

WageIndicator



in Argentina, Belgium, Brazil, Denmark, Finland, Germany, Hungary, Italy, India, Mexico, Netherlands, Poland, Spain, South Africa, South Korea, United Kingdom, United States

www.wageindicator.org

THE *WAGEINDICATOR* WEB-SURVEY: TECHNICAL DESIGN OF THE WEB SURVEY

Kea Tijdens, AIAS/University of Amsterdam, Netherlands

15/01/2007

The Questionnaire Management System (QMS)

The worldwide *WageIndicator* web-survey is managed in the Netherlands. The websites are hosted on three servers in the Netherlands, the USA and India. The survey has a sound multilingual Questionnaire Management System (QMS) that was totally renewed in the early half of 2005 to facilitate a worldwide web survey. A next update took place early 2007.

The QMS is implemented in a Plone environment, using a Zope/Python based CMS. The QMS consists of a maintenance module for the datasets, a maintenance module for the presentation layer and a module for the selection process (the so-called search tree). The QMS has a codebook mode, presenting the content of the presentation layer, i.e. the questionnaire, except for the content of the search trees. The implementation was based in an Eclipse environment and is based on Java, Struts, JSP and Maven. Its management and data layers are password protected. Changes in the QMS are made on a test-server and are uploaded on the production server on request. The QMS allows for uploading and downloading questionnaire related information.

For the search tree application a management web application was build using Struts and Hibernate. The application uses tomcat with a MySQL database. The Socrates questionnaire engine is an Open Source project. The engine is extensively tested, both by the current research team as by the public at large, visiting the website.

A multi-country and multi-lingual questionnaire

The Master questionnaire is the core of the QMS database. It holds all questions in the database. For each participating country the QMS has one or more language versions, for example es_ES for Spain or fr_BE and nl_BE for Belgium. Each *WageIndicator* website offers the web survey only in one language. Countries with more than one language employ at least one website for every language. Within a

country, the questions and responses in the questionnaires in different languages are exactly similar.

Multi-country surveys require country-specific questions. Therefore, the QMS allows questions to be switched on or off per country. For example, the question about commuting distance in kilometers is switched off for the UK, whereas the question about commuting distance in miles is switched on.

The question 'Do you have a mini-job?' is switched on for Germany and for Hungary only, as this phenomenon does not exist in other countries. This question is in the English language present in the Master-version and in the German language in the locale de_DE and in the Hungarian language in hu_HU.

If a question is switched off in the locale, it will not be shown in the locale's web survey, even if a translation is available. If a question is switched on, but no translation is available, the question is also not shown in the web survey. However, it is the data-team's policy that this does not happen. The major argument is that the QMS downloads are used for the data cleaning. The downloads show the on/off switch, but not the translations. Once a question is translated, but nevertheless is switched off, it remains in the locale. In a country with two or more languages, the same questions are switched on. Otherwise, data-analyses could not be performed at the level of countries.

Short, long and category questionnaires

In January 2005, Germany, UK and to a minor extent Spain and Poland expressed their desire to reduce the number of questions, because visitors had complained that the questionnaire was too long. Most questionnaires took 20-25 minutes to complete. Therefore, some questions were switched off for these countries. By the end of 2005, it turned out that in India and Brazil a short questionnaire was needed due to slow telecommunicating connections too.

At the same time, because an increasing number of online partners of *WageIndicator* wanted to build in the questionnaire in their website, a short questionnaire was needed. It was decided to offer the full questionnaire in the national *WageIndicator* websites and a short one for media partners, job agencies or job vacancy partners. The short questionnaire is reduced to the obligatory questions only, assuming that the visitor's employment status is 'employee'. This questionnaire can be completed in 5 minutes.

Until now, the *WageIndicator* web survey does not allow respondents to partially complete a survey, stop, and finish at a later time, although this is a convenience that helps increase response rates. Provided funding, this feature can be implemented.

Within a country and a locale, specific surveys can be held, aiming at a specific target group for specific research purposes. This is called a category questionnaire. In India, a category questionnaire of the web survey addresses IT-staff only. Within the locale en_IN, this one is identified as 'IndiaIT'. In the *WageIndicator* project, a few paper-based surveys have been held too. These surveys are also classified as a category survey. Hungary for example surveyed a representative sample of 5,000 individuals in the labor force, using a paper-based version of the questionnaire. Within the locale hu_HU, this survey is identified by the category 'Paper Version HU'.

The questionnaire units in the QMS

Questions are the basic entities in the questionnaire. Per question, the QMS consists of the units, shown in the Table.

The units per question in the QMS

UNIT	Description
ITEM	holds the text of the question.
HINT	includes the text of an instruction to a question, if available.
RELEVANT	includes the relevance rules per question. The relevance rules facilitate the routing through the questionnaire. The main routing is based on the first question about employment status, contst . Additional routing is based on scattered items. If for example a respondent has no children, no items follow about the age of the children.
REQUIRED	indicates that a question is obligatory. In a web-based survey, it is easy to make all questions obligatory. The visitor cannot proceed to the next page unless an answer is ticked. Yet, this will definitely decrease the number of completed questionnaires and increase the number of unreliable answers. In the <i>WageIndicator</i> questionnaire, only a limited number of questions are obligatory. Questions are obligatory for six reasons. First, for questions needed for the calculation of hourly wages, thus the questions about gross and net wage, payment period, allowances, working hours and working days per week. Second, for questions needed to make the Salary Check, such as tenure, education, gender, firm size, supervisory position or region. Third, for questions needed in many statistical analyses, such as age or household composition. Fourth, for questions needed for weighting the dataset. Fifth, for questions that are critical in the routing of the questionnaire. Sixth, for questions needed to give instructions to the respondents, for example when ticked 'I am a posted worker', an alert pops up, telling that the questionnaire should be completed for the current workplace, and not for the paying company.
VALUE	indicates the numerical values of the response items.
RANGE	frames the answer into a range of values, for example weekly working hours should be between 0 and 80.
CONSTRAINT	tests for inconsistent relationships, f.e. the year of re-entering the labor market after a break must be larger than the year leaving the labor market.
ALERT	reflects the text shown when the respondent does not pass the test for inconsistent relationships, f.e. 'Is your net wage larger than your gross wage?'. For technical and psychological reasons, the number of alerts is minimized.
CHOOSE	calls for a search tree. In the past years, search trees have been developed, offering the respondent a choice from long, detailed lists of occupations, industries, collective agreements, countries, regions, and trade unions. Two features increase the user-friendliness of the search tree. It allows the web-visitor to go easily back-and-forth in the search tree. In each tier and in each language the list of items is sorted alphabetically, allowing for an easy search. A search tree must meet contradictory demands. For reasons of user-friendliness a minimum number of words is required. The less words, the less likely that drop-out occurs and the more likely the respondent will provide a reliable answer. Yet, particularly for occupation and industry, visitors have to be able to identify their specific industry or occupation, and may not be able to identify broader concepts covering their particular enterprise. Will for example a respondent identify her employing company - a hairdresser – as part of the retail sector? This requires detailed lists of items.
TYPE	The unit TYPE reflects that the answer consists of a text area. The questionnaire employs four text boxes, relating to occupation, industry, wage periods and comments on the survey. They are of the type 'If you want to specify your occupation more accurately, please do so here', or 'If you have any comments on the questionnaire, please do so here'.
DATATYPE	The unit DATATYPE reflects that the answer consists of an amount xs:integer used for radio-button answers xs:string used for the text boxex xs:workweek used for hours per week xs:time used for hours per day xs:amount used for wages xs:decimal used for bonuses xs:boolean used for check boxes xs:amount-no-decimal used for number of supervisees if >9 xs:zipcode used for postal code

Response formats

Responses appear in a different form. A limited set of answer categories is used, shown in the Table.

Response formats in the QMS

UNIT	Description
radio-buttons	whereby only one answer can be ticked, for example Yes_No_I don't know
check boxes	for a multiple response question whereby several items can be ticked, for example the question how people found their job, i.e. though a newspaper announcement, family, a temp agency, a traineeship, etceteras
open-ended questions	inviting the respondent to type the response, using letters, for example for the question inviting visitors to describe their occupation in greater detail
amount boxes	allowing for respondents to fill in amounts as for wages, income and bonuses
time format	used for questions about working hours
drop down menus	used for measuring calendar years

The *WageIndicator* web survey currently does not use any of the following response formats, though this may change in the future.

- Visually anchored response categories or the ability to click the label itself to make a selection rather than clicking on the actual response button
- A constant sum indicator, for example used for questions asking to distribute a 100 points to a number of items or hours over a 24 hours day, whereby commonly the total points used and the remaining total are displayed
- Any visual or auditive features.

Response library

The QMS employs an response library, which is used for questions with similar responses, for example the response-sets Yes_No or Yes_No_DontKnow. The response-set YEARS_NOW_1950 generates a drop-down menu for all calendar years ranging from the year of survey to 1950, which is for example used for the questions about employment history.

A response set guarantees that the responses are similar in all questions calling for this set. Any change in the response set ensures that all questions that call for it still have the same answers. Response-sets allow for easy updating, for example the calendar years that need to be updated each year. Responses with country-specific lists, such as education and language spoken at home, are also stored in the answer library, because it facilitates easy updating.

Grid questions with random items

A grid question presents multiple items to be evaluated along a single set of response categories in the header. The stylized questionnaire shows the start (Matrixitem) and the end of the grid (End_Matrix). Five five-point Likert scales are available for grid questions, as is shown in the Table. All scales range from low (value 1) to high (value 5), including an option 'Not applicable (-8)'. The responses are stored in the response library. The middle categories (values 2,3,4) are not

presented in the Internet mode of the questionnaire, because the words do not fit the size of the line above the items, due to lay-out constraints.

The responses to five grid questions

Lowest value		Highest value
Strongly decreasing (1)	- - -	Strongly increasing (5)
Fully disagree (1)	- - -	Fully agree (5)
Never (1)	- - -	Always (5)
Highly dissatisfied (1)	- - -	Highly satisfied (5)
Not at all important (1)	- - -	Very important (5)

Initially in the *WageIndicator* web survey, a 4-item grid question was thought optimal for the response, aiming to prevent respondents with small screens from scrolling. However, gradually the extra 'Next' click was judged to be more important. Since 2006, therefore 8 instead of 4 items were used in a grid question. More questions on a screen are preferred. We assume a maximum at 8 items per grid question. Items should preferably not exceed 100 characters.¹

The large response on the web-survey allows for randomizing items, and still having sufficient respondents per item. Therefore, the QMS offers the possibility to present randomly up to 8 items from a pool of 8+ items. By doing so, this enlarges the scope for data analyses without having a longer questionnaire. Per country, the item pool and the number of items shown on the screen can vary. A particular set of randomly chosen items is shown for a couple of hours and is then replaced by another set of randomly chosen items. In this way problems due to back-and-forth behavior of web-visitors are avoided.

Follow-up questions are not possible in a randomized item pool. The item '... do you use the Internet' in the pool cannot be followed by an item 'I like using the Internet', if this item is only asked to respondents who have ticked 'yes' for Internet use.

'Don't know'

Offering 'Don't know' or 'Not applicable' as a response option is much discussed, because high percentages don't know reduce the number of observations for analyses considerably. Some researchers therefore take the standpoint that offering the 'Don't know' or 'Not applicable' response option should be minimized. The *WageIndicator* team however does not support this point of view. The main argument is to prevent break-off, when not offering a 'Don't know' or 'Not applicable' response option. In addition, statistical software can help to find and understand patterns in the 'Don't know' answers.

Client side paradata

The survey questionnaire was equipped with the client side paradata software. It registers the time when an item is clicked. The paradata is kept separate from the dataset.

¹ We assume that longer items have higher percentages of user missing, though we have not yet investigated this assumption.