Swipe, Snap & Chat: Mobile Survey Data Collection Using Touch Question Types and Mobile OS Features

Trent D. Buskirk, Joey Michaud and Ted Saunders
MAPOR, 2014
What did the Rotary say to the Tissue Roll?
The focus of this study was on native use of touch and technology related features of tablets and smartphones and how responses from questions using these features may vary between smartphone, tablet and PC respondents.

Respondents were recruited from Research Now’s Online Panel and were invited to participate using either PC, Tablet or Smartphone (natively).

Survey was administered via the web using an active mobile browser survey optimization approach (Buskirk and Andrus, 2010)

Survey questions covered Auto Insurance Satisfaction (~10 minutes / 60 questions).

Study was fielded between April 4 – April 11, 2014
The study contained three specific experiments:

- **Experiment 1**: Manipulated Scale Type (Standard/Slider); Scale Length (5/11pt Likert) and Scale Numbering (Yes/No)

- **Experiment 2**: Manipulated List Style (Radio/Dropdown) and List Length (short, medium and long)

- **Experiment 3**: Involved Voice To Text Prompting or Not.
Experimental Design Overview

Sample Provider Recruited Participants by native mode (device)

~1200 Respondents per Mode

Randomized to factors of Experiment 1
Randomized to factors of Experiment 2
Randomized to factors of Experiment 3

Randomized to factors of Experiment 1
Randomized to factors of Experiment 2
Randomized to factors of Experiment 3

Randomized to factors of Experiment 1
Randomized to factors of Experiment 2
Randomized to factors of Experiment 3

Randomized to factors of Experiment 1
Randomized to factors of Experiment 2
Randomized to factors of Experiment 3
For this section of Experiment 1, respondents from each device were randomized into one of two answer input types: Text (standard) and Slider (with Start Outside of Scale). This answer presentation was used for the open ended question:

How many miles per year do you drive the car that you personally drive most often? Provide your best estimate. If you drive more than 50,000 miles per year, use 50,000.

Are the response distributions for miles driven different between slider and text entry?

We also captured the screen orientation for all smartphone respondents to determine if screen orientation would moderate the affect of input type on answer distributions.
Experiment 1 – Part 1: Results

No Significant Differences in the Response Distribution of Miles Driven between Slider and Text Entry for Mobile Respondents: MWU=167,048.5; p-value=0.093
For smartphone respondents, the average miles driven (sqrt scale) was significantly different by Screen Orientation ($F_{1,1185}=2.093; p\text{-value}=0.008$), consistently across Question Presentation Method (Text vs. Slider)($F_{1,1185}=0.777; p\text{-value}=0.378$)

The miles driven entered in portrait orientation was on average, 53 units [95% CI for difference was (7,142)] higher than that entered in landscape orientation.
In this experiment respondents from each device type were randomly assigned to have three service preference questions presented using either:

- **Radio buttons**
- **Dropdown/Roller Bars**

There were a total of 3 questions in this portion of the experiment and each question asked about preferred method of contact for (1) Getting a quote, (2) account service and (3) making a claim.

An identical set of 6 response options were available for each of the three questions.
The primary metric for comparison was the number of times among the three questions for which one of the first three answer options was selected (primacy rates).

Primary comparisons were made across the experimental groups: Drop Down/Roller Bar group versus Radio Button group (for smartphone respondents)

The sample sizes across the two conditions were virtually the same.

A secondary question asks whether or not the phone operating system (iOS vs Android) and browser type modify the association between primacy distributions and scale type.
Experiment 2 – Part 1: Results – Comparing Primacy Effects across Dropdown and Radio Button Question Types for Smartphone Respondents

No Significant Differences in the distribution of the number of Experiment 2, Part 1 questions with Primacy Indications across the two question presentation styles: radio vs. dropdown/rollerbars

(Row Means Differ Statistic: \( \chi^2(1) = 0.07; \) Exact p-value = 0.80)

**Total Primacy Indication = number of questions for which one of the first three answer choices was selected**
Experiment 2 – By Operating System and Browser Type

Radio buttons

- All Devices/Browsers
- iPhone
- Android
- Chrome/Firefox

Drop-down

- Safari Browser
- Default Browser
- In person with an agent
- On the phone with an agent
- E-mail with an agent
- On the phone with Customer Service
- Insurance company website
- Mobile phone application
Browser Type and Phone *may* Moderate Scale-Type And Primacy Effects

**No Significant Differences in number of Questions on which one of the 1st 3 options chosen across two Styles:**
Row means differ statistic: $X^2(1)=0.43; \ p-value=0.51 \ (n=738)$

**No Significant Differences in number of Questions on which one of the 1st 3 options chosen across two Styles:**
Row Means Differ statistic: $X^2(1)=0.22; \ p-value=0.64 \ (n=264)$

**Significant Differences in number of questions on which one of the 1st 3 options chosen across two Styles:**
Row Means Differ statistic: $X^2(1)=5.58; \ p-value=0.02 \ (n=195)$
Experiment 3: Overview Voice-to-Text (V2T) Test

In this experiment, tablet and smartphone respondents were randomized into V2T Prompt and No V2T Prompt Groups.

This experiment applied to two open-ended questions about Auto Insurance Choice and Opinion on Monitoring Devices for Auto Insurance Discounts.

All respondents, regardless of group received the first open ended question.

Those in the V2T Prompt Group received a brief description about V2T and a visual.

A follow-up question asked respondents in the V2T prompt group how they entered their response to the open ended question.
Experiment 3: Overview, Cont.

All respondents were later asked a second open ended question about their opinion about using monitoring devices to earn auto insurance discounts.

Respondents in the V2T Prompt Group received an additional instruction:

> When answering this question, please try to use the voice-to-text feature on your [tablet/phone]. Tap in the answer box and then tap the microphone key on your [tablet/phone] to start recording your answer. It looks like this: 🎤

A follow-up question about how respondents entered their open ended response was also asked of the V2T prompt group.

> Initially we were planning to auto-detect the V2T usage; but the technology to auto detect the use of this feature across multiple devices, browsers and operating systems was not available at the time of our fielding.
A majority of users answered the first open ended question (97%) by using their keyboard only.

After prompting, about 46% who previously typed used V2T and 2/3 of those using V2T before prompting used it after prompting.

Overall, the proportion using V2T after prompting was significantly higher than the proportion before prompting (e.g. 47% after vs. 3% before; McNemar’s Test p-value < 0.00001).
Among smartphone respondents assigned to the V2T Prompt Group:

Significant differences were noted in the distribution of the number of words entered for open ended question 2 between respondents using V2T and those who Typed (Mann-Whitney U=31,042.5; p-value<0.001)
Experiment 3: Follow-Up

Respondents in the V2T Prompt group who used the V2T feature to answer the second open ended question were asked an additional follow up question:

Do you think you gave a longer answer using voice-to-text than you would have if you had typed it in?

35% of Smartphone Respondents answered YES
And Voice-to-Text doesn’t work all that well for many survey respondents.

And it’s not quite like it is in the movies...

3 Main Themes About Using Voice-to-Text for Survey Data Entry (open-ended ?s) emerged from respondents’ comments about the V2T feature among those in the V2T Prompt Group.
Voice To Text: The Respondents Have Spoken...

**Location prohibits talking**

“I would use it if my spouse wasn't trying to sleep next to me.”

**Current Location is too loud**

“I'm at work too loud to voice-to-text.”

“Wrong lots of times.”

“Current Location is too loud”

“‘I’m on the quiet car on my train and talking is frowned upon.’

“My background has way too much going on I would need to use it when I'm alone in a calmer location.”

“Doesn’t Work Well”

“I have a Texas accent and Siri has trouble understanding me.”

“My words are never interpreted correctly and I have to type anyway.”

“Current Location is too loud”

“‘My background has way too much going on I would need to use it when I'm alone in a calmer location.’

“Location prohibits talking”

“I am on the quiet car on my train and talking is frowned upon.”
Will appear in a special issue of *Methods, Data, Analyses (MDA)*, covering collection of survey data using mixed-devices

published by GESIS ([www.gesis.org/mda](http://www.gesis.org/mda)) and to appear Fall/Winter, 2015

- Detailed comparisons across modes
- Additional outcomes from Experiment 1 including a within-person cross-over design with scale switching (e.g. radio to slider and vice versa)
- Expanded Discussion of Camera Upload outcomes
THANK YOU!

TBuskirk@m-s-g.com
314-695-1378
www.m-s-g.com

@trentbuskirk
For the second component of Experiment 2 we looked at the relationship between:

**List Style** -
- Radio
- Drop-down/Barrel Roll

**List Length**
- Short – 6 Choices
- Medium – 11 Choices
- Long – 21 Choices

Occasionally, someone from your insurance company may need to call you. Generally, what is the best time for them to call you on a weekday?

Respondents from each device were randomized to one list style and one list length independently. Here we focus on SMARTPHONE respondents, exclusively.
Respondents were given various times to choose from 8:00 AM to 6:00 PM within each of three list groups:

- **Short (6-Choices):** every two hours from 8AM until 6PM
- **Medium (11-Choices):** every hour, on the hour from 8AM until 6PM
- **Long (21-Choices):** every half hour from 8AM until 6PM
To analyze the likelihood for time slot selection across the two factors (list type and list length), we collapsed all responses from the three lists into time blocks:

- (1) 8AM < 12 Noon; (2) 12<2PM; (3) 2<4PM; (4) 4<6PM; (5) 6PM

11 Choice List: 8AM, 9AM, 10AM and 11AM → Time Slot 1
21 Choice List: 8AM, 8:30AM,...11:00AM, 11:30AM→Time Slot 1

Treating time-slot as an ordinal variable (time since morning), a proportional odds model was fit with list style, list length and their interaction as predictors.

We restrict this analysis to Smartphone Respondents.
Generally, the longer the list, the more likely a choice from an earlier time slot is selected, compared to a later time slot, consistently across list style.

No differences noted in odds of selecting earlier (rather than later) time slots between Radio and Drop Down List Presentation.
Significant differences in the odds of selecting earlier times compared to later times were found between long list and short list smartphone respondents but not between medium and short list smartphone respondents.

Respondents assigned to the long list were 1.49 (95% CI: 1.16-1.91) times more likely to select an earlier time slot on the list compared to respondents assigned to the short list.

Respondents assigned to the medium list were 1.27 (95% CI: 0.996-1.63) times more likely to select an earlier time slot on the list compared to respondents assigned to the short list.
Miles Driven (sqrt) Consistently Higher when entered in Portrait Mode, Consistently across Scale Types
ANOVA Table for relationship between U1 and Scale-Type (Exp1 Factor) and LandScape Answered SCALE IS SQRT!

ANOVA Results for U1

Based on Main Effects Model.
With only landscape factor
Overall Satisfaction with Auto Insurance Provide – In Two Ways: Within Person Cross-Over Design*

5 Point Slider 2 Standard- o1

5 Point Standard 2 Slider- o1
### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
<th>25th</th>
<th>50th (Median)</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>v9_words</td>
<td>1198</td>
<td>11.35</td>
<td>11.004</td>
<td>0</td>
<td>84</td>
<td>3.00</td>
<td>9.00</td>
<td>16.00</td>
</tr>
<tr>
<td>v9_chars</td>
<td>1198</td>
<td>57.06</td>
<td>57.226</td>
<td>0</td>
<td>413</td>
<td>16.00</td>
<td>42.00</td>
<td>79.00</td>
</tr>
<tr>
<td>Assignment to V2T prompting and Use of V2T Feature</td>
<td>1143</td>
<td>1.7052</td>
<td>.81128</td>
<td>1.00</td>
<td>3.00</td>
<td>1.0000</td>
<td>1.0000</td>
<td>2.0000</td>
</tr>
</tbody>
</table>

#### Ranks

<table>
<thead>
<tr>
<th>Assignment to V2T prompting and Use of V2T Feature</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>V2T Prompt Group: No V2T Used</td>
<td>292</td>
<td>252.81</td>
<td>73820.50</td>
</tr>
<tr>
<td>V2T Prompt Group: Used V2T</td>
<td>257</td>
<td>300.21</td>
<td>77154.50</td>
</tr>
<tr>
<td>Total</td>
<td>549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V2T Prompt Group: No V2T Used</td>
<td>292</td>
<td>254.41</td>
<td>74288.00</td>
</tr>
<tr>
<td>V2T Prompt Group: Used V2T</td>
<td>257</td>
<td>298.39</td>
<td>76687.00</td>
</tr>
<tr>
<td>Total</td>
<td>549</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Test Statistics

<table>
<thead>
<tr>
<th></th>
<th>v9_words</th>
<th>v9_chars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>31042.500</td>
<td>31510.000</td>
</tr>
<tr>
<td>Wilcoxon W</td>
<td>73820.500</td>
<td>74288.000</td>
</tr>
<tr>
<td>Z</td>
<td>-.3497</td>
<td>-.3242</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.000</td>
<td>.001</td>
</tr>
<tr>
<td>Monte Carlo Sig. (2-tailed)</td>
<td>.000&lt;sup&gt;o&lt;/sup&gt;</td>
<td>.002&lt;sup&gt;o&lt;/sup&gt;</td>
</tr>
<tr>
<td>99% Confidence Interval</td>
<td>Lower Bound</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Upper Bound</td>
<td>.001</td>
</tr>
<tr>
<td>Monte Carlo Sig. (1-tailed)</td>
<td>.000&lt;sup&gt;o&lt;/sup&gt;</td>
<td>.001&lt;sup&gt;o&lt;/sup&gt;</td>
</tr>
<tr>
<td>99% Confidence Interval</td>
<td>Lower Bound</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Upper Bound</td>
<td>.000</td>
</tr>
</tbody>
</table>

---

a. [DEVICETYPE] Device type = Mobile/Smartphone
b. Grouping Variable: Assignment to V2T prompting and Use of V2T Feature
c. Based on 10000 sampled tables with starting seed 1795812538.
Description of the Three Experiments

**Experiment 1**

- **Scale Type**: Standard
- **Scale Length**: 5 Points
- **Scale Numbering**: Yes, Numbers + Anchors
- **Slider Start**: Outside

**Experiment 2 – 2 parts**

- **List Type**: Radio Buttons
- **List Length**: 6 choices

**Experiment 3**

- **Voice To Text**: Prompt Provided
- **Voice To Text Length**: No Prompt
### Demographics of Respondents by Mode

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Laptop</th>
<th>Tablet</th>
<th>Smartphone</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Male</td>
<td>73.9%</td>
<td>42.2%</td>
<td>38.3%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Percent BS or More</td>
<td>62.5%</td>
<td>68.5%</td>
<td>60.2%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent Married</td>
<td>75.9%</td>
<td>71.5%</td>
<td>60.4%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Percent Full-Time</td>
<td>26.6%</td>
<td>61.6%</td>
<td>72.2%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Percent Income &lt;$75K</td>
<td>44.0%</td>
<td>41.7%</td>
<td>56.5%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Median Age/Range</td>
<td>66 (19-93)</td>
<td>40 (19-83)</td>
<td>31 (18-79)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Median # Prior Surveys Last 30 Days</td>
<td>5.0</td>
<td>5.0</td>
<td>4.0</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>