Survey Quality Predictor

A survey quality prediction system for continuous questions used in survey research

Lydia Repke

Lukas Birkenmaier

Cem Eryar

Invalid Date

Introduction

The Survey Quality Predictor (SQP) 3.0 is an open-access system for predicting the reliability, validity, and quality of survey questions for continuous latent variables based on their formal and linguistic characteristics (such as the properties of the answer scale and the administration mode). SQP is grounded in decades of methodological research and enables survey designers and researchers to evaluate and enhance the quality of their measurement instruments. By systematically linking the characteristics of survey items to their measurement quality, SQP provides a practical framework for improving data quality in survey-based research (Felderer et al., 2024).

Relying on SQP's prediction power and its day-by-day extending database, researchers can effectively utilize its features in at least three different ways (Felderer et al., 2024):

- During the questionnaire development process, researchers can consult the database, predict the qualities of different item versions, and make informed comparisons.
- SQP enables researchers to compare different language versions of the same item by analyzing the coded characteristics. This analysis helps identify potential methodological differences between the source items and their translations, thus supporting quality control of translations in cross-cultural surveys.
- SQP's quality predictions can be used in substantial analyses to correct for measurement errors of items capturing subjective concepts (Andrews, 1984; Saris et al., 2022; Saris & Revilla, 2016).

In this tutorial, we will explain the conceptual framework of SQP, demonstrate its function through an example, and conclude with a discussion on its benefits, limitations, and potential for future development. Overall, this tutorial aims to ensure that survey designers and researchers can harness SQP to create better surveys, mitigate potential biases, and enhance the overall quality of their research.

SQP

The Survey Quality Predictor (SQP) is a software tool developed to predict the measurement quality of survey items for continuous latent variables. It operates by analyzing survey items' formal and linguistic characteristics (GESIS – Leibniz Institute for the Social Sciences, 2024). SQP's development is rooted in decades of methodological research, beginning in the 1980s when Willem Saris and his colleagues conducted a series of multitrait-multimethod (MTMM) experiments. These experiments aimed to quantify the reliability and validity of survey items and disentangle random and systematic errors, ultimately improving the quality of survey instruments

(Saris & Gallhofer, 2014). The insights from these early experiments laid the foundation for SQP's predictive capabilities.

The first version of SQP was released in 2001 (Saris, 2001). This version operationalized the findings from MTMM research into a practical tool that allowed users to code survey items and predict their reliability and validity. While groundbreaking, SQP 1.0 had limitations, such as a relatively small database on which the prediction was based and a lack of advanced features (GESIS – Leibniz Institute for the Social Sciences, 2024). A decade later, SQP 2.0 introduced significant advancements, including an expanded database and a new prediction algorithm. In 2015, SQP received a usability makeover, improving the user experience as version 2.1 (Saris et al., 2022). The release of SQP 3.0 in 2022 marked a major milestone in survey research methodology. This version included thousands of survey items coded in 28 languages for 33 countries, making it particularly valuable for cross-cultural and multilingual studies like the European Social Survey (ESS). Its updated prediction algorithm accounted for interactions between item characteristics, offering more accurate reliability and validity estimates. SQP 3.0 also introduced new features that improved and enhanced the user experience (GESIS – Leibniz Institute for the Social Sciences, 2024).

SQP has evolved into a comprehensive tool that supports survey researchers in designing highquality questionnaires, ensuring cross-cultural comparability, and improving the quality of survey questions. Its integration of predictive modeling, and user-friendly features positions SQP as an indispensable resource in modern survey research.

The Impact of Survey Question Design and the Role of SQP

:::: {style="text-align: justify;"}

The decisions made in designing survey questions significantly impact respondents' answer behavior, as variations in question formulation, response scales, and contextual features influence how individuals interpret and respond to survey items. For instance, Saris et al. (2022) illustrate this with an example of measuring trust in the economy. They compare the effects of three different response scale designs on the question:

Examples

(Q1) On the whole, how satisfied are you with the present state of the economy in Britain?

4-point bipolar scale:

- 1. Very satisfied
- 2. Satisfied
- 3. Dissatisfied
- 4. Very dissatisfied

4-point unipolar scale:

- 1. Not at all satisfied
- 2. Satisfied
- 3. Rather satisfied
- 4. Very satisfied

Numeric scale (0-10):

Very dissatisfied										Very satisfied
0	1	2	3	4	5	6	7	8	9	10

:::::: {style="text-align: justify;"}

The results revealed significant differences in response distributions and correlations with related variables based on the scale used. For instance, numeric scales tended to produce higher correlations with other trust measures, whereas bipolar and unipolar scales generated distinct response patterns. These variations stem from the way survey questions influence respondents' cognitive processing.

Bipolar scales emphasize contrasts between extremes, potentially encouraging more polarized responses, while unipolar scales direct attention to a single dimension, which may reduce polarization but introduce other biases. In contrast, numeric scales convey a sense of precision and neutrality yet require respondents to map their attitudes onto a numeric continuum. These subtle but impactful differences highlight the critical role of thoughtful survey design, especially for subjective constructs like trust or satisfaction (Saris et al., 2022).

Therefore, SQP is essential for ensuring high-quality survey research, as it systematically addresses the impact of question design on measurement outcomes. By predicting the reliability and validity of survey questions, SQP helps researchers identify potential biases and unintended effects introduced by variations in scales, wording, or structure. For example, as demonstrated by differences between numeric, bipolar, and unipolar scales, minor design changes can profoundly influence respondents' cognitive processing and, consequently, the data collected. This is particularly critical for subjective measures like trust, satisfaction, or attitudes, where nuanced design decisions can significantly affect data quality and the validity of subsequent analyses.

Introduction to the Coding Process in SQP

The SQP Website provides researchers with a comprehensive tool to evaluate and enhance the quality of survey questions systematically. At the heart of this process lies the coding of survey items, where specific characteristics of questions are analyzed to generate predictions about their reliability and validity. The SQP website offers an intuitive interface that facilitates both the coding process and access to its extensive database of survey items.

To use the tool's full functionality, researchers must register and create an account on the platform. Once logged in, they can interact with a searchable database containing a vast collection of survey items from various studies, languages, and countries, add new survey questions, and code them to get quality predictions.

GESIS Leibniz Institute for the Social Sciences	Surve Predie	ey Quality ctor	✓ Resources	∨ Ab	oout SQP	✓ Access	Database
Database	È						
Filter by 🚯							
Study	Quest	tion owner 🗸 🗸	Language	✓ C	ountry	~	Quality prediction
Advanced filters							~
Applied filters							
Studies:	Question owner:	Language:	Country:	Quality p	rediction: Te	ext:	Characteristics:
					clear filter	s	apply filters
Legend of quality:	M My codings	A Authorized codings	0 Other users' coding	rs (clear filter	s	apply filters
Legend of quality:	My codings	Authorized codings Study	Other users' coding Language Court			s Actions	

The database allows researchers to explore existing items, apply filters, and use advanced search options to focus on specific topics or coding characteristics. For instance, users interested in survey items coded for "centrality" or measuring concepts like "political efficacy" can refine their search accordingly.

As mentioned above, the platform further enables researchers to contribute to the database with new survey items by entering key details, such as the question text, response options, and metadata about the study. After adding an item, users proceed to the coding stage, where they thoroughly analyze the survey question's characteristics. This process involves evaluating features like the response scale, linguistic complexity, and the formulation of the request for an answer.

Description
The topic of the assertion that one wants to measure using this
survey item, determined by the research goal.
The concept one wants to measure, classified as one of the basic
concepts distinguished in SQP.
Relates to the domain choice; identifies sensitive or delicate items
that can bias responses.
Measures the familiarity of respondents with the topic, linked to the
domain choice.
Refers to the time period mentioned in the request, e.g., present,
past, or future.
The basic formulation of the request, e.g., direct, indirect, or
omitted in item batteries.
Requests starting with words like who, what, when, where, how,
etc., or their translations.
Requests may be formulated as interrogative, imperative, or
declarative.
Identifies requests where responses can be ordered, e.g., low to high
or vice versa.
Determines whether requests are balanced (mentioning both poles)
or unbalanced.

List of Characteristics

Variable Name	Description
Presence of	Encourages responses, e.g., 'Please, tell me' or 'We would like to
encouragement	ask you'
Emphasis on	Focuses on subjective opinion, e.g., 'What do you think about?' or
subjective opinion	'According to you'
Information	Includes information on others' opinions, e.g., 'Some people are
regarding others'	against nuclear energy while others support it.'
opinions	
Use of stimulus or	Survey items with a stimulus (noun/words) or statements (complete
statement	sentences).
Absolute or	Determines whether respondents provide absolute or comparative
comparative	judgments.
judgment	
Response scale:	The type of response options, e.g., two-category, more-step
basic choice	procedures, open-ended, etc.
Response scale	Characteristics like number of categories, maximum values, labels,
characteristics	and scale range.
Don't know option	Indicates whether a 'Don't know' option is present.
Interviewer	Instructions provided to interviewers, e.g., about showcards or visual
instruction	aids.
Respondent	Instructions provided to respondents, usually in imperative or polite
instruction	forms.
Extra information or	Additional information or definitions provided for clarity, though
definition	not strictly necessary.
Knowledge provided	Defines whether definitions, explanations, or both are provided.
Introduction	Introduces the survey topic or related questions, often at the
available	beginning.
Linguistic	Covers linguistic features of introduction, request, and answer scale,
characteristics	e.g., number of sentences, words, or abstract nouns.
Showcard used	Indicates whether a showcard is used for response options or
	additional assistance.
Showcard	Details characteristics of showcards, e.g., scale orientation, labels, or
characteristics	numbers.
Computer-assisted	Indicates whether the mode of data collection is computer-assisted.
Interviewer	Specifies whether the mode is interviewer-administered or
	self-administered.
Visual or oral	States whether the questionnaire is visually presented or orally read
presentation	to respondents.
Position	Indicates the position of the survey item within the questionnaire.

Although coding between 30 and 60 characteristics may seem daunting initially, the platform provides clear instructions and contextual help to guide users through the process.

Once the coding is complete, SQP generates a prediction of the item's measurement quality, including its reliability, validity, and potential method effects. Furthermore, the platform includes the function to compare codings across different survey items or variations of the same item, highlighting how design choices can influence measurement quality. SQP also supports transparency and replicability by enabling users to document their own survey questions and coding decisions. These features are particularly beneficial for longitudinal or comparative research, where consistent question design is critical.

For a detailed explanation of the coding process and additional resources, see the full SQP guideline here: SQP Manual.

Practical Application of SQP

To illustrate how measuring the same construct can vary depending on the type of question and the answer scale used—and how these variations can lead to different quality prediction scores—we examine a question on economic satisfaction from Round 4 of the European Social Survey (ESS). You can click here to access the questions in the example below.

Questions	Question #1	Question #2
Study	ESS Round 4	ESS Round 4
Question name	HS7 / testc7 / Political satisfaction: country's economy	HS19 / testc19 / Political satisfaction: country's economy
Country/Country prediction	United Kingdom	United Kingdom
Language	English English	
Introduction	Now we want to ask you a few questions about politics and government.	
Request for an answer text	On the whole how satisfied are you with the present state of the economy in (Britain/the UK)? Please use this card where 0 means dissatisfied and 10 means satisfied.	Now please tell me to what extent you agree or disagree with each of the following statements. On the whole I am satisfied with the present state of the economy in (Britain/the UK).
Answer options	0 Dissatisfied 1 2 3 4 5 6 7 8 9 10 Satisfied	1 Agree strongly 2 Agree 3 Neither disagree nor agree 4 Disagree 5 Disagree strongly
User	authorized coding	authorized coding

- The first version "HS7 / testc7 / Political satisfaction: country's economy" measures the construct using a numeric rating scale that ranges from "0 Dissatisfied" to "10 Satisfied." The question asks, "On the whole, how satisfied are you with the present state of the economy in (Britain/the UK)?" This scale captures various degrees of satisfaction along a continuum, providing respondents with a range of options to express their attitudes.
- The second version "HS19 / testc19 / Political satisfaction: country's economy" measures the same construct using a 5-point Likert scale with fully labeled response options. The question states, "To what extent do you agree or disagree with the following statement: On the whole, I am satisfied with the present state of the economy in (Britain/the UK)." Here, response options range from "1 Agree strongly" to "5 Disagree strongly," categorizing responses into distinct levels of agreement or disagreement.

Codings	Coding of Question #1	Coding of Question #2
> Domain	National politics	National politics
> Domain: national politics	Economic/financial matters	Economic/financial matters
> Concept	Feeling	Feeling
> Social Desirability	A bit	A bit
> Centrality	Rather central	Rather central
> Reference period	Present	Present
> Formulation of the request for an answer: basic choice	Direct request	Indirect request
> WH word used in the request	WH word used	WH word used
> 'WH' word	How (extremity)	How (extremity)
> Request for an answer type	Interrogative	Imperative
> Use of gradation	Gradation used	Gradation used
> Balance of the request	Unbalanced	Balanced or not applicable
> Presence of encouragement to answer	No particular encouragement present	Encouragement present
> Emphasis on subjective opinion in request	No emphasis on opinion present	No emphasis on opinion present
> Information about the opinion of other people	No information about opinions of others	No information about opinions of others
> Balance of the request	Unbalanced	Balanced or not applicable
> Presence of encouragement to answer	No particular encouragement present	Encouragement present
> Emphasis on subjective opinion in request	No emphasis on opinion present	No emphasis on opinion present
> Information about the opinion of other people	No information about opinions of others	No information about opinions of others
> Use of stimulus or statement in the request	No stimulus or statement	Stimulus or statement is present
> Absolute or comparative judgment	An absolute judgement	An absolute judgement
> Response scale: basic choice	More than two-categories scales	More than two-categories scales
> Number of categories	11	5
> Labels of categories	Partially labelled	Fully labelled
> Labels with short text or complete sentences	Short text	Short text
> Order of the labels	First label negative or not applicable	First label positive
> Correspondence between labels and numbers of the scale	Ne Medium correspondence	Low correspondence
> Theoretical range of the concept bipolar/unipolar	Theoretically bipolar	Theoretically bipolar
> Range of the used scale bipolar/unipolar	Bipolar	Bipolar
> Symmetry of response scale	Symmetric	Symmetric
> Neutral category	Present	Present

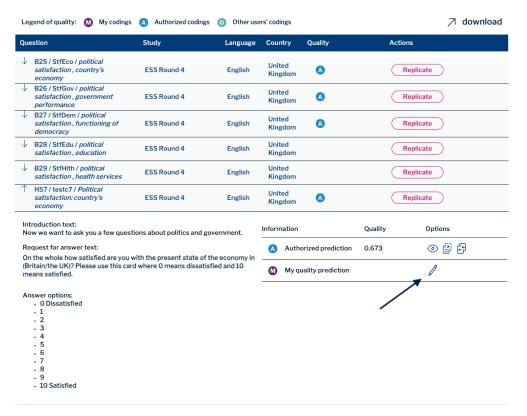
Using SQP, both survey items were coded to capture their design characteristics, revealing notable differences. Shared attributes included the domain, classified as "National Politics," and the concept of "Feeling," specifically focusing on satisfaction with the economy. However, variations emerged in other areas. The numeric scale question was coded as using a symmetric, theoretically bipolar scale with partially labeled categories, emphasizing both extremes and intermediate values. In contrast, the Likert scale question was coded as having fully labeled categories, a balanced structure, and an indirect request format. Additional differences included the presence of encouragement to respond in the Likert scale version, which was absent in the numeric scale question. These distinctions, highlighted in the coding interface in blue, reflect design nuances that can influence respondents' interpretations and responses.

Quality prediction	Prediction of Question #1	Prediction of Question #2
Reliability = 1 - random error (r ²)	0.811	0.667
Validity = 1 - method effect (v ²)	0.829	0.826
Quality = reliability (r ²) x validity (v ²)	0.673	0.552 ←
Reliability Coefficient (r)	0.901	0.817
Validity Coefficient (v)	0.911	0.909
Quality Coefficient (q)	0.820	0.743
Method Effect Coefficient (μ)	0.413	0.417
Reliability Coefficient Interquartile Range	(0.880, 0.940)	(0.790, 0.860)
Validity Coefficient Interquartile Range	(0.900, 0.970)	(0.880, 0.970)

SQP generated quality predictions for both types of questions, highlighting the impact of various design choices. The numeric rating scale question achieved a reliability score of 0.811 and a validity score of 0.829, resulting in an overall quality score of 0.673. In contrast, the 5-point Likert scale question received a reliability score of 0.667 and a validity score of 0.826, leading to an overall quality score of 0.552. These results demonstrate that while both scales have similar validity scores, their reliability scores differ significantly, resulting in better overall quality for the 11-point scale. This comparison underscores the importance of aligning question design with research objectives to optimize data quality.

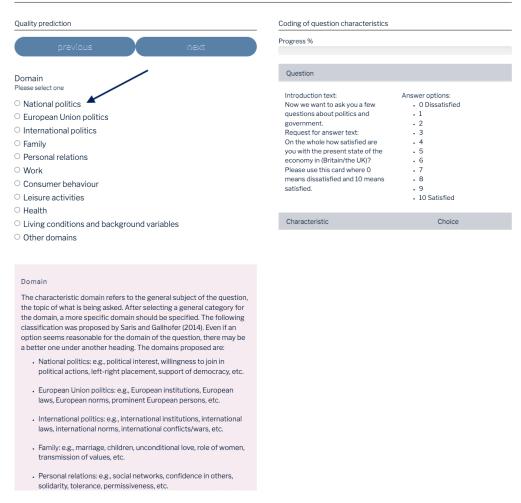
A Step-by-Step Guide to Hands-on Application

Step 1: To begin creating your own quality prediction, locate the question in the survey item list. Simply click the pen icon next to the relevant question to open the coding interface. This will allow you to code the question's characteristics and predict its measurement quality.



Step 2: Once you've accessed the coding interface, you will start by selecting the appropriate characteristics for the question. For example, since the question about satisfaction with the economy pertains to national politics, select "National politics" from the list of options. A helpful pink box appears below the coding options,

providing detailed explanations for each characteristic. Use this guidance to ensure you understand the options and make accurate selections. After completing this step, click "Next" to proceed to the next characteristic.



ESS Round 4 / HS7 / testc7 / Political satisfaction: country's economy / United Kingdom / English

Step 3: Repeat the process for each characteristic of the survey item. As you progress, the progress bar at the top of the screen will indicate how much of the coding is complete. On the right-hand side, your previous answers are displayed, enabling you to review and make corrections if necessary. If you realize an error or need to adjust a response, simply go back to the corresponding step and update your selection. Continue coding until the progress bar is fully completed, ensuring that all characteristics have been coded accurately.

Code question

Quality prediction	Coding of question characteristics				
previous next	Progress %	K			
Emphasis on subjective opinion in request Please select one	Question				
 No emphasis on opinion present Emphasis on opinion present 	Introduction text: Now we want to ask you a few questions about politics and government. Request for answer text: On the whole how satisfied are	Answer options: • 0 Dissatisfied • 1 • 2 • 3 • 4 • 5 • 6 • 7 • 8 • 9			
Emphasis on Subjective Opinion in Request A request may or may not provide an emphasis on the subjective opinion of the respondent.	you with the present state of the economy in (Britain/the UK)? Please use this card where 0 means dissatisfied and 10 means satisfied.				
 Emphasis on opinion present: occurs when requests use words that seek to obtain the subjective opinion of the respondent. For example: 	Characteristic	• 10 Satisfied Choice			
"Please, give us your opinion about"	> Domain	National politics			
"What do you think about?"	> Domain: national politics	Economic/financial matters			
"According to you, what is the"	> Concept	Feeling			
"In your opinion"	Social Desirability	A bit			
"To what extent would you say the government is doing a good job?"					
	> Centrality	Rather central			
 No emphasis on opinion: occurs when requests do not use specific words that seek to obtain the subjective opinion of the 	Reference period	Present			
respondent.	Formulation of the request for an answer: basic choice	Direct request			
	> WH word used in the request	WH word used			
	> 'WH' word	How (extremity)			
	> Request for an answer type	Interrogative			
	> Use of gradation	Gradation used			
	> Balance of the request	Unbalanced			
	> Presence of encouragement to answer	No particular encouragement present	t		

ESS Round 4 / HS7 / testc7 / Political satisfaction: country's economy / United Kingdom / English

Step 4: By following these steps, you can generate a reliable quality prediction for the survey item, gaining valuable insights into its measurement properties.

Conclusion

The Survey Quality Predictor (SQP) is a powerful tool for improving survey research by enabling a systematic evaluation of question design. It can predict the reliability, validity, and quality of survey items based on their formal and linguistic characteristics. This capability is invaluable for researchers looking to enhance their measurement instruments. By linking these characteristics to measurement quality, SQP provides actionable insights for creating better questionnaires, supports quality control in survey research, and helps reduce potential biases when measuring subjective concepts such as attitudes and opinions.

Despite its advantages, SQP has some limitations. Its dependence on user coding can introduce subjectivity in interpreting the characteristics of survey questions. While the platform provides authorized codings for numerous survey questions and offers detailed guidance to minimize this risk, the quality of predictions can still vary depending on the user's expertise. Additionally, the coding process—which requires assessing up to 60 characteristics for each item—can be time-consuming for those unfamiliar with the tool. Furthermore, although SQP's database is extensive, it is not exhaustive, and the quality of predictions may not fully capture the nuances of all survey contexts or emerging methodologies.

Looking ahead, the potential for SQP's development is promising. Expanding the database to include more languages, regions, and survey design choices would enhance its applicability, particularly in underrepresented contexts. Incorporating large language models to automate parts of the coding process could alleviate the cognitive load on users and improve consistency. Additionally, integrating SQP with newer survey methodologies and adaptive designs, along with continuously updating its algorithm to reflect advancements in survey research, will help maintain its relevance in a rapidly evolving field. By addressing these challenges and embracing innovation, SQP can continue to be an indispensable tool for survey researchers around the world.

- Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48(2), 409–442. https://doi.org/10.1086/2688 40
- Felderer, B., Repke, L., Weber, W., Schweisthal, J., & Bothmann, L. (2024). Predicting the Validity and Reliability of Survey Questions. Open Science Framework. https://doi.org/10.3 1219/osf.io/hkngd
- GESIS Leibniz Institute for the Social Sciences. (2024). User Manual for SQP 3.0. GESIS Leibniz Institute for the Social Sciences. https://sqp.gesis.org
- Saris, W. E. (2001). SQP: Survey quality predictor.
- Saris, W. E., & Gallhofer, I. N. (2014). Design, Evaluation, and Analysis of Questionnaires for Survey Research (2nd Edition). Wiley.
- Saris, W. E., Oberski, D., & Weber, W. (2022). The Quality of Survey Questions for Continuous Latent Variables: Your Guide to the SQP Database & Predictions. Independently Published.
- Saris, W. E., & Revilla, M. (2016). Correction for Measurement Errors in Survey Research: Necessary and Possible. Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 127(3), 1005–1020. https://doi.org/10.1007/s11205-015-1002-x