



Google Brand Studio

Grading the Grids: What Works and What Doesn't Using Paradata to Assess Response Quality and Usability

Mario Callegaro, Yongwei Yang, and Natalie Rojowsky-Kessel: Google
Amy Hill, Marni Hirschorn, Rich Timpone, and Cecile Carre: Ipsos

September 2017



The problem we are trying to solve

Grids, tables or matrixes are probably the most challenging type of questions to show on different devices

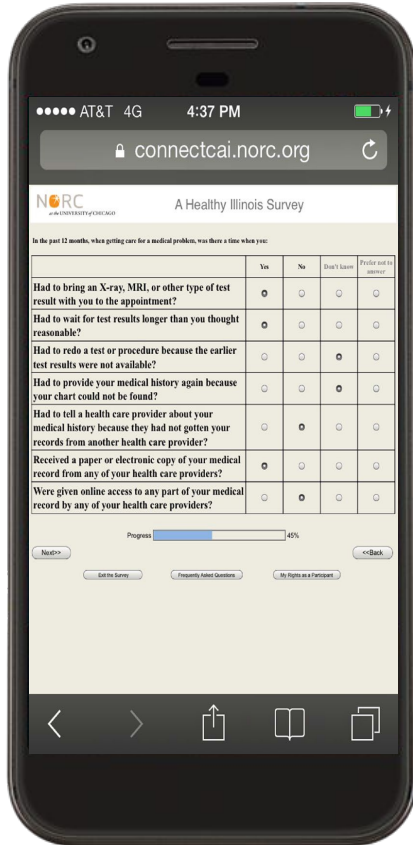
Over the years, different options have been tested, although no firm conclusion have been made in the literature

“We lack systematic research where potentially negative effect (e.g. lower data quality) of tables [...] would be compared with actual disadvantages (e.g. increased time and length) of a series of single radio button questions”

Callegaro, Lozar Manfreda & Vehovar, 2015, p.82

6 Ways To Show Grids on Smartphones

As is



Stern, Sterrett,
& Bilgen, 2016



Wenz,
2017

Stacking or banking

Thomas, Barlas,
Graham, & Subias,
2015

How important are the following
in deciding what beverage to
DRINK BETWEEN MEALS?

Select one answer from each row
in the grid

It is low in calories		
Not important	Important	Extremely important
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It is refreshing		
Not important	Important	Extremely important
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

It is a good value for the money		
Not important	Important	Extremely important
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

World Trade Center Health Registry
2015 Health Survey

Help Log Out

Q30. Have you ever been told by a doctor
or other health professional that you had
any of these conditions? Note: Cancer is
covered later in this survey.

Hypertension, or high blood pressure

Yes	No
<input type="radio"/>	<input type="radio"/>
<input type="button" value="Clear Answer"/>	

High cholesterol

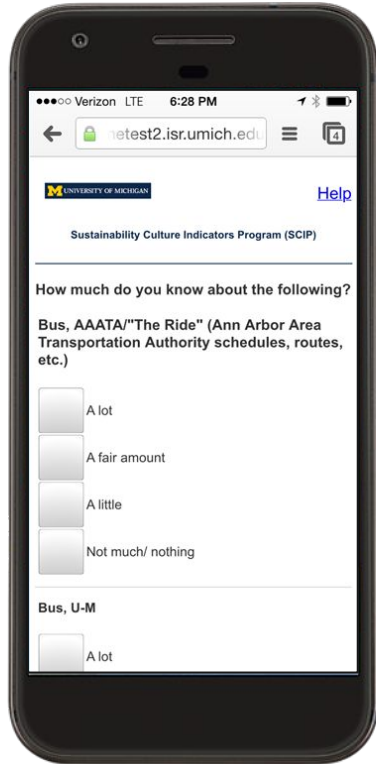
Yes	No
<input type="radio"/>	<input type="radio"/>
<input type="button" value="Clear Answer"/>	

Angina, or angina pectoris

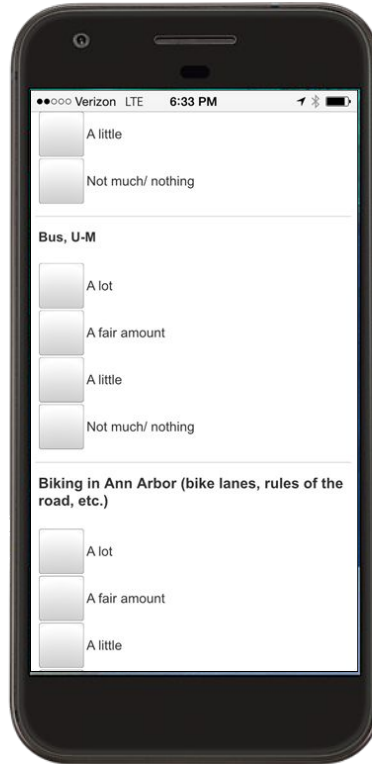
Yes	No
<input type="radio"/>	<input type="radio"/>

Richards, Powell,
Murphy, Nguyen &
Yu, 2016

Single questions with scrolling

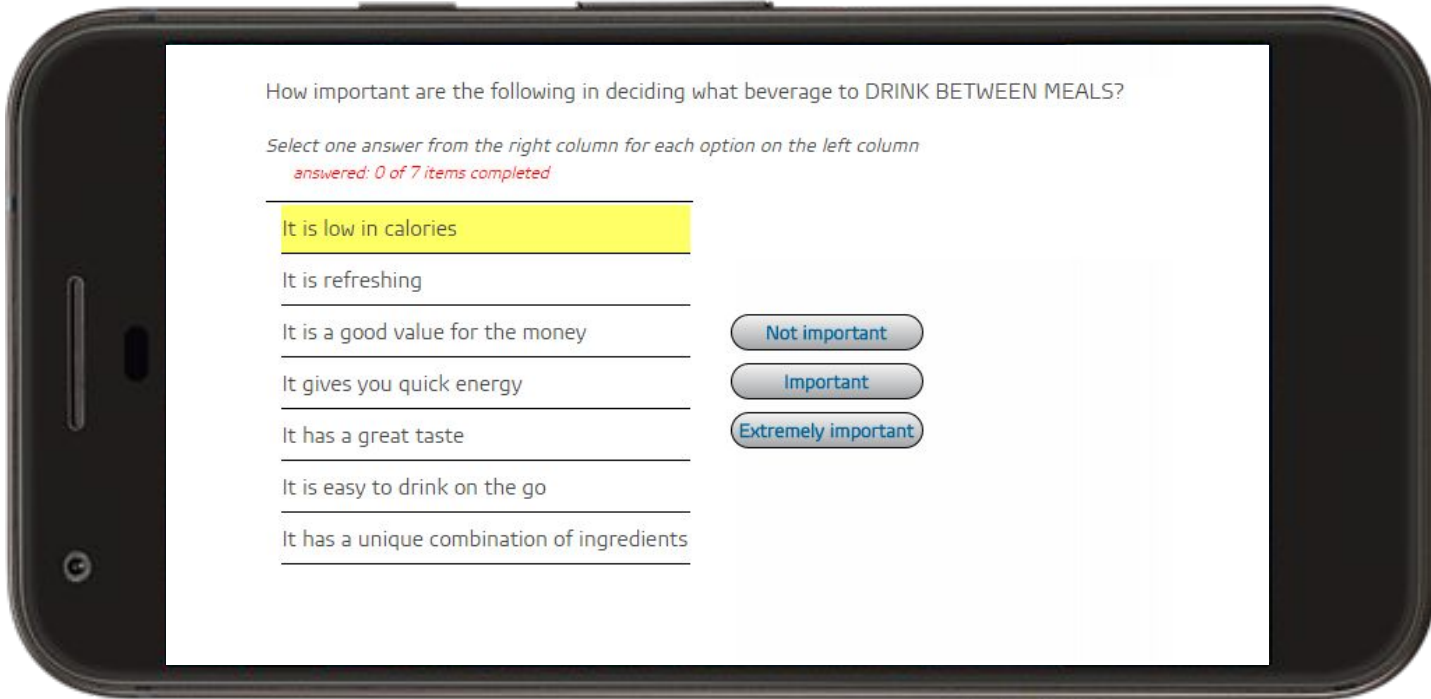


Couper,
2016



McGeeney,
2015

Two columns – responses on right (GFK)



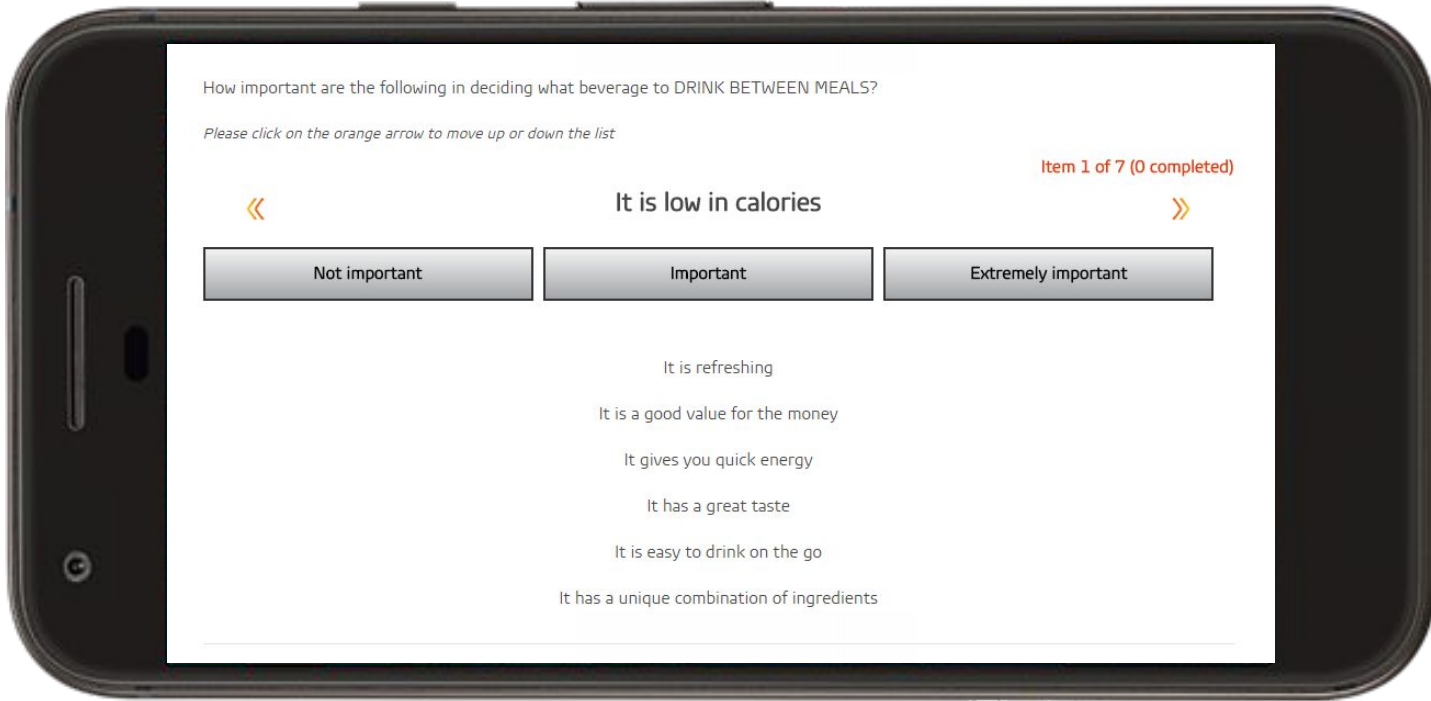
How important are the following in deciding what beverage to DRINK BETWEEN MEALS?

Select one answer from the right column for each option on the left column

answered: 0 of 7 items completed

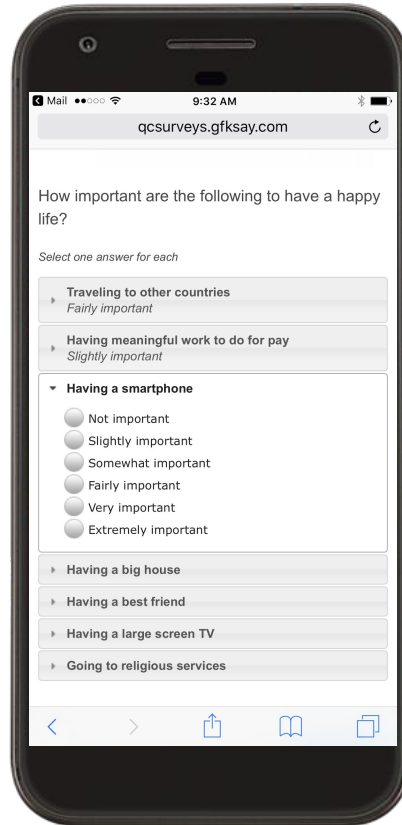
It is low in calories	
It is refreshing	
It is a good value for the money	Not important
It gives you quick energy	Important
It has a great taste	Extremely important
It is easy to drink on the go	
It has a unique combination of ingredients	

Focal element or progressive (GFK)



Accordion or collapsible/unfolding grid (GFK)

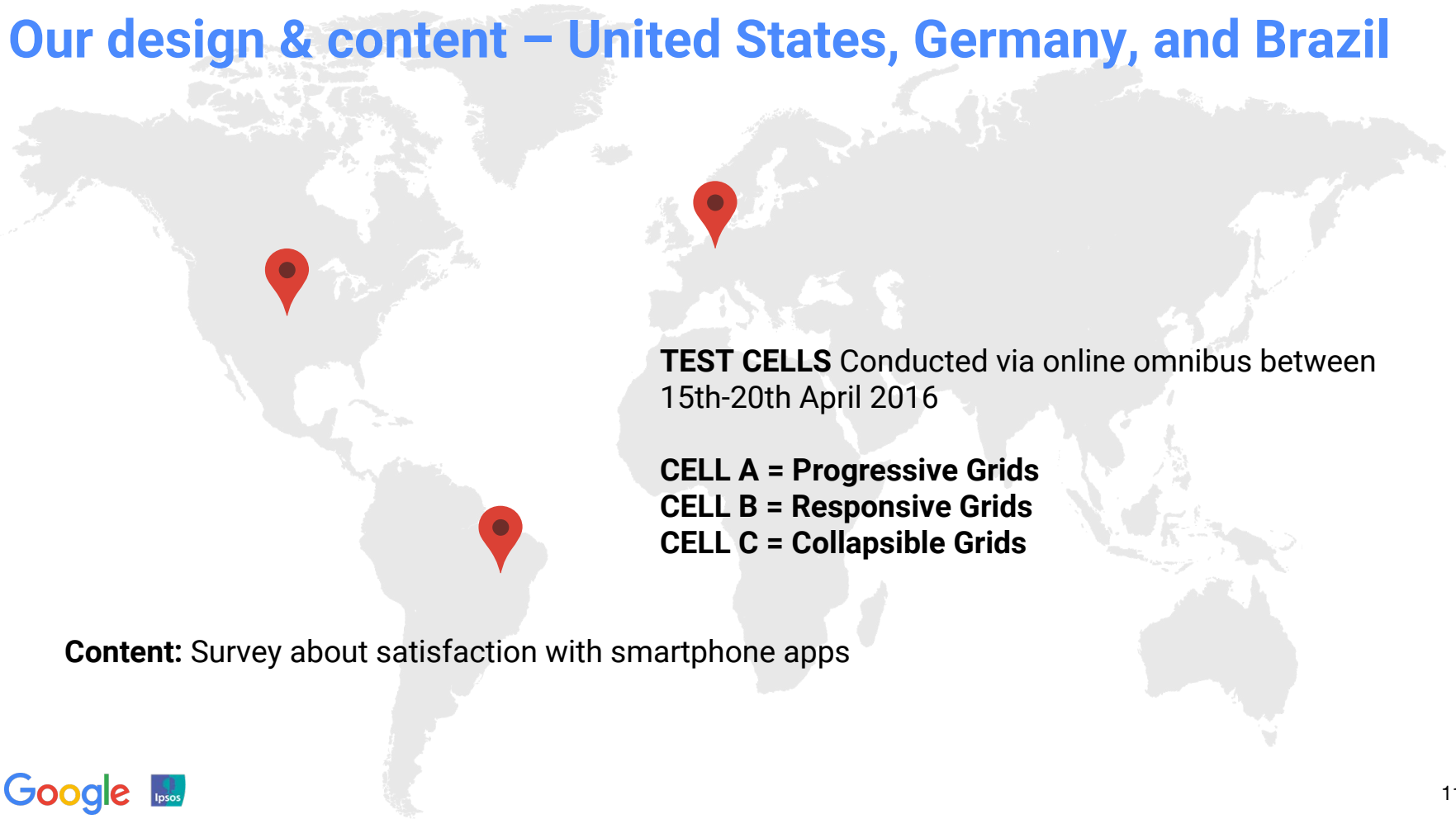
The GFK accordion opens the next available item automatically (auto advance)



Buttermore, Balas &
Thomas, 2017

Ipsos Experiment #1

Our design & content – United States, Germany, and Brazil



TEST CELLS Conducted via online omnibus between
15th-20th April 2016

CELL A = Progressive Grids
CELL B = Responsive Grids
CELL C = Collapsible Grids

Content: Survey about satisfaction with smartphone apps

Device

The majority still complete on a laptop/desktop and Brazil is the highest in smartphone completion. Switching is minimal.

The remainder of results from this test are filtered to Desktop completes only.

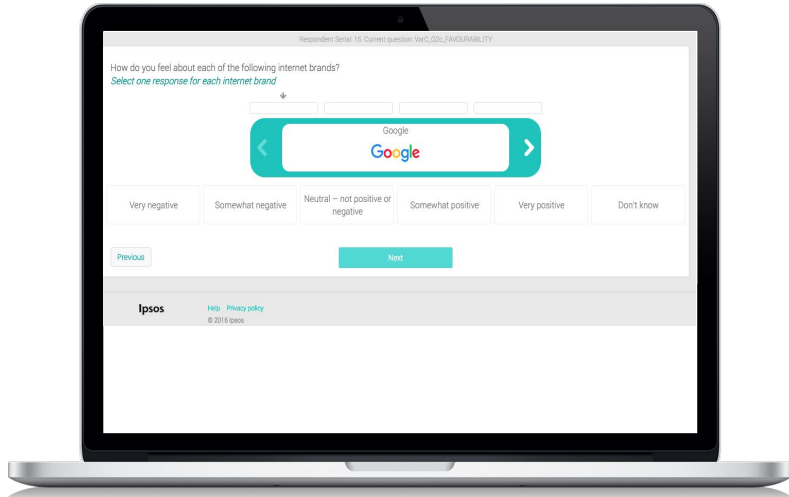
Final Device	BR	DE	US
Laptop/PC	79%	87%	84%
Smartphone	19%	5%	9%
Tablet	2%	8%	7%

Device Switched	Incidence	n=
BR	0%	1
DE	0.8%	17
US	0.5%	10

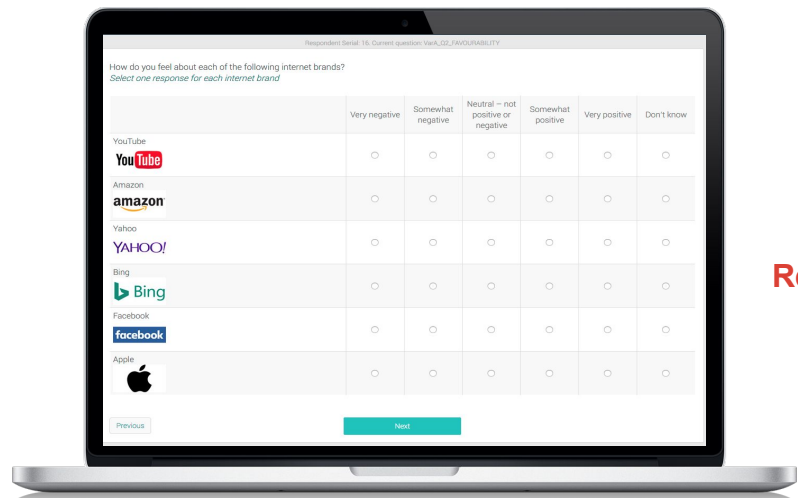
No quotas were applied by device for this test.



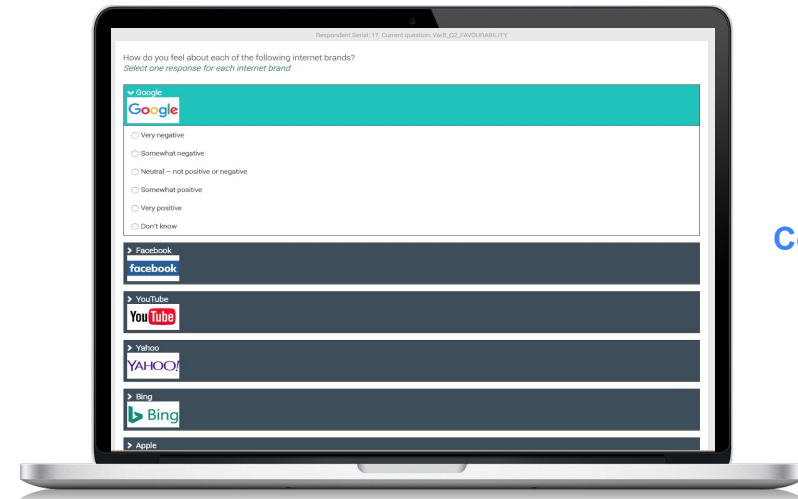
Visual Design: Desktop



Progressive grid
(with auto-advance)



Responsive grid



Collapsible grid

Quality indicator: % of item nonresponse prompts Favorability (Likert scale)

Collapsible grids are significantly more likely to have errors than both other Test Cells in all markets.

No errors were seen at all for Progressive and Responsive grids.

US	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	100%	96%
1 error	-	-	3%
2 errors or more	-	-	1%
DE	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	100%	88%
1 error	-	-	10%
2 errors or more	-	-	2%
DE	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	100%	90%
1 error	-	-	6%
2 errors or more	-	-	4%

Quality indicator: % of item nonresponse prompts

Google Mission (Discrete Analog)

Collapsible grids again are most likely to have errors.

While Responsive grids have more errors than Progressive grids, Progressive grids with auto-advance are designed to be error free.

US	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	98%	95%
1 error	-	2%	4%
2 errors or more	-	-	1%
DE	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	96%	95%
1 error	-	4%	4%
2 errors or more	-	-	1%
BR	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
No errors	100%	95%	91%
1 error	-	4%	5%
2 errors or more	-	1%	4%



Survey length: Time To Complete

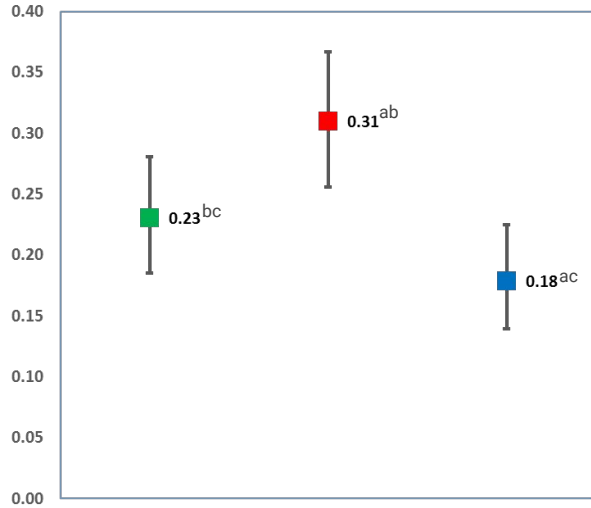
Responsive Grids are consistently the fastest Test cell.

US	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
Favorability (per brand)	30s	23s	34s
Google Mission	46s	35s	50s
DE	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
Favorability (per brand)	31s	23s	37s
Google Mission	45s	36s	52s
BR	Cell A Progressive Grid	Cell B Responsive Grid	Cell C Collapsible Grid
Favorability (per brand)	41s	29s	43s
Google Mission	51s	49s	65s

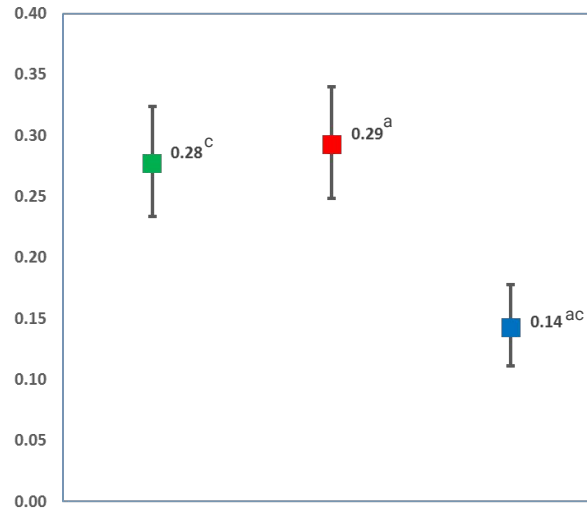
Quality indicator: Straightlining in the Prequel Study

Probability of Straightlining – PC/Tablet Platform models with age and gender controls; plus education in the US and Germany

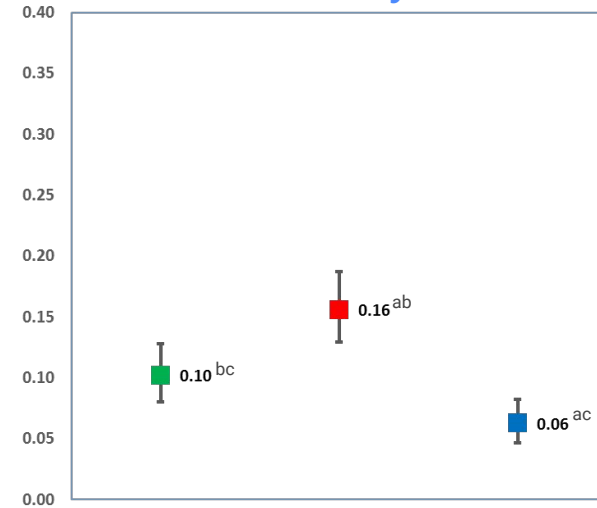
United States



Brazil



Germany



Progressive Grid



Responsive Grid



Collapsible Grid



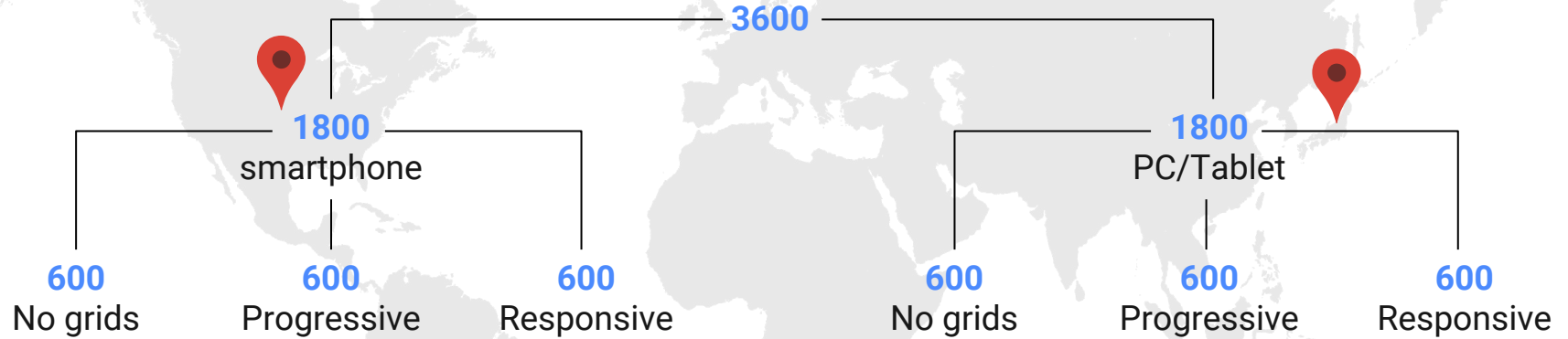
Collapsible grid (Ipsos) vs. accordion grid (GFK)

The difference between the Ipsos collapsible grid and the GFK accordion grid was that the Ipsos grid did not automatically open the next item (auto advance).

This reason helps explain the high amount of nonresponse prompts in our study.

Ipsos Experiment #2

Our design & content – United States and Japan

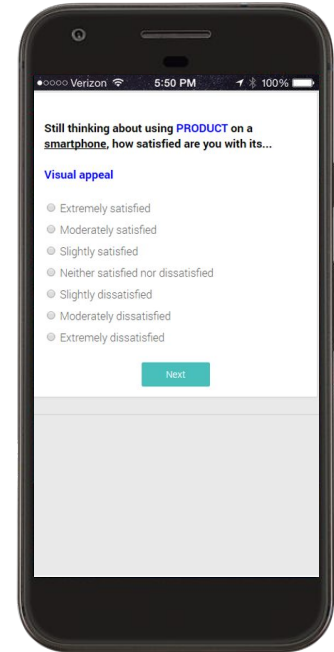
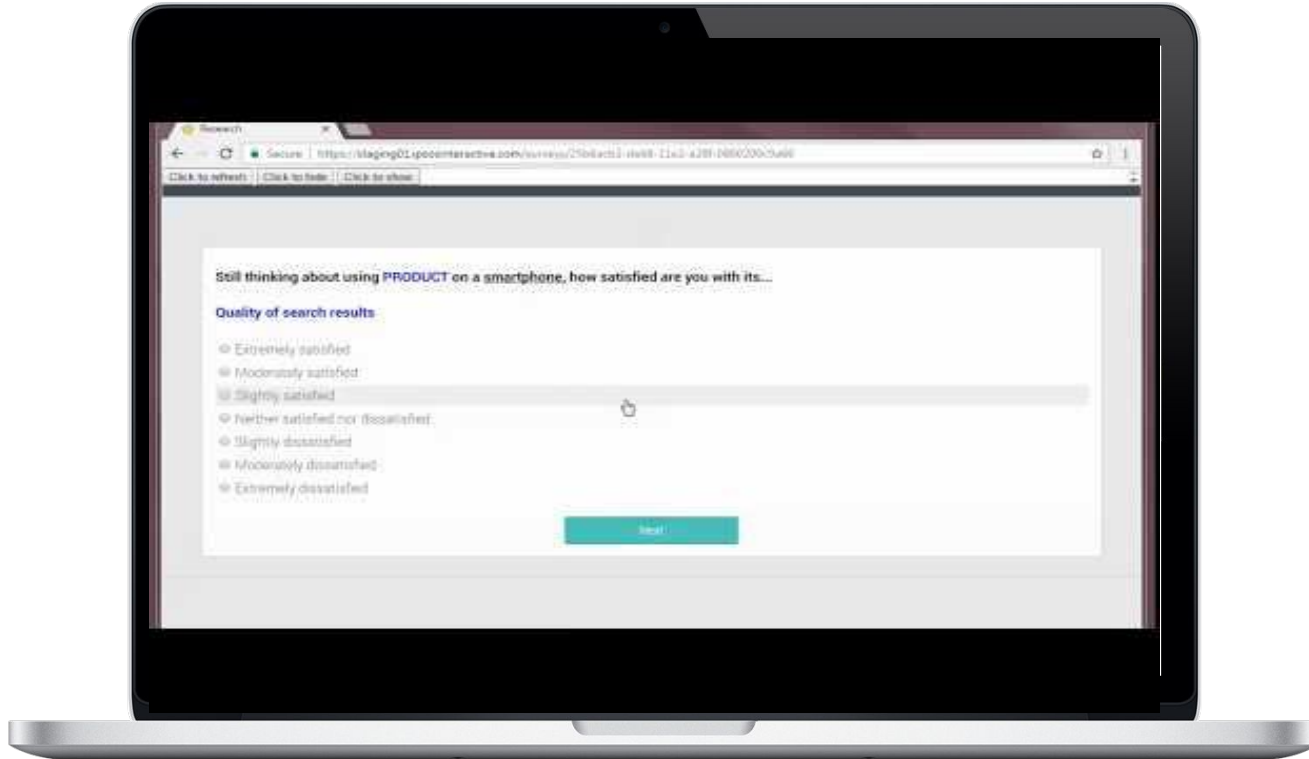


Content: Survey about satisfaction with smartphone apps

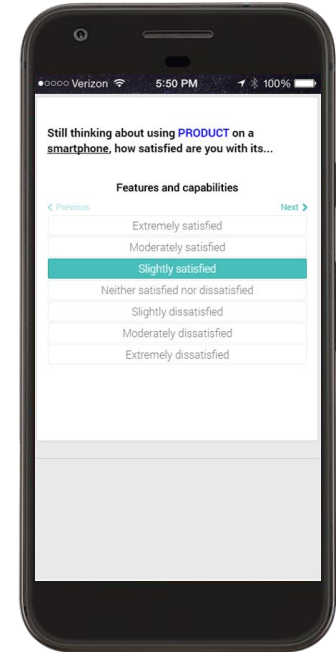
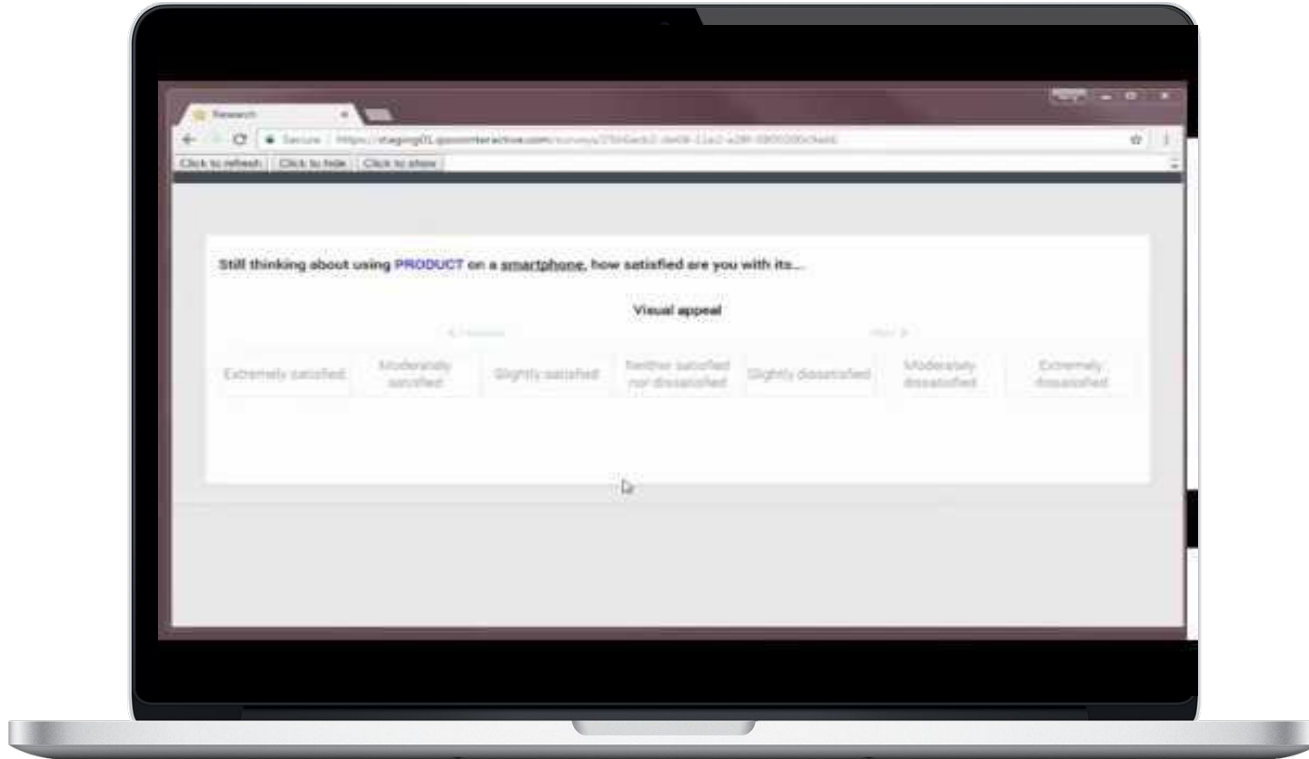
Respondents opted-in to the survey and were recruited through online panels

Visual Design for our Experiment

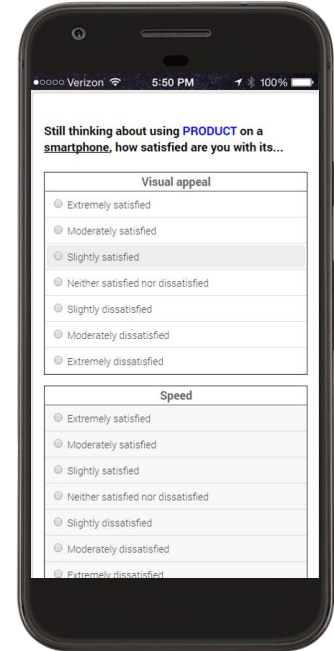
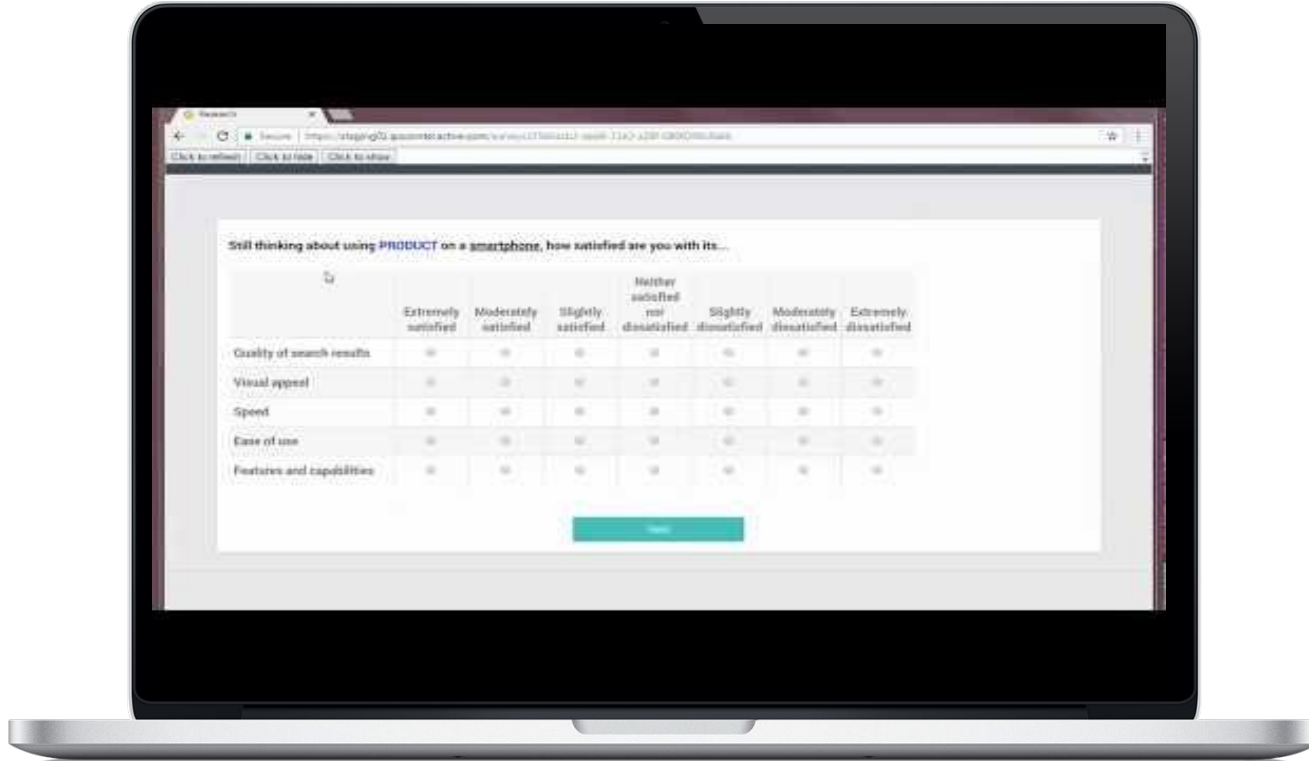
No grids: single item per screen



Progressive grids



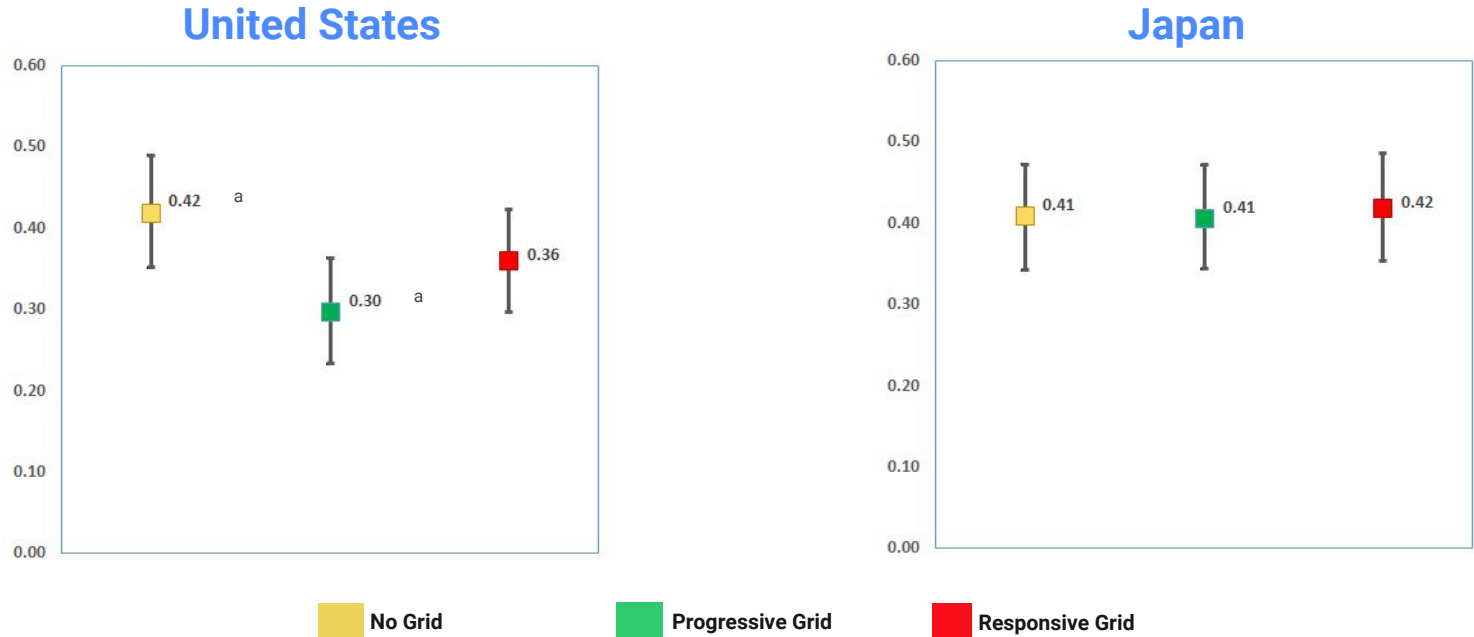
Responsive grids



Impact on Data Quality

Quality indicator: Straightlining

Probability of Satisfaction Straightlining – PC/Tablet Platform

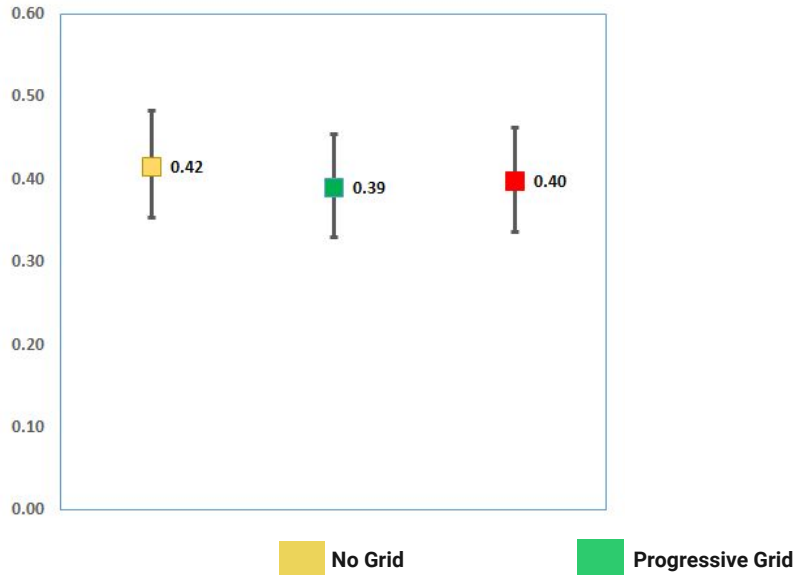


^aStatistically significant difference at 95% confidence level
Probabilities and distributions simulated from logit models including demographic controls based on 5,000 bootstraps.
Base: Total Completes, US weighted analysis N= 1994 Japan unweighted analysis N=2,261
Counts represent the total respondents whose 1st product evaluated had 5 questions for consistency.

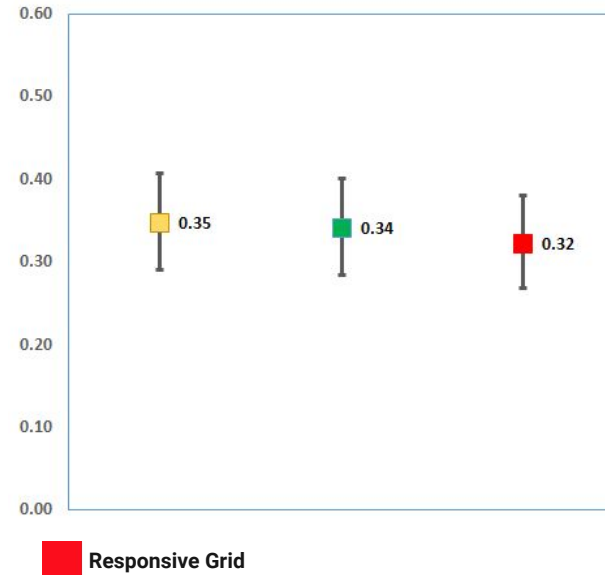
Quality indicator: Straightlining

Probability of Satisfaction Straightlining - Smartphone Platform

United States



Japan



Probabilities and distributions simulated from logit models including demographic controls based on 5000 bootstraps.

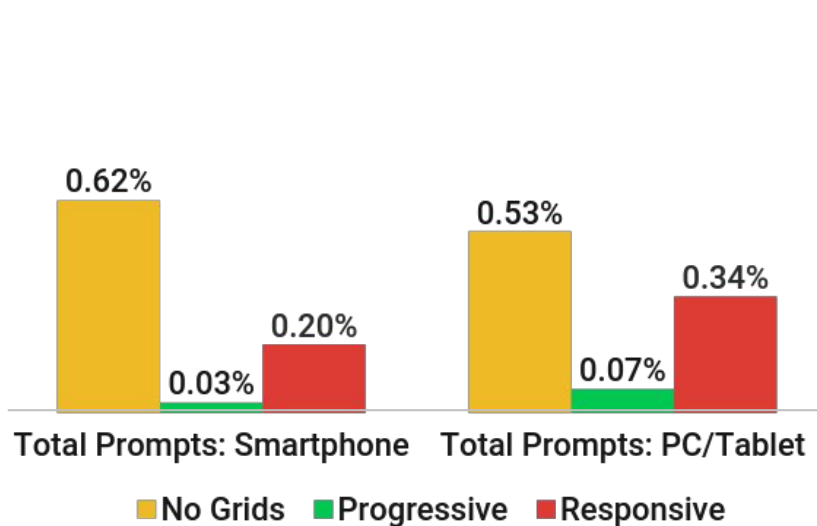
Base: Total Completes, US weighted analysis N= 1994 Japan unweighted analysis N=2261

Counts represent the total respondents whose 1st product evaluated had 5 questions for consistency.

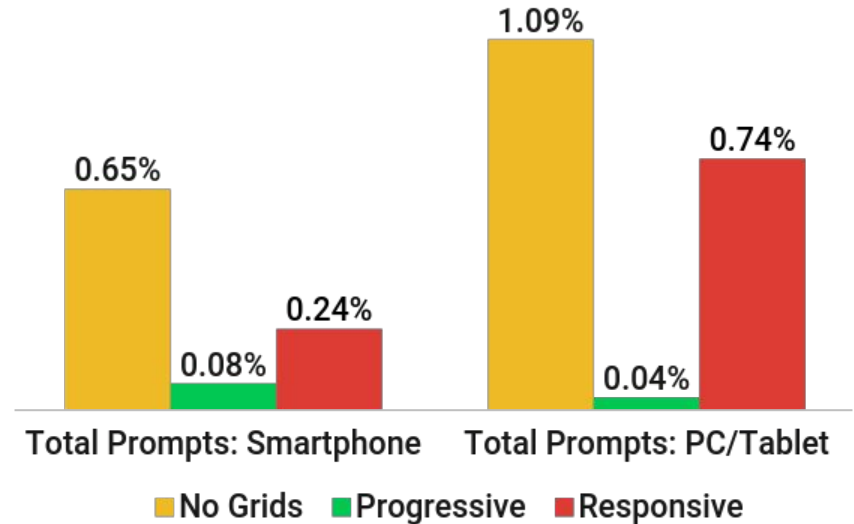
Quality indicator: % of item nonresponse prompts

Single questions per screens (no grids) had more item nonresponse prompts regardless of device

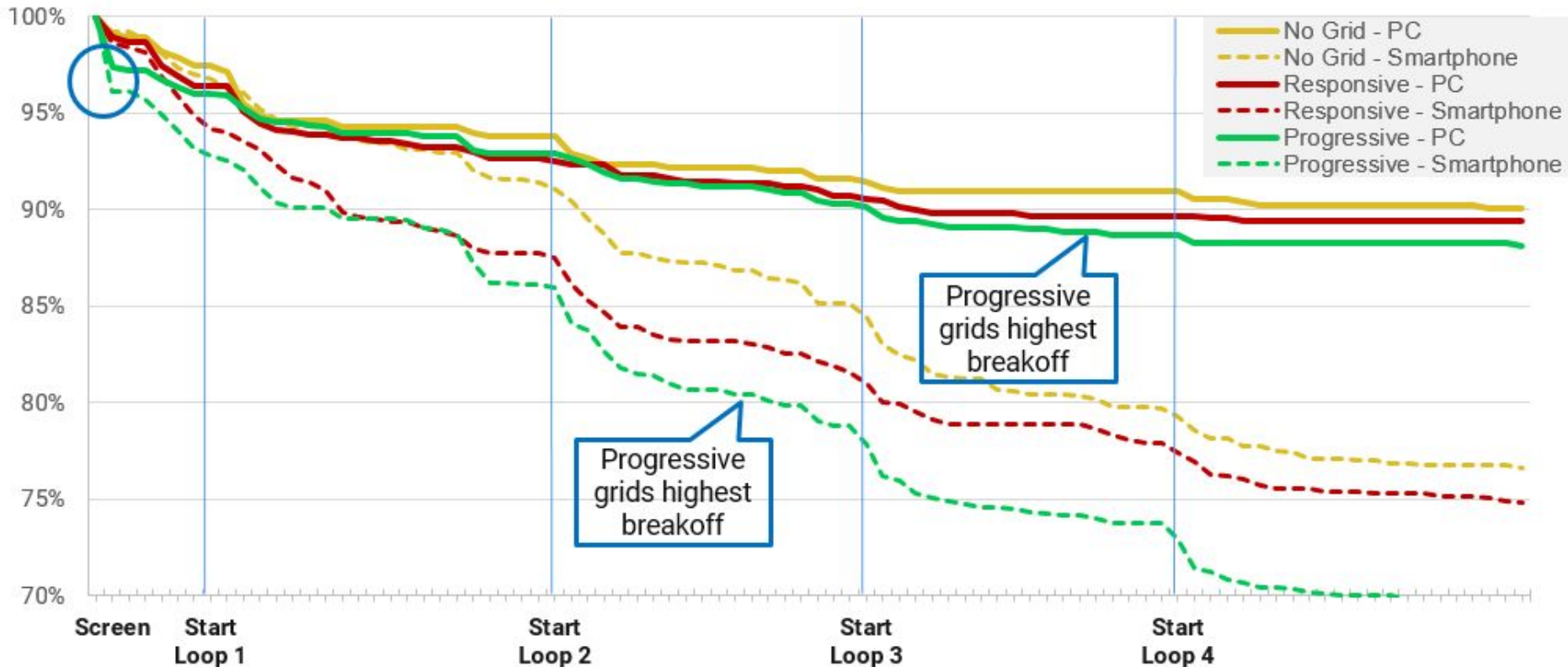
United States



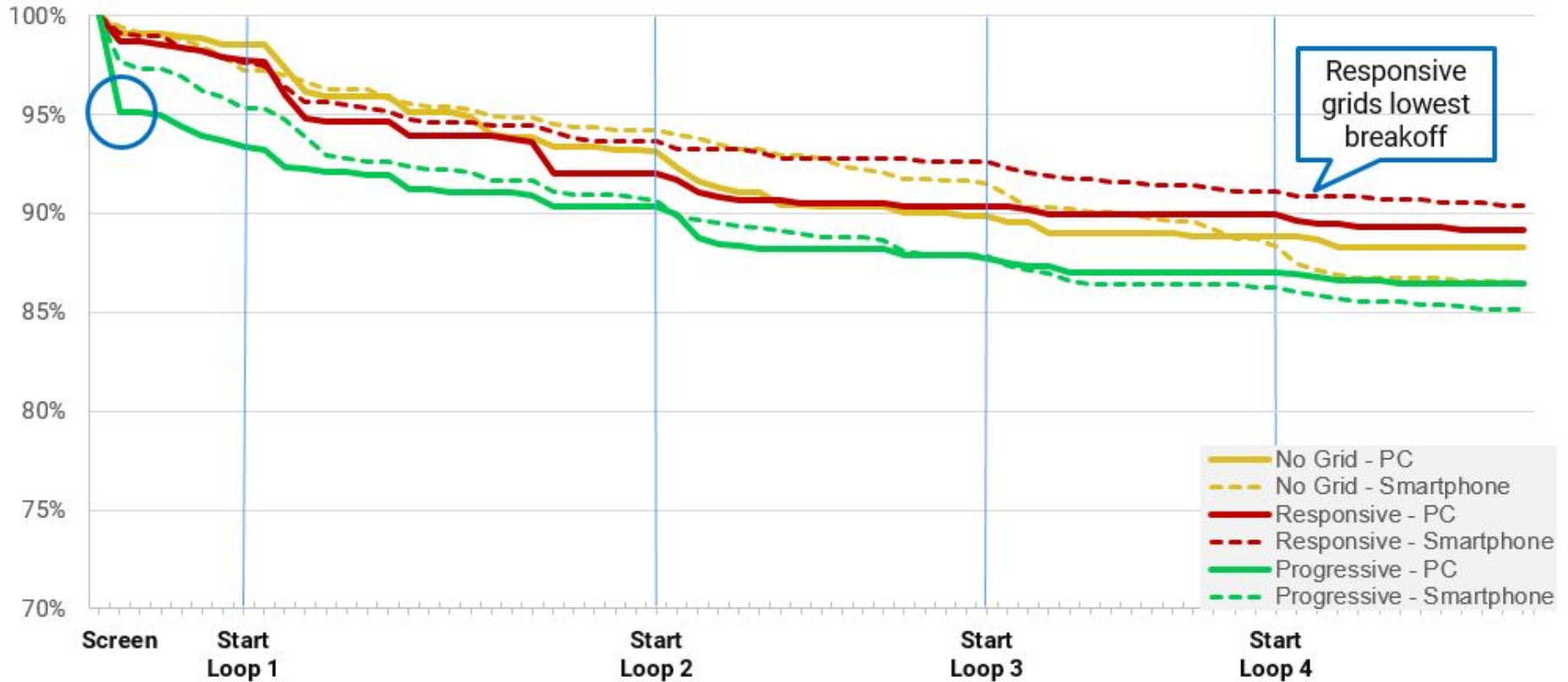
Japan



Quality indicator: Breakoff by grid condition – United States (%)



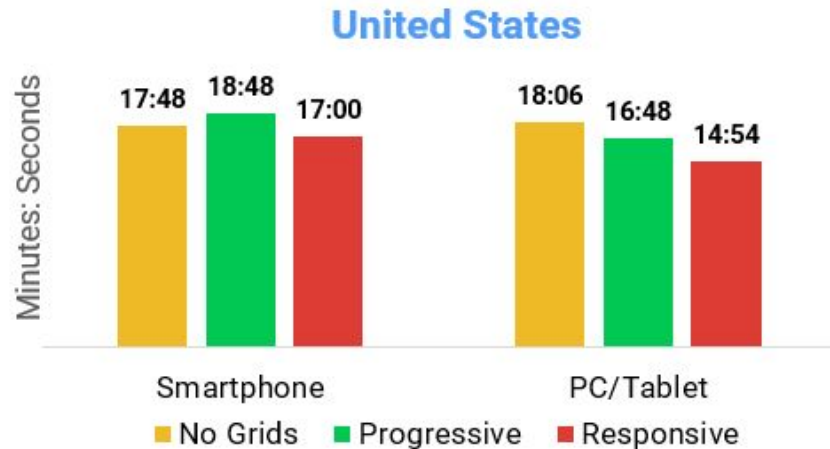
Quality indicator: Breakoff by grid condition – Japan (%)



Median survey length

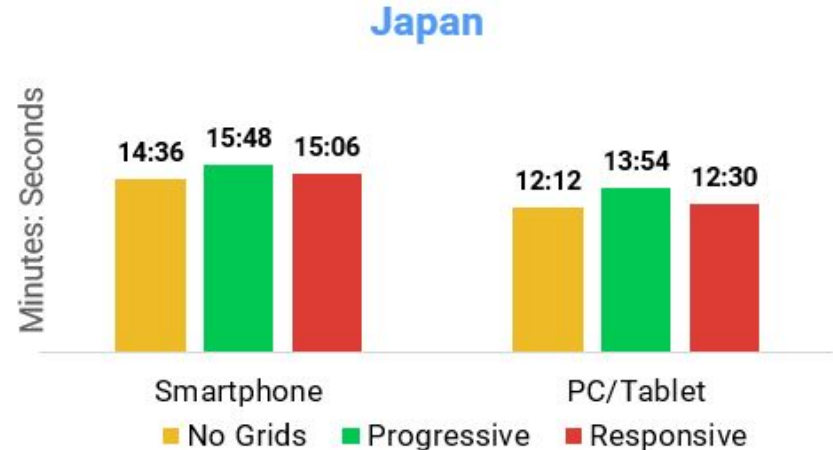
In both markets, PC/Tablet respondents were faster than their smartphone counterparts

The progressive grids required more time to complete for smartphones in the U.S.



Smartphone χ^2 (2): 12.83**

PC/Tablet χ^2 (2): 20.06**



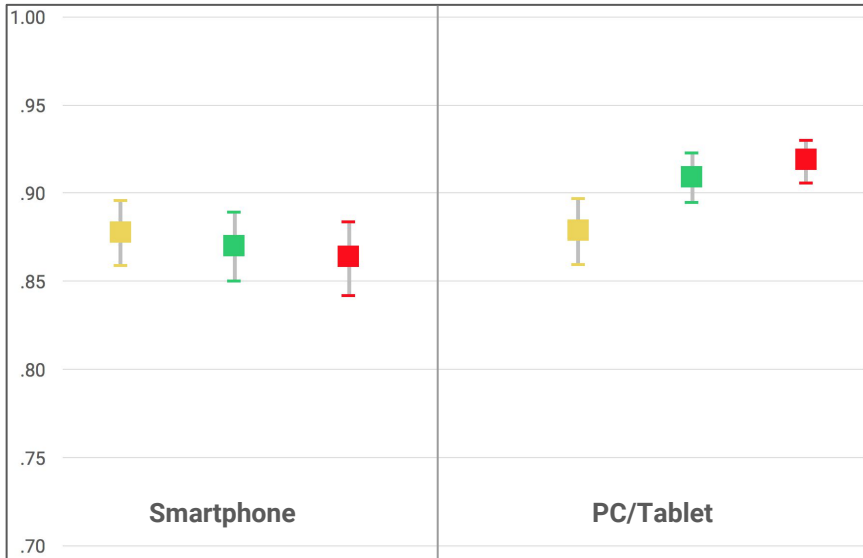
Smartphone χ^2 (2): 3.31

PC/Tablet χ^2 (2): 7.55*

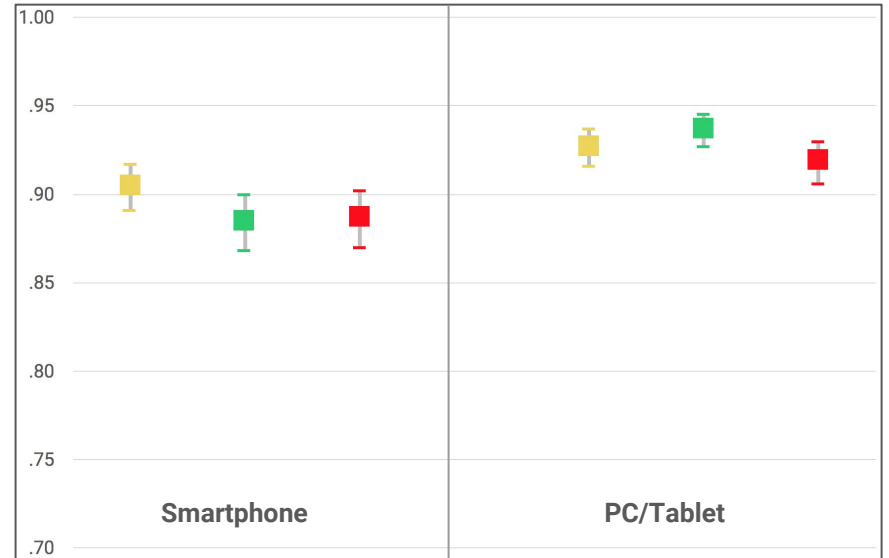
Cronbach Alpha

Differences were negligible ...

United States



Japan



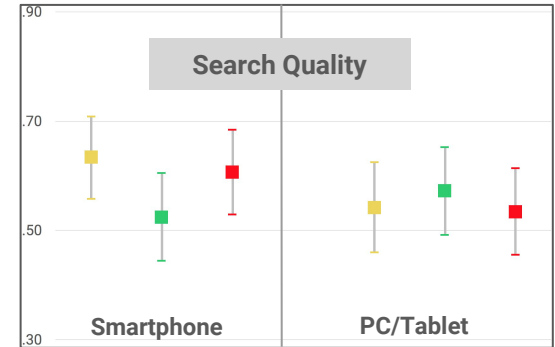
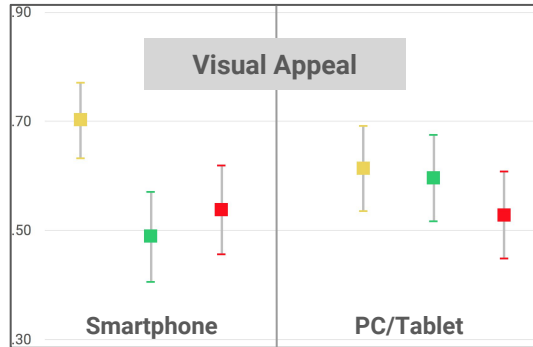
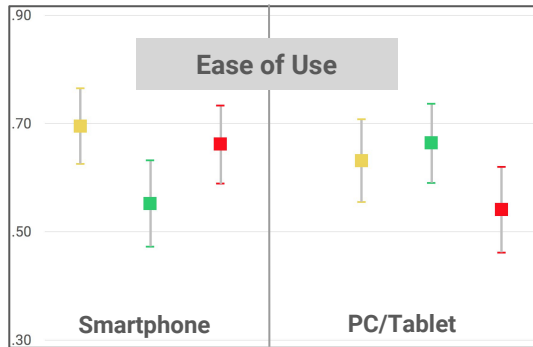
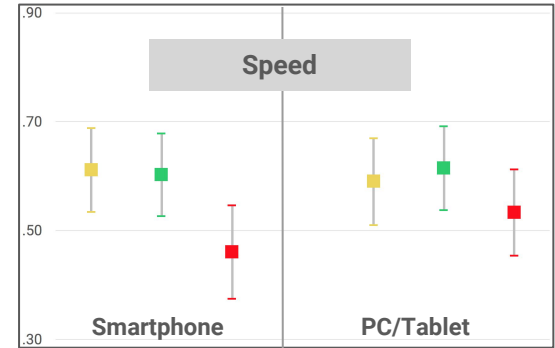
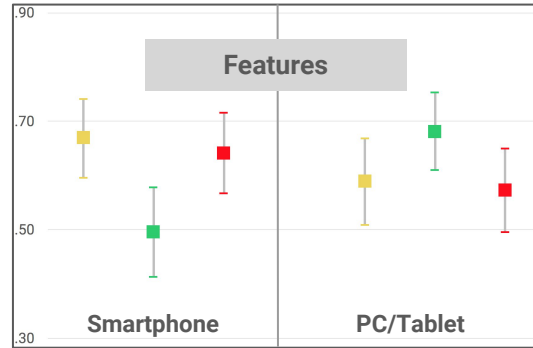
No Grid

Progressive Grid

Responsive Grid

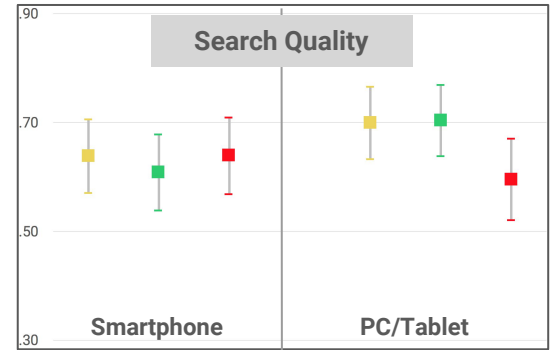
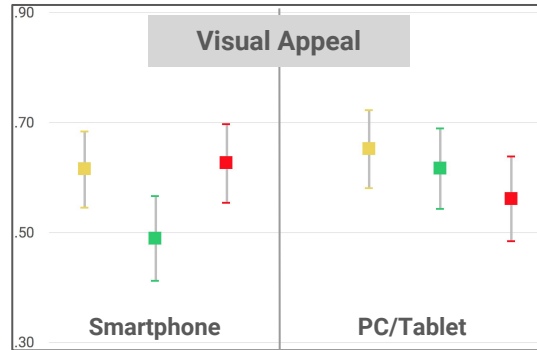
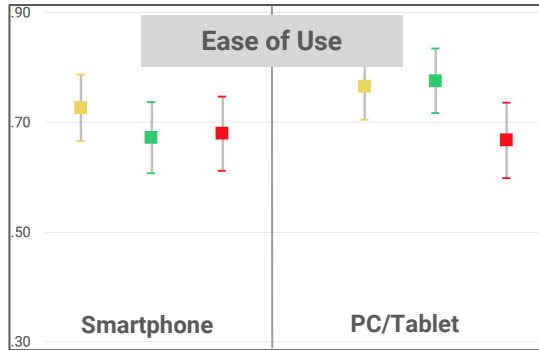
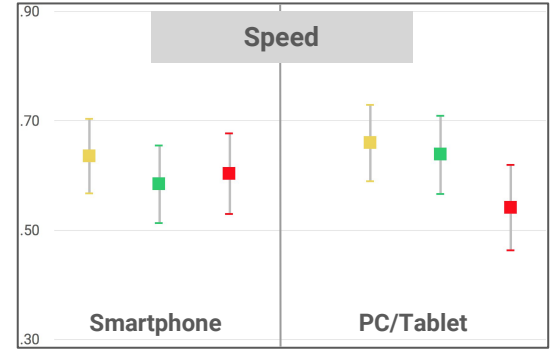
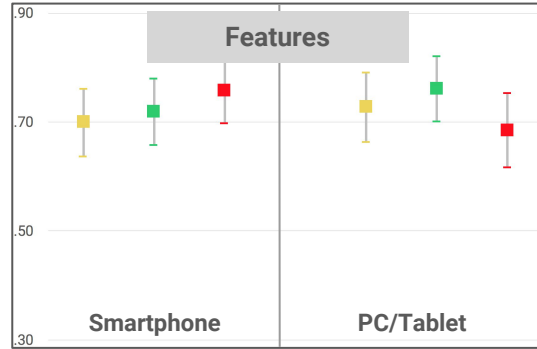
Correlations with Overall Satisfaction – United States

No clear pattern



Correlations with Overall Satisfaction – Japan

No clear pattern



Conclusions



Conclusions

- ✓ The no grid condition (one question per screen) resulted in the lowest breakoff-rate in U.S. and not necessarily the longest completion time
- ✓ Progressive grids consistently produced the highest amount of breakoffs across platforms and countries
- ✓ Progressive grids had the lowest amount of straightlining but only in the US – PC/Tablet condition
- ✓ Responsive grids are the fastest to complete on PC and Tablets and had the lowest breakoff-rate in Japan
- ✓ The relationship across variables is not affected systematically by the presentation of the grid

Future Research

Redo some analyses:

- median time
- breakoff rates
- relationship among variables

with multivariate models to fully explore patterns of data and control for demographics, attrition, and other characteristics related to using smartphones

Set up a research agenda on grids based on theory and usability principles



How many rows or columns in a grid?

SurveyMonkey experiment (Grady & Liu, 2017)

Manipulated grids:

Rows: 5, 10, 20

Columns: 3, 5 and 7

Outcome:

Lowest breakoff rate and highest subjective satisfaction: 5 rows

No statistically significant differences among the number of columns

Google Brand Studio

*thank
you!*

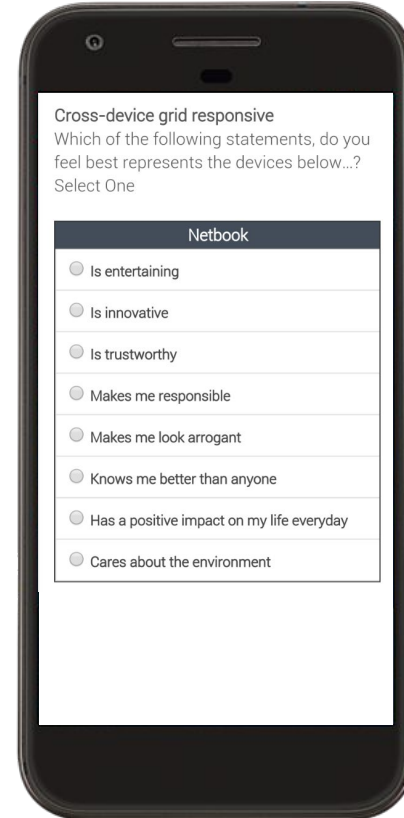
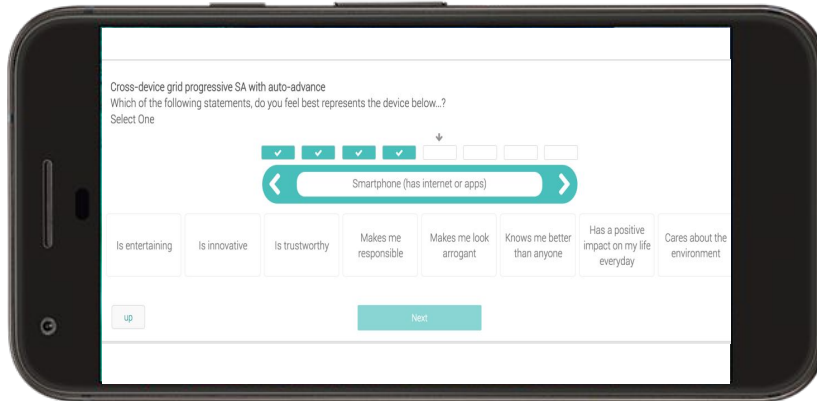
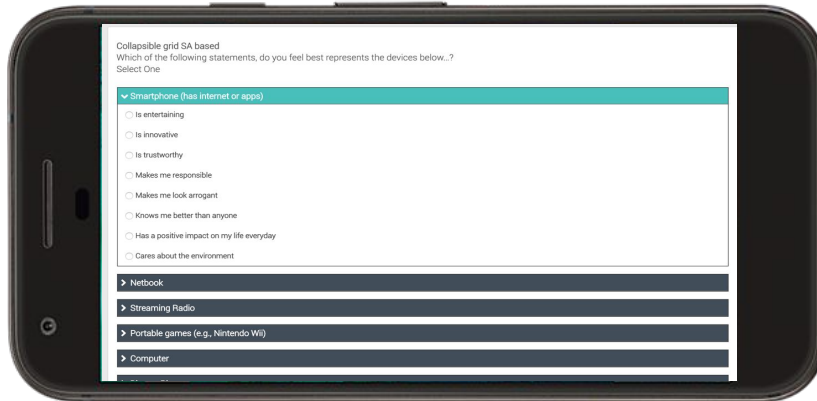
Mario Callegaro Ph.D.

Senior Survey Research Scientist

callegaro@google.com

Appendix

Smartphone view Experiment #1



No Grids: single item per screen - JP



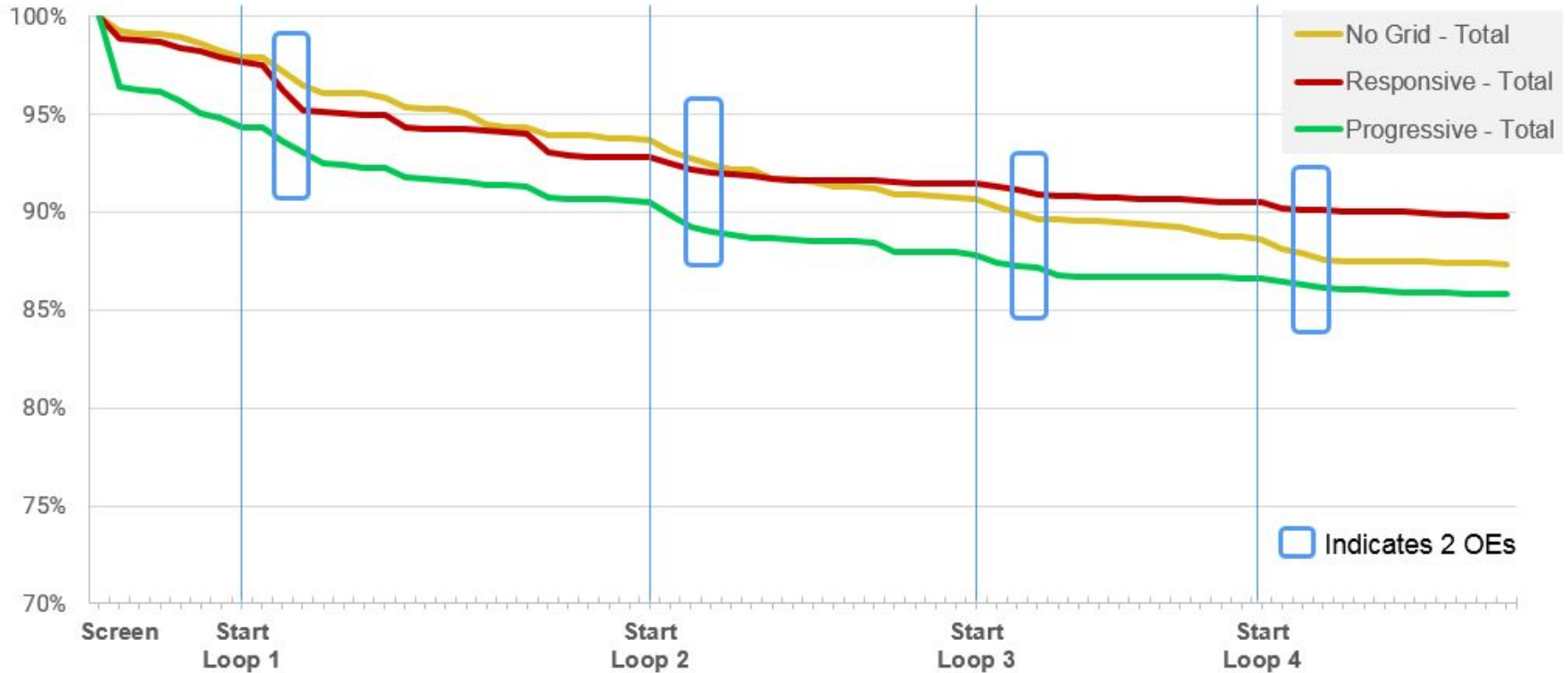
Progressive Grids - JP



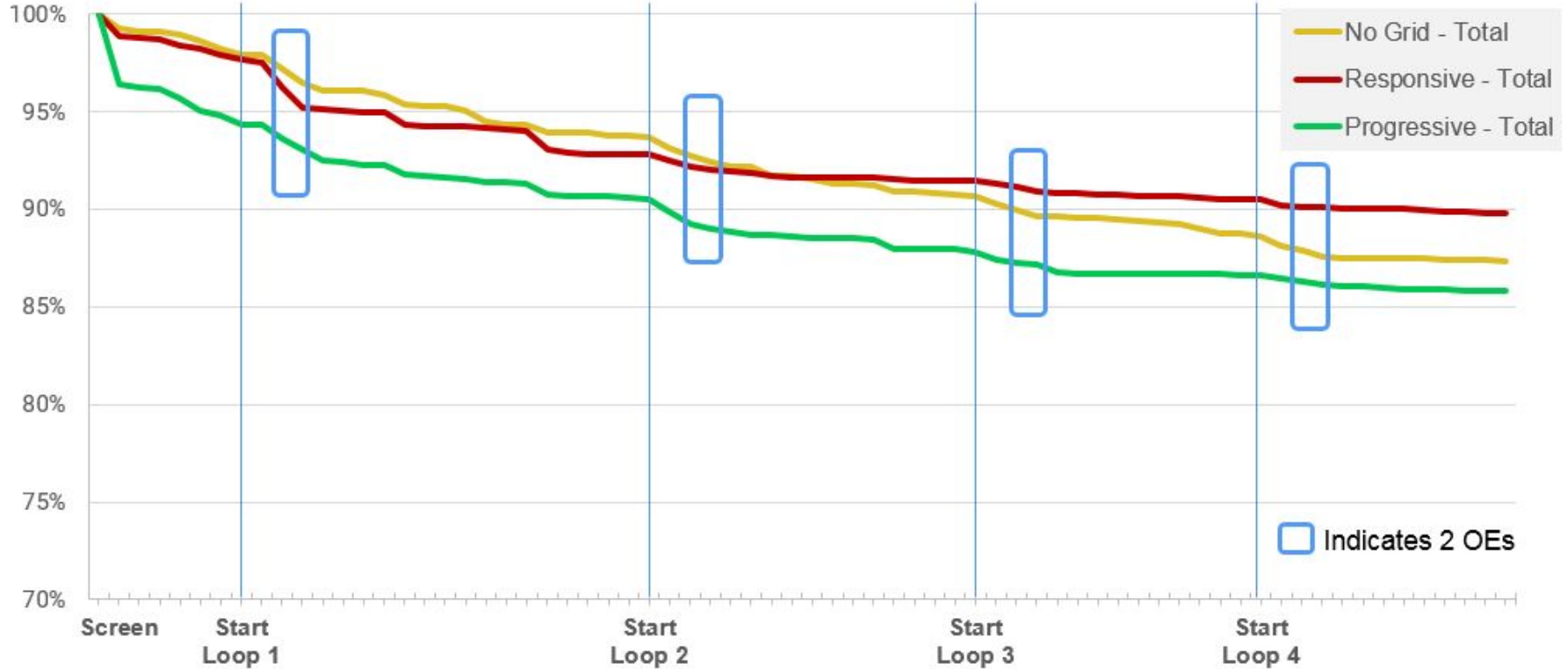
Responsive grids - JP



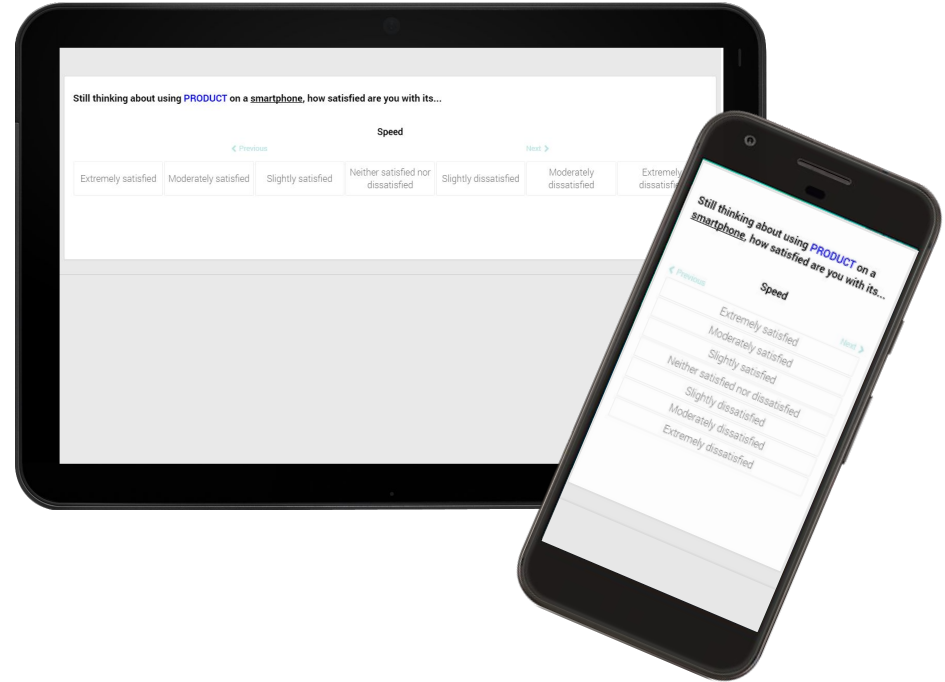
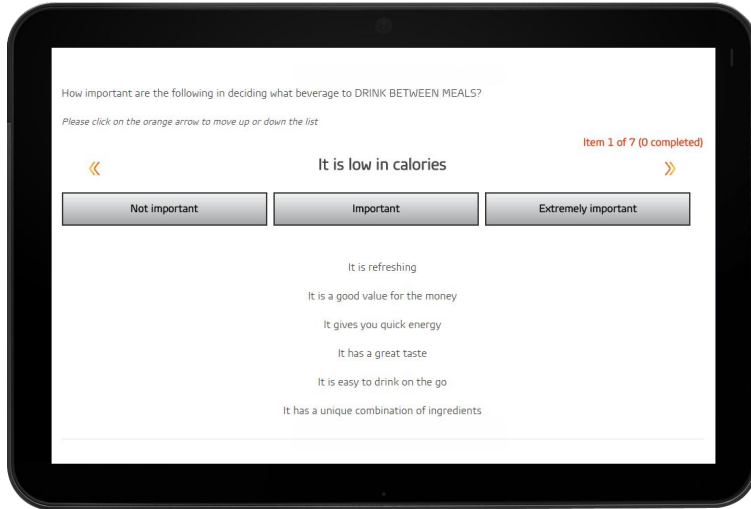
Quality indicator: Breakoff by grid condition – US (%)



Quality indicator: Breakoff by grid condition – Japan (%)

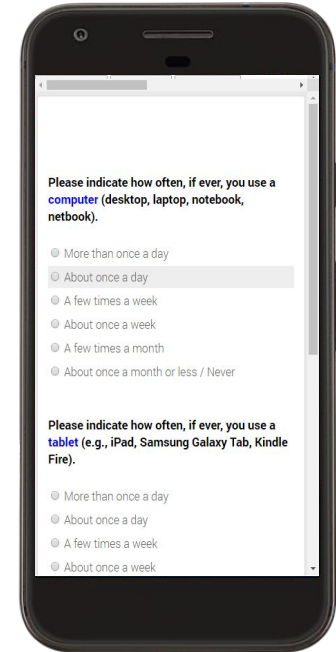
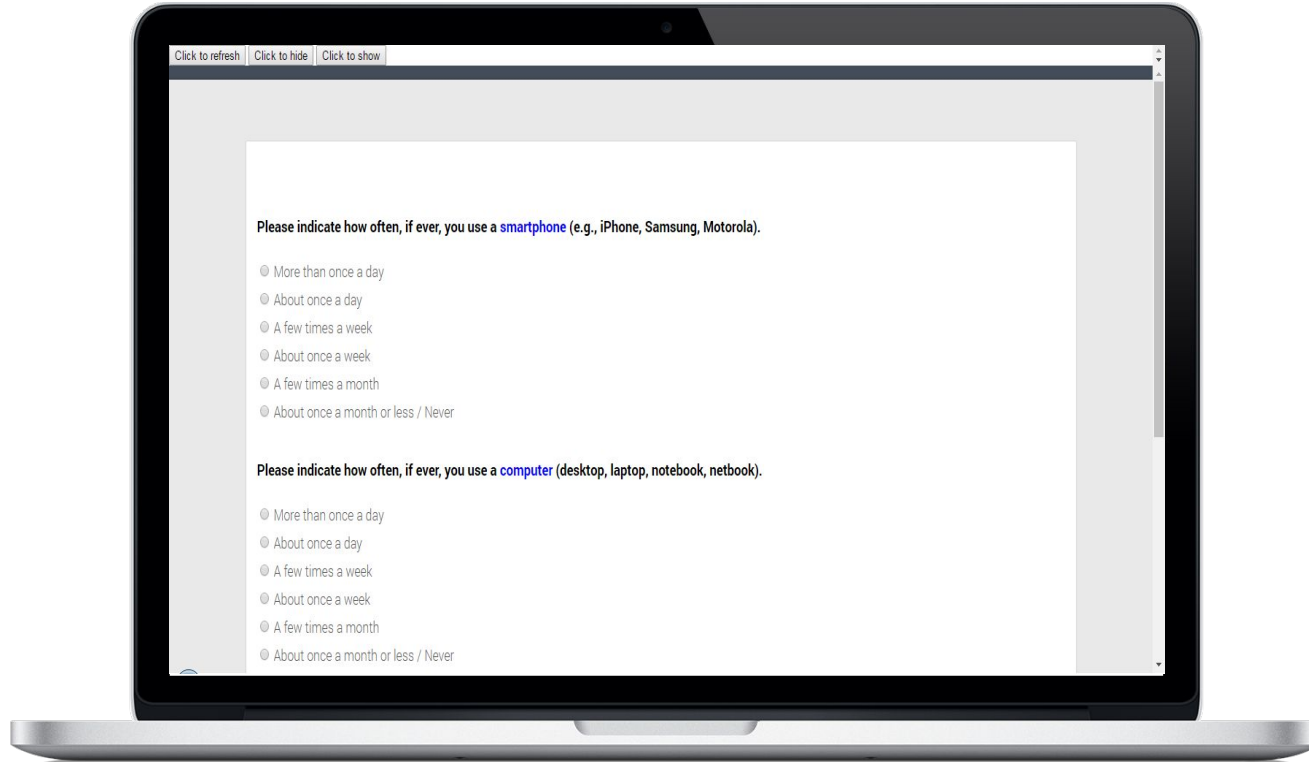


Comparison of Focal Element vs. Progressive grid

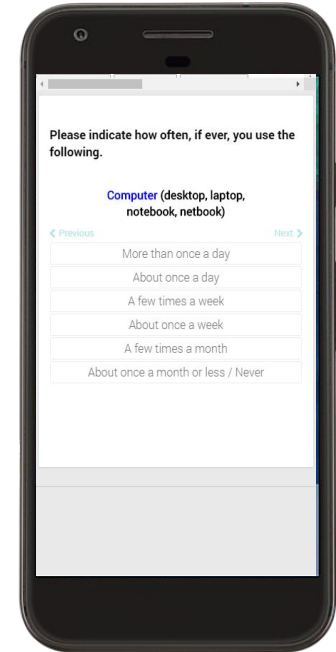
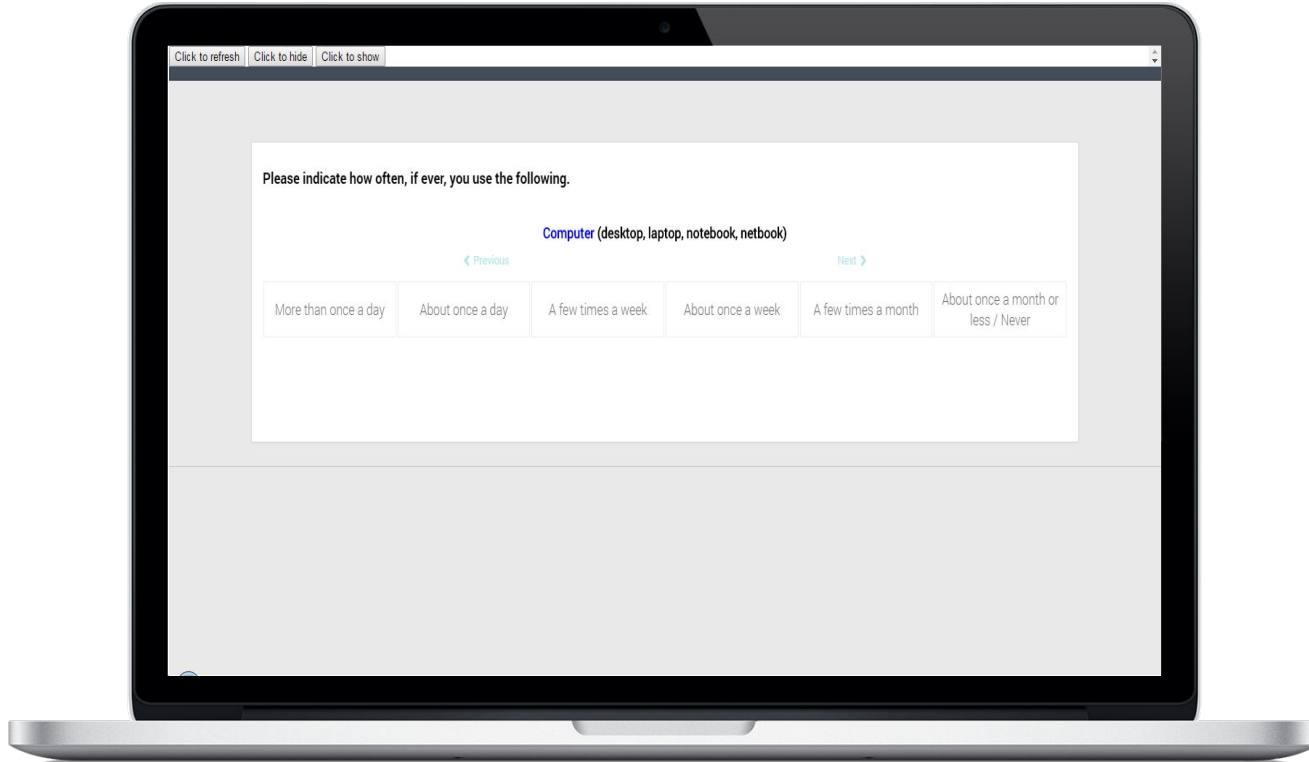


Thomas, Barlas, Graham,
& Subias, 2015

Initial exposure – US, no grid



Initial exposure – US, progressive grid



Initial exposure – US, responsive table

Click to refresh | Click to hide | Click to show

Please indicate how often, if ever, you use the following.

	More than once a day	About once a day	A few times a week	About once a week	A few times a month	About once a month or less / Never
Computer (desktop, laptop, notebook, netbook)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Smartphone (e.g., iPhone, Samsung, Motorola)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tablet (e.g., iPad, Samsung Galaxy Tab, Kindle Fire)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Next

Please indicate how often, if ever, you use the following.

Computer (desktop, laptop, notebook, netbook)

- More than once a day
- About once a day
- A few times a week
- About once a week
- A few times a month
- About once a month or less / Never

Smartphone (e.g., iPhone, Samsung, Motorola)

- More than once a day
- About once a day
- A few times a week
- About once a week
- A few times a month
- About once a month or less / Never

Impact of Grids

First Difference Calculations for Demographic Groups

US Analysis

	First Difference Calculations		
	Change in prob.	95% Conf Int.	
Female	0.046*	0.000	0.092
Age (25-53)	0.057*	0.009	0.105
Black	0.056	-0.017	0.126
Hispanic	-0.049	-0.118	0.024
College Grad	-0.051*	-0.098	-0.005

Japan Analysis

	First Difference Calculations		
	Change in prob.	95% Conf Int.	
Female	-0.019	-0.061	0.023
Age (25-53)	0.058*	0.012	0.102
College Grad	-0.025	-0.069	0.017

Models Used for Calculations

Logit Models Used to Simulate Probabilities and Distributions

US Analysis

Variable	Coefficient
Smartphone	-0.013
Grid1	-0.543**
Grid2	-0.249
Phone*Grid1	0.435+
Phone*Grid2	0.171
Female	0.189*
Age	0.008*
Black	0.229
Hispanic	-0.209
College Grad	-0.218*
Constant	-0.651**

Japan Analysis

Variable	Coefficient
Smartphone	-0.263+
Grid1	-0.013
Grid2	0.041
Phone*Grid1	-0.014
Phone*Grid2	-0.156
Female	-0.079
Age	0.009*
College Grad	-0.106
Constant	-0.732**

LR χ^2 (8): 32.03**

Wald χ^2 (10): 31.60**

** p<.01, * p<.05, + p<.10

Base: Total Completes, US weighted analysis N= 1994 Japan unweighted analysis N=2261

Counts represent the total respondents whose 1st product evaluated had 5 questions for consistency.



Grids in general: summary of the survey research literature

“The major advantage of matrix questions is the efficient use of space [...] these types of questions are relative difficult for respondents since so much text is presented on a single screen” (Toepoel, 2016, p. 148)

“We lack systematic research where potentially negative effect (e.g. lower data quality) of tables [...] would be compared with actual disadvantages (e.g. increased time and length) of a series of single radio button questions” (Callegaro, Lozar Manfreda & Vehovar, 2015, p.82)

“Minimize the use of matrices, and when they cannot be avoided, minimize their complexity” (Dillman, Smyth & Christian, 2014, p. 368)

Our experiment

Sample description and time of data collection and countries

United States Field 8/17 – 8/30/16

Total sample size: 3,600 per country

- 600 per cell per country
- 600 x 2 = 1,200 per grid condition per country
- 600 x 3 = 1,800 per platform condition per country

Randomly assign respondents to one of the three grid conditions

Japan Field 8/26 – 9/7/16

Total sample size: 3,600 per country

- 600 per cell per country
- 600 x 2 = 1,200 per grid condition per country
- 600 x 3 = 1,800 per platform condition per country

Randomly assign respondents to one of the three grid conditions

References

Couper, M. P. (2016, March). *Grids versus item-by-item designs on smartphones*. Presented at the General Online Research (GOR) Conference, Dresden, Germany.

Buttermore, N., Bargas, F. M., & Thomas, R. K. (2017, May). *Expanding alternatives: The accordion grid as an alternative to the traditional grid format*. Presented at the 72nd Annual conference of the American Association for Public Opinion Research, New Orleans, LA.

Grady, R. H., & Liu, M. (2017, May). *What's the best size for matrix-style questions in online surveys?* Presented at the 72nd Annual conference of the American Association for Public Opinion Research, New Orleans, LA

McGeeney, K. (2015, July 1). *Tips for creating web surveys for completion on a mobile device*.

Randall, T. K., Barlas, F. M., Graham, P., & Subias, T. (2015). *What they can't see can hurt you: Improving grids for mobile devices*. Presented at the CASRO Digital Research, Nashville, TN.

Richards, A., Powell, R., Murphy, J., Nguyen, M., & Shengchao, Y. (2016). Gridlocked: The impact of adapting survey grids for smartphones. *Survey Practice*, 9(3).

Stern, M., Sterrett, D., & Bilgen, I. (2016). The effects of grids on web surveys completed with mobile devices. *Social Currents*, 3(3), 217–233.

Wenz, A. (2017, April 26). *Completing web surveys on mobile devices: Does screen size affect data quality?* Institute for Social and Economic Research. University of Essex.