SQP USERS' MANUAL

2017

Table of contents

The Survey Quality Predictor (SQP) is an online software which allows predicting the measurement quality of survey questions. SQP has two main functionalities:

1. It is an extensive open-source database of survey questions and quality predictions built up through the collaboration of the users.

The SQP database contains a wide range of survey questions concerning many different topics in many different forms and languages. The database is organized in studies which group related questions. Each question has attached information about the country and language of administration. Users can explore the database and search for their questions of interest. The quality of questions in the database can already be available from other users, but users can also obtain their own quality prediction by coding the characteristics of the question.

2. It is a coding system of formal and linguistic characteristics of survey questions, through which a prediction of reliability, validity and quality can be generated for more than 20 countries.

This prediction is based on a meta-analysis of the relationships between the quality estimates of survey questions obtained through multiple Multitrait-Multimethod (MTMM) experiments and the formal and linguistic characteristics of the questions in those experiments.

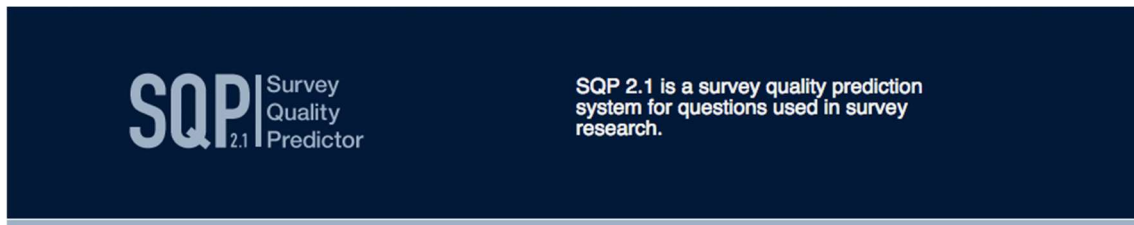The information provided by SQP is particularly useful to:

- consult, create, compare and evaluate questions in the SQP database,
- improve the design of questionnaires using the information about the quality of different question formats,
- and to correct for measurement error in the substantive analyses, after the data has been collected.

1. BECOMING A USER

Users should register and login to SQP to make full use of the software. Registration requires the specification of a user name, a verification e-mail, name and last name, and a password.

Because SQP is an open software, it is strongly recommended using the SQP Demo for testing and learning how to use it. The Demo interface is accessible from the login page.

Screen 1: SQP login

2. USING THE DEMO

The SQP Demo provides users with the main functionalities of SQP. Demo users can consult existing questions and quality predictions, can create new questions, and can code the characteristics of their own questions (or existing ones) to obtain their own quality predictions. The differences between the Demo and SQP are:

- In the Demo all created studies, questions and codes for the characteristics of questions are deleted periodically.

- The complete database of questions cannot be found in the Demo. This is only available for registered SQP users. However, the Demo provides a study which contains a sample of the type of survey questions that can be found in the SQP database. In this study, named as 'SQP Illustration', users can find a set of questions with MTMM estimates and authorized quality predictions.

Registering to SQP is recommended to access the full database and preserve users' work, once they are familiar with the software and the coding procedure of the characteristics.

3. GETTING STARTED

Once registered and logged into SQP, users access to the Home page. As illustrated in Screen 2, the Home page provides the different possibilities related to the database.

Screen 2: SQP Home

The SQP database contains many questions added by users on a wide range of topics and surveys, countries and languages. Many of the questions in the database have been already coded and their quality prediction is available. To start, users have three possibilities. First, users can follow the link 'Go to the database' to access it and search for questions (see more in Section 4). Second, users can introduce their own questions in the database following the link 'Add a new question' (see more in Section 5). And third, users can ask the SQP team to download information from the database by sending an email to: sqp@upf.edu.

4. THE DATABASE

The list of all available questions is provided in the SQP Database.

Screen 3: SQP Database

As presented in Screen 3, the database provides each question by row. The first column, 'Question', provides the questions' name (in navy blue), the short name (in a lighter blue colour), and a brief definition of the concept being measured by the question (in italic). The second column, 'Study', specifies the study where the questions belong. The following columns provide the 'Language' and the 'Country' of administration of the questions, and the last column, 'Quality', provides a summary of the type of quality information, if available.

Whether the quality prediction available belongs to the user himself or to another user is indicated by the keys �֍ and Ⓤ, respectively. The key Ⓐ indicates that the available quality prediction has been obtained by trained users and authorized by the SQP team. The MTMM quality estimates are not obtained through the SQP prediction system. Instead, most of them are used in SQP as part of the analysis together with the characteristics of the questions, which allows the quality of new questions to be predicted. All questions with MTMM estimates are identified by the key Ⓜ and they also have authorized quality predictions. Furthermore, the key Ⓒ indicates if the question prediction belongs to a different country than the one presented in the column 'Country'. The countries for which SQP can provide a prediction are limited. The countries for which the current version of SQP can provide a prediction are the following: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Netherlands, Norway, Poland, Portugal, Slovenia, Slovakia, Spain, Switzerland, Sweden, Ukraine, United Kingdom and United States. If a question belongs to a country not on this list, a country for the prediction, a 'Prediction country', is necessary to get a prediction of its quality. For more information, see Limit 1.

4.1. Searching questions in the database

Once on the SQP Database screen, if users are interested in a specific question or set of questions, they can use the filter options provided on the left-hand side of the database list (see Screen 3).

If users know of the existence of a specific question in a study or are only interested in a certain language or country, then they can add to the drop-box filter options the name of the study, the language and/or the country to obtain a shorter list. For instance, imagine users are interested in a question asked in the United Kingdom within the study 'ESS Round 1'. Users should select from the options: 'All Studies = ESS Round 1' and 'All countries = United Kingdom', as in Screen 3.1. With these details, the list has now been reduced to 319 questions.

Screen 3.1: Using drop-down filter options in SQP

If users are interested in a specific topic, they can use the 'Containing text' box for looking for specific wording in any part of the question (i.e. the name, the short name, the concept or the question texts). For instance, imagine users are interested in questions related to 'political efficacy'. Users should search for words related to the topic to get a shorter list of questions. Screen 3.2 illustrates that by searching for the words "political efficacy", the previous list of 319 questions has been reduced to 11 questions.

Screen 3.2: Using the containing text filter in SQP

This list shows that 9 out of the 11 questions belong to MTMM experiments. This means that information about the results of these experiments is available for each of these questions. These 9 questions have also already been coded and authorized by the SQP team, meaning that a reliable prediction is available for the users. Furthermore, the first question in the list, question B2, has another quality prediction which is made by another user.

Users interested in questions with available predictions or questions with MTMM estimates can use the selection criteria 'Only with Predictions' or 'Only with MTMM' to get a shorter list.

4.2. When quality information is available for questions in the database

If users choose a question from the database that has already information about its quality, the details of the predictions will already be available. For example, by selecting question B2 (in Screen 3.2) users will get Screen 3.3, in which a summary of the question details, its text and an overview of the quality information are presented.

Screen 3.3: Summary of question details and its available quality information

Screen 3.3 shows how question B2 was formulated and what is the quality obtained by the two different predictions and the MTMM estimate. Finding differences in the quality obtained for the same question is possible. First, differences between the MTMM estimates and the SQP predictions can occur because the quality estimates obtained through MTMM experiments are the result of a one point in time analysis, and the SQP predictions are based on the results of many of these analyses. In general, the SQP quality predictions are more reliable. For more information, see FAQ 2. Second, differences between other users' predictions and the authorized ones can occur because of differences in the choices made by each user when coding the characteristics of the question. Since the authorized predictions have been obtained by trained users under the control and supervision of the SQP team, their codes and predictions are more reliable. For more information, see FAQ 3.

Whenever the quality prediction of a question of interest is available but not authorized, it is recommended double-checking the coding before extracting any conclusions from it. The coding choices affect the prediction results; therefore, users should do the coding carefully.

From Screen 3.3 users can also obtain a more detailed presentation of the quality information following the link 'View Prediction Detail'. For example, by following the link on the authorized prediction users access to Screen 3.4.

Screen 3.4: SQP quality details

Screen 3.4 shows the breakdown of the quality prediction and MTMM estimate into reliability and validity. The values provided in this screen are the squared values of the quality coefficients. Users can get the information about the quality coefficients, as shown in Screen 3.5 following the link 'View Quality Coefficients'.

Screen 3.5: SQP quality coefficients details

Screen 3.5 shows the prediction and the MTMM coefficients for the reliability, validity and quality. Moreover, it presents the interquartile range and predicted standard error for the predictions and the 95% confidence interval for the MTMM results. For more information about the meaning of reliability, validity and quality, see FAQ 1.

Moreover, users can consult how the characteristics of the chosen question have been coded to obtain that quality prediction, following the link 'View Prediction Codes', and can also consult which suggestions are provided to improve the current quality prediction obtained, following the link 'View Potential Improvements'. More information about the codes and the potential improvements is provided in Sections 6 and 7, respectively.

4.3. When quality information is not available for questions in the database

The codes obtained by other users cannot be modified. A user willing to make different choices in the coding to obtain a different quality prediction should start the own coding. Going back to the summary of questions details (see Screen 3.6), users can start a new coding following the link 'Code question to create my own quality prediction'.

Screen 3.6: Summary of question details and its available quality information

How to code the characteristics of a question is explained in Section 6.

5. ADDING A NEW QUESTION TO THE DATABASE

Users can add their own questions to the database. Users can access the questions' creation screen following the link 'Add a new question' in the SQP Home (see Screen 2) and the SQP Database (see Screen 3).

To create a new question in the database, users should first specify its details, as shown in Screen 4.

Screen 4: Creation of new question details

These details refer to the study to which it belongs, the country in which the survey question is asked, and the language in which it is written.

- Users should choose a study which is meaningful for the question being created. Preferably, users should choose a study which has already been created by him or her, rather than adding questions to studies, which are neither his or her own nor related to their question. However, before creating a new study, users should also check that their questions have not been already introduced in the database by other users, and avoid duplicates by using those already created questions..

- SQP can provide a prediction for the following countries: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Ireland, Netherlands, Norway, Poland, Portugal, Slovenia, Slovakia, Spain, Switzerland, Sweden, Ukraine, United Kingdom and United States. Users willing to add a question belonging to a country not on the list, should select a 'Prediction country', that is the country which is most similar in terms of linguistic and cultural characteristics, to obtain a quality prediction.

- Users can select any language from the list since languages do not affect the prediction. The language should be selected accordingly to the language in which the question has been administrated to respondents.

- Users should specify a name for the question, which can be the name of the question in the questionnaire, a short name, which can be the name in the dataset, and a short description of the concept being measured. These will help other users to identify questions of interest within the database.

To add a new survey question, users should specify the text of the question in the correct boxes. SQP provides the users with three different boxes for the question text: the 'Introduction Text' box, the 'Request for Answer Text' box and the 'Answer Options' box.

Screen 5: Introduction of new question text

Introduction Text:
If Present - Text in the question used to introduce the concept of the question. Such as: "Now I am going to ask you about..."

Request for Answer Text:
Text in the question that requests an answer such as: "Please select the option....", "How much time..."

Answer options:
Answer options or numbers in the answer scale. One option per line.

Save Question

As illustrated in Screen 5, in the first box, users should write the introduction text of the question, if there is any. Introductions usually serve to introduce the concept of the next question(s). For example:

"Next I will ask you some questions about politics."

In the second box, users should write the text of the question, i.e. the request for an answer. This text should only include the text that is actually read to or by the respondent, thus, interviewer instructions should be left out. For example, a request for an answer would be:

"Could you tell me to what extent you agree or disagree with the way democracy works? Use this card to answer."

Finally, in the third box, the answer options text should be introduced. If it is a close-ended response scale, users should introduce one option per line. However, if it is an open-ended question, users should leave the box empty. Continuing with the example, the answer options could be introduced as follows:

"1. Completely disagree

2. Disagree

3. Neither agree nor disagree

4. Agree

5. Completely agree"

Only if the "Don't Know" option is explicitly provided, users should add the "Don't Know" as an extra option. However, if the "Don't Know" option is implicitly used or not used at all, then it should not appear in the answer options' box. In any case, it will be later coded as a formal characteristic of the response scale.

Once the question is saved, users can start coding the question. How to code the characteristics of a question is explained in section 6.

6. Coding the characteristics of a question

Users should code the characteristics of the selected or created question, following the Coding instructions (Survey Quality Predictor, 2017[1]).

The coding page, as shown in Screen 6, presents a summary of the question in the left-side and the starting of the coding process at the right-side of the screen.

Screen 6: Starting the coding procedure

The question summary includes the questions' creation details, the introduction (if present), the request for answer and the answer options' texts. These texts are just a reminder and should not be the object of coding. Coders should base their codes on the questionnaires and showcards (if used). For example, the interviewer instructions and the "Don't Know" option would hardly ever appear on the screens, therefore users should rely on the information in the questionnaires to see whether there are instructions or not.

Once the users start the coding process, SQP will provide the characteristics that need to be coded based on the question and previous coding decisions. For example, if users specify that there is no introduction present, SQP will not ask about the linguistic characteristics of the introduction. The characteristics are presented one after the other. Users need to answer all presented characteristics to finish the coding.

As illustrated in Screen 7, each characteristic and its possible choices appear above the question text, on the left-hand side of the screen. To keep track of the users' coding decisions, SQP provides the progress of the coding on the right-hand side of the screen.

---

[1] Survey Quality Predictor. (2017). SQP Coding Instructions. Universitat Pompeu Fabra. Barcelona, Spain. Retrieved from: http://sqp.upf.edu/media/files/sqp_coding_instructions.pdf

Screen 7: Coding in progress

The coding list in Screen 7, indicates that 6 characteristics have already been coded, among which are: the 'Domain', the 'Concept', the 'Social Desirability', the 'Centrality' and the 'Reference period' of the question. The user should now indicate in the left-hand side the current characteristic, 'Formulation of the request for an answer: basic choice', whether the request is formulated in an indirect or direct way, or if no request is present in this question.

6.1. The list of characteristics

The number of characteristics to be coded depend on the type of question. Table 1 provides an overview of the characteristics users will be asked to code in SQP, and a short description of each.

Table 1: Summary of SQP basic characteristics	
Characteristics	**Short Description**
Domain	The domain is the topic of the assertion that one wants to measure using this question. It is determined by the research goal.
Concept	The concept that they want to measure should be classified in one of the basic concepts distinguished on the list.
Social desirability	Social desirability relates to the choice of the domain. Identifies sensitive/delicate/irritable questions, which can bias the responses obtained.
Centrality	Centrality is also directly connected with the choice of the domain. It measures the familiarity of the respondents with the topic.
Reference period	Is again also connected with the research topic. The time period mentioned in the request can be: present, past or future.

Formulation of the request for an answer: basic choice	Identifies if a request is formulated as a direct or indirect request or if there is no request present, which means that the request (in this case the stimuli or statement) belongs to a battery of questions (except for the first item of the battery which will be either a direct or an indirect request introducing the battery).
WH word used in the request	Identifies questions that use words like: 'who', 'which', 'what', 'when', 'where' and 'how', 'to what extent', 'to what/which degree' or 'whether'. Or translations of them in other languages.
Request for an answer type	Identifies if a question is formulated in an interrogative, imperative or declarative form.
Use of gradation	Identifies requests that indicate responses that can be ordered from low to high or from high to low.
Balance of the request	Identifies leading questions. A request is Balanced when it contains both possible answer poles and Unbalanced when just one pole is mentioned.
Presence of encouragement to answer	Identifies leading questions. A request is Balanced when it contains both possible answer poles and Unbalanced when just one pole is mentioned.
Emphasis on subjective opinion	Identifies an emphasis on the opinion of the respondent about something, like: 'Please give us your opinion about…', 'According to you…', 'What do you think about…', etc.
Information regarding the opinion of other people	Identifies when opinions of other people are given in the request, like: 'Some people are against nuclear energy while others are in favour of it…'
Use of stimulus or statement in the request	Identifies batteries of questions. A stimulus in a question can be a noun or a combination of nouns. A statement in a question consists of complete sentences.
Absolute or comparative judgement	Identifies if the respondent has to compare two events or things. For example: 'Are you feeling better than last year or not?'
Response scale: basic choice	Identifies what types of answer options are provided: 1. Categories, 2. Frequencies, 3. Yes/ No scales, etc.
Response scale characteristics	Number of categories, frequencies and maximum possible value
	Labels of the categories
	Labels with short or long text
	Order of the labels
	Correspondence between the labels and the numbers of the scale
	Theoretical range of scale bipolar/ unipolar
	Range of the used scale bipolar/ unipolar
	Symmetry of the response scale

	Neutral category
	Number of fixed reference points
Don't know option	Identifies whether there is a Don't know option.
Interviewer instruction	If an interviewer administrates the survey, interviewer instructions will often be present, regarding which card to use or how to continue.
Respondent instruction	An instruction to the respondent is often present in imperative requests. However, these instructions explicitly ask the respondent to do something.
Extra motivation, information or definition available	Identifies if there is an extra sentence introducing a motivation, other information or a definition of something.
Introduction available	Identifies the presence of an introduction, which mainly serves to initiate the topic of the request to the respondent.
Linguistic characteristics of the introduction, the request for an answer and the answer scale	Number of sentences in the introduction
	Number of words in the introduction
	Number of subordinate clauses in the introduction
	Request present in the introduction
	Number of sentences in the request
	Number of words in the request
	Total number of nouns in the request for an answer
	Total number of abstract nouns in the request for an answer
	Total number of syllables in request
	Number of subordinate clauses in request
	Number of syllables in answer scale
	Total number of nouns in answer scale
	Total number of abstract nouns in answer scale
Show card used	Identifies the use of Show cards. These are sometimes used during the interview to show the response options or to assist in explaining the question.
Show card characteristics	Horizontal or vertical scale
	Overlap of text categories
	Numbers or letter before answer categories
	Scale with numbers or numbers in boxes
	Start of the response sentence on the show card
	Question on the show card
	Picture provided
Computer assisted	Identifies the mode of data collection: if the interview is computer-based or not.

Interviewer	Identifies the mode of data collection: if it is a personal interview or a self-administered questionnaire.
Visual or oral presentation	It identifies if the questionnaire is self-administered (visual) or interviewer-administered (oral).
Position	Identifies the position of the question in the questionnaire.

6.2. Help while coding the characteristics of a question

SQP provides coding Help screens, such as the yellow box presented in Screen 7.1, which explain every characteristic and its choices in detail.

Screen 7.1: Coding Help screens

6.3. Coding completed

To complete the coding of the characteristics of a survey question and obtain a quality prediction, users will need to code between 30 and 60 characteristics, from a total of 73. At first, this may seem rather tedious. However, as soon as users become familiar with the characteristics and their meanings, the coding procedure becomes simpler and faster. As illustrated in Screen 7.2, SQP will display a message when the coding is completed.

Screen 7.2: Coding complete

By selecting 'Get quality prediction', users obtain predictions of the reliability, validity and quality of the question, as illustrated in Screen 7.3. By selecting 'Back to question list' users go back to the screen which displays the list of filtered questions, as shown in Screen 3.2.

Screen 7.3: SQP quality details

In the predictions overview, users will find the information regarding the reliability, validity and quality of the predictions and those from the MTMM results (if available). Moreover, users have the option to 'View Quality Coefficients' (see Screen 7.4). These are the square root of the quality indicators on the previous screen.

Screen 7.4: SQP quality coefficients details

In Screen 7.4, SQP also provides the uncertainty of the estimates by indicating their interquartile ranges and the standard errors. Moreover, SQP allows users to view the potential improvements that can be made in the survey question to obtain a higher quality. The potential improvements feature can be accessed from the link 'View potential improvements' on the bottom left hand side of Screens 7.4 and 7.5.

7. Potential improvements

The potential improvements feature suggests to users which characteristics of the survey question could be changed to obtain a prediction of higher quality, as illustrated in Screen 8.

Screen 8: View potential improvements tool, 20 first characteristics

In the context of designing a survey question, users should consider the purpose of the question, and evaluate the suggestions critically when deciding whether they apply to the survey question.

This feature presents first a selection of characteristics which are most important, in terms of affecting the prediction, for all questions. To obtain the complete list of suggested improvements, click on 'Evaluate all remaining variables' (in Screen 8). The complete list consists of 63 characteristics that can be used to improve the question, as illustrated in Screen 9.

Screen 9: View potential improvements tool, all characteristics

Question	
B2 / **PolCmpl** / *Internal political efficacy , complicated*	
ESS Round 1 United Kingdom - English	

Request for Answer Text:
How often does politics seem so complicated that you can't really understand what is going on? Please use this card.

Answer options:
• Never
• Seldom
• Occasionally
• Regularly
• Frequently

- Hide Question Text

Quality Prediction

⊕ My Quality Prediction

Reliability = 1 - random error r^2 0.662

Validity = 1 - method effect v^2 0.977

Quality = reliability (r^2) x validity (v^2) q^2 **0.646**

Potential Improvements

All Variables

Variable	Potential Max Quality by Change in Variable	
Number of categories	0.679	+ 0.03
Total number of words in request	0.675	+ 0.03
Correspondence between labels and numbers of the scale	0.662	+ 0.02
Knowledge provided	0.658	+ 0.01
Neutral category	0.655	+ 0.01
Request for an answer type	0.653	+ 0.01
Number of subordinate clauses in request	0.652	+ 0.01
Start of the response sentence on the visual aid	0.652	+ 0.01
Don't know option	0.651	+ 0.01
Total number of nouns in question	0.651	+ 0.01
Presence of encouragement to answer	0.651	+ 0.01
Labels of categories	0.651	+ 0.01
Order of the labels	0.650	
WH word used in the request	0.649	

Selected Variable

Click on a variable on the left to see the potential change in the quality that could be produced by a different value for that variable.

Detailed information about the specific change to be made is obtained by clicking on the characteristic of interest, as illustrated in Screen 10. Then, on the right hand side, users obtain information about how to get the largest improvement in quality regarding the selected characteristic .

Screen 10: Potential improvement suggestion

Question	
B2 / PolCmpl / *Internal political efficacy , complicated*	
ESS Round 1 United Kingdom - English	

Request for Answer Text:
How often does politics seem so complicated that you can't really understand what is going on? Please use this card.

Answer options:
- Never
- Seldom
- Occasionally
- Regularly
- Frequently

- Hide Question Text

Quality Prediction

✿ My Quality Prediction

Reliability = 1 - random error	r^2	0.662
Validity = 1 - method effect	v^2	0.977
Quality = reliability (r^2) x validity (v^2)	q^2	**0.646**

Potential Improvements

All Variables

Variable	Potential Max Quality by Change in Variable	
Number of categories	0.679	+ 0.03
Total number of words in request	0.675	+ 0.03
Correspondence between labels and numbers of the scale	0.662	+ 0.02
Knowledge provided	0.658	+ 0.01
Neutral category	0.655	+ 0.01
Request for an answer type	0.653	+ 0.01
Number of subordinate clauses in request	0.652	+ 0.01
Start of the response sentence on the visual aid	0.652	+ 0.01
Don't know option	0.651	+ 0.01
Total number of nouns in question	0.651	+ 0.01
Presence of encouragement to answer	0.651	+ 0.01
Labels of categories	0.651	+ 0.01
Order of the labels	0.650	
WH word used in the request	0.649	

Selected Variable

Number of categories (ncategories)

Current choice: **5**

Choice	Average what if-prediction	
2	0.638	- 0.01
4	0.640	- 0.01
5	0.646	
7	0.651	+ 0.01
11	0.679	+ 0.03

Number of categories

Enter the number of categories of the answer options. For example: SQP will provide a 'Suggested value' for this particular characteristic. In the example, the total number of categories is... more >>

The effects on the quality that SQP calculates are independent of other related changes that the tool suggests. This means that the effects presented do not consider related characteristics that may also change when following the improvement suggestions. For instance, increasing or reducing the number of points in a scale may lead to other changes in the characteristics related to the answer options, such as the labels or the presence of a neutral category. These consequences are not immediately considered. Therefore, it makes sense to reformulate the question and test the new question in the same way as indicated before.