

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





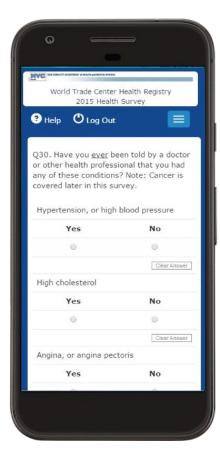
Stern, Sterrett, & Bilgen, 2016

Google 🌆

4

Stacking or banking





Richards, Powell, Murphy, Nguyen & Yu, 2016



Thomas, Barlas,

2015

Graham, & Subias,

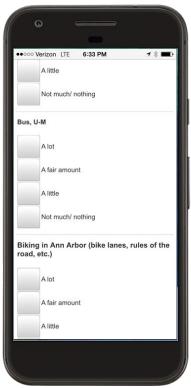
Single questions with scrolling

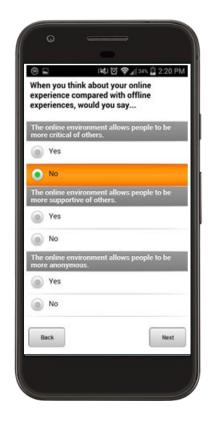
	• Verizon LTE 6:28 PM 7 %	
÷	anetest2.isr.umich.edu =	4
M	INIVERSITY OF MICHIGAN	elp
	Sustainability Culture Indicators Program (SCIP)	
How	much do you know about the followin	g?
	, AAATA/"The Ride" (Ann Arbor Area sportation Authority schedules, route	s,
	A lot	
	A fair amount	
	A little	
	Not much/ nothing	
		-
Bus,	U-M	

Couper,

Google Iss

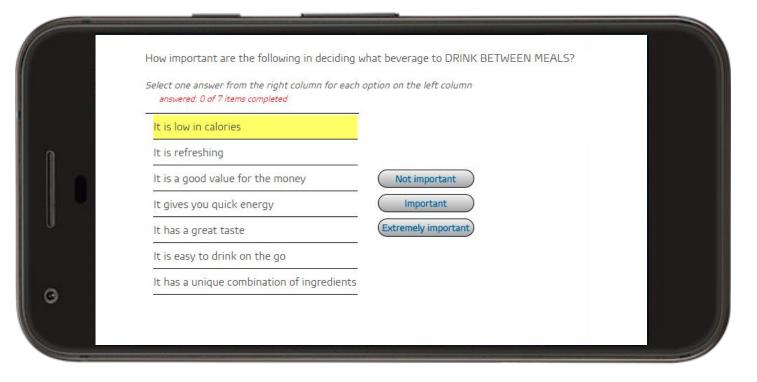
2016





McGeeney, 2015

Two columns – responses on right (GFK)





Thomas, Barlas, Graham, & Subias, 2015

Focal element or progressive (GFK)

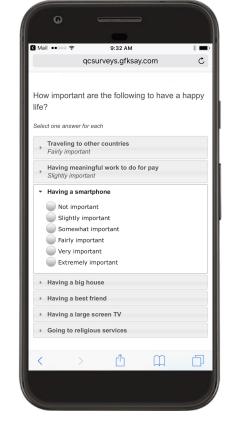
1		ng what beverage to DRINK BETWEEN MEALS?	
Pleas	e click on the orange arrow to move up o	r down the list	Item 1 of 7 (0 completed)
	*	It is low in calories	»
	Not important	Important	Extremely important
		It is refreshing	
		It is a good value for the money	
		It gives you quick energy	
		It has a great taste	
		It is easy to drink on the go	
		It has a unique combination of ingredients	



Thomas, Barlas, Graham, & Subias, 2015

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 <u>Desktop</u> 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

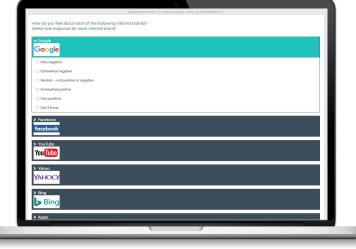
	< Google >				
Very negative	Somewhat negative	Neutral – not positive or negative	Somewhat positive	Very positive	Don't know
Previous		Ne	st		
Ipsos	Help Privacy policy © 2016 (page				

	Very negative	Somewhat negative	Neutral – not positive or negative	Somewhat positive	Very positive	Don't kr
YouTube You Tube						
Amazon amazon						
Yahoo YAHOO!						
Bing Bing						
Facebook facebook						
Apple						
Previous	Ne					

Responsive grid

Progressive grid (with auto-advance)

Google 🔜



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.

Goodle

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%

Significantly different at 95% confidence (Test Cell vs Control). Red is lower, green is higher ¹⁴

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 Progessive grids, Progressive grids with auto-advance are designed to be error free.

(Joodle

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	BR Cell A Cell Progressive Grid Respons		Cell C Collapsible Grid
No errors	100%	95%	91%
1 error	-	4%	5%
2 errors or more	_	1%	4%

Significantly different at 95% confidence (Test Cell vs Control). Red is lower, green is higher

Survey length: Time To Complete

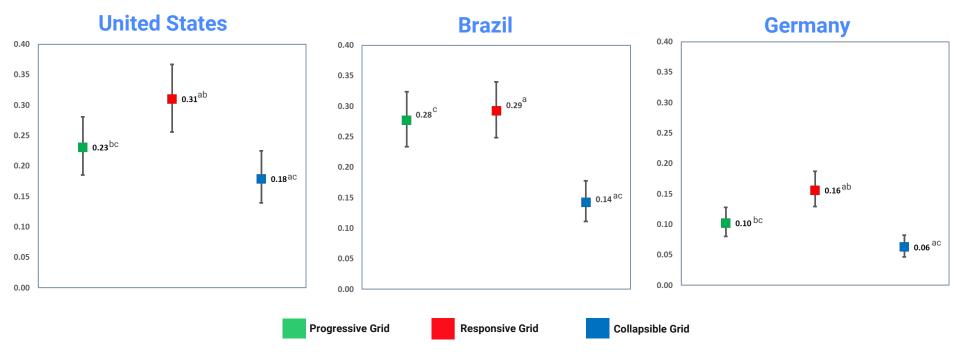
Responsive Grids are consistently the fastest Test cell.

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





^{als-}Statistically 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 n= 1,756, Brazil n=1,600, Germany n=2,996

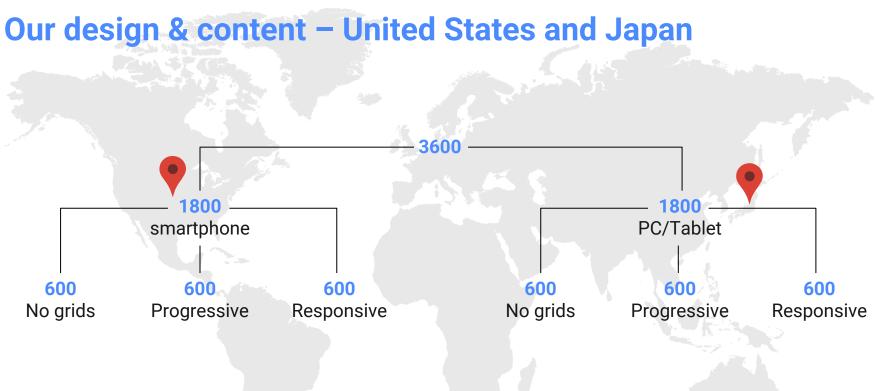
Collapsible grid (Ipsos) vs. accordion grid (GFK)

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



Content: Survey about satisfaction with smartphone apps

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

Google

Visual Design for our Experiment

No grids: single item per screen

energi 🔹 🔹 👘 👘 👘 👘 👘 👘 👘 🕹 👘 👘 🕹 👘 👘 👘 👘 👘 👘 👘 👘 👘 👘 👘 👘 👘	●occo Verizon
Still thinking about using PRODUCT on a smartphone, how satisfied are you with its Duality of search results	Visual appeal



Progressive grids

								Fi ∢ Previous
Still thinking about	using PRODUCT on a	a smartphone, how setia	fied are you wi	th its				
		1043	and append					
			ail appeal					Neith
					11 P			
Contrain section.	Moderately assorbed		transferd of	International States	Moderatury descalation	Ecouraly domainshad		
		Þ						
		ur .						



100%

Next

Responsive grids

								-
Still thinking about using P	Estrainely	Moderatoly	how satisfy	Hetthey satisfied rest	Sightly	Moderatory		
Guality of search results.							*	
Visual append								
Speed							- 10	
Case of use								
Features and capitalities								

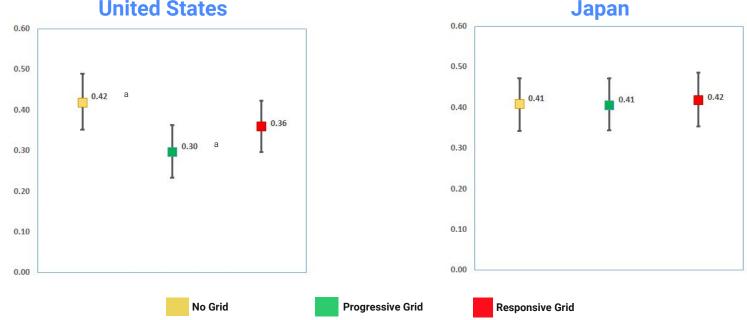


24

Impact on Data Quality

Quality indicator: Straightlining

Probability of Satisfaction Straightlining – PC/Tablet Platform



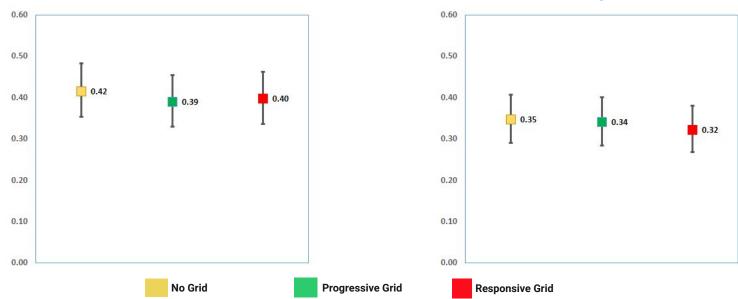
United States



^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



Japan

United States



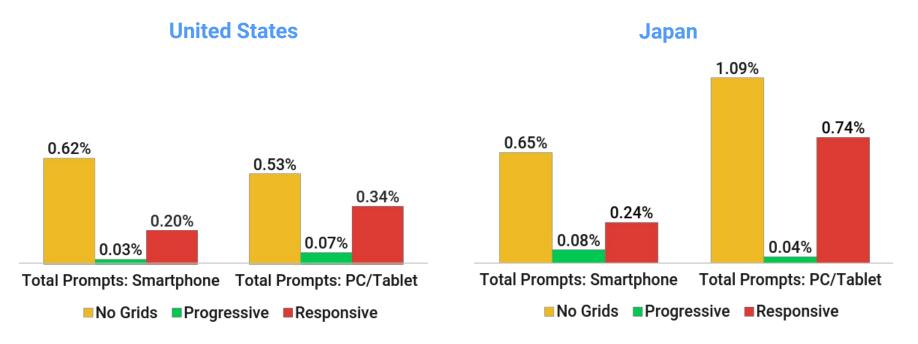
Probabilities and distributions simulated from logit models including demographic controls based on 5000 bootstraps.

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

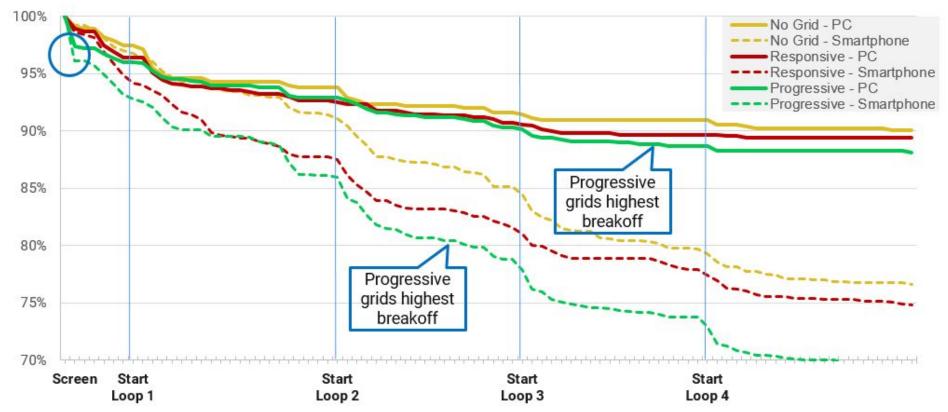
Single questions per screens (no grids) had more item nonresponse prompts regardless of device





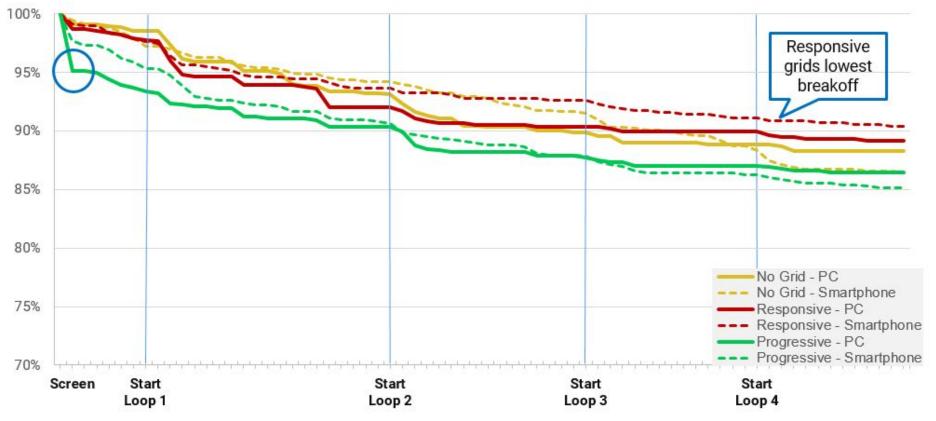
% of error prompts experienced out of the total number of opportunities for error (number of stubs displayed)

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





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

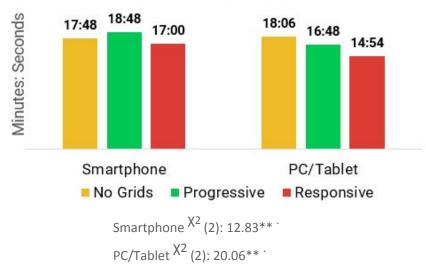




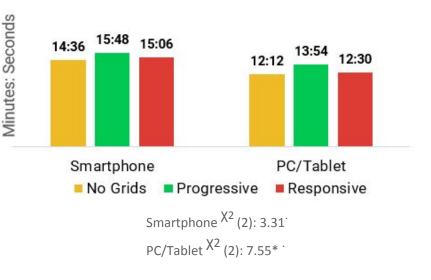
Median survey length

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.



United States



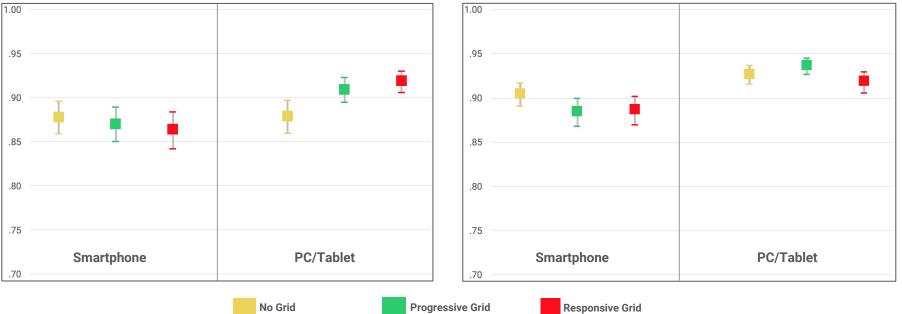
Japan



Cronbach Alpha

Differences were negligible ...

United States

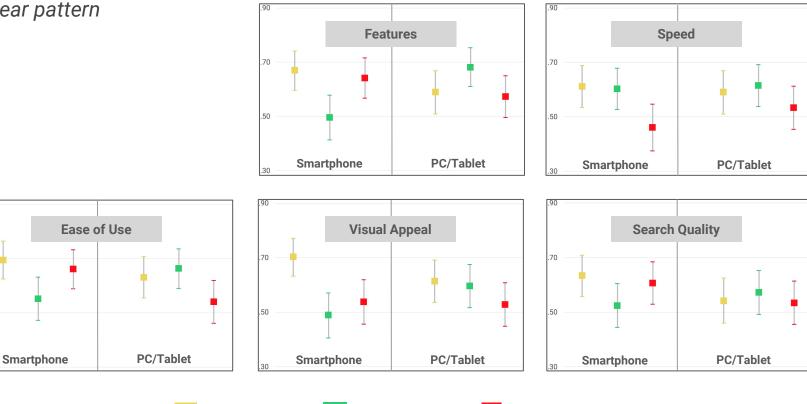


Japan



Correlations with Overall Satisfaction – United States

No clear pattern





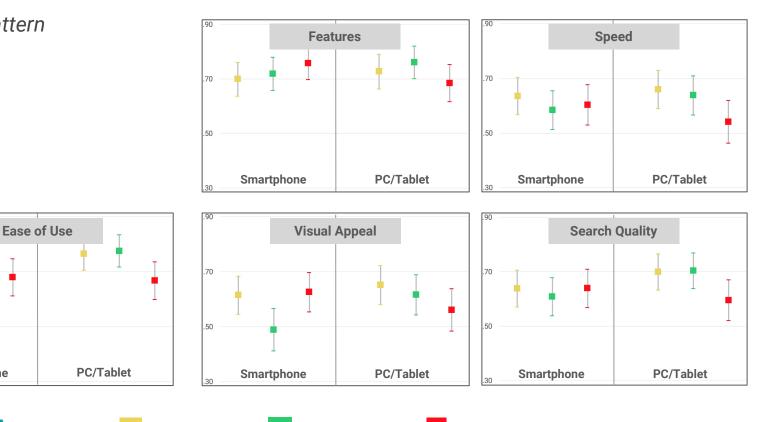
90

.70

.50

Correlations with Overall Satisfaction – Japan

No clear pattern





Smartphone

.70

50

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





thank you!

Mario Callegaro Ph.D.

Senior Survey Research Scientist

callegaro@google.com

Appendix

Smartphone view Experiment #1

	Which of the foll Select One	owing statements,	do you feel best rep	resents the device	s below?				
	✓ Smartph <u>one</u> (has internet or apps	.)						
	 Is entertaining 								
	 Is innovative 								
	 Is trustworthy 								
	 Makes me res 	ponsible							
	 Makes me loo 	k arrogant							
	C Knows me bet	ter than anyone							
	O Has a positive	impact on my life ever	yday						
	 Cares about the second s	e environment							
	> Netbook								
	Streaming Ra	<i>s</i> .							
3			_						
·	 Portable gam 	es (e.g., Nintendo W	ii)						
	> Computer								
-									
-									
	Cross-device grid	progressive SA wit	th auto-advance						
			th auto-advance lo you feel best repre	esents the device l	pelow?				
				esents the device l					
	Which of the follow				pelow?				
	Which of the follow			esents the device l					
]	Which of the follow			× ×					
	Which of the follow			× ×	•				
	Which of the follow			Smartphone (ha	as internet or apps)		Has a positive		
	Which of the follow			Smartphone (ha	s internet or apps)	Knows me better	Has a positive impact on my life	Cares about the	
•	Which of the follor Select One	wing statements, d	io you feel best repre	Smartphone (ha	as internet or apps)	Knows me better than anyone		Cares about the environment	
0	Which of the follor Select One	wing statements, d	io you feel best repre	Smartphone (ha	s internet or apps)		impact on my life		
0	Which of the follor Select One	wing statements, d	io you feel best repre	Smartphone (ha Makes me responsible	sinternet or apps) Makes me look arrogant		impact on my life		
	Which of the follor Select One	wing statements, d	io you feel best repre	Smartphone (ha Makes me responsible	s internet or apps)		impact on my life		

Google 🔜

	ne following statements, do you presents the devices below? Netbook
Is enter	rtaining
Is innov	vative
Is trust	worthy
Makes	me responsible
Makes	me look arrogant
Knows	me better than anyone
🔍 Has a p	positive impact on my life everyday
Cares a	about the environment

No Grids: single item per screen - JP

Cite B Security ("High of the security of the securit	りき続き、 <u>スマートフォンで</u> び 用についてお考えください。
	検索結果の的確さ
引き続き、スマートフォンでのPRODUCTの使用についてお考えください。	◎ 非常に満足である
開発および性能	◎ 満足である
	 ○ やや満足である ○ どちらとも言えない
◎ 林澤に直定したふ	◎ こうこも言えない
0 #2063	◎ 不満である
0 255262200 6	◎ 非常に不満である
しやれる運行動画	
※平淵26-8	х –
0.0000000000000000000000000000000000000	



Progressive Grids - JP

-								<	
51888+ <u>A</u>	-1-242C09800	UCTOBURE OUT	BEASCON.						非常に満足
			見た目の信好よさ						राजार रंठमेल्रे
	. 4.6				±>`.				どちら
DECERTOR.		PREEZE.	20525830	opfaras.	7,872.0	自然に不満である			toto;
Press Car		T. C. State States	5.693	Coll Contractor	0.000	THE STREET			不
			D-						非常(



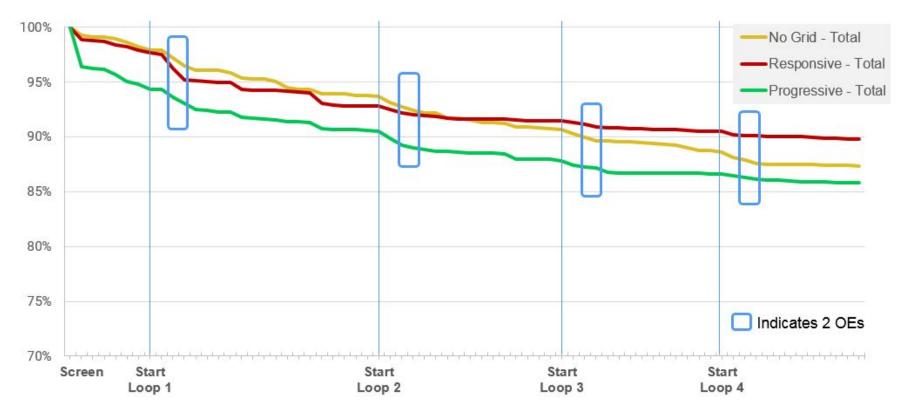
1 ∦ 100%

Responsive grids - JP

feile Cleateres C	link he stew.	Withward 966 1142 638 (2002)00546	A(引き続き、スマートフォンでの
				用についてお考えください。
				見た目の格好よさ
20450 77-	トフォンでのPRODUCTの他用に			◎ 非常に満足である
316446+ AX.	C2322 COMODOCIONOINC	- Juic Base Codes		◎ 満足である
	# 5 #			◎ やや満足である
	** ** ** **			どちらとも言えない
	にやこやに 調測者も不不不			◎ やや不満である
	22226888			◎ 不満である
	ててたたたでであるある			◎ 非常に不満である
	0000000	h		検索結果の的確さ
用能および住間	0.0000			◎ 非常に満足である
検索構現の内積さ	6.0.0.0.0.0.0.0			◎ 満足である
見た日の柿好よさ				◎ やや満足である
無いやすき				◎ どちらとも言えない
彩雕透绘				◎ やや不満である
AD-REALW.	0 0 0 0 0 0 0		1	◎ 不満である

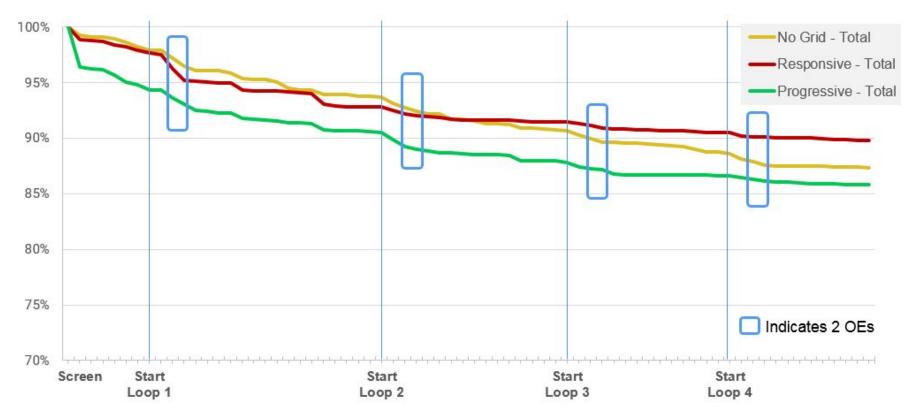


Quality indicator: Breakoff by grid condition – US (%)



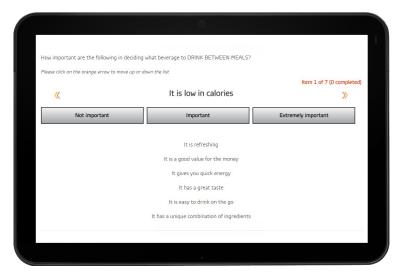


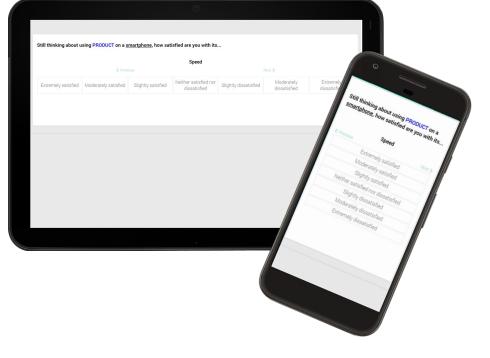
Quality indicator: Breakoff by grid condition – Japan (%)





Comparison of Focal Element vs. Progressive grid





Thomas, Barlas, Graham, & Subias, 2015



Initial exposure – US, no grid

0
efresh Click to hide Click to show
Please indicate how often, if ever, you use a smartphone (e.g., iPhone, Samsung, Motorola).
More than once a day
About once a day
Adout once a day A few times a week
About once a week
A few times a month
About once a month or less / Never
Please indicate how often, if ever, you use a 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



Initial exposure – US, progressive grid

G000

ase indicate how often, if ever, you use the	e the following.		follow	indicate how often, ing.
	Computer (desktop, laptop, notebook, netbook)			Computer (deskt notebook, ne
< Previous		Next >	< Previ	us More than on
			ut once a month or	About onc
fore than once a day About once a day	day A few times a week About once a week	A few times a month	less / Never	A few time
				About onc
				A few time

Initial exposure – US, responsive table

Please indicate how often, if ever, you use	the following.							Please indicate how often, if ever, y 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, notebool More than once a day About once a day
Computer (desktop, laptop, notebook, netbook)	0	0	0	0	0	0		 A few times a week About once a week
Smartphone (e.g., iPhone, Samsung, Motorola)	0	0	0	0	0	0		 A few times a month About once a month or less / Never
Tablet (e.g., iPad, Samsung Galaxy Tab, Kindle Fire)	0	0	0	0	0	0		Smartphone (e.g., iPhone, Samsung, I
								More than once a day About once a day
		Next						A few times a week
								About once a week
								A few times a month
								About once a month or less / Never



First Difference Calculations for Demographic Groups

US Analysis

Japan Analysis

	First Difference Calculations					
	Change in prob.	95% Co	onf 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			

	First Difference	e Calcula	ations
	Change in prob.	95% Co	onf 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



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.

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**

Wald X2 (10): 31.60** .

Japan Analysis

Coefficient
-0.263+
-0.013
0.041
-0.014
-0.156
-0.079
0.009*
-0.106
-0.732**

LR ^{X2} (8): 32.03** [·]



** 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

Google 🌆

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.

