Talk about a small, qualitative study Westat did as an initial start at exploring SMS or short message service or texting as a method for collecting data.

Thanks my colleagues …. And Fred Conrad who helped with the initial design of the application specs.

Motivation for exploring SMS

- Clinical applications, behavior modification programs that use SMS with great success
- Very successful commercial applications (Donate Haiti; American Idol)
- Response rates, especially with young adult population, are increasingly difficult to maintain
  - 23% households are cell-only (Jan – June 2009, Wireless Substitution Report)
  - Estimates of SMS use is high, especially for young adults (60%-65% for 18-29)
- Studies using SMS as survey reminder prompt show improvements in response rates (Steeh, 2009)

Westat

Clinical applications using SMS technology report great success with the technology in terms of improving behaviors, such as influenza vaccinations and remembering to take medications for a chronic illness such as asthma. In those case, even when the prompt is posed as a question, the answer isn't as important to the objective as the targeted behavior. We've all also heard of some very succesfull uses of SMS technology in some commercial venues – Donate Haiti and American Idol.

Simultaneously as we see these successful uses of SMS technology, survey organizations are experiencing greater challenges in maintaining response rate, and particularly with certain populations. These are also the populations that tend to be cell users, and by some estimates a large proportion of those same people are SMS users. So it seems a natural leap to incorporate this technology into our survey designs and data collection protocols. In fact, some experimental studies that have used SMS as a prompt to complete a survey, report higher response rates from the group receiving the SMS prompt.

So there is evidence to suggest survey researchers can make use of this technology.

Other SMS applications differ from a survey context

- Clinical applications:
  - Prompt rather than data collection
  - Some pre-existing relationship, prior personal contact
- Commercial applications:
  - One question, one text response
  - Motivation to respond
  - Language use less formal…
  - Sources of survey error – not as central to their focus

Westat

That being said, there are differences between these successful uses of the SMS technology and a typical household survey.

Clinical applications – as I mentioned, the objective is often something other than data collection. Similar to the existing uses in survye applications, the objective of the SMS is to prompt a certain behavior. Direct data collection is not the goal. In these types of applications, there is ususally some preexisting relationship or a prior personal contact in which an SMS contact is discussed.

In commercial applications, the SMS owner has a personal motivation, or perceives a personal benefit in engaging in a SMS interaction. Voting for the next American Idol or ordering that pizza for delivery now. Even with those interactions, the exchange is short one or two text per party. The language is not as formal or as intentional or studied as survey questions. And sources of survey error are not a central concern.

So these particular characteristics that contribute to making these applications work reflect characteristics of the SMS technology

3

Design has to work within SMS features

- Intended as a <u>S</u>hort <u>M</u>essage <u>S</u>ervice
- Response is not inherently linked to question – "Stateless system"
  - Example: Two Yes/No questions
  - Design must create a link between Question and Response
- Dictionary per question - like an IVR
- Limit on number of characters per text

Westat

4

SMS stands for Short Message Service….. Not Survey Methods Service, not Serious Measurement Service. The functionality is intended for short one or two exchange interactions.

- That's in part a result of the fact that SMS is a "stateless system" The response is not inherently linked to the question. It took me a really long time to wrap my head around what that means, so hopefully I can explain in a way that makes sense. Let's use the example of 2 consecutive Yes/No questions.

Are you male? Enter Yes or No.   Answer: No Your response gets stuck in a queue and takes longer than you expect to get a response. So you send "No" again. Which just happens to correspond with the system sending out the next question:

Do you wear glasses? Enter Yes or No   Answer: Yes So the system get's what it thinks is a No in response to the second question just because of the timing factor. Then a minute later it gets a Yes. Now the system has three answers and doesn't know how to interpret two of them.

The SMS study the military did with their recruitment specialists and reported at AAPOR last year, and they reported that this same problem with linking responses to answers "jumbled data" caused them to have to drop several hundred cases from their study.   What it means is that your design needs to create a unique link between each question and it's response cateogories. So the approach cannot be like an IVR where you use the same digits as response categories across question.

On the other hand, in order to 'understand' the responses, each text has an associated dictionary filled with valid and expected responses to an outgoing text.   This is like an IVR.

And one of the other challenging characteristics with SMS, especially for a survey application, is the limit on the number of characters per text. The limits vary by providers, so anywhere between 120 characters and 160 characters. How thee providers handle texts outside of that range vaires as well.

4

Your survey goes here….

Westat

And wrap all this in a tiny 2 x 2 in square screen covered with fingerprints. …

This text contains 61 characters. An example of a text string closer to about 130 is: "It is important for the Westat Survey of SMS Users to hear from you. The survey only has 3 questions. Please reply A to start."   This is 126 characters

Applying 'successful' design to a survey

- Establish prior contact, relationship
- Keep it short, simple
- Motivate the response
- Think about measurement error specific to a stateless technology

Westat

6

Establish prior contact -if not only for the obvious reason of getting the person approval to send them a text, which they may in fact get charged for depending on the type of plan they have.

Those other applications keep the number of exchanges to a minimum, and we'll talk in a minute about why that's very important for reasons outside o the potential costs incurred with each text sent or received.

Motivate the response – can't promise them their favorite dancer will win, but you can offer other types of incentives, as with many surveys.

Test design: Participant identification

- Selected 200 Westat remote telephone interviewers
- Invited participation via email with a link to a screener (129 responded):
  - Cell phone for personal use?
  - Send and receive text messages?
  - Cell phone number?
  - Cell service provider?
  - Frequency of texting?
- $25 incentive to participate

Westat

So with that as a general background about SMS technology, let's talk a little about the design of our exploratory study.

Test SMS Application Design

- Intro sequence to build encouragement for slow responders
- 2 – 4 questions after introduction texts
- Question formats:
  - Only two response choices, or standard numeric (date)
  - Consistent # characters for each response choice
  - Response choices and auto-complete
  - Unique choices across questions

Test SMS Application Design (continued)

- Included clerical review for strings of 5+
- Included some 'error handling'
  - Entry errors, time-out prompts

- No texts sent between 9pm and 9am

Test Procedures: SMS Data collection

- Message asking to start survey
- SMS survey application begins
- All who completed received a $25 electronic gift card code as final text
- When sent gift card, also sent an email with link to debriefing questionnaire
- Sample of NR - telephone debriefing

After invited with the email invitation, and got their agreement to receive texts from us there was a week or sometimes longer delay before we stated to send out messages.   So the very first message was a message asking them to start.

With this one design and implementation approach our objectives focused primarily on feasibility of collecting survey data using SMS. In addition we collected debriefing data to speak to the usability of the application.

Some Preliminary Results

Can we send, receive and track?

Status	Count
Complete	100
Partials*	10
Nonresponse	16
Known ineligibles	3
TOTAL	129

AAPOR(1) Response Rate: 79%
AAPOR(2) Response Rate: 87%

Who responded and who didn't?

Usage	Respondents 100	Nonrespondents 16	Partials 10	Total 126[a]
Low (< 1 week)	28	8	5	41
Medium (weekly)	17	2	1	20
High (daily)	55	6	4	65

[a]: 3 people did not provide response to usage question

Westat

14

Guessing that we'd get better cooperation from regular text users. Of respondents, those who completed , almost ¾ were regular text users. Not much difference in terms of text usage among NR and partials.

How quickly did they respond?

Time when respondents started to respond to survey	
Within first hour	34
More than an hour but on first day	15
On the 2nd or 3rd day	17
After the 3rd day	30

Westat
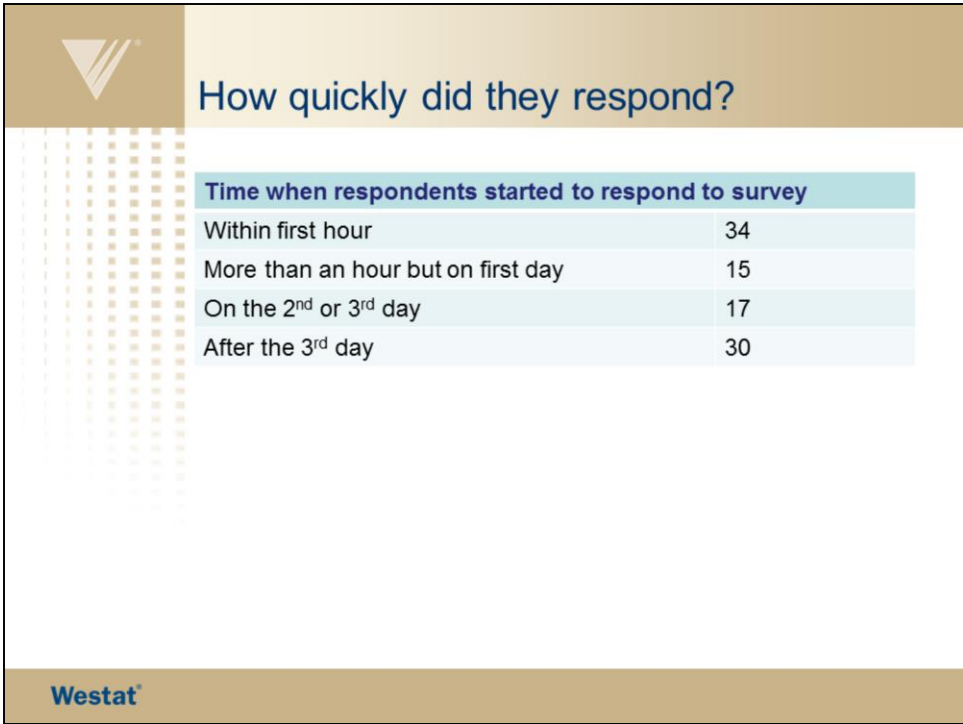
First day, in this case is the first 4 – 5 hours because of 9 pm cut-off time. So some of the 17 would move-up, and that takes us at over 50% respond in the first 4 or 5 hours – much like web survey response times. Differs from mail which requires a slightly longer time line to account for physical movement of the paper across post offices.
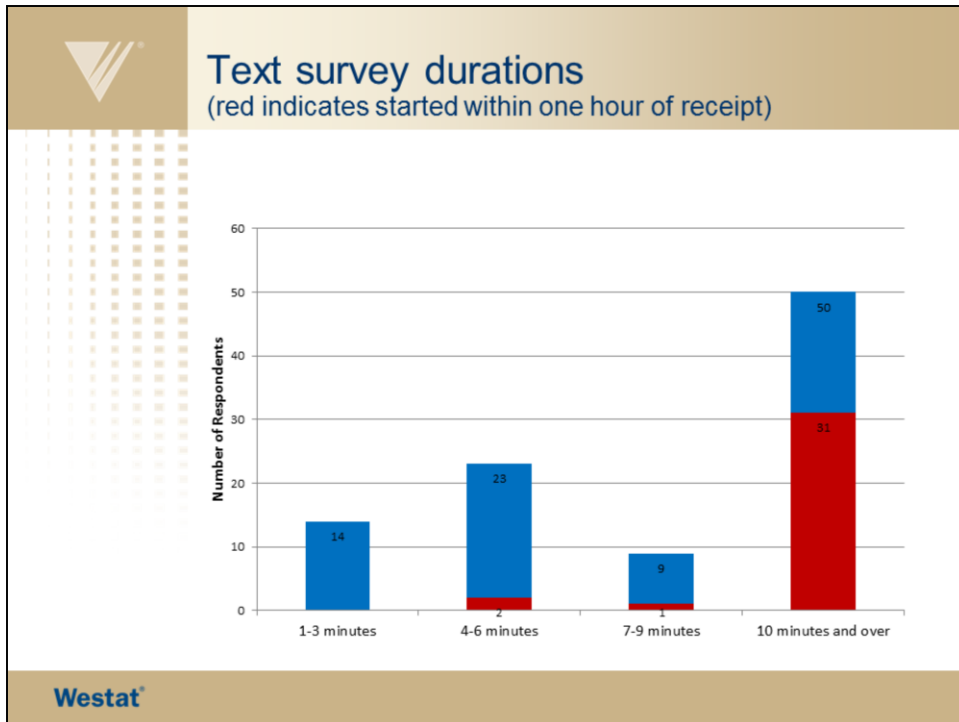
These data surprised me and suggest we probably have some learning to do with this technology. With this really short survey that only included 2 to 4 questions plus an intro, about ½ of the respondents took over 10 minutes to complete it, but just a little over 10% finished in 1-3 minutes.

Of those who required 10 minutes or more to complete the survey, over half of them actutally initiated their first response soon after we sent the first text message – back-log on the system, load/volume issues

Another way to look at this though is to look at response time at a question level compared to the estimates of completion time for the whole survey. The table at the bottom shows the median time before getting a response to a question once we sent the question. So respondents were very responsive to the texted questions. We on the other hand seemed to encounter some difficulty in processing the responses, resulting in a fairly lengthy delay between the response to one questin and the receipt of the next question.

Other Minor Technical Issues

- Anticipating and handling of special characters

  - 11 respondents/partials entered extra text (e.g., "THANKS", "Hello")

  - 9 respondents/partials included special characters (e.g., ☺)

- Affects SMS response time – complicated by stateless system

Westat

18

Some of those delays were due to other unanticipated technical issues.

Polite respondents texting thank you at the end of their response – clericial review

Reasons for Nonresponse

- Debriefings with non-respondents
  - Reported never received text message

  - Burden:
    "for me to spell out a word was going to be a pain in the neck" – low freq texter

  - Cost:
    did not have text messages included in provider plan – uncertain if extra cost would apply

Usability Feedback - Positive

Characteristic of SMS Survey	%
Thought "very easy" to read questions	93
Felt response choices corresponded "very well" with question	88
Felt time between texts was "about right"	78
Reported an 'error free' exchange (*9 of 11 reporting errors from high usage group*)	89
Rate overall experience as "very or somewhat positive"	92

Westat

20

Felt time between text was about right – 78% BUT I thought we have evidence to the contrary. But maybe this perception also reflects a characteristic of SMS users. Since there more usual interactions using TXT are probably short interactions, without a lot of back and forth, they do have "breaks" between texts. Texts from Sally are handled, and then maybe an hour later you get one from you Mom.   So their expectation is different than what we expect based on our normal automated data collection methods paradigm such as with a Web instrument.

Usability of Application (continued)

- **Some challenges reported**
  - Low frequency users:
    - Finding letters on keypad
    - Case uncertainty
    - Would prefer another response mode

  - High/mid frequency users
    - Time delay between texts
    - Prefer one letter/number response (mid users)
    - Encountered unexpected errors (high users)
    - Interruption to their usual usage

Westat

21

Not to allow you to think we have completely rose-colored glasses, there were some challenges reported back as well, and they differed a bit by frequency of usage.

Summary of results

- Good response rate (about 80%)
- Most perceived as easy to complete (92%)
  - High frequency texters tend to get more errors

- Room to improve
  - Diminishing delays, reduce durations
  - Handling extra/special characters
  - Design to be more universally appealing across all usage levels

Westat 22

Room to improve –

-While not everyone seemed to find the delays problematic, others did. We can do thinks to imrpove our own response time, and a lot of that has to do with techniques for handling special characters.

What questions do we have for next time?

- Error handling mimicking other automated methods – sometimes caused problems

- Methods to address stateless system:
  - Limit number of questions– but how short?
  - Best way to make response unique – entry burden, intuitive link, etc

- Limits on kinds of questions?
  - Small screen, character limits - comparability to other modes
  - Measurement error properties – unknown

Wendy Hicks
WendyHicks@Westat.com

THANK YOU!