Schriftenreihe der ASI – Arbeitsgemeinschaft
Sozialwissenschaftlicher Institute

Stefanie Eifler · Frank Faulbaum *Hrsg.*

Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung

Springer VS

Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute

Herausgegeben von
Frank Faulbaum
Universität Duisburg-Essen

Paul Hill
RWTH Aachen

Birgit Pfau-Effinger
Universität Hamburg

Jürgen Schupp
Deutsches Institut für
Wirtschaftsforschung e.V. Berlin
(DIW)

Jette Schröder
GESIS – Leibniz-Institut für
Sozialwissenschaften, Mannheim

Christof Wolf
GESIS – Leibniz-Institut für
Sozialwissenschaften, Mannheim

Stefanie Eifler · Frank Faulbaum
(Hrsg.)

Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung

Springer VS

Herausgeber
Prof. Dr. Stefanie Eifler
Katholische Universität
Eichstätt-Ingolstadt
Eichstätt, Deutschland

Prof. Dr. Frank Faulbaum
Universität Duisburg-Essen
Duisburg, Deutschland

Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute
ISBN 978-3-658-15833-0 ISBN 978-3-658-15834-7 (eBook)
DOI 10.1007/978-3-658-15834-7

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer VS

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist Teil von Springer Nature
Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Inhalt

Mess- und Methodeneffekte

Vorwort

Nicht zuletzt befördert durch den Artikel von Edith de Leeuw im Journal of Official Statistics (2005) haben Mixed-Mode-Designs in der Umfrageforschung in den letzten zehn Jahren an Verbreitung gewonnen. Hintergrund dieser Entwicklung sind vor allem die Reduktion von Umfragekosten und die Reduktion von Ausfällen (Unit Nonresponse) und damit die Bias-Reduktion. Allerdings wurde rasch deutlich, dass man sich mit der Einbeziehung verschiedener Modes in das Erhebungsdesign vor allem zwei Arten von Effekten einhandelt, nämlich Effekte auf die Messung und die Datenqualität bzw. den Messfehler und Selektionseffekte bzw. Effekte auf die Zusammensetzung der Stichprobe (vgl. Voogt und Saris 2005; Vannieuwenhuyze und Loosfeldt 2013). Zu den klassischen Komponenten eines Modes - Administrationsform, Erhebungstechnologie und Übertragungskanal (akustisch, visuell) - ist der Aspekt der Nutzung verschiedener Geräte innerhalb eines Modes hinzugekommen (vgl. Toepoel und Lugtig 2015) und hat sich weiter diversifiziert. Access-Panels, insbesondere Online-Access-Panels, haben darüber hinaus durch den Zugang zu teilnahmebereiten Probanden die Durchführung von Umfrage-Experimenten und den Test innovativer Erhebungsmethoden erleichtert. Obgleich Mixed-Mode-Designs oder Designs mit mehreren Geräten (devices) spezifische methodische Probleme mit sich bringen, deren Minimierung ein Anliegen der Umfrageforschung sein muss, tragen weiterhin Erkenntnisse über Anwendungen einzelner Modes zum besseren Verständnis auch von Mixed-Mode-Designs bei.

Alle oben erwähnten Themen werden auch in den Beiträgen zu diesem Band angesprochen, der aus einer Tagung der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) e.V. und der Sektion „Methoden der empirischen Sozialforschung“ der Deutschen Gesellschaft für Soziologe zum Thema „Mixed-Mode-Surveys“ am 06./07. November 2015 in Köln hervorgegangen ist. Im ersten Beitrag dieses Bandes beschäftigt sich Michael Bosnjak mit den möglichen Auswirkungen von Modes auf die Datenqualität, wobei er sich auf Ergebnisse von Metanalysen stützt. Im zweiten Beitrag widmet sich Annette Scherpenzeel der Anwendungsvielfalt methodisch innovativer Erhebungsmethoden unter Nutzung des niederländischen LISS-Panels (vgl. Das 2012). Heinz Leitgöb thematisiert in seinem Beitrag die Zusammensetzung der Modes aus verschiedenen Komponenten und widmet

sich der Frage, wie Modeeffekte in die Effektbeiträge ihrer Komponenten zerlegt werden können. Die beiden folgenden Beiträge von Annelies Blom/Jessica Herzing und von Karl-Heinz Reuband beschäftigen sich mit dem Einsatz von Modes in der Teilnehmerrekrutierung. Danach berichten Denis Saßenroth/Jürgen Schupp sowie Jennifer Allen/Patrick Schmich über Probleme und Effekte, die bei Transformationen von Single-Mode-Designs in Mixed-Mode-Designs auftreten. Volker Hüfken widmet sich anschließend der Frage, wie sich Modes auf die Wahl von Abstufungen auf einer Rating-Skala auswirken. Matthias Sand berichtet über Ergebnisse einer Studie über Effekte bei der Validierung von Mobilfunknummern durch verschiedene Anbieter. Der Band endet mit einem Methodenvergleich zwischen zwei Erhebungsinstrumenten für sensitive Fragen, der Randomized-Response-Technik und dem Crosswise-Modell, verfasst von Dirk Enzmann.

Unser Dank an dieser Stelle geht an alle Autorinnen und Autoren für ihre Kooperationsbereitschaft, die diese Publikation ermöglicht hat. Ein besonderes Dankeschön richten wir an Frau Bettina Zacharias (GESIS) für die professionelle Anpassung der überarbeiten Beiträge an die Vorgaben des Verlags.

Eichstätt und Duisburg im August 2016

Stefanie Eifler & Frank Faulbaum

Literatur

De Leeuw, E.D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21,* 233–255

Toepoel, V., & Lugtig, P. (Hrsg.). (2015). *The collection of survey data using mixed devices.* Special issue of methodes, data, analyses (mda), Volumes 9, 2. Mannheim: GESIS.

Vannieuwenhuyze, J.T.A., & Loosfeldt, G, (2013). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods and Research, 42,* 82.104.

Voogt, Robert J.J., & Saris, W.E. (2005). Mixed mode designs: finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics, 21*, 367-388.

Mode-Vielfalt, Mode-Komponenten und Datenqualität

Mixed-Mode Surveys and Data Quality

Meta-Analytic Evidence and Avenues for Future Research

Michael Bosnjak
GESIS – Leibniz Institute for the Social Sciences
and University of Mannheim

1 The evidence-based movement and the use of meta-analysis in survey methodology

It has often been noted that survey methodology is inherently pragmatically oriented (e.g., Bosnjak and Danner 2015, p. 309; Goyder 1987, p. 11): How to sample and recruit respondents to reduce coverage and sampling errors, how to operationalize concepts to reduce measurement error, and how to minimize the differences between those who responded from those who did not on all variables of interest (i.e., nonresponse bias) are generic guiding questions in survey methodology (Dillman et al. 2014; Groves et al. 2011). Accordingly, survey methodology is aligned towards generating a body of knowledge to support decision-making about designing, preparing, implementing, and post-processing survey-based projects. In doing so, survey methodology is structurally similar to other disciplines who are committed to generating the best empirical evidence and using it to guide actions. The very first and most notably among the scientific disciplines championing such a view were the health sciences: Evidence-based medicine, defined as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sacket et al. 1996, p. 71), has decisively contributed to replace anecdotes, case study insights, and theoretical reasoning by focusing on the findings from systematically synthesized, high-quality experimental research. Some of the crowning achievements evidence-based

S. Eifler und F. Faulbaum (Hrsg.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*, Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, DOI 10.1007/978-3-658-15834-7_1

medicine yielded were establishing the Cochrane Collaboration[1] to independently collate and summarize clinical experiments, setting methodological and publication standards[2], building infrastructures for pre-registering experiments[3], developing and updating guidelines for clinical practice[4], and developing knowledge resources and courses for teaching evidence-based medicine[5]. In the early 1990s, other disciplines started to follow their example giving birth to scientific movements such as Evidence-based Education (Pring and Thomas 2004), Evidence-based Management (Rousseau 2012), Evidence-based Criminology (Farrington et al. 2003), and Evidence-based Software Engineering (Dybä et al. 2005).

A key characteristic of all those committed to the evidence-based paradigm is the belief that empirical findings should be classified by their epistemological strength. This central tenet is realized by prioritizing the strongest types of empirical evidence, namely meta-analysis of randomized experiments, and using them to guide decision making. In a nutshell, meta-analysis can be described as a set of statistical methods for aggregating, summarizing, and drawing inferences from collections of thematically related studies. The key idea is to quantify the size, direction, and/or strength of an effect, and to cancel out sampling errors associated with individual studies. Consequently, meta-analytic findings are typically characterized by a larger precision and validity than any individual study.

In contrast to meta-analytic evidence, which is commonly regarded as most conclusive, study types classified as ´mid-level quality´, - according to the evidence-based paradigm - should be treated with due caution when drawing conclusions. Such mid-level evidence encompasses sporadic randomized experiments, quasi-experiments not yet synthesized in a quantitative fashion, and findings from study types not allowing for causal inference, such as narrative reviews, correlational studies, and case reports.

While survey methodology has produced a plethora of (quasi-)experimental findings, the use of meta-analytic techniques to systematically summarize their outcomes is still underutilized. A bibliographic database search using Web of Science and Google Scholar performed in March 2016 has identi-

1 http://www.cochrane.org/
2 http://prisma-statement.org/
3 http://www.alltrials.net/
4 http://www.guideline.gov/
5 http://www.cebm.net/

fied about fifty meta-analyses relevant for survey methodology. This overall output matches the number of published meta-analysis in a top-tier psychology journal such as Psychological Bulletin within just two annual volumes. Therefore, survey methodology appears remarkably reluctant to apply existing quantitative tools to synthesize (quasi-)experimental evidence systematically. A considerable share of the available meta-analytic findings are on the effects of survey modes and their combinations.

The overall aim of this contribution is to briefly summarize the available meta-analytic evidence on the implications of mixed-mode surveys for survey data quality, and to identify gaps in research in order to provide survey methodologists a starting point to perform research syntheses in the future. Accordingly, the following chapter will outline what the available meta-analyses tell us about the effect of modes on measurement-related and representation-related errors. Meta-analyses focusing exclusively on one single mode, that is not taking a comparative perspective (e.g., de Leeuw et al. 2007; Cook et al. 2000; Green and Hutchinson 1996), or those using statistical synthesis procedures not belonging to any established meta-analytic methodology (e.g., Goyder 1985, 1987), will not be considered.

2 (Mixed) Mode Effects: What does meta-analytic evidence tell us?

For much of the 20th century, mixing survey modes was a rare exception (Dillman and Messer 2010). To optimize costs, response rates, and measurement quality, it became increasingly common to combine multiple modes of data collection (Groves et al. 2011, Chapter 5.4). For instance, by starting with the least expensive mode (e.g., Web), and then moving to more expensive ones to collect data from nonrespondents (e.g. mail, telephone, or even face-to-face interviews), costlier approaches are successfully applied to fewer cases, optimizing the overall expense for a survey project. In longitudinal surveys, starting with expensive face-to-face interviews as an attempt to maximize initial cooperation rates, and then moving to more reasonably priced Internet-based data collections, became common practice in probability-based panel surveys (e.g., Blom et al. 2016; Bosnjak et al. 2016). Moreover, the type of survey questions asked are believed to require different modes for subsets of items (de Leeuw 2005; Dillman et al. 2014, Chapter 11): Interviewer-administered surveys are assumed to be ideal to gain cooperation and to ask non-obtrusive

questions. To reduce item-nonresponse and to yield valid answers for sensitive questions, follow-up self-administered surveys seem most appropriate.

Remarkably, what has developed as best practice appears only partly supported by the available meta-analytic evidence. Table 1 provides a heuristic framework, classifying different types of meta-analytic research questions, which are instrumental in informing survey methodologists about the impact of (mixed) modes on two generic sources of errors pertaining to the Total Survey Error concept (Biemer 2010; Groves 2004), namely measurement-related and representation-related errors and biases.

The first type of meta-analytic research question is on the synthesized effect within various survey modes in head-to-head comparison studies (′Mode Effects′ in Table 1). Here, primary studies varying the survey mode as the independent variable, and assessing its effect on measurement-related and/or representation-related dependent variables, are aggregated. The specific outcome variables considered here were, for instance, the average mode-specific degree of social desirability bias (e.g., de Leeuw and van der Zouwen 1988; Richman et al. 1999; Gnambs and Kaspar 2016), response rate differences between modes (e.g., Hox and de Leeuw 1994; Lozar Manfreda et al. 2008), and differences in terms of item-nonresponse (de Leeuw 1992).

The second type of meta-analytic research question is on relating two distinct mixed-mode designs to the two pertinent quality outcome dimensions (′Mixed-Mode Effects′ in Table 1). Within this category, the overall effect of mode-preferences are of interest to survey methodologists, that is meta-analytic findings focusing on studies summarizing the impact of allowing respondents to choose between two or more response modes offered concurrently (′Mode Choice Effects′ in Table 1; e.g., Medway and Fulton 2012). In addition, another noteworthy type of mixed-mode effect involves combining modes sequentially, either varied to follow-up on nonrespondents, or varied according to data collection occasions in multiple-wave surveys (′Serial Mode Combination Effects′ in Table 1; Shih and Fan 2007).

Within the cells of Table 1, the available meta-analytic studies are listed chronologically. Taken together, most of the 17 meta-analyses published are on mode effects, only three are on mixed-mode effects. In the two subsection to follow, we will briefly summarize what the available meta-analytic findings suggest, before we conclude with sketching avenues for future research.

Table 1 Summary of available meta-analyses on (mixed) mode effects

Type of meta-analytic research question	Total Survey Error Dimension Considered	
	Measurement-related errors and biases (including proxies)	Representation-related errors and biases (including proxies)
Mode Effects	▪ Telephone vs. face-to-face: Social desirability bias (de Leeuw & van der Zouwen 1988) ▪ Mail vs. face-to-face; Mail vs. Telephone: Response validity, Social desirability, Response similarity (de Leeuw 1992) ▪ Computerized self-administered vs. paper-and-pencil and face-to-face: Disclosure of sensitive behaviors (Weisband & Kiesler 1996) ▪ Computerized self-administered vs. paper-and-pencil, Computerized self-administered vs. face-to-face: Social desirability (Richman et al. 1999) ▪ Computerized self-administered vs. paper-and-pencil: Social Desirability (Dwight & Feigelson 2000) ▪ Computerized versus paper self-administration: Social desirability (reporting of sensitive behaviors) (Tourangeau & Yan 2007)	▪ Telephone vs. face-to-face: Response rates (overall and item-nonresponse) (de Leeuw & van der Zouwen 1988) ▪ Mail vs. face-to-face vs Telephone Item-nonresponse, Response rates (de Leeuw 1992) ▪ Mail, Phone, face-to-face: Completion rate (Hox & de Leeuw 1994) ▪ Web vs. other: Response Rates (Lozar Manfreda et al. 2008) ▪ Web vs. Mail: Response Rates (Shih & Fan 2008) ▪ (Face-to-face, Mail, CATI)*Incentive: Response-Rates (Mercer et al. 2015)

Table 1

Type of meta-analytic research question		Total Survey Error Dimension Considered	
		Measurement-related errors and biases (including proxies)	Representation-related errors and biases (including proxies)
Mode Effects		▪ CATI vs. other: Positivity bias (Ye, Fulton & Tourangeau 2011) ▪ Computerized self-administered vs. paper-and-pencil: Social Desirablity (Dodou & de Winter 2014) ▪ Computerized self-administered vs. paper-and-pencil: Disclosure of sensitive behaviors (Gnambs & Kaspar 2015) ▪ Web-based vs. paper-and-pencil: Social desirability (Gnambs & Kaspar 2016)	
Mixed-Mode Effects	Mode Choice Effects	▪ No meta-analytic studies published.	▪ Web and Mail: Response Rates (Shih & Fan 2007) ▪ Mail and Web: Response Rates (Medway & Fulton 2012) ▪ Mobile Web and Web: Breakoff Rates (Mavletova & Couper 2015)
	Serial Mode Combination Effects	▪ No meta-analytic studies published.	▪ Web and Mail: Response Rates (Shih & Fan 2007)

Mode effects and measurement-related survey quality

Previous meta-analytic research on mode effects have primarily focused on the degree of social desirability as a measurement-related survey quality indicator. Depending on the operational definition of social desirability, three distinct eligibility criteria to identify studies about the impact of mode can be employed, each discovering unique patterns of findings: A first approach focuses on studies utilizing validated social desirability scales and instruments to detect lie tendencies within two or more modes. The two most recent meta-analyses about such studies have found no substantial differences between Web-based and paper-based modes of survey administration (Gnambs & Kaspar, 2016, based on 30 experimental studies), and computer-assisted and paper-and-pencil administration (Dodou and de Winter 2014, based on 51 studies that included 62 independent samples), paralleling the findings from older meta-analyses. Tourangeau and Yan (2007) estimated the mean effect size across 10 studies that compared computer- and paper-based administration of sensitive questionnaires. The computer mode included not only self-administered surveys but also interactive voice responses and audio computer-assisted self-interviews. Only 4 of the 10 studies included social desirability scales, and all 4 of these studies were also included in Richman et al. (1999) and Dwight and Feigelson (2000). No significant difference between computer-assisted self-administration and the paper-and-pencil mode was found across these 4 studies. Dwight and Feigelson (2000) also explored whether this zero-effect was robust over time. The authors conclude that older studies have in fact found evidence of less socially desirable responses for computer-assisted self-administration compared to paper-and-pencil administration, but this effect could not be replicated in more recent studies. However, when comparing computerized self-administration with interviewer-administered modes (e.g., face-to-face surveys), social desirability bias scale estimates are about .5 standard deviations larger for interviewer administration (Richman et al. 1999).

In the second approach, socially desirable responding is viewed as deliberate over-reporting of favorable characteristics, such as conscientiousness, and underreporting of unfavorable traits, such as neuroticism. Gnambs and Kaspar (2016) have pooled ten experimental mode comparisons between Web-based and paper-and-pencil administration of the Big Five personality traits, and 28 studies comparing modes on self-reported psychopathological conditions. No

substantial differences were found between the two self-administered modes considered.

A third approach defining social desirability involves underreporting of sensitive behaviors, such as illicit drug use, delinquency, sexuality, and victimization. A recent meta-analysis having employed this conceptualization of social desirability by Gnambs and Kaspar (2015) revealed that computer-assisted surveys resulted in prevalence rates of sensitive behaviors that were about 1.5 times higher than comparable reports obtained via paper-and-pencil questionnaires; for highly sensitive issues, such as sexuality and illicit drug use, this mode effect was even larger. A re-analysis of six papers on mode-specific differences for self-reported sensitive behaviors summarized in Tourangeau and Yan (2007, Table 3), which also encompasses interactive voice responses and audio computer-assisted self-interviews, corroborates the finding that computer-assisted modes yield on average about 0.2 standard deviations larger self-reported frequency measures of sensitive behaviors compared to paper-and-pencil administration. If self-administration modes (computerized or mail surveys) are compared with interviewer-administered modes, such as telephone surveys (de Leeuw 1992, p. 32), or face-to-face surveys (de Leeuw 1992, p. 31; Weisband and Kiesler 1996), reporting of sensitive behaviors is consistently lower in interviewer-administered modes. Only two published papers report comparisons within the category of interviewer-administered modes (telephone and face-to-face; de Leeuw and van der Zouwen 1988; de Leeuw 1992, p. 28), but neither detects substantial differences in terms of reporting sensitive behaviors.

Measurement quality indicators other than social desirability have only received marginal consideration in past meta-analyses. De Leeuw (1992) found no substantial mode differences between mail, face-to-face, and telephone surveys for factual questions which could have been checked against public records. Based on 18 experimental comparisons, Ye, Fulton and Tourangeau (2011) found that telephone respondents are more likely to select the most extreme positive option than Web, mail, or IVR survey respondents, but not respondents in face-to-face interviews. The differences found in choosing the most extreme option amounted to 6-11 percentage points on average.

Taken as a whole, the meta-analytic evidence suggests that within each of the two classes (1) self-administered modes (computerized modes, including Web-based surveys; paper-and-pencil administration, including mail surveys) and (2) interviewer-administered modes (face-to-face surveys; telephone surveys), there are no substantial differences in terms of social desirability bias,

except for the propensity to report sensitive behaviors. The largest prevalence rates for sensitive behaviors are typically found in self-administered computerized survey settings, followed by paper-and-pencil administration. Interviewer-administered modes typically yield the lowest prevalence rates for sensitive behaviors.

For other, less frequently investigated measurement quality indicators, the few available meta-analytic findings point to comparable valid answers to factual questions between mail, face-to-face, and telephone modes. Moreover, the most positive category is chosen more often in interviewer-administered (CATI) surveys compared to self-administered ones.

Mode effects and representation-related survey quality

Despite the fact that response rates are in general loosely related to nonresponse bias (Groves and Peytcheva 2008), the majority of primary studies about the impact of mode on representation-related outcome variables are using response rates as outcomes. These primary studies have been summarized within six meta-analyses, allowing to infer the following ranking: Face-to-face surveys typically achieve the highest response rates (de Leeuw and van der Zouwen 1988; de Leeuw 1992; Hox and de Leeuw 1994), telephone surveys the next highest (de Leeuw and van der Zouwen 1988; de Leeuw 1992; Hox and de Leeuw 1994), followed by mail surveys (Hox and de Leeuw 1994), and Web-based surveys at the bottom of the response rate league (Lozar Manfreda et al. 2008; Shih and Fan 2008).

Specifically, Hox and de Leeuw (1994) found, based on 45 primary studies that the average completion rate is about 70% in face-to-face surveys, 67% in telephone surveys, and 61% in mail surveys. For Web surveys, Lozar Manfreda et al. (2008) estimated, based on 24 studies reporting 45 experimental mode comparisons, a lower response rate of about 11% on average compared to other survey modes. 27 out of 45 experimental comparisons in Lozar Manfred et al. (2008) involved Web versus mail comparisons. In both Hox and de Leeuw (1994) and Lozar Manfreda et al. (2008), the average effects were not homogeneous, that is, they were moderated by specific study characteristics. Most notably, in Hox and de Leeuw (1994), response to face-to-face and telephone surveys went down in the period they covered (1947 to 1992), and the response to mail surveys went up slightly. In Lozar Manfreda et al. (2008), the difference between Web and other modes was moderated by the number of contacts: The more contact attempts to recruit nonrespondents, the larger the discrepancy between Web and other modes. A recent study by Mercer et al.

(2015) explicitly addressed the moderation effects between modes and incentives on response rates. Based on 55 experiments containing 178 experimental conditions, Mercer et al. (2015) found that while prepaid incentives increased response rates in face-to-face, mail, and telephone surveys, postpaid incentives only increased response rate for two interviewer-administered modes. Moreover, the dose-response relationships proved to be mode-dependent in the prepaid condition: The expected increase in response rate for a $1 prepaid incentive is 6 percentage points for mail surveys; promising the same amount over the phone yields an expected improvement of only 1 percentage point.

Only a few meta-analyses have addressed representation-related outcome variables other than response rates: De Leeuw and van der Zouwen (1988), and de Leeuw (1992) did not find mode difference for the amount of item-nonresponse between mail, face-to-face, and telephone surveys.

Mixed-mode effects and representation-related survey quality

The three meta-analytic research summaries about mixed-mode effects (Mavletova and Couper 2015; Medway and Fulton 2012; Shih and Fan 2007) on representation-related survey quality indicators have exclusively compared self-administered modes and used response rates and breakoff rates as outcomes.

Medway and Fulton (2012) found, based on 19 experimental comparisons between either a mail survey or a survey in which participants were given the choice of responding by mail or by Web, that providing a concurrent Web option in mail surveys lowers response rates: Doing so decreases the odds of response by 12.8 percent as compared to a mail-only survey.

Shih and Fan (2007) have synthesized both experimental as well as non-experimental findings reported in 52 studies. From those studies that offered respondents options for Web and mail surveys simultaneously, the overall response rate was 25%. In those studies where respondents were sent mail surveys first and offered the option for Web survey mode in follow-up reminders, the overall response rate was 42%. Studies where respondents were sent Web surveys first and then offered the option for mail survey mode in follow-up reminders showed an overall response rate of 47%.

When optimizing response rates is an issue, the findings reported in Medway and Fulton (2012) and Shih and Fan (2007) suggest not to offer Web and mail modes concurrently. Instead, they should be combined serially.

A recent meta-analysis by Mavletova and Couper (2015) based on 39 independent samples suggest that allowing respondents to choose the visual format in mobile Web surveys (PC versus mobile optimized Web surveys) decreases breakoff rates. If respondents have an opportunity to select their preferred visual format, the odds of breakoff are decreased by 0.62 ($p < 0.05$) compared to the surveys in which respondents are initially assigned to a mobile web survey mode. Furthermore, mobile-optimized web surveys decrease the odds of breakoffs among mobile respondents by 0.71 compared to non-optimized web surveys.

3 Conclusions and avenues for future research

The meta-analytic findings presented provide a number of recommendations for survey practice. To reduce social desirability bias and the tendency to endorse positive statements, (computer assisted) self-administered modes should be used. To gain cooperation, interviewer-administered surveys should be considered. Because item-nonresponse rates do not seem to differ substantially between modes, optimizing the completeness of survey data from actual respondents is not an issue of mode, but some other factors which are out of scope here. All these recommendations are partly based on meta-analysis performed more than 20 years ago. In the meantime, numerous new primary studies emerged, calling for updated research syntheses. Such research updates decisively contribute to explore the robustness of findings across time and would be a valuable addition for survey methodologists.

The following survey implementation variants should, according to the meta-analytic findings available, be abandoned: (1) Offering mode choices for the initial invitation to participate in mail and Web surveys, because this survey implementation strategy decreases response rates. Furthermore, (2) prepaid incentives should be preferred over promised ones. The use of prepaid incentive value denominations should be tailored towards known mode-specific findings regarding their effectiveness. In other words, incentive-response rate elasticities appear to be mode-specific, a relationship which might deserve more research attention.

Surprisingly, despite the prevalence of documented mixed-mode surveys, no meta-analytic evidence seems to be available on combining self-administered and interviewer-administered data collection modes. Future quantitative summaries on such issues would yield valuable additional recom-

mendations for optimal mode combinations. Another striking shortcoming is the non-existing meta-analytic evidence regarding mixed modes effects for interviewer-administered surveys. Filling these blind-spots appears most promising.

Finally, there is a lack of meta-analytic studies on the impact of (mixing) modes on estimates of biases in terms of measurement and representation, not just on proxy variables or partial elements of bias, such as response rates (Groves and Peytcheva 2008).

References

Biemer, P.P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly, 74(5)*, 817-848.

Blom, A.G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review, 34(1)*, 8-25.

Bosnjak, M., & Danner, D. (2015). Survey participation and response. *Psihologija, 48(4*), 307-310.

Cook, C., Heath, F., & Thompson, R.L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement, 60(6)*, 821-836.

Bosnjak, M., Das, M., & Lynn, P. (2016). Methods for probability-based online and mixed-mode panels: Recent trends and future perspectives. *Social Science Computer Review, 34(1)*, 3-7.

De Leeuw, E.D. (1992). *Data Quality in Mail, Telephone and Face to Face Surveys*. TT Publikaties, Plantage Daklaan 40, 1018CN Amsterdam.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21(2)*, 233-255.

De Leeuw, E. D., & van der Zouwen, J. (1988). *Data quality in face to face interviews: A comparative meta-analysis. Telephone survey methodology*. Russell Sage Foundation, New York.

De Leeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys a meta-analysis. *Public Opinion Quarterly, 71(3)*, 413-443.

Dillman, D.A. & Messer, B.L. (2010). Mixed-mode survey. In J.D. Wright & P.V. Marsden (Eds.), *Handbook of Survey Research (2nd edition)* (pp. 551-574). San Diego, CA: Elsevier.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.

Dodou, D., & De Winter, J.C.F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*, 487-495.

Dwight, S.A., & Feigelson, M.E. (2000). A quantitative review of the effect of computerized testing on the measurement of social desirability. *Educational and Psychological Measurement, 60(3)*, 340-360.

Dybä, T., Kitchenham, B.A., & Jorgensen, M. (2005). Evidence-based software engineering for practitioners. *Software, IEEE, 22(1)*, 58-65.

Farrington, D.P., MacKenzie, D.L., Sherman, L.W., & Welsh, B.C. (Eds.). (2003). *Evidence-based crime prevention*. Routledge.

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior research methods, 47(4)*, 1237-1259.

Gnambs, T., & Kaspar, K. (2016). Socially Desirable Responding in Web-Based Questionnaires A Meta-Analytic Review of the Candor Hypothesis. Assessment, Online first: http://dx.doi.org/10.1177/1073191115624547

Goyder, J. (1985). Face-to-face interviews and mailed questionnaires: The net difference in response rate. *Public Opinion Quarterly, 49(2)*, 234-252.

Goyder, J. (1987). *The silent minority: Nonrespondents on sample surveys*. Westview Press.

Groves, R.M. (2004). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.

Green, K.E., & Hutchinson, S.R. (1996). *Reviewing the Research on Mail Survey Response Rates: Meta-Analysis*.

Groves, R.M., Fowler Jr, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2011). *Survey methodology (Vol. 561)*. John Wiley & Sons.

Groves, R.M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. *Public Opinion Quarterly, 72(2)*, 167-189.

Hox, J.J., & De Leeuw, E.D. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys. *Quality and Quantity, 28(4)*, 329-344.

Lozar Manfreda, K.L., Bosnjak, M., Berzelak, J., Haas, I., Vehovar, V., & Berzelak, N. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *Journal of the Market Research Society, 50(1)*, 79.

Mavletova, A. & Couper, M.P. (2015). A meta-analysis of breakoff rates in mobile web surveys. In D. Toninelli, R. Pinter &, P. de Pedraza (Eds.), *Mobile research methods. Opportunities and challenges of mobile research methodologies* (pp. 81-98). London, UK: Ubiquity Press.

Medway, R. L., & Fulton, J. (2012). When more gets you less: a meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly, 76(4)*, 733-746.

Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly, 79(1)*, 105-129.

Pring, R., & Thomas, G. (2004). *Evidence-based practice in education*. McGraw-Hill Education (UK).

Richman, W.L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84(5),* 754.

Rousseau, D.M. (2012). *The Oxford handbook of evidence-based management*. Oxford University Press.

Sackett, D.L., Rosenberg, W.M., Gray, J.M., Haynes, R.B., & Richardson, W.S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312(7023)*, 71-72.

Shih, T.H., & Fan, X. (2007). Response rates and mode preferences in web-mail mixed-mode surveys: a meta-analysis. *International Journal of Internet Science, 2(1)*, 59-82.

Shih, T.H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field methods, 20(3)*, 249-271.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133(5)*, 859.

Weisband, S., & Kiesler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. Proceedings of the ACM CHI 96 Human Factors in Computing Systems Conference April 14-18, 1996, Vancouver, Canada, pp. 3-10. http://www.sigchi.org/chi96/proceedings/papers/Weisband/sw_txt.htm

Ye, C., Fulton, J., & Tourangeau, R. (2011). More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly, 75(2)*, 349-365.

Mixing Online Panel Data Collection with Innovative Methods

Annette Scherpenzeel
Munich Center for the Economics of Aging (MEA)

1 Introduction[1]

Mixed mode studies in survey research have been around since a long time. Whenever a new mode of data collection emerges, it has been combined with existing traditional modes of data collection, and compared in terms of resulting response rates, response bias, data quality and measurement error (Couper 2011, for general overviews of traditional mixed mode studies see de Leeuw 2005 and De Leeuw et al. 2008). The first studies focused on comparisons of the traditional modes paper-and-pencil and face-to-face interviews. These modes were, in turn evaluated again after telephone interviewing evolved. The next stage was a series of studies comparing the effect of computer assisted forms of telephone interviewing (CATI) and face-to-face interviewing (CAPI). Since the emergence of web interviewing (CAWI), most mixed mode studies have compared this mode with the foregoing modes (for overviews of such studies, see Lozar et al. 2008; de Leeuw and Hox 2011).

The most recent comparative mode studies focus on the comparison of web interviewing on computers versus mobile devices, such as smartphones and tablets (see, for example, Antoun 2015; de Bruyne 2015; Lugtig and Toepoel 2016). Furthermore, technological development continues to offer ways of

1 We thank CentERdata for enabling the studies reported in this paper, in particular Josette Janssen, Salima Douhou, Maarten Streefkerk and Joris Mulder who organized the complex field work for all studies, CentERdata's director Marcel Das, and Maurice Martens, Iggy van der Wielen, Bart van Nieuwburg and Lennard Kuyten who developed the innovative apps and software.

S. Eifler und F. Faulbaum (Hrsg.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*, Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, DOI 10.1007/978-3-658-15834-7_2

data collection that go beyond asking survey questions, for example by continuously tracking people's behavior. It is this development, the mixing of interviewing with completely new ways of measuring behavior and feelings, which is the subject of the present paper. We will illustrate the possibilities of new mode mixtures with studies that have been conducted in the LISS panel, a probability based online panel which was especially developed for scientific research and for experimentation. The purpose of the paper is to not only demonstrate the interesting and truly innovative ways of data collection which have become possible, but to also evaluate the problems encountered and the resulting response rates and data quality. We will discuss the different ways in which the new techniques were combined with the normal panel interviews using the framework of mixed-mode designs given by De Leeuw and Hox (2011). They distinguish four main groups of mixed-mode design, which can be applied to traditional mixed-mode studies as well as to the most recent multi-mode web surveys:

1. Contacting by Different Modes: to reduce nonresponse, e.g. Advance letter, reminders, screening
2. Another Mode for Specific Questions in the Questionnaire: e.g. Self-administered part for sensitive questions
3. Different Modes for Different Respondents: same questions in different modes, to reduce nonresponse or noncoverage. E.g.: for non-internet group; for nonrespondents; respondents in different countries
4. Alternating Modes in a Longitudinal Design. E.g. first wave face-to-face, second wave Internet

We present the applications of new techniques and innovative modes within the framework of this categorization in order to show the purposes which such techniques can fulfill in survey data collection.

2 Mixed modes in probability based online panels

In the past decade several online panels for scientific purposes have been set up, based on true probability samples. In the United States, Knowledge Networks and RAND have for example built such panels. In Europe, scientific online panels have been established in different countries, such as the Longitudinal Internet Studies for the Social Sciences (LISS) Panel in the Netherlands, the German Internet Panel (GIP), the Longitudinal Study by Internet for