agreed or strongly agreed that Design B was easy to use (scored >2)
- Score for Design B was statistically significant and higher than Design A ( $p=0.002$ ).

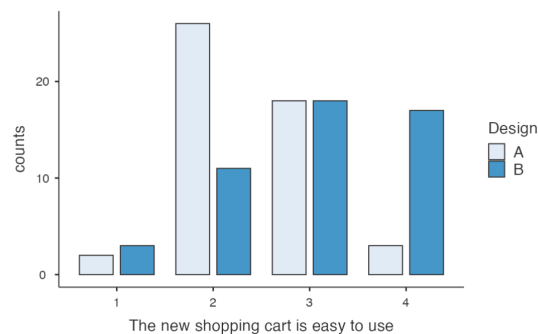


Fig viii. Graph showing Design B scored higher than Design A for ease-of-use.

## Both A/B preferred over current app

As illustrated in Figure ix, both designs A/B compared well with the current 'live' design:

- n=111 agreed or strongly agreed that Design A was preferred over the current design (scored >2)
- n=143 agreed or strongly agreed that Design B was preferred over the current design (scored >2)
- Score for Design B was statistically significant and higher than Design A ( $p<0.001$ ).

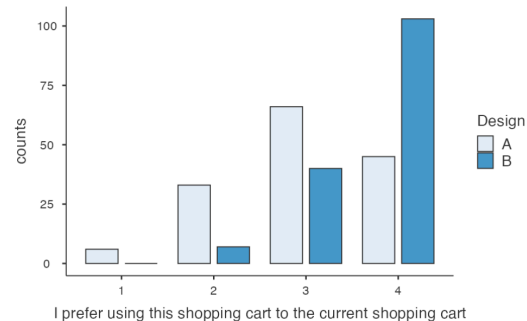


Fig ix. Graph showing both Design A and Design B compared well to the current design.

## Design B preferred over Design A

Comparing responses for designs A and B, there was a statistically significant difference:

- Design B scored higher than Design A for ease-of-use ( $p=0.002$ )
- Design B was strongly preferred over the current design and to a greater extent than Design A ( $p<0.001$ )
- In discussion with internal stakeholders, it was agreed that both findings were relevant and useful in determining a design direction
- Therefore of customers sampled, Design B was preferred to a greater extent than Design A.

Q3: The new shopping cart design was easy to use (Likert-scale 'Strongly disagree'=0 to 'Strongly agree'=4)

Q4: How does the new design compare the current shopping cart? (Likert-scale 'Much worse'=0 to 'Much better'=4)

**Conclusion: Design B was preferred to a greater extent than Design A and the current 'live' design**

# Sentiment - A

After rating Design A, participants were asked if there was anything they particularly liked or disliked

## Very poor rating

*"Not streamline."*

*"It seems more difficult to navigate."*

*"It is slower."*

*"Too many screens had to get through."*

## Poor rating

*"Don't like the white background."*

*"I dislike how plain and boring it looks."*

*"Too much space."*

*"Too many pages to get through."*

## Good rating

*"Easy to do and understand."*

*"It seemed a lot clearer than the current."*

*"Larger icons makes it easier to read."*

*"Writing looked bigger."*

*"It's different but easily adaptable."*

## Very good rating

*"Easy to use and layout was clear."*

*"Easy to see."*

*"Easy to navigate."*

*"I like it better than the one I use now."*

*"Looks streamlined."*

Q5: Is there anything you particularly like or dislike about the new design? (Optional free text)

**Very few negative comments mentioned the larger number of steps as a reason for negative feedback on Design A**

**Some negative comments mentioned lower information density and white space as reasons for negative feedback on Design A**

**Most comments were positive. A clear design and improved accessibility and legibility were mentioned as reasons for positive feedback on Design A**

# Sentiment - B

After rating Design B, participants were asked if there was anything they particularly liked or disliked

## Very poor rating

*"It's not familiar to me. When I buy I don't always have time to fuss around learning how to use stuff."*

*"I'm used to the current way and would be disappointed if it changed."*

## Poor rating

*"Detail is down too low on screen."*

*"Current page has the items shown higher in the screen which is better."*

*"I dislike that I had to scroll to find items."*

*"Too much scrolling around."*

*"Had to move the page down to see."*

## Good rating

*"Good to be able to scroll through all items from one to the other."*

*"Has all the info you need to compare."*

*"Was easy to use and sort by type."*

*"I liked the list broken up by product type."*

## Very good rating

*"There seems to be more information at the ready."*

*"Easy to see all your items."*

*"I like that you can filter the list."*

*"It's a better looking format."*

Q5: Is there anything you particularly like or dislike about the new design? (Optional free text)

**Very few negative comments mentioned adapting to change as a reason for negative feedback on Design B**

**Some negative comments mentioned the unfiltered (fully expanded) page length as a reason for negative feedback on Design B**

**Most comments were positive. Usefulness of information, the compare function and the filter and sort functions were mentioned as reasons for positive feedback on Design B**



# Summary

A discussion with key stakeholders was facilitated to discuss results and identify opportunities for action

Finding	Interpretation	Action
Features, usability and navigation are reasons given for negative feedback on the current app (see page 13)	Current design negative feedback is consistent with previous research and Customer Support surveys, indicating opportunities to improve user experience generally, and specifically to improve the features, usability and navigation of the shopping cart	Stakeholders agree to proceed with redevelopment of the shopping cart as a high priority work package
Design B was preferred to a greater extent than Design A and the current 'live' design (see page 14)	Both designs scored well. Both designs were preferred over the current design. Design B is more likely to address usability issues than Design A, however Design A has some design elements that were well regarded that may be incorporated into Design B in some way	Stakeholders agree to proceed with Design B, noting some opportunities to investigate how small visual design elements from Design A might be incorporated into Design B without changing the journey
A larger number of steps and low information density reasons for negative feedback on Design A (see page 15)	Design A dividing the journey into more steps distributed smaller amounts of information over more screens, which negatively impacted the perception of speed and a less streamlined process	Stakeholders agree to proceed with Design B higher information density and fewer navigation steps
Some negative comments mentioned the unfiltered (fully expanded) page length as a reason for negative feedback on Design B (see page 16)	Design B comments were mostly positive. Usefulness of information, the compare function and the filter and sort functions were mentioned as reasons for positive feedback on Design B. However some negative comments on page length indicate that the new filter controls were overlooked by some	Stakeholders agree to proceed with Design B filter controls noting some opportunities to investigate how to build awareness of new controls
Some negative feedback indicated that lack of familiarity would cause initial reaction to change (see page 16)	Both designs were preferred to the current design, however some negative feedback on the impact of change in general, rather than a specific design, indicates opportunities to onboard customers with the new design	Stakeholders agree to investigate either/all of: In-app help popups; A direct email marketing campaign of new features; Video help and customer support pages on the website; Customer Support team training
Cells A/B had a similar mix of customer value, engagement, tenure, gender, ages and device usage (see page 10/11)	Recruitment was valid and screened appropriately to represent a mix of customers demographics and behaviours. Cells A/B were appropriately similar for the purpose of comparison.	Stakeholders agreed that confidence in results was not likely to be impacted by the sample

# Appendix

# Frequently asked questions

This section contains follow-up questions arising from a discussion of results with key stakeholders

## How the preferred design was determined

Non-parametric tests are a suitable analysis method for an A/B evaluation of two designs from Likert-scale data. Mann-Whitney and Wilcoxon rank analysis has been used to determine statistically significant preferences.

The Mann-Whitney U test is performed on two independent cells (two separate groups participants with similar make up of characteristics) to determine how likely a selected value from Cell A will be less or greater than a selected value Cell B:

- Merged data from two cells is sorted from lowest value to highest value, and then ranked by values (lower value get rank 1, the second rank 2, etc)
- Two tailed analysis was performed on ranked values
- Rank ties were treated as the average of the ranks for the entire group using continuity correction.

## How certain the findings were

Design B median Likert-score is statistically significant and higher than for Design A.

Analysis was undertaken to find statistically significant differences between two cells A/B n=150 at 95% confidence.

The sample size used in this study is appropriate to get a 'quick read' and inform an urgent decision for the development team. Confidence was reported on the two samples of n=150 (300 total).

Priority was given to recruiting current, engaged, higher value customers, meaning that the sample n=300 does not include low value, new customers, former customers and non-customers that would be expected in a population sample.

Therefore, findings were carefully interpreted as findings of the sample, and were not overstated as representative of a population and no inference has been made about the larger population.

# Interpretation


## Interpretation of Mann-Whitney U test results


Shorthand	Observation	Interpretation
$P < 0.05$	Medians are the same and distributions different and a statistical difference	It is likely that one of the two groups has more positive sentiment than the other. There was a statistically significant difference between the two groups. Since the medians are the same it was most likely the difference in the shape of the distributions of the two groups. Inspect for a positively skewed distribution (more favourable responses) to determine the preferred option
$P < 0.05$	Medians different and distributions different and a statistical difference	Need to inspect data carefully to draw conclusions which could be a positive skew in either group distribution – typically, more positive sentiment is favoured
$A = B$	Medians are the same and distributions the same and no statistical difference	No significant difference (null hypothesis is accepted)
$A = B$	Medians are the same and distributions different and no statistical difference	Any difference is not significant (null hypothesis is accepted)
$A = B$	Medians different and distributions the same and no statistical difference	Any difference is not significant (null hypothesis is accepted)
$A \neq B$	Medians different and distributions the same and a statistical difference	The medians can be reported as statistics of the difference between the groups / cells. There was a statistically significant difference between the two groups. Since the medians are the different compare medians for a more favourable response to determine the preferred option
$A = B$	Medians different and distributions different and no statistical difference	Any difference is not significant (null hypothesis is accepted)


This page left intentionally blank

# Thank you

Get in touch and let's chat about your next project

 [www.davehone.com](http://www.davehone.com)

 [dave@thinkingabout.com.au](mailto:dave@thinkingabout.com.au)

 0410 762 999