ABSTRACT

High quality social-science surveys in general and surveys in travel behavior in particular are paramount for preparing decision-making processes, implementing integrative transport planning approaches and for transportation modelling. Quantitative and qualitative indicators are necessary to determine the accuracy of the survey results. Based on an interdisciplinary overview of existing quality assessment approaches, this paper develops a framework for systematically monitoring and assessing the quality of household travel surveys (HTS). 24 indicators are derived and classified along the three areas of Collection Quality, Content Quality, and Application Quality. For each indicator, the desired direction for improving quality and units for its measurement are given. No composite indicator is derived in order to keep the developed framework transparent, flexible and adaptable to each individual case study. The functionality of the developed framework is demonstrated with the help of two exemplary applications. The provided framework is applicable to traditional mixed-method HTS but also to innovative approaches such as smartphone-based tracking of travel behavior and also to non-transport social science surveys.

1. INTRODUCTION AND METHODOLOGICAL APPROACH

1.1 Background and Objectives
Household travel surveys (HTS) have a long tradition and are regularly conducted in many countries at least on national level but often also on regional and municipal levels. Surprisingly, there is neither a commonly accepted definition of “survey quality”, nor exists a framework for measuring this quality so far (Stopher & Jones, 2008; Lyberg, 2012).

HTS are a special case of surveys in the field of transportation with several specific characteristics: (1) Data are collected on different levels (households, persons, trips, and
sometimes stages); (2) major parts of the collected data refer to a predefined reporting day; (3) the survey content is complex (travel diaries) leading to a high response burden; and (4) underreporting is a major issue (particularly for trips). All these specific characteristics need to be considered when defining quality of HTS (Stopher & Jones, 2008; Armoogum, 2014).

This paper for the first time provides a framework of HTS quality. The first goal of the paper is to identify quality indicators that are able to cover all main quality aspects of HTS comprehensively. Therefore, this paper provides a systematic research of existing approaches, compares advantages and disadvantages of different quality concepts, and derives a set of quality indicators specifically for HTS. The demonstration of the practicability of the developed concept is a further goal of this contribution.

1.2 Methodology
Starting with a general definition of quality, requirements and expectations, different perspectives of looking at quality are presented. Afterwards, the current state of the art in terms of survey quality is compiled by carrying out a systematic desk research. On the one hand, general approaches to assess the quality of a survey such as the Psychometric Properties of Tests and Measures, based on the Classical Test Theory (CTT), or the Total Survey Error (TSE) concept are introduced briefly. On the other hand, quality criteria from documents (QCD) such as guidelines, recommendations, codes of ethics, and best-practice examples are analyzed. Typical examples of such documents are introduced and compared. In that context, material from Europe and the U.S. as well as documents both for surveys in general and for HTS are included.

Based on the results of the desk research in Section 2, a set of quality indicators for HTS is derived in Section 3. The indicators are grouped into three main quality areas. Section 4 gives a brief insight into two examples of the application of the indicators in order to evaluate the survey quality. A first assessment of strengths and weaknesses as well as possibilities and limitations of the application of the presented concept of quality is also included in this part of the paper. The paper ends with some brief conclusions in Section 5. Figure 1 summarizes the methodological approach of this contribution.
2. DEFINITION OF QUALITY

2.1 General Definition of Quality
An often cited, concise and vivid definition comes from Juran & Gryna (1970), who generally classify quality as "fitness for use". With this in mind, ISO 9000 standard defines quality as "the degree to which a set of inherent characteristics fulfils requirements". In this context, "inherent" is understood "as a permanent characteristic" (and thus measurable) and "requirements" as "requirement or expectation that is stated, generally implied or obligatory" (CEN, 2005). The requirements may be set by "different interested parties [and] for example, in a document" (CEN, 2005). According to ISO 9000, there is not an individual criterion (such as price or material quality) that is decisive for the quality of a product or service, but the fulfillment of the agreed requirements as a whole. To express to what extent the requirements are met, quality can be used "together with adjectives such as poor, good or excellent" (CEN, 2005). The approach – derived from ISO 9000 – corresponds to (1) a customer- or user-oriented quality understanding according to Garvin (1984). In addition, four further possible views are defined by Garvin (1984): (2) transcendent (quality as a subjective experience), (3) product- or service-oriented (quality as fulfillment of requirements for the product), (4) manufacturer or supplier-oriented (quality as fulfillment of production requirements), (5) value-oriented (quality as the optimum of costs and benefits).

2.2 Quality of Surveys
So far, neither a commonly accepted definition of “survey quality”, nor a framework for measuring this quality is available (Stopher & Jones, 2008; Lyberg, 2012). However,
various concepts exist in different disciplines that might contribute to the development of a quality concept for surveys in general and for HTS in particular. In what follows, three concepts are described that are of special relevance.

2.2.1 Classical Test Theory (CTT) – Psychometric Properties of Tests and Measures
According to the CTT, a measured score results from the addition of "true score" (error-free score) and an “error score” (1st axiom). Another four axioms describe the relationships between true score and error score as well as between the error scores among each other (Gulliksen, 1950; Lord & Novick, 1968). On this basis, the quality of a measurement or a measuring instrument can basically be determined by three key criteria (Wittenborn, 1972; Baumgartner & Jackson, 1995):

Full objectivity (sometimes also "neutrality") of a measuring instrument is given if two users of the instrument achieve the same measurement result, i.e. the measurement is independent of the respective person carrying out the test. Perfect reliability of a measuring instrument would be given if a score can be measured without the influence of an error score, thus repeated measurements (by the same person) with the same measuring instrument lead to the same results. Since this ideal case does not occur in practice, the degree of reliability describes the detection accuracy or trustworthiness of the measuring instrument. It does not matter if the measured score is the true value. Validity of a measuring instrument can be assumed if the instrument measures exactly what it should measure, i.e. the measured score meets the true score. Validity is thus the central and most important requirement for a measurement and indicates the extent to which the measuring instrument is suitable for the intended measurement purpose.

2.2.2 Concept of “Total Survey Error” (TSE)
Sample surveys are subject to a number of error types. The “total survey error” basically consists of the two fields of "sampling errors" and "nonsampling errors". (Weisberg, 2005; Groves & Lyberg, 2010).

Sampling errors result from the fact that not the complete population, but only a random sample is surveyed. The sampling error corresponds to the difference between the value of a target variable in the population and in the sample and thus quantifies the accuracy of the estimation. The size of the sampling error can usually be calculated or at least estimated.

The calculation or estimation of nonsampling errors is much more complex. Being systematic, they lead to a wrong estimation of the target variables (“bias”). Nonsampling errors are classified into five types:

- **Frame errors** (also coverage errors) occur when elements of the population have no chance to be sampled. They may also occur if actual sampled units do not belong to the population of interest.
Unit-nonresponse errors occur when sampled units do not participate in the survey.

Item-nonresponse errors result from skipped questions.

Measurement errors are caused by incorrect values that result from general inaccuracy or malfunction of measuring instruments (or survey methods) or rather conscious or unconscious mistaken statements of the respondents.

Processing errors (also postsurvey errors) can occur in the entire process of (usually electronic) data handling and analyzing.

Systematic frame and unit-nonresponse errors lead to selectivity. If certain groups of participants are over- or under-represented in a random sample, structures can be corrected by weighting if appropriate population information is available. However, if there are systematic losses within these groups (missing not at random – MNAR), individuals may show a significantly different behavior, and the use of parameter estimates (e.g. arithmetic mean, unit value) leads to biased values for the population.

Item-nonresponse, measurement, and processing errors might compensate each other according to the axioms of classical test theory if they occur unsystematically. If they occur systematically, biased results are the consequence.

The concept of consistently decomposing the total error into its components, analyzing it, minimizing it and thus reducing the total survey error, had its beginnings in the 1940s, starting from initial systematization of errors in sample surveys (Deming, 1944). The process of preparation, conduction, and post-processing of a survey is interpreted as a sequence of successive (and connected) decisions that are made in such a way that the resulting error remains as small as possible. In this context, the term "total survey design" was derived as a survey design that aims particularly at error control and error prevention.

2.2.3 Quality Criteria from Documents (QCD) – Guidelines, Recommendations, Codes of Ethics, and Best-Practice Examples

There are several national and international organizations, which published diverse documents that are all meant to ensure the quality of surveys.

Codes of ethics are usually issued by professional associations and contain general rules in the form of instructions for correct (in the sense of ethical) behavior. The “International Code of Social Research”, published jointly by the “International Chamber of Commerce” (ICC) and the “European Society for Opinion and Marketing Research” (ESOMAR), defines eight basic principles from which 14 articles are derived for market research (ICC/ESOMAR, 2008). Article 4 (“Transparency”) and Article 11 (“Publication of Results”) are of special importance for the definition and assessment of survey quality. Both can in principle be interpreted in such a way that the extensive documentation of all work steps and a far-reaching transparency towards the client and (when the results are published) the public are an important basis for "ethically correct" market and social
research. However, like the other articles in the code, the phrasing is terse and rather vague.

The “Code of Professional Ethics and Practices” is the equivalent code of the “American Association for Public Opinion Research” (AAPOR). The document has only seven pages and is divided into the main sections: (1) Principles of Professional Responsibility in Our Dealings with People (meaning both respondents and customers/clients), (2) Principles of Professional Practice in the Conduct of Our Work, and (3) Standards for Disclosure (AAPOR, 2015).

Both examples do not provide a definition of survey quality. However, they assume that the quality of a survey is ensured if the professionals responsible for carrying out the survey work in accordance with the provided ethical standards. The measurement or calculation of the quality of a survey is not possible based on these documents.

Various standards and guidelines aim in a similar direction. The “Quality Criteria for Survey Research” were published by the “German Research Foundation” (DFG) and deal with the whole workflow of preparation, implementation and processing of a survey (Kaase, 1999). The quality requirements for each work step are described by the presentation of "best practices" (desirable procedures from a purely methodological point of view) and "good practices" (acknowledged and acceptable procedures also with regard to costs and effort). In this way, a "range of quality" is defined. It enables the involvement of private-sector survey institutes, so that the formulated quality requirements can be presented as a broad consensus in science and practice.

Under the heading “Best Practices for Survey Research”, the AAPOR website lists twelve recommendations, improving to a "quality survey”. Each recommendation is described with a text that contains in some cases specific information (for example regarding the information to be disclosed), but in other cases it remains more general ("sending advance letters [...] and sending reminders or making follow-up calls"). Another requirement is to balance survey methodology and the related costs with the error effects ("Use designs that balance costs with errors"). In this sense, a survey is a “quality survey” if the cost-benefit ratio is convenient (AAPOR, 2018). With this in mind, for cost-intensive surveys, other quality requirements apply than for low-cost surveys.

A further approach is taken by the “European Statistics Code of Practice” of the “European Statistical System” (ESS, a partnership of Eurostat, the Statistical Office of the European Union, and the National Statistical Institutes of the EU member states), which formulates the quality principles for European statistics and provides them with indicators (ESS 2017). These are supplemented by the “Quality Assurance Framework”, also published by the ESS, which describes "activities, methods and tools" that should enable practice-oriented implementation of the principles (ESS 2015). The principles refer to the
“institutional environment” (e.g. “Commitment to Quality”), “Statistical processes” (“Sound methodology”, “Appropriate Statistical Procedures”, “Non-excessive Burden on Respondents”, and “Cost effectiveness”) also "Statistical output" (e.g. "Relevance", "Accuracy and Reliability", “Timeliness and Punctuality", "Coherence and Comparability", and "Accessibility and Clarity"). Therefore, the understanding of quality is relatively broad and, in addition to content and organizational aspects, also takes into account the needs of respondents, data users and taxpayers.

The transition from internal (that apply for one sector or country) to interdisciplinary accepted quality standards is achieved by the implementation of the ISO 20252 standard “Market, opinion and social research – Vocabulary and service requirements”, which allows the certification of appropriate services (ISO, 2012). In addition to the explanation and definition of a total of 69 terms, requirements for quality management, project planning, and survey methodology (timetable, sample, questionnaire, etc.), data collection, data management, and data processing as well as reporting are defined. Quality is not defined by content requirements, but by rules for organization and implementation. In contrast to the sector-specific standards and guidelines that rely on the ethical and scientific conducting of the researchers and organizations involved, compliance of the ISO 20252 standard should be ensured through documentation and control. Even if the relevant requirements are met, this does not necessarily lead to high quality. Rather, improving internal processes, creating consistent standards towards customers, and a responsible presence in the public should be perceived as "high quality" for the perception of certified companies and the entire market research industry (Köster, 2007).

All the above described options for assessing different quality aspects of surveys in general apply also to HTS. In addition to these general options, recommendations specifically for HTS exist:

In Germany, the “Recommendations for traffic surveys” (EVE), published by the “Road and Transportation Research Association” (FGSV) "are intended as an aid [...] to ensure a sufficient standard of quality. It contains advice and concrete measures to meet the respective minimum standards” (FGSV, 2012). In the document, however, quality is treated more in the sense of "data quality". The recommendations are based on the assumption that survey quality is assured once the formulated requirements are met.

In Austria, the “Handbook for Mobility Surveys” (KOMOD), published by the Austrian “Federal Ministry for Transport, Innovation and Technology”, defines HTS quality and standards, which should form the basis of the call for tenders of conducting a HTS (BMVIT, 2011). The chapter on quality assurance lists concrete steps as the basis for achieving high quality surveys such as the implementation and documentation of pre-tests and plausibility checks, effective control of the survey institute, the preparation of interim and final reports on quality assurance. These aspects of quality assurance are repeatedly
addressed elsewhere in the document and go beyond understanding quality as mere "data quality".

Finally, the report "Survey Harmonisation with New Technologies Improvement (SHANTI)", which was compiled as part of a cooperative project funded by the European Union, provides an overview of the HTS landscape in Europe. The document contains recommendations on how to conduct HTS in order to obtain comparable mobility indicators across Europe (Armoogum, 2014).

2.2.4 Synthesis

The transferability of the CTT-based psychometric properties of tests and measures (objectivity, reliability, and validity) to social science surveys is discussed controversially. The criticism is differentiated into aspects, which concern (1) the application of the properties in the practice, (2) fundamental doubts against the consideration of interviews as "test" or "measurement", and (3) the general rejection of the CTT, whereby the existence or recognizability of a true score is fundamentally questioned (Hambleton, 1991; Weichbold, 2009). Even if transferability is assumed, the psychometric properties can only be operationalized to a very limited extent for quality assessment in survey research. Nevertheless, of course, the goal is to carry out objective, reliable and valid HTS, even if the “ground truth” is not known.

Looking at the various documents from survey research and practice on the one hand as well as the types of errors on the other, a number of links among these sources become apparent. Thus, some of the requirements and recommendations from QCD are directed to methods and principles which are intended to minimize the known error influences in accordance with TSE. It is therefore obvious to derive a comprehensive set of indicators from the combination of QCD and TSE, which reflects both the practical relevance of the QCD and the theoretical foundation of the TSE and at the same time covers all areas of quality.

3. DERIVATION OF A QUALITY INDICATOR SET (QIS) FOR HOUSEHOLD TRAVEL SURVEYS

In line with the approach of combining QCD and TSE as well as the objective of comprehensively covering the field of survey quality comprehensively, three main quality areas were identified by the consolidation of the various evaluated sources:

- The extent to which it is possible to avoid frame errors and unit-nonresponse errors, to estimate possible consequences and to minimize them is the subject of the first area. In the following, this is referred to as **Collection Quality**.
- All efforts to gather, process, and analyze the data without errors, comparable, with the highest possible accuracy, and to estimate possible content errors (item-nonresponse, measurement, and processing errors) are summarized in the second area. This is called **Content Quality**.
• The third area, **Application Quality**, includes all methodological-organizational aspects of a survey and thus focuses particularly on the survey implementation.

Based on the analyses of literature in Section 2, the three quality areas are qualified with specific indicators in the next step. The indicators are intended to facilitate the assessment and evaluation of survey quality as well as the comparison of different HTS survey concepts in terms of quality.

The following tables contain the selected indicators first for the Collection Quality (Table 1), second for Content Quality (Table 2), and third referring to the Application Quality (Table 3). For each indicator, the strategic direction is indicated. For example, in terms of quality, a "high" Response Rate, a "low" Proxy Rate and, if possible, "none" of Selectivity improve survey quality.

The first indicator in the area of Collection Quality is **Coverage**. A high Coverage means a low level of over- or under-coverage of the target population in the sample. The **Response Rate** (share of completed households out of the gross sample) and the **Contact Rate** (share of contacted households out of the gross sample) should be kept as high as possible. In contrast, the **Break-Off Rate**, the share of non-completed households that started to reply but did not complete the survey, should be as low as possible. **Selectivity**, meaning structural differences between responding and non-responding households, should be absent and could be measured by the weighted deviation of two distributions. The same direction is intended with **Item Nonresponse** (missing answers or data in completed households).

<table>
<thead>
<tr>
<th>Indicators of Collection Quality</th>
<th>Direction</th>
<th>Possible Measurement Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Coverage</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>1.2 Response Rate (Unit Nonresponse)</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>1.3 Contact Rate</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>1.4 Break-Off Rate</td>
<td>Low</td>
<td>[%]</td>
</tr>
<tr>
<td>1.5 Selectivity</td>
<td>None</td>
<td>[Average Deviation]</td>
</tr>
<tr>
<td>1.6 Item Nonresponse</td>
<td>None</td>
<td>[%]</td>
</tr>
<tr>
<td>1.7 Response Time</td>
<td>Low</td>
<td>[Days]</td>
</tr>
<tr>
<td>1.8 Proxy Rate</td>
<td>Low</td>
<td>[%]</td>
</tr>
<tr>
<td>1.9 Nonresponse Bias</td>
<td>None</td>
<td>(Corresponding Item Unit)</td>
</tr>
</tbody>
</table>

*Table 1 – Indicators for Assessment of HTS Collection Quality*
The duration between mailing of announcement letter (or a questionnaire) and the completion of the household is defined as **Response Time** and should be short. Nevertheless, proxy-interviews are quite common in HTS, the **Proxy Rate** (the share of proxy-reported information) should be kept small. Systematic behavioral differences between responding and non-responding households, the so-called Nonresponse Bias are an important and risky topic and should be minimized.

The first indicator in the second area, Content Quality, is **Statistical Accuracy**. It stands for the degree of conformity between the estimate and the true parameter value in the population. The Accuracy should be high. Another important indicator is the **Error Rate of Raw Data**. It is defined as the share of error-suspicious households during raw data checking and is intended to be low. A high **Success Rate of Geocoding** means a high share of successfully geocoded destination points. The **Accuracy of Estimated Quantitative Values by Participants** should kept as high as possible, meaning that there is only small over- or under-estimation of metric variables (e.g. distances or durations). A high share of households that avail support options (e.g. telephone hotlines, web chats or FAQ websites) might serve the quality of the survey. The indicator is named **Availing of Support**. Biases caused by measurement errors resulting from the selected survey method (or methods), called **Method Effects**, are a sensitive topic. A typical HTS issue is the share of (intentionally or non-intentionally) **Non-Reported Trips**. Finally, in this quality area, Comparability has to be considered. In that context, it stands for the degree of suitability for comparing survey results with earlier or other surveys and should be ensured.

<table>
<thead>
<tr>
<th>Indicators of Content Quality</th>
<th>Direction</th>
<th>Possible Measurement Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Statistical Accuracy</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>2.2 Error Rate of Raw Data</td>
<td>Low</td>
<td>[%]</td>
</tr>
<tr>
<td>2.3 Success Rate of Geocoding</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>2.4 Accuracy of Estimated Quantitative Values by Participants</td>
<td>High</td>
<td>(Corresponding Item Unit)</td>
</tr>
<tr>
<td>2.5 Availing of Support</td>
<td>High</td>
<td>[%]</td>
</tr>
<tr>
<td>2.6 Method Effects</td>
<td>None</td>
<td>(Corresponding Item Unit)</td>
</tr>
<tr>
<td>2.7 Non-reported trips</td>
<td>Low</td>
<td>[Trips]</td>
</tr>
<tr>
<td>2.8 Comparability</td>
<td>High</td>
<td>(Qualitative)</td>
</tr>
</tbody>
</table>

Table 2 – Indicators for Assessment of HTS Content Quality
<table>
<thead>
<tr>
<th>Indicators of Application Quality</th>
<th>Direction</th>
<th>Possible Measurement Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Contact Frequency</td>
<td>High</td>
<td>[Number]</td>
</tr>
<tr>
<td>3.2 Interview Duration</td>
<td>Low</td>
<td>[Minutes]</td>
</tr>
<tr>
<td>3.3 Cost-Effectiveness</td>
<td>High</td>
<td>[Ratio]</td>
</tr>
<tr>
<td>3.4 Documentation and Transparency</td>
<td>High</td>
<td>(Qualitative)</td>
</tr>
<tr>
<td>3.5 Accessibility of Micro Data</td>
<td>High</td>
<td>(Qualitative)</td>
</tr>
<tr>
<td>3.6 Completeness</td>
<td>High</td>
<td>(Qualitative)</td>
</tr>
<tr>
<td>3.7 Actuality</td>
<td>High</td>
<td>[Years]</td>
</tr>
</tbody>
</table>

Table 3 – Indicators for Assessment of HTS Application Quality

The area of Application Quality contains the indicator **Contact Frequency**. It should be understood as the number of postal letters sent to respondents, telephone calls, etc. It should be high as long as it has a positive effect on other indicators (e.g., the Response or Contact Rate). The **Interview Duration**, the time needed for answering a suitable telephone interview or filling out an online questionnaire should be short. **Cost Effectiveness** is another important indicator. It stands for the benefit-cost ratio of the survey and should be high.

The indicators **Documentation and Transparency** (the degree of accessibility of methodological information and metadata) as well as **Accessibility of Micro Data** (the degree of accessibility of micro data for secondary analyzes) are important especially for scientific users. The degree of representation of the reality’s complexity, denoted as **Completeness**, and the degree of representation of the reality’s timeliness, denoted as **Actuality**, both should strive upward.

The compilation of the indicators cannot be exhaustive. From the authors' point of view, however, all areas that are essential for the assessment of survey quality are covered by indicators. A prioritization of the indicators is not intended. From the authors’ viewpoint, this would hardly be possible, since such a prioritization would depend very much on the specific framework conditions of each individual survey. To account for the indicators and/or to aggregate to a composite indicator, all effects would have to be monetized. This is theoretically possible, is attempted in the mentioned total-survey-error concept, but practically not feasible – and not necessary – for the presented indicator set.

In survey practice, it will rarely be possible – and necessary – to apply all the above listed indicators at the same time. Data is often not available for all indicators. Specific indicators can be excluded if no differences are expected e.g. because of similar framework conditions.
Multicollinearity, meaning substantial correlations between two or more indicators, is not expected as the indicators cover very different fields. However, certain interactions cannot be ruled out or are even obvious (e.g. Response Rate vs. Proxy Rate, or vs. Cost Effectiveness). The authors consider that as unproblematic, since the indicators are not offset against each other or directly used in statistical modelling.

4. EXAMPLES OF APPLICATION
The set of indicators, divided into three quality areas, has been applied in two cases so far. Hubrich (2017) published about causes of errors and the willingness to participate in HTS. Final outcome of the research are recommendations for quality improvements of HTS from a methodological point of view. The recommendations are based on the evaluation of several options for action. The evaluation focused on specific analysis groups of potential respondents. Altogether, 14 of the 24 indicators of the above presented QIS were applied to access the quality of a large series of HTS at municipal level in Germany. Out of the area of Collection Quality, the indicators Response Rate, Break-Off Rate, Selectivity, Item Nonresponse, and Response Time could be used. The indicators Statistical Accuracy, Error Rate of Raw Data, Success Rate of Geocoding, Accuracy of Estimated Quantitative Values, Availing of Support, and Method Effects were analyzed to provide assessments of Content Quality. Finally, Contact Frequency, Interview Duration, and Cost Effectiveness were the indicators out of the area of Application Quality that could be included. In conclusion, recommendations for future HTS were established; these were divided into a cross-group section as well as a section for specific groups of potential respondents.

Hubrich & Wittwer (2014) and Hubrich et al. (2017) compared the traditional HTS approach where trip data is collected for all individuals in each household with an ITS (individual travel survey) where information of daily trips are collected for selected individuals only (minimum one). In the first step, advantages of an ITS in terms of Statistical Accuracy were highlighted (Hubrich & Wittwer, 2014). In the second step, on the basis of two surveys (HTS 2013 and ITS 2016 as a follow-up survey), the comparative evaluation focused on the indicators Response Rate and Selectivity (Collection Quality), Error Rate of Raw Data (Content Quality), Contact Frequency and Interview Duration, (Application Quality). Finally, in the third step, a scenario technique was used to estimate the implications for Cost-Effectiveness, taking into account the indicators examined in the first two steps.

5. DISCUSSION AND CONCLUSION
This contribution provides an interdisciplinary overview of the state of art in quality assessments of surveys in general and of surveys of travel behavior in particular. Based on this, a concept of quality indicators is developed and structured into a Quality Indicator Set (QIS). The derived indicators are divided into the three areas (1) Collection Quality, (2) Content Quality, and (2) Application Quality. For each indicator, the desired quality (target attainment/goal achievement) and possible measurement units are provided.
The fact that prioritization and mathematical compensation of the indicators is not provided, can certainly be discussed. At this point, however, the simple applicability of the indicators should be in the foreground. The consideration that some of the selected indicators can be eliminated for certain survey circumstances should also be understood in this sense. Naturally, depending on the specific application, it must be ensured that no essential indicators are neglected or that proper consideration is given to the most important indicators.

As is illustrated by the examples briefly presented in Section 4, from the authors' point of view, all indicators are very well suited to handle different evaluation tasks with regard to the suitability and quality of HTS. Also a fundamental transferability to other types of surveys and for different topics is seen.

Overall, the developed framework for measuring HTS quality provides for the first time a systematic approach for evaluating and comparing HTS quality. It can be the basis for setting up standards for tendering and monitoring field work procedures, as well as for disclosing, processing and analyzing data. It is applicable to traditional mixed-method HTS but also to innovative approaches such as smartphone-based tracking of travel behavior. The developed framework can be the basis for setting up quality management procedures that ensure reliable, valid and thus also comparable HTS data as a core input for transport modelling, planning and policy making.

REFERENCES


