



Student Paper Series

Framing Service Quality

By Florian Birk (MPP 2013), Amani Joas (MPP 2013),
David Rinnert (MPP 2014) & John Schilling (MPP 2013).

Framing Service Quality

A Comprehensive Indicator for Measuring Service Quality in Public Administration: Case Study of Bürgerämter in Berlin

By Florian Birk (MPP 2013), Amani Joas (MPP 2013), David Rinnert (MPP 2014) & John Schilling (MPP 2013).

Conveners: Prof. Dr. Kai Wegrich & Jan Tiessen (PROGNOS)

FOREWORD

‘Quality of government’ is an elusive concept. We all seem to know what it means, but despite a plethora of studies by high-powered think tanks and international organisations and the emergence of a range of indicators, a continuous confusion about core questions remains. Critical issues include the inclusion and exclusion of variables, the design, weighting and aggregation of indicators and the validity of the selected indicators in general. While the fascination with governance indicators lies partly in the aggregation of a range of different indicators into a ranking of ‘good governance’ or ‘sustainable governance’, this over-ambition is the root of many of the flaws of these indicator studies; another one is the ambition to compare a range of countries, if not all of them.

The following paper was developed in the context of the project course ‘Mapping the State’, offered as part of Hertie School’s Master of Public Policy programme. In this course, students were tasked with critically reviewing existing indicator studies and developing their own indicator design on selected aspects of quality of government (note: government, not governance!). Our practice partner in this course, Jan Tiessen from PROGNOS, had a keen interest in our students’ work. The long-term ambition is to advance some of the ideas to inform visual maps that illustrate the regional distribution of quality of government and public service. Small project groups focussed on selected aspects of quality of government, i.e. accountability, productivity, bureaucracy and service quality. All four student projects wrestled with the issue of how to boil-down general concepts of ‘quality of government’ towards something measurable, while still contributing to the wider concerns relating to quality of government. All four groups presented innovative ideas to measure a number of aspects associated with quality of government: The productivity group measured the productivity of garbage collection services in Berlin and compared this with other German cities. The accountability group developed and tested a framework for measuring the transparency and accessibility of information provided by US public school districts (transparency & accessibility). The bureaucracy group measured the direct costs and red-tape involved in an information request processed by finance ministries in four countries. And, finally, the service quality group developed and tested an indicator design for measuring the quality of services of citizen-focussed services in Berlin (so-called “*Bürgerämter*”, citizen offices).

In order to address the thorny issue of measuring the quality of services to the citizens, the authors of the paper at hand selected the Berlin *Bürgerämter* as their empirical site. They engaged with the problematic of ‘subjective’ (perception-based) versus ‘objective’ measures of service quality. They did so by developing a framework to measure ‘soft’ and ‘hard’ dimensions of the concept. In pre-testing their questionnaire design intended to capture the ‘soft’ dimensions of service quality, the group visited four Berlin *Bürgerämter* to conduct exit-poll type of interviews with users of the services. The contact with the real-world provided a sobering lesson concerning what is doable and what is not doable in such a survey. The ‘hard’ measures – ranging from opening hours to waiting times and proximity to public transport – were used to compare and rank Berlin 40 *Bürgerämter* (with striking differences across them).

The results of this study are made publicly available in this working paper. The reader might wonder why such a study has – according to the best of our knowledge – not been done before. This study is not only a model work of Hertie Students and the school more widely – rigorous methodological work that is of both academic and practice-related relevance. It deserves a broad readership. I can only congratulate the students on their achievement and wish that the publications of the paper sparks the interest of some institutions to get back to our students, because more research in this field is badly needed!

Kai Wegrich

Executive Summary

Trying to get a customer at a Berlin *Bürgeramt* ('citizen center') to partake in our **survey concerning the quality of administration services**, he stopped for a moment and replied: "Service quality? Are you kidding me? I have been waiting here for 5 hours and these people are racist. Don't ask me about service quality!"

This frustrated customer of a government service provides surprising insight into the question of what people expect from a good service: **They want to get what they came for. They want to get it quickly; and they want to be taken seriously while being treated with esteem and courtesy.**

Accordingly, we define service quality as a pure outcome/output concept, which is only concerned with the actual quality of a service, while it is not concerned with the inputs or processes by which it came about. **Service quality in customer-based government services matters greatly to the perception of the quality of government**, as it is here where citizens get the chance to directly experience government effectiveness and judge the way the administration is using its resources. Rendering high service quality, while being an end in itself, further has instrumental value as it allows citizens to develop trust in its government. This characteristic should not be underestimated.

In this paper **we built a framework, which allows for a valid and reliable assessment of the service quality of customer-focused government services.** While defining service quality is a difficult task, operationalizing and measuring such an elusive concept can be done optimally, however, never perfectly due to a number of problems. Given the 'soft' nature of service quality, most prior research has focused on survey-based methods, which while being essential have serious flaws owed to this method's inherent lack of reliability. Following suggestions by prior researchers in the field, **we have developed a unique assessment framework, which combines a lean survey based on the state-of-the-art SERVQUAL questionnaire with a set of 'hard' variables adding the reliability, which previous approaches lack.**

We have constructed an indicator framework, which rests on two columns, namely a 'soft' survey based indicator and a 'hard' indicator based on objective variables (**Part 2**). The indicator was initially developed, tested and re-adjusted for measuring service quality in Berlin's 40 *Bürgerämter* (**Part 4**). However, it can be tweaked in such way that it becomes applicable to most customer-based government services. The 'soft' part of this indicator was developed on the lines of the five SERVQUAL dimensions that help to score services: tangibles, reliability, responsiveness, assurance and empathy (detailed definitions below). While the original SERVQUAL questionnaire with its 44 questions has the advantage of drawing an exhaustive picture of a customer's appreciation of a service, it has the inherent problem of impracticability: getting a large amount of government service customers to fill-out such a long questionnaire is extremely difficult and prohibitively costly, which is why **we slimmed the survey to a mere seven-question survey.** This allows for a comprehensive coverage of all important dimensions, while having the advantage of offering a feasible research basis for those trying to measure service quality.

On the 'hard' dimension, we have picked a set of eleven variables, which unequivocally add/subtract from service quality, while having the advantage of being measured in a straightforward manner. **While the operationalization and weighting of these variables is open to academic and professional debate, we are confident to have followed a transparent and clearly comprehensible process (Part 3).** We acknowledge, however, that these variables will have to be adjusted when other services are being measured.

All in all, the Service Quality Indicator for Customer-Based Public Services (SQI) should be regarded as a starting point for a comprehensive and easily constructible framework to assess service quality in the good governance arena. We hope that this indicator framework will be perfected to the point that it can contribute to a better overall measurement and therefore improvement of government services.

Service Quality Indicator (SQI) - First Results

The SQI consists of two measures (hard and soft), which are both normalized onto a scale of approx. -3 to +3. In our pretest for the Berlin Bürgerämter (BA), BA "A" received a top score of 2.01 (anonymized, real data) while BA "B" received the lowest score of -1.3. The winner BA "A" had the longest and most convenient opening hours and by far the shortest waiting time of 8 minutes on average.

Our survey-based soft indicator scores seem to correlate with the hard indicator scores given that BA "C" received a hard|soft score of (0.8|0.2); BA "D" scored (-0.4|-0.3) and BA "E" scored (0.5|0.1). We closely examined 40 BAs on their hard measure scores and took the time to conduct surveys in three BAs, which helped us to get a clear impression of what a high quality service environment looks like. After having conducted the pretests, we are confident that the final SQI framework depicts service quality as well as it is possible and is therefore a valuable evaluative tool.



Table of Contents

Executive Summary.....	1
Table of Contents.....	3
List of Figures and Tables.....	4
1. Introduction.....	5
2. Service Quality Indicator Framework.....	6
2.1 The ‘Soft’ dimension: Towards a questionnaire based on SERVQUAL.....	6
2.2 The ‘Hard’ dimension of service quality: Operationalization of variables.....	8
3. Towards a comprehensive method for a <i>composite</i> service quality indicator:	
Methodology.....	10
3.1 ‘Best Practice’ – UCM, Factor Analysis or weighted aggregation?.....	11
3.2 The SQI data composition method – strengths and weaknesses.....	12
4. Pretest results and data analysis: Lessons learnt.....	16
4.1 Survey pretest results.....	16
4.2 First results: ‘Hard’ indicator.....	22
4.3 Mapping the final indicator scores: possibilities.....	24
5. Conclusions and outlook: Opportunities for expansion of the framework.....	25
5.1 E-governance as a crucial approach.....	25
5.2 Conclusion.....	26
Bibliographical references.....	28
Annex.....	30

List of Figures and Tables

- Figure 1 – Composition & Methodology of the SQI – ‘Soft’ and ‘Hard’ Dimension
- Figure 2 – SQI ‘soft’ dimension pretest - overview
- Figure 3 – Bürgeramt ‘Hard’ Indicator Score – Overall Overview
- Figure 4 – Mapping the SQI dimensions – possibility
- Figure 6 – First Draft of the Questionnaire
- Figure 7 – Second Draft of the Questionnaire
- Figure 8 – Final Draft of the Questionnaire
- Figure 9 – Lessons - assessment criteria for ‘hard’ service quality measures
- Figure 10 – Service Quality Indicator – Theoretical Base for construction
- Figure 11 – Operationalization of ‘Hard’ Indicator Framework
- Figure 12 – Methodology “Test”: Our Aggregation framework compared with a Factor Analysis aggregation Framework
- Figure 13 – Methodology "Test" - Difference between Original Indicator Score and Indicator Score with change in weights
- Figure 14 – Scree Plot with the Soft Service Quality Indicator score from our pretest and the age of the interviewees
- Figure 15 – Bar plot displaying education in years and service quality score
- Figure 16 – ‘Soft’ Service Quality Indicator score and waiting time as stated by the interviewees
- Figure 17 - *Bürgeramt* ‘Hard’ Indicator Score – District Overview
- Figure 18 – Dataset for the 'Hard' Service Quality Indicator
- Figure 19 – Specific hard and soft measure scores from our pretests
- Figure 20 – Possibility of Mapping Indicator Scores

1. Introduction

Why should one focus on the quality of services in public administration when measuring good governance more generally? Acknowledging the broad and elusive character of the ‘good governance’ concept, there are many dimensions one could think of when taking stock of governmental performance. While concepts such as impartiality, transparency or productivity play a key role in the assessment of public administration, the dimension of service quality has not been a focus of good governance indicators. As a consequence, the few existing indicators trying to assess the service quality of public administration lack both a comprehensive theoretical underpinning and a sufficient methodological toolkit. Most of the existing indicators aim exclusively at measuring citizens’ perceptions towards the quality of public services. Since the quality of public services is difficult to standardize and therefore to a certain extent based on subjective perceptions (and expectations), such a ‘soft’ indicator is crucial when measuring service quality. However, an exclusively perception-based approach is vulnerable to criticism due to the method’s inherent lack of reliability. Asking customers about their perceptions simply produces very different results depending on the specific context.

Thus, both the theoretical and methodological shortcomings and the fact that service quality has been a widely neglected issue make the assessment of service quality in public administration a highly interesting and valuable task when mapping good governance regimes. This paper not only presents a newly developed and comprehensive indicator framework but also extensively tests this approach providing a valid, reliable, and easily applicable framework assessing the quality of public administration. We complemented the ‘soft’ dimension of service quality with objective (‘hard’) standards to balance the shortcomings of the first and put both dimensions to a test. The data gathered in pretests were then used to optimize the initially developed model.

In order to create a comprehensive and widely applicable model, our research focused on citizen centers (*Bürgerämter*) in Berlin. *Bürgerämter* fulfill a wide range of tasks, such as resident registrations, i.e. every citizen has to make use of their services once in a while. This allowed us to gather perceptions from a broad spectrum of people and compare them to the more ‘objective’ performance measures of these entities.

Accordingly the structure of the paper will be as follows: First, we will discuss both indicators used for the ‘soft’ and the ‘hard’ dimension. Regarding the former we develop a questionnaire based on the widely used SERVQUAL model in order to enable surveying the perceptions of the individual citizens. The latter will develop and discuss ‘hard’ performance

measures relating to service quality in the *Bürgerämter*. Chapter 3 then discusses how to approach both indicators methodologically and combine them in a reasonable way (in our new *Service Quality Indicator – SQI*). Chapter 4 tests the developed framework and looks at the lessons learnt. Finally, in the conclusion we will shed light on the opportunities and limits of the optimized framework.

2. Service Quality Indicator (SQI) Framework

2.1 The ‘Soft’ Dimension: Towards a Questionnaire based on SERVQUAL

Initially, we developed a first draft questionnaire based on the widely used SERVQUAL model (Parasuraman et al. 1985, 1988, 1991). This model identifies five key dimensions of service quality (Parasuraman et al 1988: 23), namely:

- **Tangibles:** Physical appearance, equipment, and appearance of personnel.
- **Reliability:** Ability to perform the promised services dependably and accurately.
- **Responsiveness:** Willingness to help customers and provide prompt service.
- **Assurance:** Knowledge and courtesy of employees and their ability to inspire trust and confidence.
- **Empathy:** Caring, individualized attention that the firm provides for its customers.

In the first part, employing these five quality dimensions, the SERVQUAL model uses two surveys containing 44 questions to assess the customers' expectations and perceptions regarding service quality (22 statements on expectations, 22 statements on perceptions). The existing gaps in quality of service are measured by a gap score (Gap score = (P)erception statements scores – (E)xpectation statements scores). In case that the gap score is positive, expectations are being exceeded (and vice versa) (Parasuraman et al. 1988, 1991; Ramseook-Munhurrun et al. 2010: 39).

In the second part of the surveys customers are being asked to “allocate weights to each of the five dimensions of service quality which reflect their relative importance from the customer’s perspective. Aggregating these weights in line with the size of the gaps identified in the other sections of the questionnaire allows for an assessment of the “focus” of the organization” (Donnelly and Wisniewski 1995: 18; Parasuraman et al. 1991).

However, asking customers of *Bürgerämter* 44 long questions would not only be a time-consuming task, but would also result in comparatively high cost of conducting representative customer surveys. In addition and given our practical experience, it would be very difficult to persuade the interviewees to actually participate in such a survey. Due to these limitations, there have been no broad SERVQUAL surveys beyond the boundaries of

one city or region until today. While the case study of service quality in one city or administrative district might be helpful for local authorities, it does not provide representative data for a countrywide indicator allowing for comparisons. Hence, in order to develop such an indicator, the SERVQUAL questionnaires needed to be shortened and simplified.

Our first draft questionnaire (see *Annex*, figure 6) was based on the following considerations:

First of all, we agree with Cronin and Taylor's (1991, 1994) assessment that it is questionable whether a gap score measured through SERVQUAL surveys can be regarded as a valid proxy for service quality because “the SERVQUAL scale appears at best as an operationalization of only one of the many forms of expectancy-disconfirmation” (Cronin and Taylor 1994: 127). In addition, the proxy “expectation” is too ambiguous and one cannot control for interfering variables in the measurement of expectations with a gap score. Furthermore and particularly in regard to the conducting of surveys, a single questionnaire which only focuses on customers' perceptions seemed to be more feasible in its application than asking customers also about their expectations beforehand. Hence, customers' perceptions remain the best, i.e. the only measurable reflection of service quality whereas customers' expectations should not be part of our conceptual model.

Initially, we reduced the 44 questions to 12 questions, 2-3 questions for each of the five dimensions mentioned above (see *Annex*, figure 6, parts 1-5 of the questionnaire). In order to do so, we translated some of the questions provided by the original questionnaire by Parasuraman et al. (1991: 448-449) into German and left out others if their content differed from the services actually provided by a *Bürgeramt*. For instance, we deleted the question “you feel safe in your transactions with XYZ” on the assurance dimension. In contrast, questions like “XYZ has modern-looking equipment” and “XYZ's physical facilities are visually appealing” were merged into one (see *Annex*, figure 6, question 1). Although the “deletion of items” might “affect the integrity of the [original SERVQUAL] scale” and might “cast doubt on whether the reduced scale fully captures service quality” (Parasuraman et al. 1991: 445), we think the remaining questions constitute a necessary and meaningful compromise in order to have an applicable survey.

Whereas the point-allocation questions remained basically the same as in Parasuraman et al.'s questionnaire (1991: 449), we added a part asking the customers for their comments (see *Annex*, figure 6, parts 7 & 8).

The last part of our first draft questionnaire dealt with standardized personal questions (see *Annex*, figure 6, part 8), i.e. age, gender, education and employment. In addition, we

included two further questions: on the one hand, we asked for the customer's birth place because we assumed that some people from different parts of Germany might have different expectations in regard to service quality. The same consideration played a role in asking Berlin born citizens about the district in which they were born.

On the other hand we asked the customers about how long they had to wait for and if they had made an appointment beforehand (where applicable, we asked for how long they had to wait for their appointment). These questions taken together allow for a comprehensive measurement of a customer's perception of service quality, while the 'hard' measures are added in order to push our indicator beyond a purely perception-based approach.

2.2 The 'Hard' Dimension of Service Quality: Operationalization of Variables

A good indicator for service quality in government institutions needs to combine the 'soft', survey based variables with more objective or 'hard' measures. The reason for this is that such variables will provide the indicator with the reliability, which a purely survey based method would lack.

(a) The underlying variables for a reliable and valid indicator

Our indicator combines a total of 11 'hard' variables, which were chosen according to our previously explained criteria of feasibility, validity, reliability, acceptability and costs (the rating process of acceptable variables is found in Annex, figure 9). By far the biggest constraint comes with the feasibility dimension. Given that hard data that are meaningful to assess service quality (e.g. average waiting time) are not regularly published, we had to rely on previous research on the quality of *Bürgerämter* by the University of Speyer, which issued detailed reports for 2006-2008 (Klages 2006; Masser 2008). Additionally, we used information that is available online (berlin.de; bvg.de) to create meaningful variables of our own (e.g. average distance to public transport and opening hours). In the following we will present some of the variables used, while discussing the further statistical aggregation methods in chapter 3. A detailed list of all variables and their operationalization is found in Annex, figure 11).

The 11 hard variables include (1) opening hours; (2) Saturday opening hours; (3) disabled parking; (4) access via public transport; (5) distance to public transport; (6) picture machine; (7) feedback channel; (8) average waiting time; (9) ratio of minimally waiting

customers; (10) average handling time (not included in all indicator aggregations¹); and (11) average time of service completion. In practice, this approach combines continuous ordinal variables, such as the core variable of average waiting time measured in minutes, with categorical Dummy variables, such as Disabled Parking (0 for no availability, 1 for availability). Variable 2 (Saturday Opening Hours) stands out as the open hours on Saturdays per month are counted while one extra point is added if services without prior appointments are also possible. Variables 1-7 are obtained from online sources (mostly the websites of the *Bürgerämter*), while variables 8-11 are based on the research reports by the University of Speyer (Klages 2006; Masser 2008). The idea behind this approach is to capture such factors, that unequivocally can be regarded as adding to or reducing service quality, while at the same time having the advantage of being measured objectively.

The strength of these variables lies with their reliability. For example if a given *Bürgeramt* were to add disabled parking, a photo machine and decrease the average waiting time, this would immediately show up in the indicator (for methodology of data composition, see chapter 3). The obvious problem, on the other hand lies (a) with the limits and timeliness of the available data and (b) with the questionable relative validity of some variables given the concept of service quality. For example, while it should be rather clear that a new photo machine may add quality to the service of obtaining a new ID, it is not at all clear how important this is relative to the waiting time or even the vast amount of services which do not actually require a picture.

The variables can furthermore be brought into relation with the five SERVQUAL dimensions namely, reliability, assurance, tangibles, empathy and responsiveness. According to our classification, all the hard variables fall within the reliability and responsiveness dimensions, as they are either concerned with the delivery of a dependable service or with the amount of effort a *Bürgeramt* has taken to make their service easily accessible (see *Annex*, figure 11 for details). For example, the variable on opening hours best fits into the reliability dimension, since a dependable and accurate service delivery includes accessibility of the service to its customers.

¹ The variable measures the handling time of one customer. As higher handling time could mean two things (either bad service because of unnecessary long duration or better service because of in depth answers to customers questions), this variable is not included in all final 'hard' indicator scores. It could, nevertheless be used at a later stage.

(b) Variables outside the scope of a service quality indicator

After just one request to the Senate of Berlin, we obtained a large set of input data (including around 50 different variables) concerning Berlin's *Bürgerämter*. Using these data would have allowed us to include input considerations, such as the amount of professional training taken by the average employee. However, we explicitly decided not to use such variables, as this would force us to hypothesize a positive causal relationship between a certain input (e.g. amount of professional training) and the outcome of service quality. Since we are building an output/outcome focused instrument, we decided against the inclusion of such measures (see *Annex*, figures 9 & 10). However, in a next step we believe that it would be intriguing to use our indicator as a dependent variable in a linear regression and see to what extent input changes, such as increased professional training, have an effect on our measure of service quality. Using analogous reasoning we rejected to calculate performance measures such as the amount of IDs issued per employee or to use further removed outcome measures such as trust in local government, since such measures are too far removed to be unequivocally linked to the concept of service quality. Again, such measures should be used in correlation analysis with the developed indicator; however this is beyond the scope of this paper.

Taking all this into account, we have developed a credible and valuable set of objective variables, which compose the hard dimension of this indicator framework. One needs to take into account the obvious shortcomings and limitations of this approach, however these, when viewed in detail, amount to imperfections rather than overwhelming obstacles. The next section describes how we tackled the methodological issues and how the indicator works in practice.

3. Towards a comprehensive method for the composite Service Quality Indicator (SQI):

Methodology

The service quality framework outlined in part 2 aims at the development of two indicator scores ('hard' and 'soft') per district/per *Bürgeramt* in the end. But how is this achievable given the broad array of underlying variables?

The main challenge of composing one or a few indicator variables is the significant simplification of complexity of the original dataset. There are many more general problems related to the quantification of qualitative data, such as causality questions and the choice of proxy variables to explain certain issues (see Schnell/Hill/Esser 2005: 231). While a broad discussion on these wide-ranging challenges of empirical research would go beyond the scope of this paper, the following chapter aims to shed light on the specific methodological

questions at stake when quantifying, normalizing and finally aggregating the SQI framework outlined above.

Putting aside the broader questions about data evaluation, quality of composite indicators is determined by the following criteria:

1. Information available for the variables used;
2. Choice of variables and measures for the indicator;
3. Scaling of variables & mathematical aggregation method;
4. Relative importance of each variable / weighting technique;

(based on Munda/Nardo 2003: 15)

While we dealt with issues 1 and 2 in previous part, the methods to tackle criteria 3 and 4 remain to be specified and justified to lay out a service quality framework that can be applied beyond the scope of *Bürgerämter* in one city (see figure 1, p. 13, for an overview).

3.1 ‘Best Practice’ – UCM, Factor Analysis or weighted Aggregation?

How should many variables be aggregated into one or a few indicators? Taking into account the methodology of existing indicators, one finds that there is no consensus on this question. This is no surprise given the vast amount of problems one faces when aggregating large datasets into a handful of final scores. Nevertheless, there are distinct features of the methods available that need to be assessed in order to determine the best approach for our purposes and justify it.

Kaufmann and Kraay have decided to base the World Bank’s *World Governance Indicators* (Kaufmann/Kraay/Mastruzzi 2010) on the so-called Unobserved Component Model (UCM). The UCM is an aggregation method that takes into account the imprecision of measurement and lack of data for certain sources by decomposing a response series into different components. In the end, the UCM makes the WGI an indicator that is “a weighted average of data emanating from each source, weights being larger for sources that provide a more informative signal of governance” (Den Boer 2011: 12). Proponents claim that thereby, the UCM reduces errors and enhances validity (Kaufmann/Kraay/Mastruzzi 2010). However, since it is unlikely that there will be significant differences in the amount of data and responses available from the sources we are analyzing (*Bürgerämter* or, at a later stage, other public administration offices), UCM is not likely to optimize aggregation for our indicator framework but would rather complicate matters because of its complex statistical foundation. Aside from UCM, factor analysis and principal component analysis (PCA) are statistical methods that are proposed by many as a best practice tool for creating aggregate indicators

(see Nardo et al. 2005: 21). Researchers and official authorities across the globe use factor analysis to assess administration (e.g. the Social and Cultural Planning Office of the Netherlands in a public sector performance study, see Social and Cultural Planning Office 2004). Factor analysis allows for reduction of correlated variables by identifying common ‘factors’ (or ‘components’ in the case of the relatively similar PCA method) based on shared variation (UN Department on Economic and Social Affairs 2007: 33). However, despite the many advantages of factor analysis, this method is not the best choice for our framework for two reasons. Firstly, factor analysis is usually somewhat sensitive to small-sample problems (a dataset with 100 *Bürgerämter* could still be considered as being rather small) and secondly, it often extracts more than one or two dimensions of a set of variables. As the service quality framework is intended to aggregate one indicator for all ‘hard’ measures and one for all ‘soft’ measures, a different methodological approach is needed.

Aggregation through summation or multiplication of standardized variables with the use of weights is a third method that is widely used for indicators despite being strongly criticized due to its limitations (Jollands/Lermit/Patterson 2003). While there is a long list of indices utilizing such ‘simpler’ forms of aggregation (e.g. Index Of Sustainable Economic Welfare, Ecological footprint, Human Development Index), this method is subject to ongoing discussion mainly because of its comparatively high sensitivity to individual influence (e.g. through the choice of the function and the weights). In general, one of the major limitations of such aggregate indices “is the manner in which the constituent variables to be included in the index are determined” (Jollands/Lermit/Patterson 2003: 6).

Nevertheless, we think that the third method suits our purposes best for the main reason that it allows us to develop two final scores that include as much information as possible in an easily comprehensible and meaningful way. As outlined by proponents of rather straightforward aggregation methods “a complex, information-rich world requires frameworks that organise data to reveal succinct views and interrelationships” (Heycox 1999: 191).

3.2 The SQI data composition method – strengths and weaknesses

(a) Missing data and imputation

First of all, our chosen method of data composition requires - as all other methods - an analysis of the question of missing data. This is especially relevant for our ‘hard’ measures dataset since in general, data “can be missing at random (MAR or CMAR) because of malfunctioning equipment, (...) lack of personnel, but there is no particular reason to consider

that the collected data are substantially different from the data that could not be collected. On the other hand, data are often missing in a *non-random* (NMAR) fashion” (Nardo et al. 2005: 35). While in our pretest, we face NMAR data (see part 4.1), it remains to be established in future studies making use of our framework whether data is MAR or NMAR. If missing data issues would arise, future research could decide to use case deletion or imputation methods to account for these problems (see OECD 2008: 24).

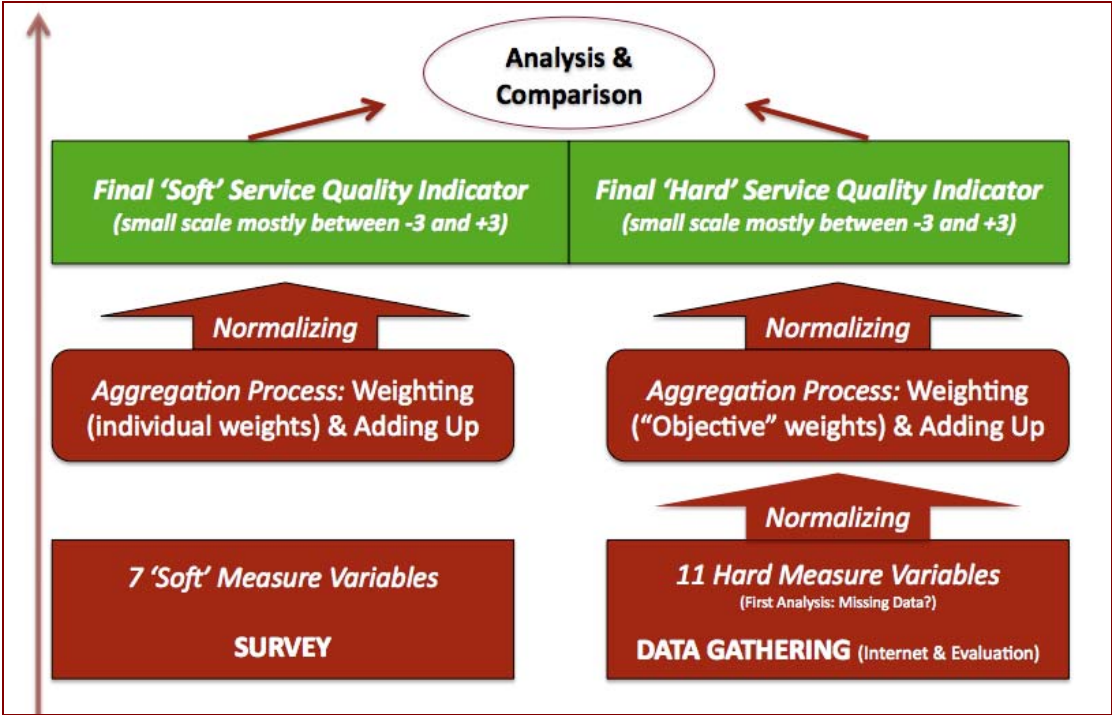


Figure 1 – Composition & Methodology of the SQI – ‘Soft’ and ‘Hard’ Dimension

(b) Normalization of variables

Before adding up the different ‘hard’ indicator variables and weighting them, normalization is needed because of the broad array of measurement units (e.g. opening hours and distance to bus/train station). While there are different methods of normalization, z-standardization is the most appropriate one for our indicator. To standardize a variable according to this method, the mean of a variable is subtracted from the value for each observation, resulting in a new mean of zero. Following, “the difference between the individual's score and the mean is divided by the standard deviation, which results in a standard deviation of one. If we start with a variable x , and generate a variable x^* , the process is”²:

² “How do I standardize variables in Stata?”, in: <http://www.ats.ucla.edu/stat/stata/faq/standardize.htm> (Access: 12/04/2012)

$$z = \frac{(x - \mu)}{\sigma}$$

Z-transformation leads to greater effects of extreme values on the composite indicator, which is not a problem for our framework since we want to spot ‘champions’ concerning service quality (e.g. extremely low average waiting time should have a strong influence in the indicator’s final score).

The survey data for the ‘soft’ indicator do not need to be normalized as the same scale is used for all variables in the questionnaire.

(c) Weighting

Simply adding up the different standardized variables would not make sense for a reliable and valid indicator because the variables are not equally important for the theoretical concept of service quality. While each interviewee herself weights the ‘soft’ measure variables, the ‘hard’ measure dataset needs specific weighting according to the relevance of different variables. In general, “[g]reater weight should be given to components which are considered to be more significant in the context of the particular composite indicator” (OECD 2003: 10). We have developed a system of weights according to the relevance of each variable for overall service quality in relation to *all* citizens using *all* different possible services (see *Annex*, figure 11). As the choice of weights might be the most vulnerable part of our methodology, we have to address the relevant problems in this regard.

The main criticism is that the individual choice of weights would be rather arbitrary and can lead to significant differences in final indicator scores depending on small changes in the applied weights. Secondly, “weights in additive aggregations necessarily have the meaning of substitution rates and do not indicate the importance of the indicator associated” (Nardo et al. 2005: 76). Thirdly, one could argue that different groups of persons would come to different conclusions about the weights that should be used for variables of our framework (e.g. disabled persons may have the opinion that access is more important than opening hours of *Bürgerämter*).

However, we think that our weighting system can be justified and upheld against these arguments. To address the first point, we have compared the scores of our ‘hard’ measure indicator with another ‘hard’ measure indicator that consists of the same data but is based on differing weights. The results show that a change of weights (e.g. changing the weight of the variable “opening hours” from 1.2 to 1.4) leads to a maximal final indicator score change of around 0.1 which does not affect the overall ranking of different districts and/or *Bürgerämter*

(see *Annex*, figure 13). The second criticism is a rather theoretical argument that could be discussed in great detail (see Gutiérrez Sanín et al. 2011). However, as the alternatives proposed by critics ('MCA' method, see Nardo et al. 2005: 76) also have their flaws (such as the dependence of irrelevant alternatives) and as there is no other method at hand to account for the theoretical assumptions of relative importance of certain variables, we can justify our approach also against this argument. Besides, we have tested the differences in indicator scores using other methods than aggregation with weights (namely factor analysis) and found that our method suits the purposes of our SQI better (see *Annex*, figure 12). Finally, we argue that our framework and the weights included aim to take into account the needs of all citizens (and thus not of one specific group). At a later stage it would, however, be possible to adopt the SQI framework to come up with scores addressing specific groups of society by adjusting the weights according to assumed variable importance.

(d) Final Composition

Taking into account the steps outlined above, our 'hard' measures indicator is finally developed in the following way:

$$\begin{aligned} \text{Hard Measures SQI} = & \text{standardized openinghours} * 1.4 + \text{standardized} \\ & \text{satopening} * .4 + \text{disabled} * .8 + \text{standardized access} * .8 - \text{standardized} \\ & \text{distance} * .4 + \text{picture} * .4 + \text{surveypossible} * .2 - \text{standardized} \\ & \text{avrgwaiting} * 1.4 + \text{standardized percwaiting} * 0.6 - \text{standardized} \\ & \text{avrgservtime} * 1.4 \end{aligned}$$

Afterwards, the resulting variable is again standardized using the z-transformation (see above). The score of the final hard measure indicator thus ranges between approximately -3 and +3. The higher a *Bürgeramt* or a district scores on this final variable, the higher the service quality will be according to the underlying variables and weighting assumptions. It has to be underlined that the final indicator score can only be interpreted in comparison to 'hard' scores of other *Bürgerämter* and does not indicate the 'objective' quality of services in an *absolute* way.

The soft measure indicator is aggregated by summing up the answer scores of all questions and individually using the weights assigned by each interviewee:

$$\begin{aligned} \text{Soft Measures SQI} = & (\text{mean of the scores in part 1 of the} \\ & \text{questionnaire} * \text{weight for part 1} + \text{mean of the scores in part 2 of the} \\ & \text{questionnaire} * \text{weight for part 2} + \text{mean of the scores in part 3 of the} \\ & \text{questionnaire} * \text{weight for part 3}) / \text{sum of all three weights} \end{aligned}$$

After standardizing the new variable, a mean ‘soft’ score for each *Bürgeramt* is calculated. Finally, the results of both datasets (survey data & ‘hard’ data) are brought together for analysis. The next section shows how we used this method in a pretest for the *Bürgerämter* in Berlin.

4. Pretest results and data analysis: Lessons learned

Having developed a reasonable methodology for the SQI, this framework has to be analyzed and tested to optimize it. Therefore, we conducted pretests for both the ‘soft’ and the ‘hard’ service quality measures. To test the soft indicator, we conducted a pretest survey according to our approach (part 2.1) in three different citizen centers (*Bürgerämter*) in the city of Berlin between 12 April and 18 April 2012. Second, we applied the ‘hard’ indicator methodology to all 40 *Bürgerämter* of Berlin.

The results of both pretests turned out to be valuable for the optimization of our framework as they gave instructive insights into the strengths and weaknesses of the developed methodology. Besides, they offered first insights into the perception of quality of administrative services in Berlin.

4.1 Survey pretest results

While carrying out a survey in three different citizen centers in Berlin with a total number of 107 interviewees does not lead to representative data (see below, figure 2), such a small sample can uncover weaknesses and shortcomings of a purely theoretically developed framework. In general, as there is consensus that samples with more than 100 cases are sufficient for the purposes of a pretest, our survey had a sufficient sample size (Schulze 2006).

Therefore, the goal of the survey was - as for the hard indicator pretest as well - not necessarily to collect reliable and representative data but rather to identify problems concerning the actual applicability of the SQI.

	Observations (Male/Female)	Avg. Waiting Time	People w/ online appointments	Avg. Age	Avg. Education (years)
Lichtenberg	40 (16/24)	88.74	7	42.45	14
Mitte	30 (16/14)	94.15	5	39.84	14.77
Zehlendorf	38 (19/19)	53.69	9	38.22	14.73
Sum	108	78.86	21	40.17	14.5

Figure 2 – SQI ‘soft’ dimension pretest - overview

(a) Problems and obstacles

When conducting a survey one has to deal with several problems and obstacles. While not all of them are avoidable, they have to be reduced as much as possible. In that sense, we found the following difficulties in the survey at stake:

- 1. Selection bias:** Since a survey mirrors necessarily only the opinions and attitudes of a sample of the whole population the respective sample should be as representative as possible for the actual group under investigation. Therefore, when conducting surveys it is important to investigate the sample for its randomness, i.e. “each subject or case [should have] [...] an equal chance of being assigned to the experimental group” (Meier et al. 2009: 47). If there were, however, any systematic differences between the sample and the actual group under investigation the survey would not be representative and would thus lack a crucial factor contributing to the validity of the results (Meier et al. 2009).

When looking at the results of our survey pretest, we recognized several selection biases. First of all, the most obvious selection bias occurred due to the time of day during which our survey was conducted. Going to *Bürgerämter* during the day *and* during the week excludes many people from the working population who might either turn up very early in the morning, late in the evening or on Saturdays. Second, most people in a hurry did not partake in our survey. They might belong to a certain group of *Bürgeramt* customers with more significant time constraints. Thirdly, the average education of the interviewees seems to be particularly high (14.5, see above, figure 2). This indicates a bias towards interviewees with higher education than the actual population mean in the respective *Bürgeramt* district is supposed to have. At the same

time, these results match with our personal experiences as those that were hardest to reach were often people with a migrant background, language problems etc.

- 2. Interviewer effect:** The pretest revealed that we as interviewers had an effect on the final result itself. In some cases it was apparent that the interviewees' answers were biased due to the presence of the interviewer and the resulting "lack of anonymity" (Frankfort-Nachmias/Nachmias 1996: 238). A crucial reason for this problem is the factor of social desirability implying that interviewed people tend to answer in a way that is - according to their individual perception - in accordance with the perceived common sense on a given issue although their actual opinion might differ. In this regard, the answers in the last section of the questionnaire were most problematic. The last section contains questions about the personal characteristics of the interviewees. Here, we suppose reasonably (by experience) that some individuals gave biased answers when being asked about sensitive issues such as age or level of education. Others, most probably for the same reason, did not answer these questions at all. This also partly explains the high average of education stated above since some people that might have been ashamed of their 'lower' educational achievements gave false answers or did not answer at all.
- 3. Leading questions:** If the questions being asked implicitly lead the interviewee to a certain answer due to the way they have been formulated (i.e. allocating scores to a statement instead of a question: "the service is good" vs. "how do you rank the service?"), the validity of the answer and hence that of the question itself must be questioned. In general, the pretest revealed some tendencies of leading questions. An indication for this might be that some questions were often answered very similarly ("strongly agree") which is - given the number of observations and the variety of possible answers - peculiar.
- 4. Understandability:** After our first pretest we have found a number of problems also in the sense that some of the interviewees had difficulties understanding the meaning of certain parts of the questionnaire. In this regard, the issue of clear, plain and proper language is at stake, given that several interviewees had serious difficulties understanding the wording of questions as well as the scaling of possible responses. Especially, questions 2.1, 2.3 and 5.3 as well as the scales (-3 to +3; and weighting

scale: allocation of 100 points) we used in the first questionnaire (see *Annex*, figure 6 – questionnaire 1) turned out to cause problems of comprehension.

- 5. Redundancy:** Finally, redundant questions are a problem in surveys, as they do not provide any added value. On the contrary, they can rather unsettle or confuse the interviewees (Schulze 2006). In our survey, we observed redundancy of some questions as several people stated that they already had given an answer to a certain issue before and were confused why they were asked about a similar issue again. This was true for questions 2.2 and 3.2 which were both understood by many to be similar to question 3.2 of the first questionnaire (see *Annex*, figure 6).

(b) What to do about it? Lessons learned

Having identified these problems, what were the next steps to further develop a final survey for the ‘soft’ dimension of the SQI? In the following section we will elaborate on the possibilities to reduce some of the problems at stake. However, one has to underline that there are limits to the optimization of the SQI survey since, in social sciences, not every intervening variable can be controlled completely.

- 1. Selection Bias:** Given the small sample size of the pretest, the problem of selection biases is likely to decrease directly when carrying out a full-scale representative survey. Moreover, at a later stage of data analysis “the probability of various degrees of influence from random error can be quantified, so that we can use confidence intervals to express the uncertainty inherent in our estimates due to sampling variability” (Schoenbach 2011). Thus, the problems of selection biases found in the pretest are expected to be minimized at a later stage of surveying.
- 2. Interviewer bias:** As long as surveys are conducted via personal interviews, interviewers will affect the interviewees. In this regard, there will always be some extent of interaction between researcher and subject. The question, however, is whether the effects of this interaction can be further mitigated. In order to minimize interviewer effects there is consensus in sociology that the interviewer can:
 - forgo rated statements;
 - prevent the presence of a third person in the interview;
 - prevent communication between different interviewees.

Since these rules were already applied during the first pretest there is not much leverage to further reduce the interviewer effect in the present survey.

Nevertheless, one might ask why we did not decide to replace the personal interview with a written one, i.e. interviewees individually write down their answers, without the presence of an interviewer. Such an approach would certainly minimize some of the problems associated with the interviewer effect. However, if a lot of interviewees are either not motivated to partake in a written survey or are not used to writing for a longer time, the oral interview might be the better fit. Besides, in an oral interview the interviewer can help to dissolve any misunderstandings (Schulze 2006). Given the heterogeneity of the citizen body and the already identified selection and language biases, the oral interview is better suited since it provides the interviewees with more certainty about the content of the questions and allows for the interviewer to discover any misunderstandings, as observed during the pretest phase.

3. Leading questions: As we have identified the words “*immer*” (meaning “always”; see question 3.2, first questionnaire; question 2.2, second questionnaire, *Annex*, figures 6 & 7) and “*stets*” (also meaning “always”; see question 4.2, first questionnaire and question 2.3, second questionnaire) as possible ‘leading words’ we erased those formulations in the final questionnaire (see final questionnaire, *Annex*, figure 8).

4. Understandability: The statements, which were identified as problematic during the pretest (see part 4.1.a.) have been deleted completely since not only the wording but also the questions themselves were misleading and turned out not to add any value. Furthermore, the scaling was replaced (new scale from 0 to 6) in the second questionnaire due to understandability issues with the SERVQUAL scaling from -3 to +3.

We also changed the weighting scale (allocation of 100 points to the different dimensions) due to similar problems. However, for measuring weights the scale we developed in our second questionnaire turned out to deliver insignificant results because it enhanced the leading character of the questions with people stating that all service dimensions would be important for them (see second questionnaire, *Annex*, figure 7, part 4). Thus, we finally created a new scale based on a much more easily understandable ranking system (‘*Platz 1*’ to ‘*Platz 4*’, see final questionnaire, *Annex*,

figure 8, part 5) at the same time allowing the interviewees to weigh the service categories according to their personal priorities.

5. **Redundancy:** The redundant questions (questions 2.2 and 3.2, first questionnaire) were deleted in the second questionnaire while the question with a similar content and the highest level of understandability was kept.
6. **Motivation:** Finally, the pretest in the first *Bürgeramt* showed how difficult it is to find and motivate people to participate in our survey. Especially in an environment like a *Bürgeramt* people lack patience, trust and time to answer questions about a certain service they have just received; mainly due to long waiting times (average for all three *Bürgerämter* in our pretest: 78.86 minutes) but also due to their inherent skepticism concerning surveys in general. Thus, having an initial questionnaire of four pages with long text passages was discouraging for many. In order to enhance participation and minimize the impression of this survey being too complicated and long-winded we decided to delete all unnecessary initial explanations. For the same reasons, and because no interviewee made use of it, a box for notes and advices (questionnaire 1, question 7) was deleted in the following questionnaire. In the end, the merging of some questions or categories and the deletion of these passages and boxes led to a questionnaire of only two pages which is much easier to grasp and understand. Overall, simplicity and understandability turned out to be highly valuable characteristics, as they severely increased the motivation to participate in our research.

It remains to be seen whether the changes outlined above improve the SQI data on a larger scale but the pretest already demonstrated that the improved version better serves the purpose of researching the service quality of *Bürgerämter* than did the initial version, which more closely matched the SERVQUAL model.

(c) Beyond validity – possible interpretation and further analysis of the ‘soft’ indicator

Despite the fact that the three conducted surveys helped to optimize the questionnaire in terms of understandability, simplicity and validity they have also provided us with useful data, which, albeit not being fully representative, reveal insights into the possible usage and interpretation of the ‘soft’ SQI.

Firstly, the indicator itself cannot be interpreted in absolute terms but has to be analyzed relatively to the other scores. For instance, our pretest data show that perceived quality of services in the *Bürgeramt* Lichtenberg/ Hohenschönhausen (LH) is the highest in relation to the other *Bürgerämter* analyzed (see *Annex*, figure 19). Taking into account that the waiting time in the *Bürgeramt* LH was on average 30 minutes higher than the waiting time in the *Bürgeramt* Zehlendorf (see above, graph 2), our pretest reemphasizes the importance of a distinction between ‘soft’ and ‘hard’ service quality measures: waiting time is an important factor of service quality per se and is thus included in the ‘hard’ indicator score; nevertheless, it is possible that customers may *perceive* the quality of services as being high (see *Bürgeramt* LH and SZ) despite long waiting times (e.g. if waiting time is not a crucial factor of service quality for them).

Secondly, the gathered indicator data point to possibilities of detailed statistical analysis at a later stage. For instance, researchers could look at a possible correlation between age and the perceived quality of service in citizen centers (see *Annex*, figure 14). Another interesting question regarding the theory of service quality would be if education matters in this regard (see *Annex*, figure 15). In general, a sufficient sample size would allow researchers to conduct in-depth multivariate analysis, i.e. linear regression with the indicator score as a dependent variable. Examining the determinants of the ‘soft’ service quality score would allow for a contribution to the general debate on administrative service perception.

4.2 First results: ‘Hard’ indicator

Looking at the ‘hard’ indicator scores (see figure 3 and *Annex*, figure 9) - which are not representative and, as outlined in part 2, partly include data from 2007 (e.g. average waiting time) - one finds several interesting results with regard to our indicator framework. Why does Treptow-Köpenick (TK1) score so high on our ‘hard’ measure indicator? And, what does the final score tell us about service quality of the respective administration?

Taking into account the underlying dataset (see *Annex*, figure 18), one finds that in *Bürgeramt* TK1 the average waiting time is astonishingly low (8 minutes), the average service time of one day is fast, it has convenient opening hours (36 hours per week), opens on every Saturday and is located very closely to a public transport stop. The indicator reflects these advantages in relation to other *Bürgerämter* in the relatively high final score. On the contrary, the *Bürgeramt* Friedrichshain-Kreuzberg 2 has the lowest ‘hard’ service quality score because of long average waiting time, extremely long service time (an average of 28.5 days) and other issues such as the fact that it does not open on weekends. Also, when taking into account the

scores of other *Bürgerämter* or districts in relation to the underlying data, there is strong evidence for a valid, reliable and acceptable ‘hard’ measure indicator for service quality, which accurately reflects its underlying considerations. The next question is then how both indicator scores can be combined and visualized in order to come up with a detailed and accessible analysis of service quality in the end.

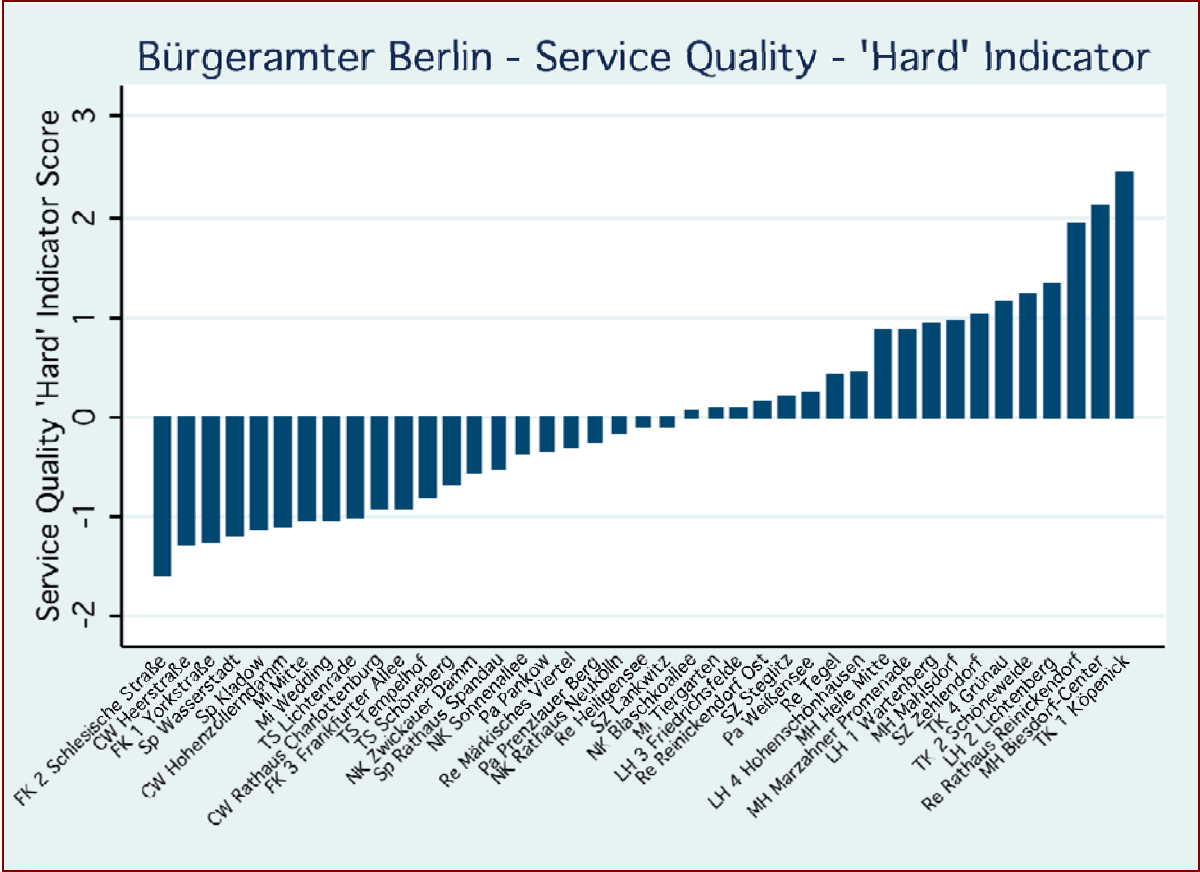
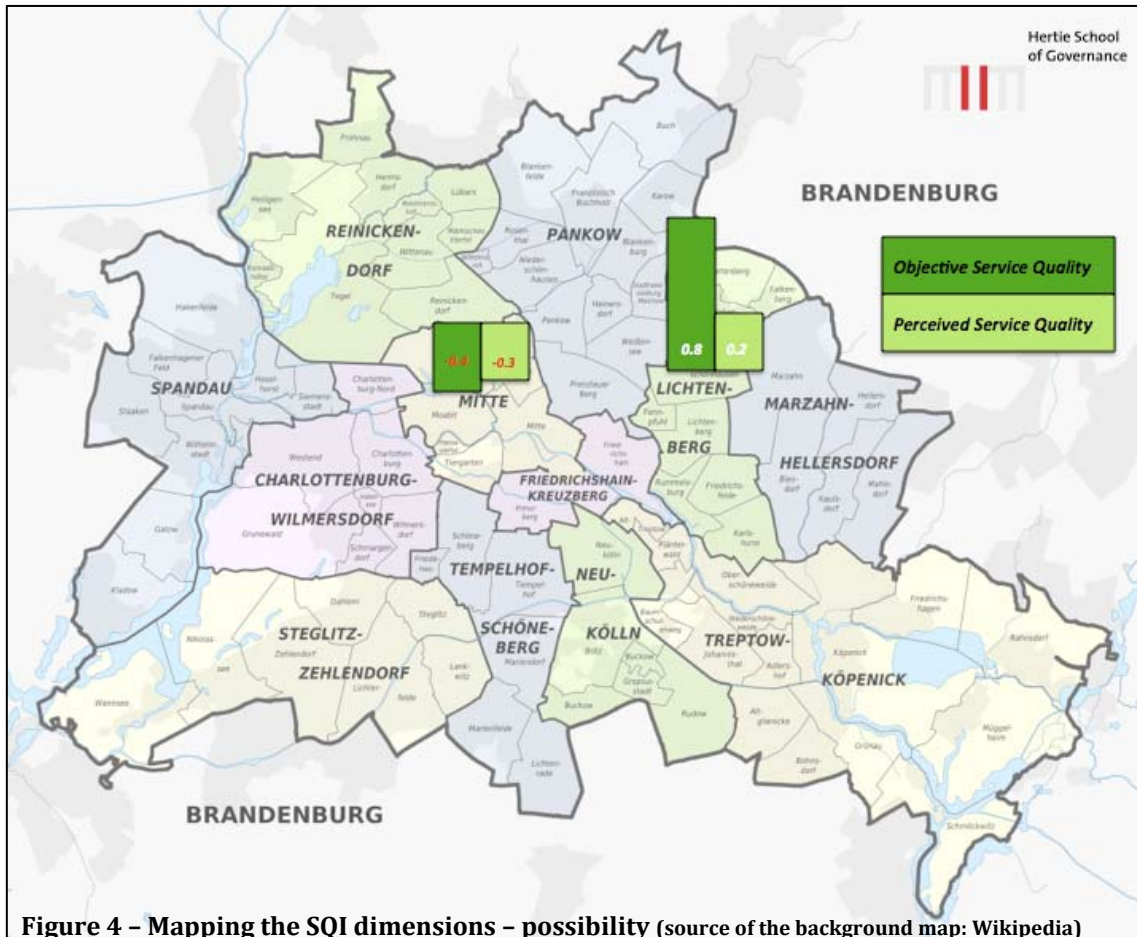


Figure 3 - Bürgeramt ‘Hard’ Indicator Score – Overall Overview

4.3. Mapping the final SQI scores: possibilities and limitations

The question of how both service quality scores could be mapped best remains to be finally settled at a later stage of research, e.g. after cross-city representative data are gathered. However, some initial ideas about possible graphical representation can already be presented using the available pretest data. First of all, it seems most reasonable to include both scores on



a geographical map (e.g. see above, figure 4; and *Annex*, figure 20), based on the data gathered in the pretests) of the respective region.

Such mapping of the final data allows for a meaningful, clear and easily understandable illustration of the results. Of course, standard illustrations such as bar graphs (see for example figure 3, p. 23) should also be used to display the indicator. Finally, more advanced graphs such as a spider-diagram could be employed to comprehensively display underlying aspects of data such as correlation between ‘hard’ and ‘soft’ measures (see below, figure 5).

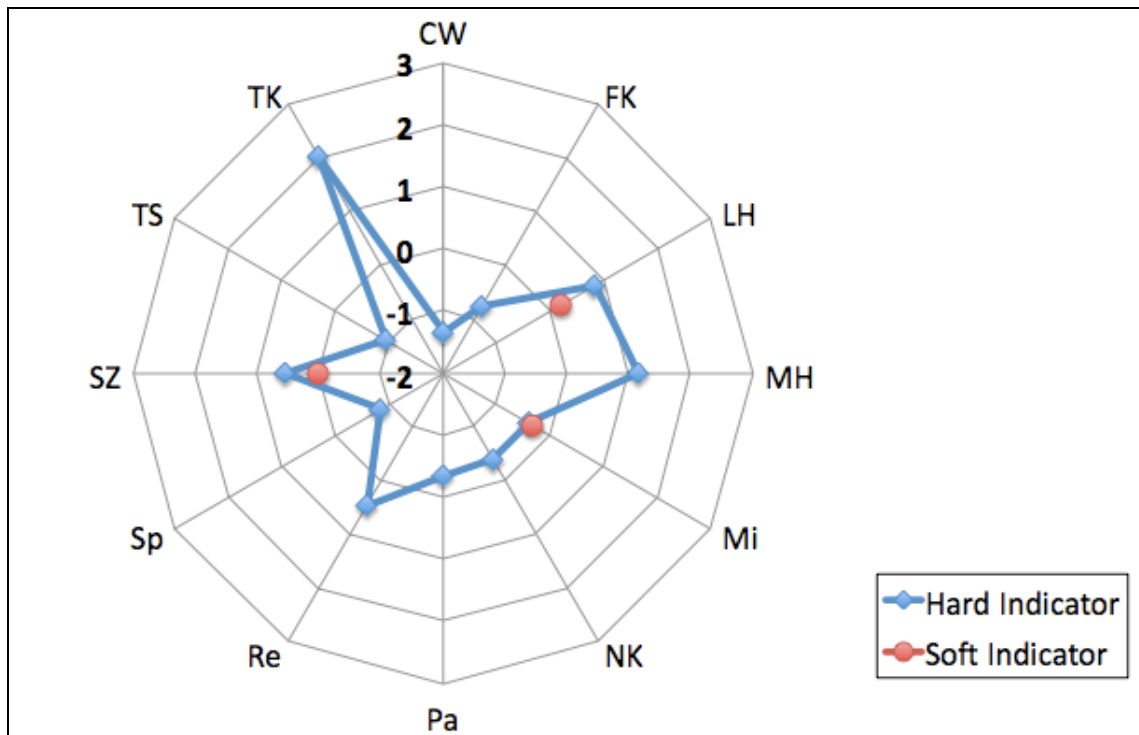


Figure 5 – Mapping the SQI – Spider-diagram based on the data gathered in the pretests

5. Conclusions and outlook: Opportunities for expansion of the framework

5.1 E-governance as a crucial approach

Given that a vast amount of service work is going to shift online, our framework would have to be supplemented with a component measuring the quality of e-governance. An internal analysis by the IT competence team of the Berlin Senate argues that in the near to medium future more than 50% of *Bürgeramt* services should be available online, which could radically increase service efficiency. Currently in 46% of all cases the customer still has to appear personally at a *Bürgeramt*. According to their internal goals, this ratio should be reduced to a mere 16% (Ganser 2007: 12-16). Given that eSignature, eVerification, ePayment and overall security is likely to be increased to a point, which makes most services available for online offer, we should quickly discuss how eService quality should be measured within our framework.

We propose a lean and simple approach, measuring eService Quality along the lines of scope (i.e. amount of services available online) and speed (i.e. amount of time necessary to complete a given service). The scope variable would be straightforward to produce: Create an exhaustive list of all services, which could possibly be offered online (estimated to be around 100 services) and check each *Bürgeramt*'s website, ticking off the ones available and ranking

them accordingly. This would only be possible across cities, as within a single city the IT structures are usually integrated which makes internal comparison impossible. The speed variable would be measured by counting the ‘clicks’ it takes to first, find the service (starting with any common search engine) and then completing the entire process. This should be done for each service in offer.

An indicator of this sort would draw a very clear and unequivocal picture of online service quality and would circumvent many of the issues, which limit the validity of the current process. However, given that in this paper, we are focusing on Berlin (where there is one website layout for all *Bürgerämter*), we could not integrate our e-Indicator framework within the actual pretest, which is why we will have to leave this work to future researchers.

5.2 Conclusion

The SQI framework allows for a comprehensive measurement of service quality in customer-based government services. In our approach, we took the previous best-practice framework of SERVQUAL and slimmed it down to increase its practicability and feasibility; while combining this survey based method with a set of eleven hard variables adding reliability, which all previous approaches lack. Our pretest results of the ‘hard’ as well as the ‘soft’ indicator support our claim of having developed a truly valuable tool for measuring service quality. However, we also have to acknowledge some limitations: First of all this indicator should be taken for what it is, namely a pure outcome/output measure. This means that it can give valuable insights into the important field of service quality; however, it should not be mistaken for an equivalent of a good governance measure, as it tells us nothing about the right inputs, processes or the efficiency, by which a service came about. Consequently, it should be used in combination with other indicators, when one wants to make meaningful claims about good governance. Secondly, this framework is and has to be imperfect due to the simple fact, that it is trying to measure something, which can neither be clearly defined nor perfectly measured. However, despite these limitations this paper outlines how researchers can succeed in measuring such an elusive concept. It would be noteworthy to do further research, which takes this framework, goes beyond our pretest and does an actual evaluation of *Bürgerämter* across German cities. Considering the costs and feasibility of this, we predict that getting the measurements for the ‘hard’ indicator should be fairly cheap and fast while conducting the ‘soft’ measure survey is somewhat more resource intensive but remains feasible: taking samples of around 600 questionnaires (around 70 man-hours) per city should suffice to provide representative data. The e-governance indicator should then be added to the

framework. Based on the results, municipalities could share best practice approaches and increase service quality on a whole. In a second step, the SQI could be adjusted and perfected to test all sorts of customer-based government services. With a larger amount of data over a period of years one could also start making claims about the correlation between the ‘soft’ and the ‘hard’ indicator and see to what extent citizens actually acknowledge the ‘objective’ improvements of services with higher personal ratings. Furthermore, such data would allow the spotting and singling-out of certain biases, which our current pretest data does not allow for. We believe that the SQI will be a valuable contribution to the field of good governance, as good choices in public administration can only be made if they are based on useful information rather than arbitrary hunches.

Bibliographical References

- Cronin, J./Taylor, S. (1992): Measuring service quality: a re-examination and extension. In: *Journal of Marketing*, 56: pp. 55-68.
- Cronin, J./Taylor, S. (1994): SERVPERF versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality. In: *Journal of Marketing*, 58:1: pp. 125-131.
- Donnelly, M./Wisniewski, M. (1995): Measuring service quality in local government. The SERVQUAL approach. In: *International Journal of Public Sector Management*, 8(7): pp. 15-20.
- Frankfort-Nachmias, C./Nachmias, D. (1996): *Research Methods in the Social Sciences*. New York: St. Martin's Press.
- Ganser, R. (2007): *Projekt Virtuelle Bürgerdienste (Online-Bürgeramt)*. Berlin: Senatsverwaltung für Inneres und Sport Berlin.
- Gutiérrez Sanín, F. et al. (2011): *Aggregating Political Dimensions: Of the Feasibility of Political Indicators*. Santa Fé: Universidad Nacional de Colombia.
- Den Boer, S. (2011): *Public sector wages and government performance. An overview of data, theory and a brief analysis*. Maastricht: Maastricht School of Governance.
- Heycox, J. (1999): *Integrating data for sustainable development: introducing the distribution of resources framework*. Novartis Foundation Symposium 220: Environmental statistics - analysing data for environmental policy. London: John Wiley & Sons.
- Jollands, N./Lermit, J./Patterson, M. (2003): The usefulness of aggregate indicators in policy making and evaluation: a discussion with application to eco-efficiency indicators in New Zealand. In: <http://www.ecoeco.org/pdf/jollands.pdf> (Access: 15/04/2012).
- Kaufmann, D./Kraay, A./Mastruzzi, M. (2010): *The Worldwide Governance Indicators: Methodology and Analytical Issues*. Washington D.C.: World Bank.
- Klages, H. (2006): *Aufbau eines Monitoringsystems „Effizienz und Effektivität“ Berliner Bürgerämter. Abschlussbericht*. Speyer: Forschungsinstitut für öffentliche Verwaltung bei der Deutschen Hochschule für Verwaltungswissenschaften.
- Masser, K.; Stallmeyer, A.; Freund, J.; Krüger, P. (2008): *Entwicklung und Erprobung eines Monitoringsystems "Effizienz und Effektivität" Berliner Bürgerämter*. Speyer: Forschungsinstitut für öffentliche Verwaltung bei der Deutschen Hochschule für Verwaltungswissenschaften.
- Meier, K.J./Brudney, J.L./Both, J. (2009): *Applied Statistics for Public & Nonprofit Administration*. Belmont: Thomson Wadsworth.
- Munda, G./Nardo, M. (2003): *On the Methodological Foundations of Composite Indicators Used for Ranking Countries*. Barcelona: Universitat Autònoma de Barcelona.
- Nardo, M. et al. (2005): *Tools for Composite Indicators Building*. Ispra: European Commission Joint Research Centre.
- Parasuraman, A./Zeithaml, V./Berry, L.L. (1985): A Conceptual Model of Service Quality and Its Implications for Future Research. In: *Journal of Marketing*, 49 (Fall): pp. 41-50.
- Parasuraman, A./Zeithaml, V./Berry, L.L. (1988): SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality. In: *Journal of Retailing*, 64 (Spring): pp. 12-40.

Parasuraman, A./Zeithaml, V./Berry, L.L. (1991): Refinement and Reassessment of the SERVQUAL Scale. In: *Journal of Retailing*, 67 (Winter): pp. 420-450.

Ramseook-Munhurrun, P./Lukea-Bhiwajee, S./Naidoo, P. (2010): Service Quality in the Public Service. In: *International Journal of Management and Marketing Research*, 3(1): pp. 37-50.

Schnell, R./Hill, P./Esser, E. (2005): Methoden der empirischen Sozialforschung. München: Oldenbourg Verlag.

Schoenbach, V. (2011): Principle of Epidemiology for Public Health. Sources of error: Selection Bias. In: <http://www.epidemiolog.net/epid160/lectures/09-SelectionBias.ppt?user=epid160> (Access: 27/04/2012).

Schulze, G. (2006): Einführung in die Methoden der empirischen Sozialforschung. In: *Bamberger Beiträge zur empirischen Sozialforschung* 1(2006), pp. 88–103.

Social and Cultural Planning Office (2004): Public sector performance: an international comparison of education, health care, law and order and public administration. The Hague: Government of the Netherlands.

OECD (2003): Composite indicators of country performance: a critical assessment. Paris: OECD.

OECD (2008): Handbook on Constructing Composite Indicators. Paris: OECD.

UN Department of Economic and Social Affairs (2007): Public Governance Indicators: A Literature Review. New York: UN Publishing.

UNDP (2009): A Guide to Measuring Local Governance. Oslo: UNDP Oslo Governance Centre.

Annex

Figure 6 – First Draft of the Questionnaire



Fragen zur Servicequalität von Bürgerämtern

Bogen Nr.:

Erläuterung: Die folgenden Aussagen zielen auf ihre Meinung hinsichtlich der Dienstleistung des Bürgeramts XY ab. Bitte markieren Sie, ob und in wie weit Sie den folgenden Aussagen zustimmen. Z.B. wenn Sie der Meinung sind, dass das Bürgeramt XY eine beliebige Eigenschaft gar nicht besitzt, dann kreuzen Sie bitte '-3' an. Falls Sie der Aussage dagegen voll zustimmen, kreuzen Sie bitte die '3' an. Wenn Sie dagegen der Aussage nur zu einem gewissen Teil zustimmen, bitten wir Sie diejenige Zahl zwischen '-3' und '3' anzukreuzen, die ihre Meinung am besten widerspiegelt. Dabei gibt es keine richtige oder falsche Antwort – wir sind nur an einer Zahl interessiert, die Ihren Eindruck bezüglich der Servicequalität des Bürgeramtes XY am besten wiedergibt.

1. In diesem Abschnitt steht das *äußere Erscheinungsbild* der Dienstleistung im Vordergrund.

	Stimme gar nicht zu						Stimme vollkommen zu
Die Ausstattung der Räume ist optisch ansprechend.	-3	-2	-1	0	1	2	3
Die ausliegenden (Info-) Materialien sind optisch ansprechend und verständlich.	-3	-2	-1	0	1	2	3

2. In diesem Abschnitt steht die *konkrete und zuverlässige Ausführung* der Dienstleistung im Vordergrund.

	Stimme gar nicht zu						Stimme vollkommen zu
Das Bürgeramt erbrachte den Service zu dem versprochenen Zeitpunkt.	-3	-2	-1	0	1	2	3
Als Sie ein Problem hatten, zeigte das Bürgeramt ein ehrliches Interesse daran, das Problem zu lösen.	-3	-2	-1	0	1	2	3
Das Bürgeramt erbrachte den erwünschten Service schon beim ersten Mal und ohne Probleme.	-3	-2	-1	0	1	2	3

3. In diesem Abschnitt steht das Verhalten der Bediensteten des Bürgeramtes sowie das schnelle und aktive Reagieren auf Ihre Bedürfnisse im Vordergrund.

	Stimme gar nicht zu						Stimme vollkommen zu
Die Bediensteten liefern schnellen Service.	-3	-2	-1	0	1	2	3
Die Bediensteten sind immer bereit Ihnen zu helfen.	-3	-2	-1	0	1	2	3

4. In diesem Abschnitt wird die fachliche Kompetenz und das Auftreten der Bediensteten bewertet.

	Stimme gar nicht zu						Stimme vollkommen zu
Die Bediensteten besitzen das nötige Wissen, um Ihre Fragen zu beantworten.	-3	-2	-1	0	1	2	3
Die Angestellten waren stets höflich.	-3	-2	-1	0	1	2	3

5. In diesem Abschnitt wird das Einfühlungsvermögen der Angestellten in Ihrer Situation bewertet.

	Stimme gar nicht zu						Stimme vollkommen zu
Das Bürgeramt bietet Ihnen bequeme Öffnungszeiten an.	-3	-2	-1	0	1	2	3
Die Bediensteten konnten Ihre individuellen Bedürfnisse nachempfinden.	-3	-2	-1	0	1	2	3
Die Angestellten schenken Ihnen persönlich Aufmerksamkeit.	-3	-2	-1	0	1	2	3

6. Fragen zur individuellen Punktevergabe

Erläuterung: Im Folgenden finden Sie fünf Eigenschaften in Bezug auf Bürgerämter und ihren Service. Wir würden nun gerne von Ihnen wissen, wie wichtig Ihnen die jeweilige Eigenschaft ist, wenn Sie die Servicequalität eines Bürgeramtes bewerten. Bitte vergeben Sie im Folgenden insgesamt 100 Punkte an diese Eigenschaften. Verteilen Sie die 100 Punkte dabei danach, wie wichtig Ihnen die jeweilige Eigenschaft eines Bürgeramtes ist. Je wichtiger Ihnen eine Eigenschaft dabei ist, desto mehr Punkte sollten Sie vergeben. Bitte achten Sie darauf, dass Sie insgesamt nicht mehr als 100 Punkte vergeben.

1. **Das äußere Erscheinungsbild der Dienstleistung (Räumlichkeiten, (Info-)Materialien).** _____ Punkte
2. **Die konkrete und zuverlässige Dienstleistung des Bürgeramtes** _____ Punkte
3. **Das Verhalten der Bediensteten des Bürgeramtes sowie das schnelle und aktive Reagieren auf Ihre Bedürfnisse.** _____ Punkte
4. **Die fachliche Kompetenz und das Auftreten der Bediensteten.** _____ Punkte
5. **Das Einfühlungsvermögen der Bediensteten in Ihrer Situation.** _____ Punkte

Gesamtzahl der Punkte: _____ 100 Punkte

7. Gibt es aus Ihrer Sicht noch Hinweise oder Anmerkungen? (Freitextfeld)

- 8. Abschließend bitten wir Sie um einige Angaben zu Ihrer Person. Diese Daten werden ausschließlich zur statistischen Auswertung benötigt und anonymisiert verarbeitet, so dass eine persönliche Zuordnung nicht**

Figure 7 – Second Draft of the Questionnaire

Fragen zur Servicequalität von Bürgerämtern

Bogen Nr.:

1. In diesem Abschnitt stehen die äußeren Umstände sowie das *äußere Erscheinungsbild* der Dienstleistung im Vordergrund.

	Stimme gar nicht zu							Stimme vollkommen zu
Die Ausstattung der Räume ist optisch ansprechend.	0	1	2	3	4	5	6	
Die ausliegenden (Info-) Materialien sind optisch ansprechend und verständlich.	0	1	2	3	4	5	6	
Das Bürgeramt bietet Ihnen bequeme Öffnungszeiten an.	0	1	2	3	4	5	6	

2. In diesem Abschnitt steht das *Verhalten der Bediensteten des Bürgeramtes* sowie das *schnelle und aktive Reagieren auf Ihre Bedürfnisse* im Vordergrund.

	Stimme gar nicht zu							Stimme vollkommen zu
Die Bediensteten liefern schnellen Service.	0	1	2	3	4	5	6	
Die Bediensteten sind immer bereit Ihnen zu helfen und konnten Ihre individuellen Bedürfnisse nachempfinden.	0	1	2	3	4	5	6	
Die Angestellten waren stets höflich.	0	1	2	3	4	5	6	

3. In diesem Abschnitt wird die *fachliche Kompetenz* und das *Auftreten der Bediensteten* bewertet.

	Stimme gar nicht zu							Stimme vollkommen zu
Die Bediensteten besitzen das nötige Wissen, um Ihre Fragen zu beantworten.	0	1	2	3	4	5	6	

4. Fragen zur individuellen Punktevergabe

Gar nicht wichtig Sehr wichtig

Die äußeren Umstände und das äußere Erscheinungsbild.	0	1	2	3	4	5	6
---	---	---	---	---	---	---	---

Das Verhalten und Auftreten der Bediensteten.	0	1	2	3	4	5	6
---	---	---	---	---	---	---	---

Die fachliche Kompetenz der Angestellten.	0	1	2	3	4	5	6
---	---	---	---	---	---	---	---

5. Angaben zu Ihrer Person: Diese Daten werden *anonymisiert* und ausschließlich zur statistischen Auswertung benötigt, so dass eine persönliche Zuordnung nicht möglich ist.

Alter	
Geschlecht	<input type="checkbox"/> Männlich <input type="checkbox"/> Weiblich
Geburtsort (wenn Berlin, dann Name des Bezirks)	
Höchster Bildungsabschluss	
Beruf	<input type="checkbox"/> Selbstständig <input type="checkbox"/> Angestellt <input type="checkbox"/> Erwerbslos/Arbeitssuchend <input type="checkbox"/> Student <input type="checkbox"/> Schüler <input type="checkbox"/> Freiwilliger (BFD,FSJ,FÖJ, Freiwilligendienste Aller Generationen) <input type="checkbox"/> Rentner
Wartezeit beim Besuch des Bürgeramtes	
Haben Sie vorab einen Termin vereinbart (z.B. online)?	
Wenn ja, wie lange mussten Sie auf einen Termin warten?	
Wenn nein, haben Sie sich über Termine informiert; die Wartezeit wäre aber zu lang gewesen?	

Vielen Dank für Ihre Hilfe

Figure 8 – Final Draft of the Questionnaire

Fragen zur Servicequalität von Bürgerämtern

Bogen Nr.:

1. In diesem Abschnitt stehen die äußeren Umstände sowie das *äußere Erscheinungsbild* der Dienstleistung im Vordergrund.

Stimme gar nicht zu	Stimme vollkommen zu
---------------------	----------------------

Die Ausstattung der Räume ist optisch ansprechend.	0 1 2 3 4 5 6
--	---------------

Die ausliegenden (Info-) Materialien sind optisch ansprechend und verständlich.	0 1 2 3 4 5 6
---	---------------

Das Bürgeramt bietet Ihnen bequeme Öffnungszeiten an.	0 1 2 3 4 5 6
---	---------------

2. In diesem Abschnitt steht das *Verhalten der Bediensteten des Bürgeramtes* sowie das *schnelle und aktive Reagieren auf Ihre Bedürfnisse* im Vordergrund.

Stimme gar nicht zu	Stimme vollkommen zu
---------------------	----------------------

Die Bediensteten sind bereit Ihnen zu helfen und könnten Ihre individuellen Bedürfnisse nachempfinden.	0 1 2 3 4 5 6
--	---------------

Die Angestellten waren höflich.	0 1 2 3 4 5 6
---------------------------------	---------------

3. In diesem Abschnitt wird die *fachliche Kompetenz* und das *Auftreten* der Bediensteten bewertet.

Stimme gar nicht zu	Stimme vollkommen zu
---------------------	----------------------

Die Bediensteten besitzen das nötige Wissen, um Ihre Fragen zu beantworten.	0 1 2 3 4 5 6
---	---------------

4. In diesem Abschnitt geht es um die *Schnelligkeit des Bürgeramtes* und um die *Wartezeit im Bürgeramt*.

Stimme gar nicht zu	Stimme vollkommen zu
---------------------	----------------------

Das Bürgeramt liefert schnellen Service.	0 1 2 3 4 5 6
--	---------------

Weitere Fragen zur Wartezeit.

Wie viele Minuten haben Sie gerade gewartet?	
Haben Sie vorab einen Termin vereinbart (z.B. online)?	
Wenn ja, wie lange mussten Sie auf einen Termin warten?	
Wenn nein, haben Sie sich über Termine informiert; die Wartezeit wäre aber zu lang gewesen? Wie viele Tage hätten Sie demnach warten müssen?	

5. Fragen zur individuellen Gewichtung. Bitte vergeben Sie Platzierungen für die einzelnen Bereiche entsprechend Ihrer persönlichen Einschätzung.

	Platz 1	Platz 2	Platz 3	Platz 4
Die äußeren Umstände und das äußere Erscheinungsbild.				
Das Verhalten und Auftreten der Bediensteten.				
Die fachliche Kompetenz der Angestellten.				
Die Schnelligkeit und Wartezeit im Bürgeramt.				

6. Angaben zu Ihrer Person: Diese Daten werden anonymisiert und ausschließlich zur statistischen Auswertung benötigt, so dass eine persönliche Zuordnung nicht möglich ist.

Alter	
Geschlecht	<input type="checkbox"/> Männlich <input type="checkbox"/> Weiblich
Geburtsort (wenn Berlin, dann Name des Bezirks)	
Wohnort (wenn Berlin, dann Name des Bezirks)	
Höchster Bildungsabschluss	
Beruf	<input type="checkbox"/> Selbstständig <input type="checkbox"/> Angestellt <input type="checkbox"/> Erwerbslos/Arbeitssuchend <input type="checkbox"/> Student <input type="checkbox"/> Schüler <input type="checkbox"/> Freiwilliger (BFD, FSJ, FÖJ, Freiwilligendienste Aller Generationen) <input type="checkbox"/> Rentner <input type="checkbox"/> Sonstiges

Figure 9 - Lessons - assessment criteria for 'hard' service quality measures. Source: own assessment & UNDP (2009)

<i>Hard Measures</i>	<i>Feasibility</i>	<i>Validity</i>	<i>Reliability</i>	<i>Acceptability</i>	<i>Costs</i>
1. Input	+	-	+	-	o
2. Process	o	-	+	-	o
3. Output	+	+	+	-	o
4. Outcome	-	+	-	+	-

Figure 10 – Service Quality Indicator – Theoretical Base for construction

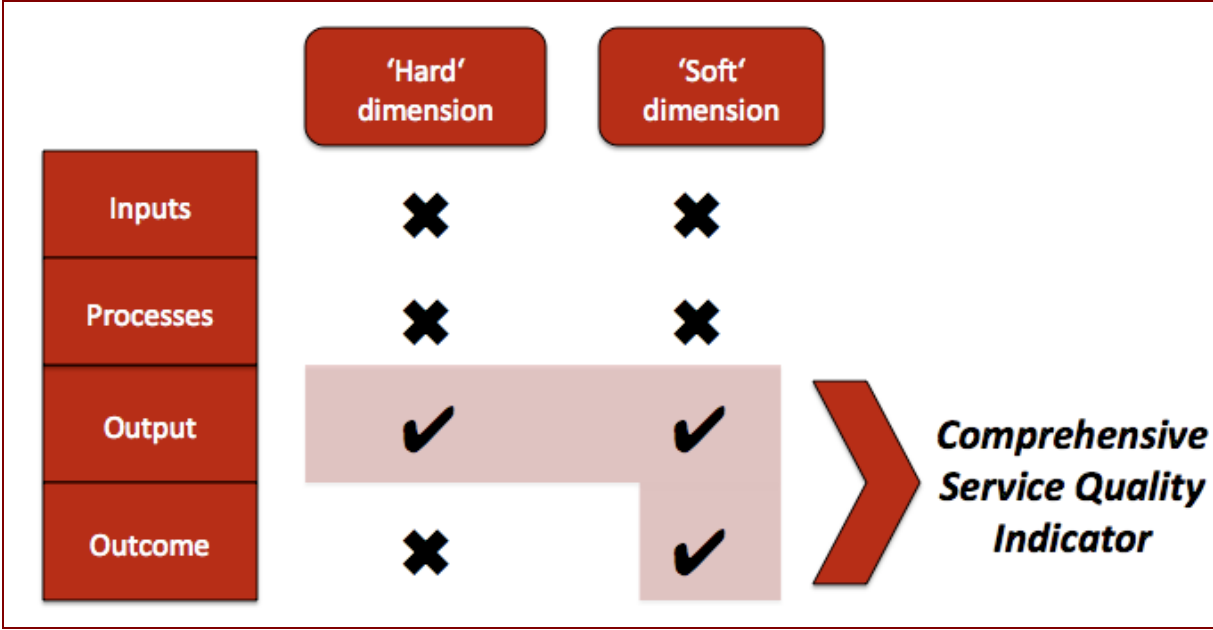


Figure 11 – Operationalization of ‘Hard’ Indicator Framework

<i>Variable</i>	<i>Measured Feature</i>	<i>Operationalization</i>	<i>Data Source</i>	<i>SERVQUAL Dimension</i>	<i>Weight used for composition</i>
openinghours	Opening Hours	Opening hours per week in minutes, plus opening hours on weekends if available.	berlin.de	Reliability	1.2
satopening	Amount of open Saturdays per month	1-4, 1 additional point if the Bürgeramt is open also to customers without required appointment.	berlin.de	Reliability	0.4
disabled	Availability of disabled parking	Dummy Variable; 0 = no, 1 = yes	berlin.de	Responsiveness	0.8
access	Amount of public transport means to reach the Buergeramt	Dummy Variable; 0 = just Bus or Tram; 1 = S-Bahn or UBahn + Bus and/or Tram	bvg.de	Responsiveness	0.8
distance	Distance to public transport stops	average distance to bus/tram, ubahn and sbahn stops in meters	bvg.de	Responsiveness	0.6
picture	Availability of picture taking within the Buergeramt	Dummy Variable; 1=yes, 0= no	Bürgeramt websites	Responsiveness	0.4
surveypossible	Availability of a customer feedback survey within the Buergeramt	Dummy Variable; 1=yes, 0= no	Bürgeramt websites	Responsiveness	0.4
avrgwaiting	Waiting time	Average waiting time per consumer in minutes (in 2007, per district)	Masser et al. 2008: 103.	Reliability	1.4
percwaiting	Amount of customers being served ‘instantly’	Percentage of people waiting less than 10minutes (in 2007, per district)	Masser et al. 2008: 35	Reliability	0.6
avrghandling	Amount of time it takes to serve a customer.	Average time for handling one customer (in 2007, per district).	Masser et al. 2008: 12	Reliability	-
avrgservtime	Amount of time it takes to receive relevant documents via mail	Average time to get “postalische Melderegisterauskünfte” in days (in 2005, per district)	Klages 2006: 36	Reliability	1.4

Figure 12 – Methodology “Test”: Our Aggregation framework compared with a Factor Analysis aggregation Framework (IPF, 1 factor extracted)

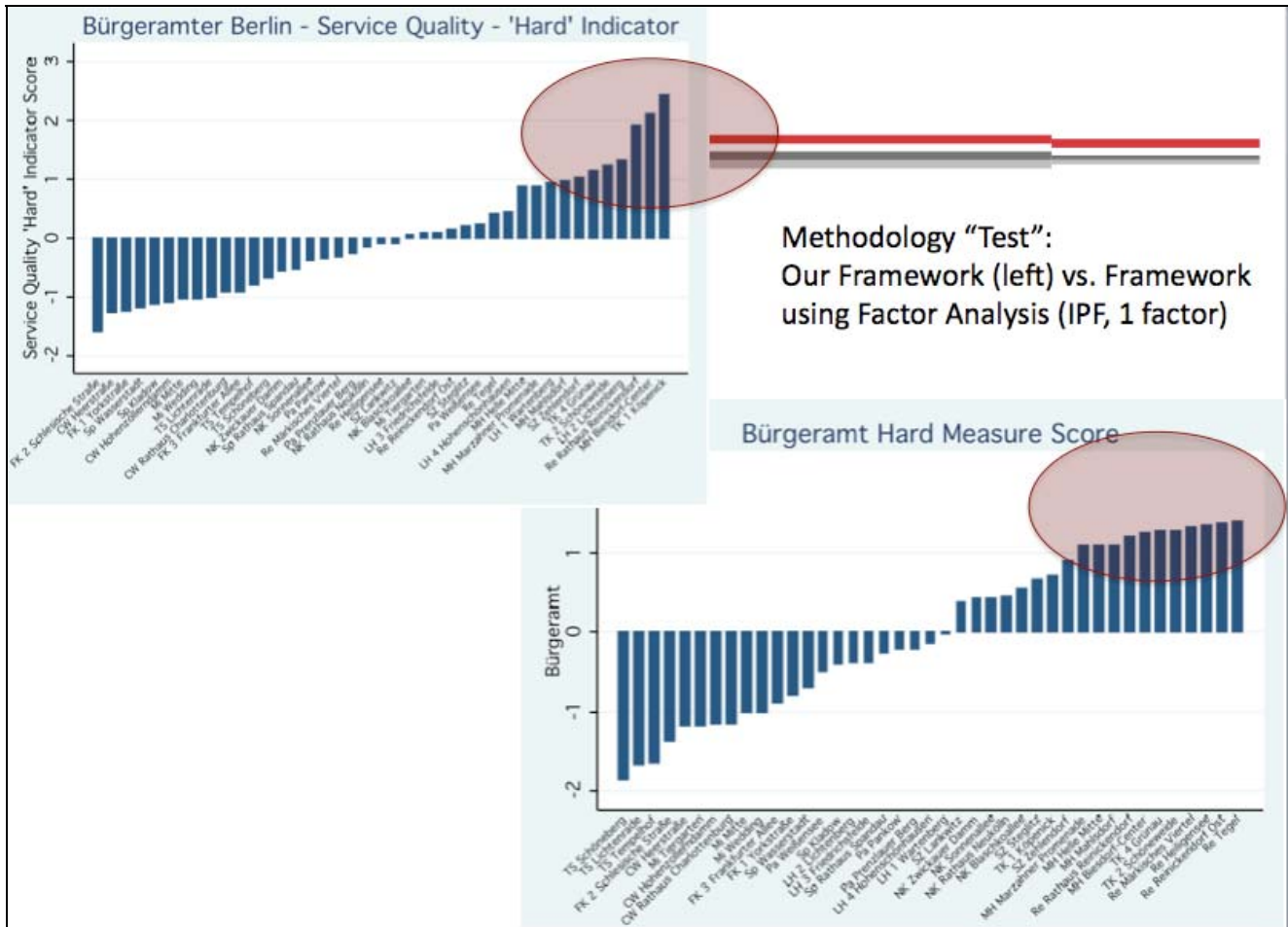


Figure 13 - Methodology "Test" - Difference between Original Indicator Score and Indicator Score with change in weights (changed weights of average service & waiting time by 0.2)

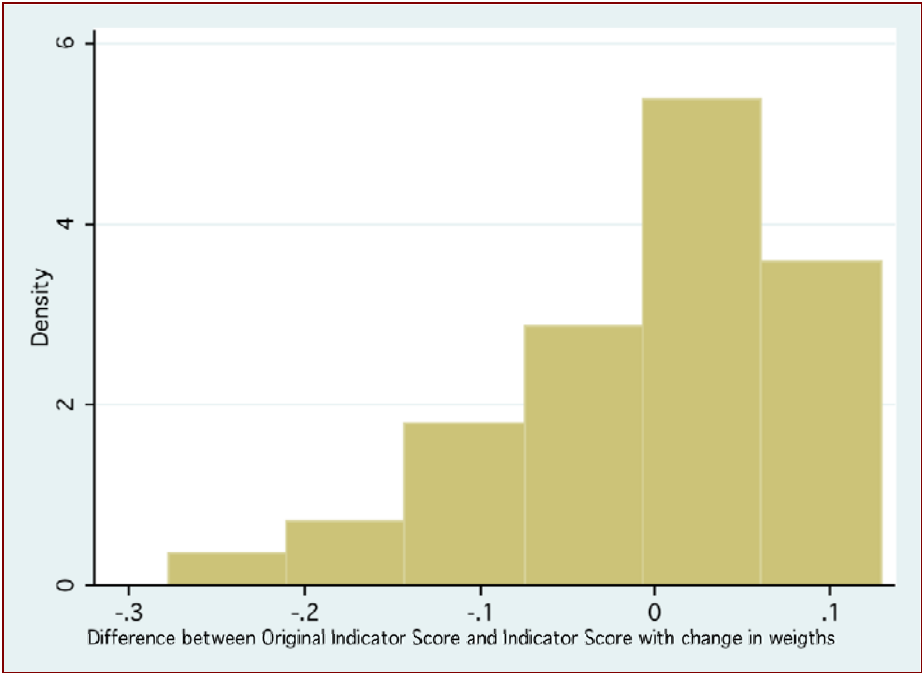


Figure 14 – Scree Plot with the Soft Service Quality Indicator score from our pretest and the age of the interviewees

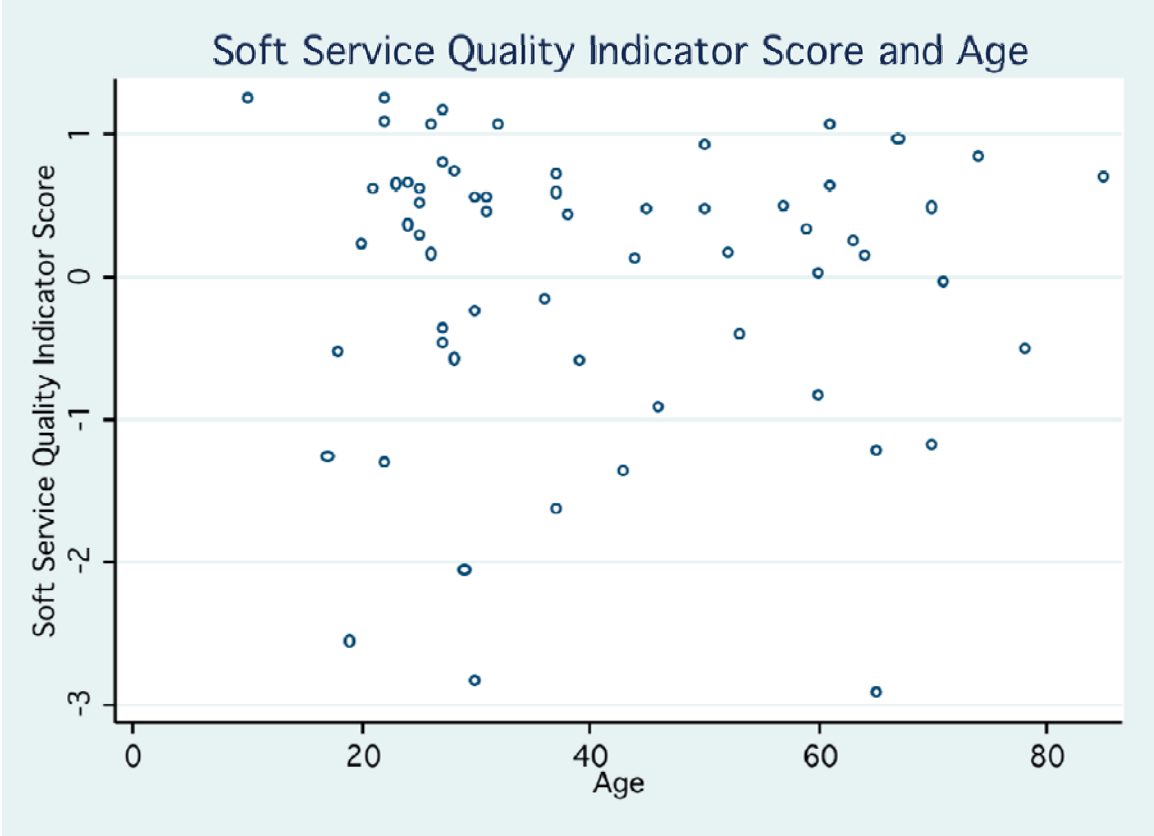


Figure 15 – Bar plot displaying education in years and service quality score (from our pretest)

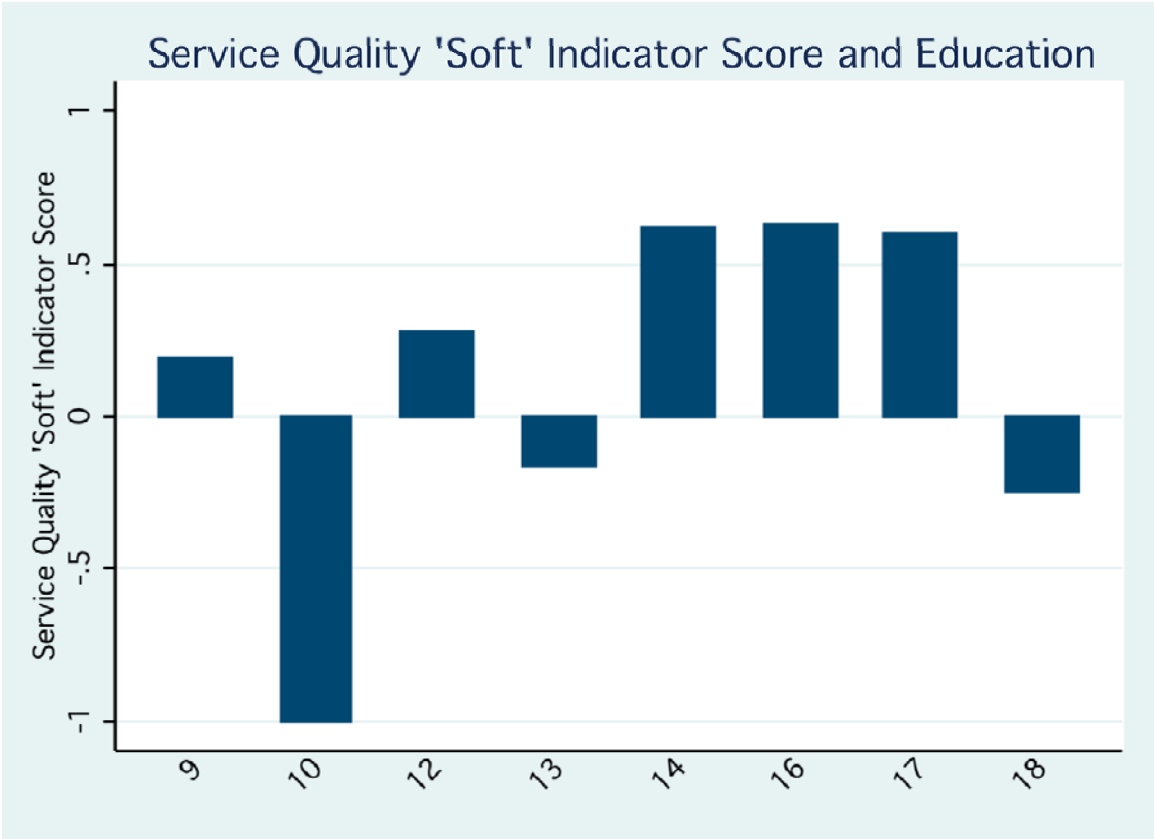


Figure 16 – ‘Soft’ Service Quality Indicator score and waiting time as stated by the interviewees



Figure 17 - *Bürgeramt* 'Hard' Indicator Score – District Overview

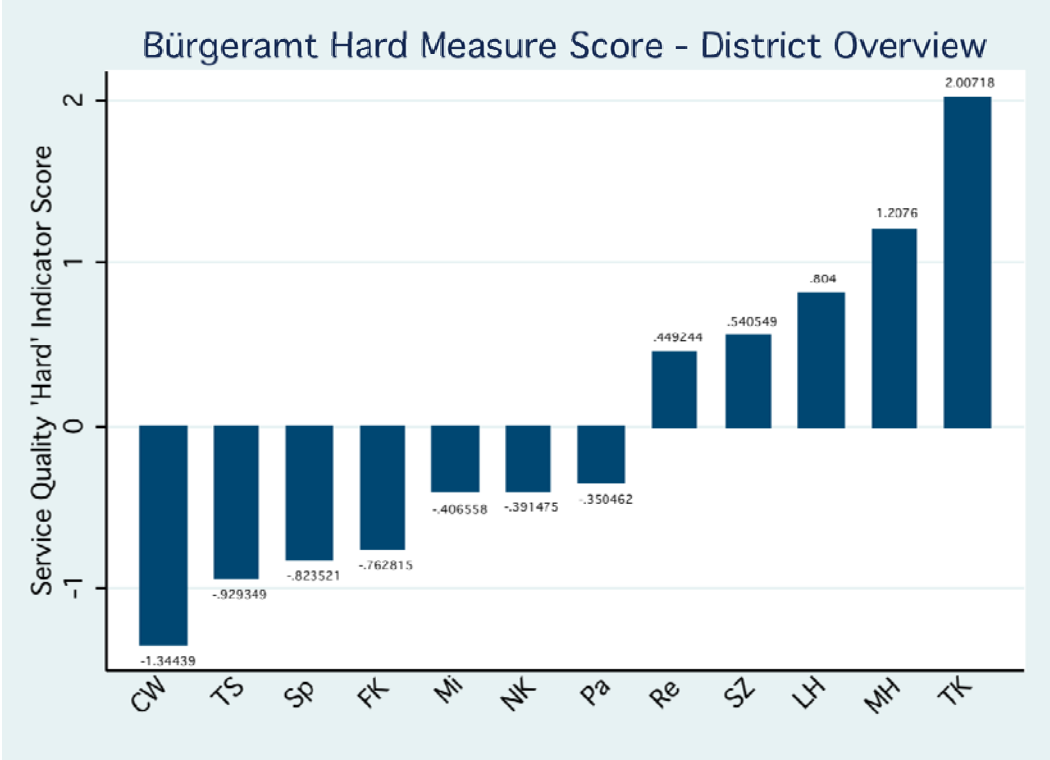


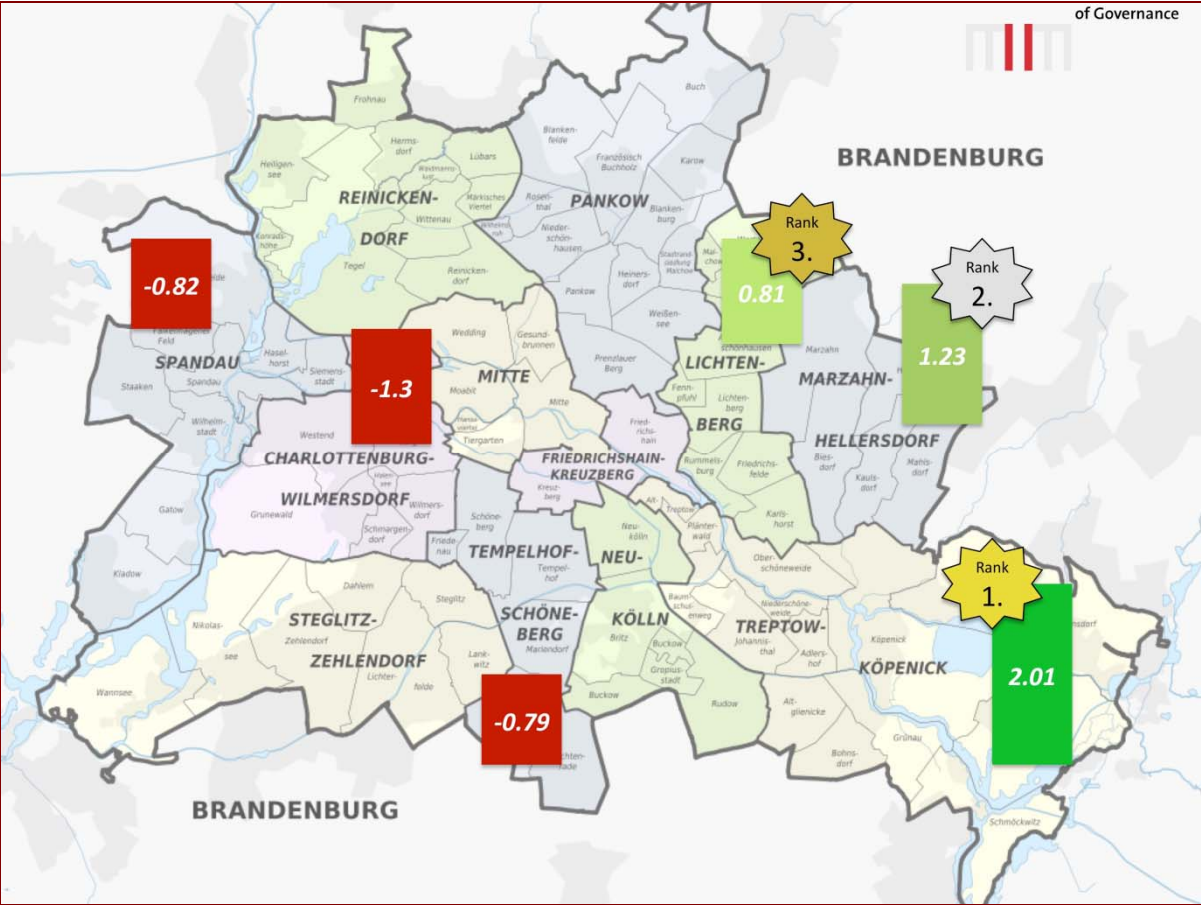
Figure 18 - Dataset for the 'Hard' Service Quality Indicator (including all the data used; TK1 as the “champion” is shown in green)

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Buergeramt	openinghours	satopening	disabled	access	distance	picture	surveypossible	avrgwaiting	percwaiting	avrghandling	avrgservtime	fulltimeper1000
3	CW Rathaus Charlottenburg	31	0	0	1	100	0	0	37	33	9	5.7	0.32
4	CW Hohenzollerndamm	31	0	0	1	200	0	0	37	33	9	5.7	0.32
5	CW Heerstraße	31	0	0	1	275	0	0	37	33	9	5.7	0.32
6	FK 1 Yorkstraße	35	0	1	1	150	0	0	32	29	9	28.5	0.21
7	FK 2 Schlesische Straße	35	0	0	0	50	0	0	32	29	9	28.5	0.21
8	FK 3 Frankfurter Allee	37	2	1	1	300	0	0	32	29	9	28.5	0.21
9	LH 1 Wartenberg	33	0	1	1	150	1	0	24	42	9	1	0.17
10	LH 2 Lichtenberg	35	4	1	1	350	1	0	24	42	9	1	0.17
11	LH 3 Friedrichsfelde	33	0	0	1	416	1	0	24	42	9	1	0.17
12	LH 4 Hohenschönhausen	33	0	1	0	175	0	0	24	42	9	1	0.17
13	MH Blesdorf-Center	35	4	1	1	350	0	1	13	73	9	3.5	0.23
14	MH Helle Mitte	31	0	1	0	200	0	1	13	73	9	3.5	0.23
15	MH Mahlsdorf	31	0	0	1	150	0	1	13	73	9	3.5	0.23
16	MH Marzahnner Promenade	31	0	0	1	200	0	1	13	73	9	3.5	0.23
17	Mi Mitte	31	0	1	1	150	0	0	38	31	9	9.2	0.21
18	Mi Tiergarten	35	4	1	1	225	0	0	38	31	9	9.2	0.21
19	Mi Wedding	31	0	1	1	150	0	0	38	31	9	9.2	0.21
20	NK Rathaus Neukölln	31	0	0	1	100	0	0	31	56	9	3	0.18
21	NK Sonnenallee	31	0	0	1	200	0	0	31	56	9	3	0.18
22	NK Blaschkoallee	31	0	1	1	200	1	0	31	56	9	3	0.18
23	NK Zwickauer Damm	31	0	0	1	300	0	0	31	56	9	3	0.18
24	Pa Pankow	31	0	1	1	400	0	0	23	40	10	3.3	0.16
25	Pa Weißensee	32	1	1	0	150	0	0	23	40	10	3.3	0.16
26	Pa Prenzlauer Berg	31	0	1	1	350	0	0	23	40	10	3.3	0.16
27	Re Tegel	31	0	0	1	100	1	0	18	70	8	8	0.19
28	Re Reinickendorf Ost	31	0	0	1	225	1	0	18	70	8	8	0.19
29	Re Rathaus Reinickendorf	36	5	0	1	150	1	0	18	70	8	8	0.19
30	Re Märkisches Viertel	31	0	0	1	450	1	0	18	70	8	8	0.19
31	Re Heiligensee	31	0	0	1	350	1	0	18	70	8	8	0.19
32	Sp Rathaus Spandau	32	2	1	1	200	0	0	45	47	9	4.5	0.21
33	Sp Kiadow	31	0	1	0	150	0	0	45	47	9	4.5	0.21
34	Sp Wasserstadt	31	0	0	0	50	0	0	45	47	9	4.5	0.21
35	SZ Steglitz	31	0	0	1	100	0	0	25	57	8	2.3	0.21
36	SZ Zehlendorf	33	2	1	1	150	0	0	25	57	8	2.3	0.21
37	SZ Lankwitz	31	0	0	0	100	0	0	25	57	8	2.3	0.21
38	TS Schöneberg	31	0	1	1	150	1	0	35	19	8	5.6	0.21
39	TS Tempelhof	31	0	1	1	150	0	0	35	19	8	5.6	0.21
40	TS Lichtenrade	31	0	1	1	250	0	0	35	19	8	5.6	0.21
41	TK 1 Köpenick	36	5	0	0	100	0	0	8	60	8	1	0.18
42	TK 2 Schöneweide	32	0	0	1	150	0	0	8	60	8	1	0.18
43	TK 4 Grünau	32	0	0	1	200	0	0	8	60	8	1	0.18

Figure 19 – Specific hard and soft measure scores from our pretests

District	Hard Indicator	Soft Indicator
CW	-1.348792	
FK	-0.7627021	
LH	0.8162766	0.2050056
MH	1.150589	
Mi	-0.4036794	-0.3240259
NK	-0.388479	
Pa	-0.3471474	
Re	0.4587671	
Sp	-0.8238795	
SZ	0.5507801	0.0084263
TS	-0.9305292	
TK	2.028796	

Figure 20 – Possibility of Mapping Indicator Scores (using the real data from our ‘hard’ measure pretest; source of the background map: Wikipedia)



The information and the opinions presented in this paper represent those of the author(s) and not of the editors or of the Hertie School of Governance. The author(s) take(s) full responsibility for the information and the opinions presented.

Hertie Student Paper Series is an online publication series of
Hertie School of Governance
Friedrichstraße 180
10117 Berlin, Germany

Telephone +49 (0)30 259219-0

Web: www.hertie-school.org/facultyandresearch/publications/hertie-school-student-paper-series

Editors: Stéphanie Novak (novak@hertie-school.org)

Ariane Goetz (goet5180@mylaurier.ca)