



# **MANAGING SURVEYS: TEN LESSONS LEARNED FROM WEB-SURVEYS**



## **WageIndicator Foundation - [www.wageindicator.org](http://www.wageindicator.org)**

WageIndicator Foundation is a non-profit NGO. It develops, operates and owns national WageIndicator websites with labour-related content, using data from its WageIndicator Salary and Working Conditions Survey, Minimum Wages Database, Collective Agreement Database, Salary Checks and Calculations, DecentWorkChecks and related Labour Law Database, and Cost of Living Survey and resulting Living Wages Database. The mission of WageIndicator is to promote labor market transparency for the benefit of all employers, employees and workers worldwide by sharing and comparing information on wages, labor law and career. WageIndicator does so by making this information freely available on easy to reach and read national websites in the national language(s), using sophisticated search engine optimization. All websites are accessible through mobile phones.

### **The author**

Kea Tijdens works with WageIndicator Foundation.

Contact: [keatijdens@wageindicator.org](mailto:keatijdens@wageindicator.org)

### **Acknowledgements**

The author acknowledges gratefully the contribution of Paulien Osse, director of WageIndicator Foundation, and all other WageIndicator colleagues and cooperating partners for expanding the Foundation's websites and its surveys to so many countries in so many years.

### **Keywords**

Web-surveys, multi-country surveys, multilingual surveys, survey management

### **Bibliographical information**

Tijdens K.G. (2020) *Managing surveys: ten lessons learned from web-surveys*. Amsterdam, WageIndicator Foundation, April.

@ 2020 by WageIndicator Foundation. All rights reserved.

WageIndicator Foundation, Email [office@wageindicator.org](mailto:office@wageindicator.org).

# Table of content

<b>Abstract</b> .....	1
<b>Ten lessons learned from web-surveys</b> .....	<b>1</b>
Introduction .....	1
Lesson 1 Defining web-surveys .....	2
Lesson 2 Software package versus own software .....	2
Lesson 3 Non-response .....	4
Lesson 4 Monitoring questionnaire intake .....	4
Lesson 5 API's versus showcards .....	4
Lesson 6 Reading versus verbal communication .....	5
Lesson 7 I versus you phrasing.....	6
Lesson 8 Generic versus personalised survey questions.....	7
Lesson 9 Counting survey questions versus characters read .....	7
Lesson 10 Managing a multi-country, multilingual survey.....	8

## Abstract

The paper reviews ten lessons learned from twenty years of managing a continuous, volunteer, multilingual, multi-country web-survey and several face-to-face surveys derived from the web-survey. The lessons address defining web-surveys, software package versus own software, non-response, monitoring questionnaire intake, API's versus showcards, reading versus verbal communication, I versus you phrasing, generic versus personalised survey questions, counting survey questions versus characters read, and managing a multi-country, multilingual survey.

# Ten lessons learned from web-surveys

## Introduction

For more than 35 years, I worked as a researcher at the University of Amsterdam. In these years I have designed and organised in total 24 printed, telephone, web-based, and face-to-face surveys and I have analysed the data and published the results. From 2000 onwards, I have been the research coordinator and survey manager of a web-survey on work and wages of the [WageIndicator Foundation](#). By February 2020 the Foundation exploits websites with information on work and wages for a large audience in [143 countries](#). Last year, the websites had almost 40 million visitors. The web-survey is continuously posted on all websites and web visitors are invited to complete the survey. In 20 years, the survey has developed from a printed survey in one country with one language, the Netherlands, to a multilingual web-survey 143 countries with in total 56 languages.

The web-survey has many face-to-face surveys derived from its questionnaire. In the early 2010's face-to-face salary surveys have been conducted in many African and Latin American countries. In 2019 in [Myanmar](#) and in 2020 in Bangladesh the Foundation has conducted country-specific face-to-face salary surveys. Since 2017, the Foundation also conducts so-called [decent-work-check surveys](#), whereby respondents' answers are directly compared to the national Labour Law to test compliance of their jobs in Indonesia, Ethiopia and Uganda. In Mexico, Facebook recruitment was used to target a very specific group of graduates of Mexican universities with the web-survey with additional specific questions for the target group. In 2020 the Foundation runs a mixed mode, multilingual salary survey in Netherlands, Spain, UK, India, Argentina and South Africa targeting [platform workers](#).

Among others, the data of these surveys is used for a [Salary Check](#), a tool for web-visitors to compare their earnings with their peers. The Salary Check acts as a mini survey of the Salary web-survey. Since the start of the surveys, more than 2 million people have completed the salary questions in all surveys jointly.

Although I'm fully aware that the management of continuous, volunteer, multilingual, multi-country surveys with many offspring surveys are a rare phenomenon, I believe that my experience can be useful for a broader audience. The objectives of this article are three-fold. First, to showcase how managing a complex survey can best be Internet-based by highlighting the fast technological developments over the past 20 years. Second, to detail how Internet-based Application Programming Interface tools (API's) can be used for long-list questions in a multilingual, multi-country context. Third, to reflect on the differences when the same survey questions are asked in a web mode compared to a face-to-face mode.

These three objectives are addressed here in the 10 lessons I've learned from running web-surveys and face-to-face offspring.

## Lesson 1

### Defining web-surveys

Whenever I present the WageIndicator web-survey most comments say that web-surveys are typically associated with undesirable volunteer or opt-in surveys. I usually reply with three arguments. The first argument is that for some social science research questions sampling frames for random sampled surveys are missing, or that random sampled surveys are beyond a reasonable budget. The second argument is that the boundaries between opt-in and random sampled surveys are blurring with the latter's response rates in quite some countries dropping below the desired 95%. The third argument is that web-surveys can address a random sampled population too, as several examples such as the [LISS panel](#) in the Netherlands shows.

My most important argument however does not address the sampling strategy, but the challenge to develop an Internet-based survey management approach when – as we do with the WageIndicator surveys - undertaking a multilingual, multi-country, multi-mode survey with a web version and face-to-face versions for selected groups. Internet-based survey management allows the survey manager to survey countries without survey agencies as the sole intermediaries between the central management and a country's survey, to manage multiple languages, and to profit from the many technological advancements made in the past years. In brief, the first lesson learned is that Internet-based surveys allow survey managers to keep control of their survey, as will be detailed in this article.

## Lesson 2

### Software package versus own software

The field of web-surveys has shown fast technological developments. Back in 2000, WageIndicator aimed to collect data on women's wages. The data was to be used to start a Salary Check where web visitors could check what their peers in the same occupation were earning. To collect wage data for enough occupations, the sample size needed to be large. Four major women's magazines in the Netherlands agreed to publish the survey in their magazines. One magazine also exploited a frequently visited website and converted the print version into a web-based form.

Unexpectedly, half of the [14,000 respondents](#) completed this web-based form.

Given this experience, WageIndicator and a [software company](#) developed a web-survey that was posted on its website in the Netherlands. The software generates the pages for the web-survey, their lay-out and the communication between respondent's device and the server where all entered data is stored, considering different browsers. To date WageIndicator is still working with this company and all its surveys run on customized software. A few times survey software packages have been reviewed, but these never met the requirements, specifically not regarding the

multilingual nature and the survey manager's possibilities to manage closely the survey.

Since 2004 a European grant allowed WageIndicator to create websites and web-surveys for seven neighbouring countries, of which one – Belgium – in two languages. Initially, an XML script had to be developed for each locale (language by country). In the 2010's, WageIndicator prepared one excel file with all source questions and answers and their translations. It was called 'mainfile'. That allowed the software company to develop one XML script for all web-surveys and translations being imported, initially through so-called PO files, but in a next stage directly from this 'mainfile'. In 2017 the 'mainfile' was extended with columns that added information needed for the dataset, such as the labels, formats, and alike, as well as start and end dates of incidental omnibus survey questions.

In 2018 calculation rules were included in the XML script, specifically aiming to calculate hourly wages and to identify outliers in wages and bonuses. In earlier years, outliers for working hours had been limited by limiting the values to be entered open number boxes and by alerting inconsistent reporting for the questions about working days per week and working hours per week. Similarly, inconsistent answers to questions about years of birth/first job/second job had been limited by adapting the drop-down lists to an adapted set of values. So, no respondent could start a first job before being born. In 2019 the software company developed a survey2csv script to ease the collection of the data from the server. Among others, it generates specific missing values for skipped questions, for breakoffs, and for questions not asked in a specific country. This script generates monthly csv data releases, which are ready for use in statistical software. Label sets are generated from 'mainfile'.

The latest technical developments refer to the development of a survey APP, used in the face-to-face surveys. Survey agencies and their interviewers are requested to use tablets or smartphone for the interviews, and to download the APP on these devices. Country-specific settings generate the appropriate questionnaire for the locale. No Internet connection is required for the interviews. Surveys can be uploaded in the central database once interviewers have WIFI access.

Other technological developments included the use of API's for long-list questions, as will be detailed in lesson 5, and the responsive design for the web-survey to adapt for smart phone use. By now, more than half of the web-survey respondents use the smartphone to complete the survey. After 20 years, I'm still happy that WageIndicator has developed its own software because that suits best the demands of the complex nature of its surveys. In summary, the second lesson learned is that Internet-based survey management is basically the only option to manage as many different surveys as WageIndicator currently does.

## **Lesson 3**

### **Non-response**

An important difference between printed, face-to-face and web-surveys relates to non-response, as this discloses in different forms. Non-response in print surveys happens when the questionnaires are not returned to the sender, because the respondent does not open the envelop in case of postal surveys, has no time or no liking to complete the survey, or just thinks that the survey doesn't fit his or her situation. Non-response in face-to-face surveys relates to survey respondents not willing to open the door for an interviewer, but if the interviewer is well-received, the chances for a break-off during the interview are little. Non-response in a web-survey are disclosed as break-off during survey completion. Any survey question that is not to respondent's liking for whatever reason can lead to break-off.

Managers of web-surveys need to take a standpoint how to cope with the dataset of incomplete questionnaires. Initially, WageIndicator decided to delete all data from questionnaires (in web-surveys called sessions) that had not been completed to the very last page, and therefore more than half of the sessions had to be deleted. By 2008, it was decided that by doing so, valuable information was deleted. Hence, policy changed and all sessions where at least the first page had been completed were included in the dataset. Any analysis, based on WageIndicator data, therefore selects the variables relevant for the publication. The third lesson learned is that the incomplete sessions require a change in the researcher's mind.

## **Lesson 4**

### **Monitoring questionnaire intake**

For specific purposes the WageIndicator web-survey incidentally acts as an omnibus survey or is converted into face-to-face surveys. This forces the data manager to track the data intake of the web or face-to-face surveys daily. The software company has built a monitoring tool, allowing survey managers and national survey agencies to monitor how many face-to-face interviews have been conducted and how many web-surveys have been completed. Survey agencies are given access to the monitoring tool to trace the performance of the interviewers and to act where needed. The monitoring tool allows a breakdown by interviewer, region, industry, occupation and a set of other variables. The fourth lesson learned is that Internet-based surveys allow to use a monitoring tool for tracking data intake.

## **Lesson 5**

### **API's versus showcards**

Since the early days of surveys, long-list questions have been asked using either an open text field with office coding or a showcard. In telephone surveys showcards need to be very short as respondents will not memorize long lists, but in face-to-face surveys the lists can be up to some 50 items. Web-surveys can facilitate lists with

many more items, because web-surveys allow respondents to use the search function or a search tree to browse the lists. For more information about long-lists variables [click here](#).

WageIndicator has built several Application Programming Interface tools (API's) for its long-list questions, among others for survey questions about occupation, education, industry, region, religion, company names and interviewer names. Any API is an excel file with four tabs. The first tab includes the codeset with the code, source label and translations. The second tab includes the structure of the codeset. In case of country-specific questions, e.g. education of region, the structure set identifies the country and then the related codes. This tab manages for example that in the survey in Indonesia only the Indonesian regions and nested cities are shown. The third tab includes mappings, e.g. the educational categories are mapped to the ISCED classification or the occupational categories are mapped into the ISCO 1-, 2-, 3- and 4-digit codes. The fourth tab includes the label set of the mapped codes. The software company has developed software that inserts the relevant parts of the excel files in the relevant language for the relevant survey question and saves the selected items for the dataset.

Some API's are large files, such as the region API with more than 21,000 lines covering all countries in the world, to meet the survey questions about the region/province and city. The source label of each province and city is in English, and for the locales they are in the national language. WageIndicator maintains an API of more than 1,700 occupational titles, an API for approx. 350 names of industries, and an API of country names, all translated in 47 languages. It also maintains an API for the 10 - 100 educational categories per country and an API for the most common religions per country. These API's cover the 143 countries and their source label is in English and for the locales they are in the national languages. For face-to-face surveys an interviewer API is maintained, allowing interviewers to identify themselves if they start a face-to-face interview. The fifth lesson learned is that API's result in good quality data without expensive office coding, and that they are useful in web-surveys and face-to-face surveys alike.

## **Lesson 6**

### **Reading versus verbal communication**

As said, the WageIndicator survey started as a printed survey, was then converted to a web-survey, and had then offspring to face-to-face surveys. The sixth lesson learned is that respondents' comprehension of survey questions and answers differs between reading and verbal communication. Per character, verbal communication takes approximately twice as much time as reading does. In 2019 WageIndicator had to convert its 15 minutes web-survey into a survey for face-to-face interviews in Myanmar. Half of the survey questions had to be cut for an interview of 15 to 20 minutes.

A second difference is that interviewers understand respondent’s level of comprehension and can adapt the survey questions accordingly, while in a web-survey no such communication exists and therefore questions should target the lowest level of comprehension. Reading survey questions is demanding for survey respondents with low literacy capacities and the number of characters in questions and answers influence both drop-out rates and response times. In the past years, the questions in the WageIndicator web-survey have been scrutinized to adapt the understanding of low-educated respondents. Where possible survey questions and answers have been shortened to the fewest words possible. For the face-to-face survey this turned out to be a good approach too.

To improve ‘communication’ with respondents the WageIndicator web-survey uses hints, e.g. for the question “Do you have children?”, a hint is shown “This includes step and foster children”. These hints have a different font and colour. In a face-to-face survey, this hint is typically not read out during the interview but is part of the interviewer training.

## Lesson 7

### I versus you phrasing

Respondents in web-surveys read the questions, whereas in face-to-face surveys questions are communicated between interviewer and respondent. This challenges the “I” versus “you” phrasing, specifically for opinion and attitude questions, and for the consent question. The figure shows two examples.

Face-to-face survey	Web-survey
Q: Can you decide on how to perform your tasks?	Q: I can decide on how to perform my tasks.
Q: Do you agree and understand that the answers you've given will be used for scientific purposes only?	Q: I agree and understand that the answers I've given will be used for scientific purposes only.

The seventh lesson learned is that a conversion of survey questions from web to face-to-face or vice versa cannot be a simple one-to-one conversion. The rule that the same questions must be asked in each survey mode should be relaxed. For the face-to-face survey in Myanmar, several web-survey questions had to be rephrased to suit the face-to-face interview setting.

## Lesson 8

### Generic versus personalised survey questions

The general opinion in survey methodology is that the phrasing of the question should be good, and translations should reflect this 'golden' phrased question. This conflicts the view that survey questions should be 'personalised' to keep the respondent's interest in the survey. Generic questions tend to address aggregated concepts. In my field of labour force surveys, it is common to ask questions about respondent's organisation. However, the word 'organisation' is a generic concept, probably not understood by everyone. In one WageIndicator survey we've identified how 'organisations' are called in the different industries. Therefore, the word 'organisation' is substituted with words such as farm, school, factory, plantation, office, hospital, and alike. We have not yet analysed the response to these questions but will do so one day. The eighth lesson points to personalisation of survey questions as a way forward that needs further study.

Personalisation also relates to addressing the survey questions to the right respondents. In labour force surveys, the distinction between dependent employees and self-employed is important, and survey questions need to take this into account by phrasing different questions to either group. For example, only employees should be asked about the relationship with their supervisor, whereas such a question should be skipped or rephrased for the self-employed. Another example regarding routing in the survey. If respondents indicate that they are employed in a small company, up to 5 employees, it doesn't make sense to ask for career possibilities in the company. WageIndicator has solved this issue by having survey questions for respondents in dependent employment and a parallel set of questions for the self-employed. Similarly, the unemployed respondents have a separate set of questions. The eighth lesson learned is that routing must be exploited to the most in order to ask only questions that are relevant for respondents, otherwise they will drop out.

## Lesson 9

### Counting survey questions versus characters read

For face-to-face surveys the fieldwork is by far the most expensive part of the survey. The price of fieldwork for a survey conducted by survey agencies usually depends on the number of questions asked. This pushes towards 'one size fits all' survey questions. For web-surveys the cost of the field work is relatively little, whereas the chance of survey breakoff is high and should be reduced. Hence, survey questions should involve as many routings as required, and thus asking many questions. For web-surveys it is important to count the number of characters respondents must read, and not the number of survey questions, because drop-out rates increase with number of characters read.

The example in the table compares two ways of asking the same question. In face-to-face surveys the choice would be in favour of the first, as it is counted as one

question. In web-surveys the expected distribution over the answers should be considered. If it is expected that most respondents will reply 'Yes', then the second question is more efficient to ask, as respondents on average must read fewer characters. Hence, the ninth lesson learned is that respondents' expected response rates should be considered in the design of survey questions for web-surveys.

Q: In the past seven days, have you worked your usual number of hours?	Q: In the past seven days, have you worked your usual number of hours?
A: Yes	A: Yes
A: No, I worked more hours	A: No
A: No, I worked less hours	Q IF NO: Why not?
A: No, I have not worked at all	A: I worked more hours
	A: I worked less hours
	A: I have not worked at all

## Lesson 10

### Managing a multi-country, multilingual survey

Managing a web-survey for so many years, I've appreciated most that I myself can manage the survey questions and answers for both the web-survey and its many face-to-face surveys. In addition, I'm able to manage its API's, and to adapt these easily when another country joins WageIndicator. I'm able to manage when translations must be added. I'm able to prepare versions of the web-survey for face-to-face surveys targeting specific groups, be it six industries in five provinces in Bangladesh or a 6-country study about platform work. Hence, the tenth lesson learned is that managing a complex survey is doable, provided it is an Internet-based survey.

In conclusion, I do hope that this article shows that web-surveys are not the worst of all survey types, but that a consistent policy towards an Internet-based survey is a way forward to keep multi-country, multilingual surveys manageable and affordable.

\*\*\*\*\*