1 Introduction

As more types of media have entered people’s homes, surveys can be conducted in ever more ways. In the thirties and forties, going door to door, or mailing surveys out, were the only options. From the seventies onwards, phone interviews became a popular alternative, thanks to the near universal adoption of the telephone. And now, as the internet is ever more ubiquitous, web-surveys are becoming an interesting option as well. Online surveys are generally considered cheaper, faster, and more convenient. In addition, they also have a potential for international reach, allow for elaborate skip-logic, and eliminate errors in data-entry.

Yet online surveys bring new problems as well, especially in terms of validity. Therefore the most important threats to the validity of both online and offline surveys will be assessed in this paper. First some limits of the paper are explicated. Then the notion of validity that is used here, will be clarified. And after that, a range of distinctions between online and offline surveys are set out across two crucial moments in conducting them: solicitation (inviting participants) and delivery (administering the actual survey). When this foundation has been laid, it will be used for categorizing and assessing threats to validity, such as low response rates, and interviewer effects. In this assessment some thoughts will be given to determining which of these issues are likely to remain a problem in the long run as well, as importance is in part a function of permanence. The paper will then be concluded with an overview of the threats, and some final words on their practical implications for survey research.

1.1 Limits

First of all, there are many types of web-surveys, and they can have many different aims. Some just serve as entertainment, others aim to aid webmasters at soliciting feedback on their specific sites, and yet others are used to mine for hypotheses, or to pilot (try out) a new questionnaire. All of these are
valid aims in themselves, but this paper will only be concerned with scientific surveys that aim to allow for inferences across large populations. A thing that follows from this, is that this paper will focus on quantitative research.

Secondly, there are many temptations in cheap, fast, online surveys: bad sampling such as posting an open invitation link on a forum, or sending out invitations to one’s whole target population (a census); bad questioning, by formulating questions without much thought, or by adding too many meaningless questions; and finally, adding in multimedia and colour indiscriminately. While these can, and often do, threaten the validity of online surveys, due diligence will be assumed here, not what is cheap, seductive, or easy. In addition, what will be evaluated here, is objective, as opposed to perceived validity (many on-line surveys are done badly, giving them a bad name).

Thirdly, there are many more threats to the validity of surveys, than can be discussed in the space of this paper. Some of these are serious threats to surveys regardless of the medium, while others have only small effects in most cases. A few examples are: bias from answers being self-reported; surveys being administered in artificial settings; ways in which options are categorized, or can constrain answers; and bias or errors in question-formulation, data-entry and analysis. These will not be further discussed. Nor will things that might amend them, such as mixed research designs, or gathering complementary data from other sources. This paper is about one-off (cross-sectional) surveys only, and mainly discusses threats that differ between online and offline surveys.

### 1.2 Validity

Validity is not a simple notion, and it, and its components, are defined differently by different authors. The definition of validity that is relied on here, is: that a survey represents what it intends and claims to represent. Then there are two sub-types of validity: external-, and internal validity. External validity refers to the validity of the survey beyond the study: its generalizability, both to the population, and across contexts. Internal validity, for surveys, refers to the rigour of measurement: that the concepts one sets out to measure, are actually measured (and completely). Validity is often further subdivided, but for reasons of space that will not be done here. Finally, validity can be contrasted with reliability: a study giving stable results across trials. Though different, validity presupposes reliability: if (sets of) questions (instruments) are not reliable indicators
of what they try to measure, they cannot guarantee that one measures what one thinks one does.\textsuperscript{40}

1.3 Online/Offline

While the distinction between the online and offline realms might seem straightforward, it is not. First of all, online is broader than the web, and also covers e-mail and Skype calls, for example. Secondly, there is a whole range of intermediate cases between online and offline. Such as: an interviewer overseeing you entering data on an iPad, which is then stored on a server; a survey on CD-rom sent through the mail; a scripted phone-interview with touch-tone data-entry; or a web-survey that includes a video-link with the interviewer?\textsuperscript{41} Are these online, offline, or both?

In an effort to clarify things, a number of dimensions can be distinguished, such as: by what means the participant is invited; whether the participant enters the data himself; whether an interviewer is present; the medium used in the survey, such as text or voice; whether answers are recorded in real-time, and so on. But even that would be too much granularity for the scope of this paper. Therefore, the two most important moments in the conduct of surveys were selected: solicitation (inviting participants), and delivery (administering the actual survey), and along these, differentiations were made between types of media (see table 1).\textsuperscript{30}

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>Offline</td>
<td>$\leftarrow$ intermediate</td>
<td>$\Rightarrow$ Online</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td>Offline</td>
<td>$\leftarrow$ intermediate</td>
<td>$\Rightarrow$ Online</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Modes of solicitation and delivery: The archetype offline survey is the door-to-door survey, solicited and administered in person (left column), while the typical online survey is at the other end of the table: a portal-site survey solicited and delivered via the web (rightmost column).

This schema accommodates for survey-techniques that are moderately complicated, such as phone solicitations for a web-survey, while still allowing for a quick overview. Another feature of dissecting things this way, is that the moments roughly correspond to the difference between external- and internal validity. This will become clearer in the following sections, as the threats to validity are discussed and classified in accordance with it (table 1).
2 Solicitation problems

2.1 Limited coverage

For a long time, the most obvious threat to external validity for web-surveys was, that large sections of the population did not have access to the internet.\textsuperscript{9,30} In 1998, only 33\% of the US population had an internet-connection at home, and in 2000, though growth was phenomenal, it was still a mere 50\%.\textsuperscript{9} This low rate of adoption was accompanied by a bias towards certain demographic groups. Households earning more than $75,000 were 20 times more likely to be online.\textsuperscript{9} Differences between web- and mail/phone surveys have been observed for other demographics (young, male, urban, educated), and for answers to various types of questions (slightly more liberal answers, if any pattern) as well.\textsuperscript{34,9} More recently, several studies have found smaller differences, or even no differences at all (for certain populations, such as students and lesbians).\textsuperscript{33,17,29,27,22}

One early solution to the problem of coverage, was that offered by the Dutch Telepanel: people were recruited by phone, and then given a computer with internet-access, if they did not already have one.\textsuperscript{9} Another approach to the problem is post-stratification. This comes down to re-weighting respondents so their weighted answers to questions on demographics, match those of the overall population.\textsuperscript{41} Naturally, this approach is not bullet-proof, as removing skews on general demographic variables only allows you to say you have excluded these known biases, not the many unknown ones (in other answers).\textsuperscript{3,31} The only real solution will come with time. As adoption continues to grow – in early 2010 the US adoption-rate has already risen to 80\% –, web-coverage will be less of a problem in the future.\textsuperscript{28} Though differences in connection-quality will likely remain, and at least for now, coverage is still a problem for e-mail and the web (see table 2).

Table 2: Coverage as a threat to external validity: A threat to e-mail and web solicitation and delivery

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td>C</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2 Lack of a sampling frame

A more fundamental problem with the internet, has to do with sampling. Random sampling is important, as without it, introducing biases is almost inevitable.\textsuperscript{3} Moreover, random samples are required for confidence-intervals and other statistics to be valid.\textsuperscript{30,13} In order to take a random sample, one first needs
The Validity of Surveys: Online and Offline

Wybo Wiersma

a list of ones population of interest, called a frame. But no such list exists for internet users. And where it is possible to randomly generate phone-numbers, this does not work for e-mail addresses. Only in very special cases is it possible to get, or compile, a list. Such as when ones population of interest consists of web-hosting companies, members of a club, or students at a certain university.

Panels again have been proposed as a solution to this problem. Either self-selected panels consisting of millions of members, such as those of Harris Interactive, or large panels (randomly) recruited by phone. For both types of panel, biases can be introduced by self-selection and panel-effects (peoples answers changing as a result of taking many surveys).

Yet a potential problem with panels recruited by phone, and with phone-surveys in general, is that land-line phone numbers (land-lines are traditionally used in phone-surveys) don’t provide a perfect frame either. Some households have multiple lines, and some people, especially students and the young, have no land-line at all. Increasingly, they have one or more mobile phones instead. So, while the sunrise internet offers no frame, the frame offered by phone-lines is becoming less solid than it used to be. It is unclear how this will play out. Yet for now, and the near future, the lack of frames is a fundamental problem for web- and especially e-mail-surveys, and a growing problem for phone-surveys (table 3).

Table 3: Lack of a frame as a threat to external validity: A fundamental threat to e-mail and web solicitation, but increasingly problematic for phone-surveys as well (lowercase letter ‘f’ indicates a modest threat).

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>f</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3 Low response rates

Another fundamental problem for online surveys, and phone surveys to a lesser degree, is a low response rate. Low response rates are problematic because the people that don’t participate, are usually different from those that do. For mail-surveys, response-rates between 40 and 70% are quite normal, while for e-mail and web-surveys, response-rates often fall (far) below 30%. Moreover, the kind of non-response that web-surveys receive, can be different from that in face to face settings. In the offline world people can refuse to participate, or walk away, while on the internet, people can read (part of) the survey, and only then decide whether they want to continue or not. And where a physically present interviewer could gently motivate somebody...
to continue filling out a survey, this is much harder to do online.

The effects of a low response-rate might be mitigated by receiving higher quality data from online surveys, as some studies have reported things such as longer answers and less item non-response for on-line surveys. Yet other studies present conflicting evidence. And much of the reported improvement, could have been due to differences in coverage as well (education level, age, etc.). Directly circumventing response-rate-problems (and frame-problems) has also been tried by inviting people to web-surveys by phone. However, these studies have had response rates that were even worse: around 10% (of internet users). The only thing that seems to have worked, so far, is offering multiple delivery modes. But even in that case, the online version was chosen by only 2 to 10% of respondents. So people still seem to have a strong preference for paper and phone surveys. Though it is likely that this problem will diminish over time (with better, and more user friendly devices).

A more pressing issue behind low response-rates, is that (e-mail) invitations are being crowded out by SPAM. This is a problem similar to that of telemarketing, which two decades earlier caused a drop in response-rates for phone-interviews. Apart from making it more difficult to harvest e-mail addresses (as people hide them), many invitation e-mails will now also end up in SPAM-filters. In addition, norms condemning SPAM can, for some people, come to cover e-mail invitations as well. One study even reports scholars having their servers hacked, after posting invitations to several news-groups. Certain professional associations, such as ESOMAR, MRA and MRS, even discourage sending invitations through e-mail. Yet the problem of overload stretches even further: people not only receive a lot of junk, they also have ever greater possibilities to engage in (more) entertaining activities (including entertaining surveys). Over time such crowding out effects could become great threats to scientific surveys across all modes, though low response rates are mostly limited to the web and phone for now (table 4).

Table 4: **Low response-rates as a threat to external validity: A threat to e-mail and web solicitation, but problematic for phone surveys as well** (lowercase letter ‘r’ indicates a modest threat).

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation Delivery</td>
<td>r</td>
<td>R</td>
<td>R</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3 Delivery problems

3.1 Controlling access

A commonly mentioned threat to the (mostly internal) validity of online surveys is, that it would be hard to control access to them.\textsuperscript{30,4} People could fill them out twice, post a link to them on a forum, or worse, even bots (computer-programs) could be stuff the ballot with bogus-answers. Traditional means of limiting control are indeed ineffective, such as limiting answers to one per IP address, and setting cookies in participants browsers. This because peoples IP-addresses change, and cookies can be removed, or circumvented by simply using another machine. Yet there is a solution that does work: providing each person in the sample with an unique password that can only be used to fill out the survey once (coupled with a properly configured server). The password can optionally be embedded in the url that is provided with the invitation, so it does not even have to be entered manually (though requiring manual entry reportedly improves the quality of responses).\textsuperscript{4,19}

Among the problems that remain, are that participants could feel insecure about whether their answers are being treated anonymously. Especially as participants can never know for sure what the survey-software records, and the password could always be used to link them back to their particular answers.\textsuperscript{34,4} Perceived anonymity is important, as it improves response-rates, and reduces social desirability biases. E-mail surveys have this problem to an even greater extent, because they are sent to, and received from, particular addresses (so identities are stored).\textsuperscript{4} Another problem that remains, but that postal mail-surveys have as well, is that security can only be achieved up to the address-level. Nothing can stop people from asking their children, or a friend, to fill out the survey. Only in-person, and phone-surveys guard against this (table 5).

Table 5: Access control as a threat to internal validity: Access can only be restricted up to the address-level for web-, e-mail, and mail-surveys (lowercase indicates a modest threat).

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>Delivery</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Interviewer effects

So called interviewer-effects: the presence of an interviewer inducing socially desired answers (or other biases), are a great threat to phone- and in-person surveys. White people are, for example, more likely to give politically correct answers, when they are being interviewed by a black person, on racial
issues. Other things that are known to be played down, besides racism, are smoking, drinking, and gambling habits, and reported sexual activities. One study has even found age-interaction-effects, with older people being more susceptible to interviewer effects. Another danger, is that an interviewer might (unknowingly) give out subtle cues about how he wants people to answer. This can happen, for example, through pronunciation, or by giving people more time for certain questions (more time is perceived as more important).

One partial solution to interviewer effects, is working with multiple interviewers, where each takes care of a limited number of participants. This should at least cancel out some of the effects that differ between interviewers, and thereby increase validity. Another way to improve things, is training interviewers. As skills do vary. Also, it could be argued that, while using interviewers introduces interviewer-effects, them being able to steer things, has many benefits as well. Both face to face, and on the phone, they can, for example, motivate people to continue, reassure participants trust in the anonymity of their responses, or detect and question bogus answers. Moreover, they can – especially face to face – pick up the respondents mood, or other contextual factors that might influence responses. In addition, interviewer-effects have been shown to be slightly smaller in phone-interviews. Some experiments even suggest that certain interviewer-effects might be mitigated with the clever use of virtual reality (such as differences in body-height, tone of voice, and strength of emotional expressions). Yet, traditional, in-person, and phone interviews, will always be affected interviewer effects (table 6).

Table 6: Interviewer effects as a threat to internal validity: Fundamental for in-person and phone surveys.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>Delivery</td>
<td>I</td>
<td>I</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3 Display effects

Finally, display effects can be seen as the online equivalent of interviewer effects. Differences in how the survey is displayed across devices, screen-sizes and operating-systems, can, in more or less subtle ways, influence how participants interpret questions. But technology can have other effects as well. A slow internet-connection can discourage people from taking a survey, or missing plugins (such as a Flash-plugin) may even make a survey inaccessible to large portions of the sam-
One workaround, is sticking to basic technology (HTML forms). Though with the use of Javascript, it has become possible to make forms appear exactly the same across operating systems, by replacing OS-specific form-buttons and check-boxes with images (as SurveyMonkey does), so there is a trade-off here. As with interviewers, digital forms also bring many benefits, such as taking care of skip-logic, validating answers, and providing drop-down selection-menus, and contextual help. In addition, they allow one to randomize the order of questions, cancelling out effects that question-order might have in paper-surveys. Moreover, multimedia might be used to motivate, or guide people (increasing response-rates). Multimedia can bias responses as well, of course, but it can be argued (though this has not been studied) that in a world where people are increasingly used to multimedia, a sober, simple design, can bias responses as well (towards sober answers). Nevertheless, on overall, the differences between browsers on personal computers seems to be decreasing (apart from screen-size). Though mobiles show a bit more variety, and differences in the way e-mails are displayed, still exist (table 7, and an overview of all threats in table 8).

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7: Display effects as a threat to internal validity: A modest threat to e-mail surveys and web-surveys (lowercase indicates a modest threat).**

<table>
<thead>
<tr>
<th>Moment</th>
<th>Person</th>
<th>Phone</th>
<th>Mail</th>
<th>E-mail</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solicitation</td>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>I</td>
<td>Cd</td>
<td>Cd</td>
<td></td>
</tr>
</tbody>
</table>

**Table 8: Overview of threats to the validity of surveys. External: coverage (C), lack of a frame (F) and low response rates (R). Internal: Access (A), interviewer effects (I), and display effect (D) (lowercase letters indicate modest threats).**

**4 Conclusion**

To conclude, the notions of on- and offline have been split across solicitation and delivery, and were divided into various types of media, along these. This schema was then used to categorize, and clarify threats to validity. It was found that, though coverage remains a threat to online surveys for now, lack of a sampling frame is a more fundamental problem for e-mail and web-surveys. Another important problem, and one that seems to be growing, is the low response-rate for online surveys.
and phone surveys, especially where caused by crowding out effects. Then for delivery, while access control is often cited as a problem for web surveys, there is a technical solution available that works up to the address-level (beyond which mail-surveys have similar problems). For in-person and phone surveys, interviewer effects are the most fundamental threats. While for online surveys, display effects can cause moderate to minor problems, especially across devices that are very different, such as personal computers and mobile phones.

The overall picture that has emerged, is that all options have advantages and disadvantages (also see table 8). Even though some options, such as soliciting participants through websites (by leaving an open link), are generally not a good idea, unless there are specific reasons for taking this route, such as contacting populations that are hard to reach otherwise. Besides this, experiments with solicitation through one medium for a survey delivered through another (such as phone to web), did not seem to work well. Thus it follows, that web surveys are mainly a good alternative when a frame can be obtained for ones population of interest (and coverage is expected to be good for them). In most other cases, in-person, phone- and mail-surveys, still seem superior: both in terms of response-rates, coverage, and the availability of frames.

The sunset over land-lines might, at some point, make the web seem more favourable in comparison. Yet only if lady fortune is with survey researchers, and the adoption-rate of smartphones increases sufficiently, it might become possible to send invitations to web-surveys by text-message. A clickable link (already supported by the iPhone) could then take people directly to the survey, in the phones browser. Over the span of a decade or more, this could make smartphones into a very powerful frame for both phone- and web-surveys. Until such times, the perfect modes for solicitation and delivery do not exist, and being aware of, and acknowledging the limits of each mode, is the best we can aspire to when designing and fielding a survey.

Bibliography


The Validity of Surveys: Online and Offline
Wybo Wiersma


32. Riegelsberger, J., & Nakhimovsky, Y., ‘Seeing the Bigger Picture: A Multi-


