TECHNICAL REVIEW OF THE GCRO QOL SURVEYS: SYNTHESIS REPORT

Mark Orkin PhD

Report

This document synthesises the findings of the GCRO Quality of Life Survey ten year technical review process. It provides the GCRO with a series of recommendations to take forward into the implementation of future Quality of Life surveys.

Written by: Mark Orkin PhD

Published by the Gauteng City-Region Observatory (GCRO), a partnership of the University of Johannesburg, the University of the Witwatersrand, Johannesburg, the Gauteng Provincial Government and organised local government in Gauteng (SALGA).

Recommended citation: Orkin, M. (2020) Technical review of the GCRO QoL surveys: Synthesis report. Johannesburg: Gauteng City-Region Observatory. Available at https://gcro.ac.za/research/project/detail/quality-life-survey-10-year-review/.

TABLE OF CONTENTS

Executive summary of recommendations

The Gauteng City-Region Observatory (GCRO) conducted an externally-chaired technical review of its flagship Quality of Life survey (QoL) to mark the ten year anniversary of this unique survey. This report provides an account of key challenges and lessons identified through the ambitious review process – including the documented revisiting of past experiences, expert workshops, and interviews of providers and participants – and attempts assessments, a synthesis, and recommendations. It focusses on four key areas: sampling and weighting, survey management, questionnaire design, and index construction; and provides recommendations for the GCRO in each.

This summary focuses on the recommendations. After some background to QoL, the first three sections each gather the major recommendations at the outset, followed by more detailed provisions. The last section of this summary is slightly different, briefly outlining the description in the main document of how the index construction was actually re-worked as the recommendation for the future.

So the recommendations in these eight pages can be read on their own. But for their rationales, the reader should not escape the main text. The QoL is a mighty, arduous, costly, multi-faceted and intricate undertaking, of great social and political import. The detailed basis for its revised implementation, feasibly for the next decade, warrants careful attention by the decision-makers concerned.

I Background

QoL is a household survey, covering Gauteng province, and has been run every two years since 2009. Sample sizes have grown over time, with the last three iterations of the survey including over 24 000 respondents. Questionnaire length has also increased over time, reaching 248 items in QoL V (2017/18), with a core of 120 questions that has been maintained over time. Content includes demographics, access to and satisfaction with basic services, political and community satisfaction and engagement, health, safety, transport, and various other topics based on government and academic interests. Data is released following a high-profile launch, historically hosted by the Gauteng Premier. It is extensively used by parts of provincial and local government, for whom GCRO also undertakes bespoke analyses and reports. The GCRO makes the data freely available for non-commercial use, and additionally conducts and widely disseminates its own analysis.

The spatially fine-grained analysis allowed by the very large sample size, the continuity of core questions, and the high data quality combine to make QoL uniquely useful. However, achieving these very attributes has come to impose unsustainable demands on both the GCRO and service providers. While the report examines sampling, field management and questionnaire design and indexing in turn, these issues are inextricably linked, and the problems arising have combined in a "perfect storm" in recent QoLs. In essence, attempting to complete a very large survey, in a short time period, in difficult field circumstances, and to a very high level of quality, resulted in service providers adopting inappropriate methods in the fieldwork or data management. These led to appreciable proportions of problematic interviews in the different QoLs. The GCRO

identified these, and required the service provider to resolve them, leading in turn to substantial overruns of time and budget.

In response, given the importance accorded to the survey, the GCRO initiated this technical review process, providing an opportunity to develop an integrated and comprehensive strategy to ensure the sustainability of the survey into the future, while protecting in its continuity and enhancing its value.

The review process entailed three day-long expert workshops, each attended by a small group of survey specialists. Workshops were preceded by an international and local scoping exercise, a number of practitioner and key informant interviews, and preparation by the GCRO of comprehensive briefing documents. In addition, two focussed statistical explorations were undertaken, of indexing options and ward characteristics. The author of this Synthesis Report was the Technical Chair of the process, participated in the discussions and nearly every interview (and was furnished with reports for the couple he could not), and participated actively in the statistical investigation and requisite re-engineering.

II Recommendations on sampling

The recommendations developed regarding sampling seek to be statistically sound, but also practical and affordable. While some degree of sampling bias and error is inevitable, these can be mitigated by careful design, well managed implementation, and suitable monitoring; but also in "letting the statistics do its work".

While there have been variations in sampling strategies across the different iterations of QoL, all (except QoL I) have been limited to the Gauteng province, based on face-to-face interviews, have sampled all of Gauteng's wards, and have used some form of multi-level randomised sampling. In addition to the more detailed recommendations covered in subsequent parts of this summary, the synthesis report makes four high-level recommendations regarding QoL sampling into the future.

As the survey is primarily funded by Gauteng Provincial Government (GPG), and due to the complexity of sampling and weighting across provinces, **it is recommended that the survey continues to focus exclusively on Gauteng province**. Notwithstanding substantial difficulties of access, **QoL should continue on a face-to-face basis**, as neither telephonic nor web-based approaches offer a viable alternative at this time for a questionnaire of QoL's scope. These alternative approaches can be tested with sub-samples, and advantage can be taken of internet and social media to support community awareness and appointment-based interviewing. **Sampling should continue to cover all 529 wards in Gauteng province.** This will protect the longitudinal continuity of the dataset, and will continue to allow for the type of fine-grained spatial analysis for which the dataset is particularly valued. Finally, **multistage stratified cluster sampling should be retained**, as clustered sampling in particular reduces cost and fieldworker stress with little loss of precision.

The balance of this section provides more detailed recommendations for each stage of the sampling process, drawing on learnings from previous QoL iterations, as well as the literature.

1. **A fixed sample size per ward should be retained.** This approach was adopted in QoL V and – subject to point (2) below – is statistically sound, providing greater consistency in the precision of estimates across the province. Additionally, it allows for easier scaling of sample size relative to budget. Cost permitting, an ideal ward-level sample size may be in the range of 40-50 interviews. However, the margin of error from even this generous per-ward sample size does not permit analysis at the ward level. Mapping at the ward level should be cognizant of this constraint. Analysis of data for groupings of wards is generally advisable.
2. **For small municipalities, which have comparatively few wards, a municipal-level "floor" or minimum is advisable, implemented through an increased sample per ward.** Otherwise, in municipalities such as Lesedi and Midvaal each with only about a dozen wards, application of the fixed ward-level sample size would result in an unworkably small municipal sample. The resulting large margins of error for these municipalities would make it hard to assess change over time in a meaningful fashion.
3. **Within wards, it is advisable to sample interview locations in clusters of 4-5, suited to completion by a team of fieldworkers in a single vehicle.** The Small Area Level of Statistics South Africa defines demarcated areas of around a hundred households, though with considerable variation. These areas are colloquially called SALs, and offer a suitable lower-level geography for the clusters of interviews. Given the variable sizes of these units, it is appropriate to select the SALs to be visited within a ward using the methodology of "probability proportional to size" of population (PPS). This gives larger units a greater probability of selection. To ensure relevant coverage, the PPS selection of SALs may additionally occur within relevant "strata" in the wards, e.g. predominant population group or dwelling type. Stratification has costs as well as benefits, however, so these must be carefully weighed by the GCRO. A previous supplier to GCRO has been GeoTerraImage (GTI), maintains and updates its database at "enumerator area" (EA) level, the geography at which Stats SA conducts its censuses. However, use of the generally slightly larger SALs is advised for QoL, as this is the lowest level at which Stats SA publicly releases boundaries and population data.
4. **For selection of individual interview points (visiting points) in the SALs, it is advised that a GIS-based dataset comprising all dwelling units and their coordinates is used as a** sampling frame. This was done in both QoL IV and V. It may be implemented using a dataset either from Stats SA or GTI. Given the massive amounts of detailed information and continually changing information, either will have imperfections. This will need to be managed by GCRO's own GIS capability in a practical fashion.
5. **A systematised process of random selection of a dwelling unit at a visiting point, and of the adult respondent at the household level, must be retained.** On arrival at a visiting point, fieldworkers will often be confronted by the reality of multiple dwelling units. They will need to list all dwellings, and then randomly select the appropriate dwelling using a Kish grid. For reasons of representivity and historical continuity, and given the extent of individual opinion content in the QoL questionnaire, interviews must then be done with a randomly selected adult resident in the selected dwelling unit. Again, the fieldworker will need to do this using an automated Kish grid, following a listing process. Control of these two levels of the process is greatly improved by their being programmed into fieldworkers' "tablets".
6. **Only one revisit should be made to secure an interview at a visiting point or with a sampled individual. If unsuccessful, the interview should be replaced by an interview at another predetermined visiting point**. While revisits are important to

minimise sampling bias, they are costly, and the international literature is reassuring that multiple revisits appear to provide little additional benefit. Consequently, limited and pragmatic management of revisits in field can reduce survey cost and difficulty. Provision of GIS-based predetermined replacement points, instead of the earlier *ad-hoc* selection of replacement points in the field, now reduces the risk of bias.

7. The realised sample needs to be weighted to match adult population distributions on selected demographics. **The process was by handled by iterative "raking" in the last two QoLs, and should be retained.** "Raking" is necessary because Stats SA does not provide the demographic statistics down to the ward level required by QoL sampling. **Additionally, the hybrid approach hitherto employed in QoLs, in which the randomly selected adult informant individuals are taken to represent households, is reasonable,** given the usage of QoL data at individual and household levels. It will additionally protect continuity. Should a different approach be deemed preferable at some time, backward reweighting of historical datasets as well as the current survey iteration would be involved.

III Recommendations on QoL survey management

At the heart of effective QoL survey management is a constructive and positive working relationship between the implementing organisation and the GCRO. Critically, GCRO must not – as it has sought to do in the most recent QoLs – micro-manage the service provider by scrutinising numerous interviews and seeking immediate resolution. Rather, the GCRO must allow the service provider initially to identify and resolve issues themselves, and report systematically on what they have encountered and done. Only when this is not appropriately done should GCRO step in. *In other words, GCRO's task is not to monitor the fieldwork, but to monitor the adequacy of the provider's doing so, and the adequacy of their corrective actions.*

The report first makes three high-level recommendations regarding survey management. The GCRO does not have either the staffing or the infrastructure to manage large-scale fieldwork – indeed, it would not be an appropriate operation for a research-oriented institution. **So it is appropriate to continue to appoint an external, private sector provider for project implementation**. There is consensus across multiple service providers that 125 fieldworkers is the maximum that can be reliably managed and supported in-field at any time. This figure, together with the desired sample size, should guide the planned duration of QoL fieldwork. **Six months should be allocated to fieldwork**, for a sample size of 18-20 000. The bias introduced by a longer fieldwork period can be minimised by ensuring fieldwork proceeds evenly across all parts of the province. Planning in advance for orderly fieldwork is much less problematic than having a shorter fieldwork period run substantially over time. Finally, given the difficulty of face-to-face fieldwork, **it is important that fieldworkers are decently remunerated, and working in a small and supportive teams. The industry standard of 4-5 fieldworkers travelling together in a small car with a dedicated supervisor should be employed.**

This section also provides a review of the GCRO's experiences with QoL to date, drawing out a series of additional more detailed recommendations. These are largely in line with standard survey best practice. In considering a project of the scope and scale of QoL, it is important to bear in mind that a degree of imperfection in implementation is unavoidable – but can be managed through weighting of the results, and by the aggregation of data required in any event.

Additionally, to repeat, a transparent and trusting working relationship between the GCRO and the service provider is critical: dealing effectively with unavoidable project contingencies benefits *both* parties.

1. The tender Terms of Reference (ToR) should stipulate that the GCRO seeks to appoint an **well-qualified private-sector service provider, i.e. with substantial prior extensive experience of similarly large-scale and rigorous probability-sample surveys in South Africa**. Clear specification of expectations in the ToR around staffing and implementation must be accompanied by a procurement process which enables an adequately detailed interrogation, face-to-face as well as on paper, of the proposals received.
2. **The software used must be adequate to the accurate implementation on tablets of the lengthy questionnaire, along with the collection of geo-coordinates and the random selection of households and respondents**. Appropriate translation of the questionnaire into additional local languages, together with validation, is essential.
3. **Fieldworker training should last at least 5 days, and include "soft skills" and testing of trainees to ensure competence.** Fieldwork experience in South Africa is that broad language and demographic matching of interviewers to respondents, where possible, improves the facility and quality of the interviews.
4. Piloting of the questionnaire should be implemented at three levels: (a) "behind the glass" observation of one-on-one interviews; (b) small-scale in-field with detailed feedback; and (c) large-scale in field with quantified assessment. **Project timelines must allow for adequate piloting**, as well as the resolution of issues identified, and consequent re-programming.
5. Sample design in line with the recommendations in the previous section.
6. Fieldwork planning must incorporate the need for **extensive efforts to achieve publicity and community entry; experienced supervisors; availability of transport; a telephonic support desk; appropriate management of substitutions; and fieldworker adaptability**.
7. Back-checks of fieldwork should include both a proportion of **telephonic back-checks, as well as immediate, random call-backs by supervisors,** and a small but spread proportion by the GCRO itself.
8. **Automated and manual back-office quality control processes in reviewing, cleaning and aggregating the data should be implemented by the service provider**. GCRO may also conduct some, primarily automated, back-office quality control processes to monitor the service provider's identification and resolution of challenges.

Finally, this section makes recommendations for internal survey management at the GCRO. While the GCRO's level of involvement in QoL survey management has grown over time, staffing and management has been ad-hoc and constrained. The need is for for engagement in procurement, preparatory work, interrogating the provider's field-work monitoring, and then data analysis and dissemination. This means that **there is continuous QoL work throughout an entire two year cycle. In addition to a full-time lead researcher, reporting directly to the Executive Director, two additional full-time posts are recommended to support him/her**: an operations manager to tackle specified organisational needs, and an assistant researcher. Within the GCRO, the establishment of an advisory grouping, to support and advise the lead researcher, is also suggested.

IV Recommendations on QoL questionnaire design and index construction

IV (a) Questionnaire design

The QoL questionnaire has historically been both long and broad. This combination holds enormous value for data users, but increases the difficulty and cost of fieldwork. The recommendations in this section aim to maintain the value of the data for a range of users while simultaneously minimising the burden on respondents and fieldworkers. They particularly take up a more appropriate questionnaire length, and how this may be attained.

The questionnaire has, over time, retained a core of approximately 120 items. Half of these contributed to the QoL index, a quarter consists of key demographic information, and the final quarter initially comprised additional content of core interest to government, such as delivery across a range of services, user satisfaction, and socio-political issues. Over time, however, much additional content has been added and little removed, yielding a total of 248 items by QoL V. Respondent and fieldworker fatigue, which compromise accuracy and decency, are both key considerations in determining appropriate questionnaire length. **A reduction in questionnaire length is recommended, to a maximum of 200 questions, or even 160-180, especially if there are resource constraints.**

Question reduction **could be attained through a reduction in core content, combined with the use of rotating modules in the balance of the questionnaire.** Revision of the QoL index (see below), and the elimination of rarely used or uninformative core variables would reduce the core to 90 questions. The balance could be distributed amongst rotating modules, user-pay modules, and topical questions of interest to internal or external researchers.

The **establishment of a clear process to guide decision-making regarding selection of questionnaire content of academic interest is recommended.** This should include a competitive application process, adjudicated by a high-level panel, and applied to GCRO researchers as well. User-pay modules should be realistically costed *pro rata,* and the GCRO should continue to work closely with government partners to develop and manage government content. Full development of all these processes will not be possible for QoL VI, but a strong start can be made, especially if necessitated by reduced grants from provincial government and the larger municipalities.

IV (a) Index construction

The Quality of Life Index based on the successive QoL surveys has been one of the key tools in summarising and communicating survey results to stakeholders and the public. The index construction has remained largely unchanged since QoL I: a simplified coding of the responses (and missing answers) to 58 selected variables ; the intuitive grouping of these variables into 10 dimensions; the generating of a summative score across equally weighted questions, for each dimension; and then granting the dimensions equal weight in a composite QoL score. This review assessed the soundness of the index as originally constructed and subsequently sustained, in relation to well-established international practice then and now. It also revisited the primary data of the last three QoLs. Deficiencies were identified in each of these stages.

The review accordingly proceeded differently in this fourth aspect of its process, towards establishing a likely way forward for the indexing of QoL VI and subsequently. It actually undertook the re-working of the index construction, starting with the QoL V data. The exercise was conducted over several months by the QoL lead researcher, a GCRO statistician, a consultant emeritus professor of statistics, and the Technical Chair. They were informed by broad discussion at the third expert workshop, a detailed seminar with a UJ academic who is a widely published international expert in indexing, the literature, and considerable experimentation among different statistical approaches to particular challenges.

The sequence and the results are presented in some statistical detail in section 4.2 of the main report. Since they are quite technical, and lengthy, they are briefly outlined here. The revised approach retains the nuance of the original data (notably by not collapsing five-point scales into "yes" or "no"). It statistically imputes missing values correctly. It then applies the widely used, applicable statistical technique of Exploratory Factor Analysis to the survey data, in lieu of subjective judgement, to guide the next three stages: (a) the grouping of the indicator questions into the indicated number of dimensions, seven rather than ten (plus a halving of the necessary indicators); (b) weighting the contribution of the indicators to each dimension; and (c) finally weighting the aggregating of the dimensions according to their empirical salience, rather than just equally, into the composite QoL index indicator.

Backwards calculation using the data from QoL IV and QoL III (which also had large and geo-referenced samples) demonstrated consistently good performance of the revised index over time, and the empirical 7-dimensional model outperforms the intuitive 10-dimensional model in the applicable statistical tests. **It is therefore recommended that the GCRO make use of this revised 7-dimension index, drawing on 33 variables, into the future.** The dimensions are Infrastructure and services; Socio-economic status; Government satisfaction; Life satisfaction; Safety; Health status; and Political engagement.

As noted at the end of this section in the report, "**the communicating of the new index from QoL VI onwards will need to be considered**. However, the familiar titles and spread of the revised seven dimensions, in relation to the original ten, will be reassuring. As previously, the main interest will be in the by-municipality scoring on the different dimensions, and overall. Comparisons with QoL V, to the extent necessary, will also focus on the revised dimensions."

1 Introduction

1.1 Background to GCRO's Quality of Life surveys (QoLs)

The Gauteng City-Region Observatory (GCRO) is a partnership between the University of Witwatersrand, the University of Johannesburg, Gauteng Provincial Government (GPG), and organised local government (SALGA). Its flagship project is the Quality of Life Survey (or QoL), a household survey which has been conducted every two years hitherto: QoL I (2009), QoL II (2011), QoL III (2013/14), QoL IV (2015/16), and QoL V (2017/18). QoL I covered some areas of the Gauteng City-Region that extend into adjacent provinces; the QoLs since then have been confined to the Gauteng province. The survey covers the province's nine constituent municipalities. They altogether include, after the last boundary revision in 2016, 529 electoral wards. The surveys have been funded in large part by the GPG, through the GCRO's core grant, with additional funding for more recent survey iterations coming from various metropolitan municipalities and provincial government departments.

Following smaller samples in QoLs I and II, the sample increased from to between 24,000 and 30,000 in QoLs III through V. This very large sample has meant that, in the last three QoLs, a minimum of 30 interviews were planned per ward, with up to 60, and at times over 100, in particular prioritised wards. The fieldwork has been outsourced each time to a private-sector provider and quality has been monitored by a small GCRO team. The GCRO team have been provided with increasingly detailed and up-to-date geo-referenced information, allowing assessment of how carefully the fieldworkers are getting to the actually sampled dwellings, and how carefully they are conducting the respondent selection and face-to-face interviews.

The core of the questionnaire is some 120 questions. Sixty were indexed from the outset into an overall QoL index covering ten dimensions of quality of life, ranging from household living conditions, community engagement, and health, through to personal safety to socio-political issues. The balance comprises a mix of questions covering areas including demographics, access to and satisfaction with basic services, household resources, and additional attitudinal questions. These questions, and particularly those forming part of the QoL index, have been sustained with only minor variations in a few instances. This continuity, coupled with the ward-level detail afforded by the sample size and distribution, has provided a unique resource for analysis and the monitoring of trends over time. Additional more detailed content, for example on highly salient topics such as transport or health, is carried from time to time. Questions have been more easily added than removed: the overall number of questions in the questionnaire has increased from 187 to 248 over the decade (although not all will be applicable to different respondents).

The results, notably of the QoL overall index and sub-index scores, are given great prominence at a high-profile launch event on completion of each survey. The launch is attended by the GPG, and historically the Gauteng Premier has spoken at the event. The GPG further disseminates results after the launch, drawing on them for publicity and also to guide decision-making. In particular, the GPG attends to the results disaggregated by municipality, indicating the municipalities' changing performance over time in delivering services.

The successive datasets have also been analysed in depth under various themes by GCRO or by researchers it has commissioned. This has yielded a range of occasional Data Briefs, substantial

Research Reports, and the occasional "Provocation", that are widely circulated to a 6000-entry mailing list. GCRO also undertakes bespoke analyses for the GPG or particular municipalities or government departments. In addition, the datasets are made publicly available for analysis by academics and other non-commercial researchers in South Africa and abroad, who have published their analyses independently and have also used the data for teaching students. The minimum sample size of 30 interviews per ward has allowed detailed and revealing GIS maps on particular themes, such as income inequality or housing.

1.2 A "perfect storm" of fieldwork management, and the ensuing Review process

The ward-level resolution enabled by the vast sample (by way of comparison, the 30,000 sample of Stats SA's General Household Survey has to cover all nine provinces), the continuity of the core questions, and the data quality achieved through the close monitoring, jointly imply that the QoLs are a unique and uniquely useful resource. However, paradoxically, achieving these very attributes has come to impose unsustainable demands upon both the successive service providers and the GCRO management, and particularly the GCRO researcher assigned to manage and support the QoL. Fieldwork has run way over deadline, providers have incurred financial or staff losses, and other GCRO research undertakings have suffered. After each iteration, both the particular service provider and the management and team members of GCRO have been cautious to come to the table for a second helping.

Key problems will be considered in some detail in the following three sections, on sampling, field management, and questionnaire design and indexing. The challenge will be that the issues, which occurred in Qols IV and V in particular, are inextricably linked across these three topics, creating a "perfect storm". In essence: in order to attempt so large a sample in the specified three-month period with stringent site and respondent identification, great difficulties of access in certain milieux, and a very long questionnaire the providers took recourse to outsourced or individualised fieldwork arrangements in QoLs IV and V, which resulted in unacceptable levels of mis-sampled or fictional interviews. Through the daily, real-time, GIS-based monitoring of fieldworkers, these problems became immediately evident to GCRO, who required that the defective interviews be redone, particularly in QoL V. This pitched the lead researcher and her managers into time-consuming and adversarial -relations with the providers, while the providers missed deadlines and incurred unbudgeted costs from having to repeat an appreciable proportion of discarded interviews. (In QoL III the key problem was a confusion of records in the database; but the need for repeat interviews, unpleasantness, delays and financial implications were similar.)

Given the importance of the QoLs, and recognising that the interlinkage among these problems would require an integrated and comprehensive solution, the GCRO Director Professor Rob Moore and his Management Committee resolved to undertake a detailed technical review of the first decade of QoLs, spanning the five implementations. The review process was convened and serviced by Dr Julia de Kadt, with the experience of having managed aspects of QoL V. The aim is:

> *to support the GCRO in re-balancing sample size and distribution, field processes, questionnaire composition, duration of data collection, and available resources in such a way*

as to facilitate survey sustainability over the next ten year, while also protecting as far as possible the continuity and value of the survey.[1]

The review process has been thorough. It has centred around three day-long workshops, in January, February and May 2019. These events were facilitated by Dr Tara Polzer-Ngwato, and including invited survey specialists from around the country. They were preceded by a scoping exercise and practitioner interviews, served by extensive prior briefing documents and presentations from GCRO and some of the specialists, and followed by written summaries of the discussions for the author. Concurrently, detailed interviews were held with the last four QoL fieldwork service providers, three GCRO researchers who had managed QoLs, and GCRO Research Director Mr Graeme Gotz who had played key roles in all five QoLs. The author of this Synthesis Report was the Technical Chair of the process, participated in the discussion and nearly every interview, and was furnished with reports for the couple he did not. The participants in the interviews and events are recorded in Appendix G.

In addition, two focussed statistical explorations were undertaken, of indexing options and ward characteristics. The former has proved unexpectedly far reaching, having to correct data coding, and reconstitute the poorly intra-correlated sub-indexes that were installed for QoL I and sustained subsequently.

An essential feature of the review has been the high quality of the substantial briefing documents for participants in the three workshops, prepared by Dr de Kadt with assistance from Ms Kiera Crowe-Pettersson. Lightly edited versions of the respective briefs on sampling and on field management are included as Appendices A and B (respectively supporting Sections 2 and 3 of this Report). They are for readers who may wish more historical detail on, and detailed diagnoses of, the approaches adopted by the service providers.

A brief on questionnaire design and indexing is also attached as Appendix C, supporting Section 4. These two aspects of the QoLs have historically been handled within GCRO itself, rather than through a service provider, but have also received thorough attention in the Review. Relevant portions or insights from all the briefs will be recognised as closely informing the discussion. In particular, Table 1 in the sampling brief (Appendix A), was invaluable in summarising for comparison the key sampling and fieldwork arrangements for the successive QoLs.

A final word is necessary on the focus of this particular Review, as defined on p. 8 above. In seeking to improving the sustainability of QoLs, discussion of the sampling and fieldwork of the QoLs, and the scope of the questionnaires – in the Sections 2, 3, and 4.1 – perforce cover the overall survey endeavour. However, over the decade the *coverage of themes* in the survey has been greatly increased, by greatly expanding the number of questions put to respondents. This is discussed in Section 4. Accordingly, the annual analysis and reporting of QoL survey results much more extensive: as presented in the QoL Report prepared for each launch of the results, and also in many of GCRO's other research products. The take-up by users, of these products and also the data, has grown correspondingly. The scope, quality and use of these wider outputs are beyond the remit of this Review. They have been considered by the panellists in the five-yearly overall Institutional Review of GCRO a whole, which was also conducted in 2019, as well as through a separate GCRO workshop focusing on aspects of QoL data sharing and use. So this

[1] "GCRO's Quality of Life Survey: Sampling workshop briefing document", p. 1, attached as Appendix A.

Review's consideration of the outputs of QoL can be limited to assessing the prevailing intuitive calculation of the 10-dimension QoL index in Section 4.2.2; and then presenting in Section 4.2.3 the considerable process that was undertaken for the Review, of correcting the data-handling, and deriving and calculating an empirically-based 7-dimension QoL index.

2 Sampling: towards a sound and sustainable solution

Variations in arrangements among previous QoLs have highlighted key sampling issues, involving both prior design and in-field protocols: sampling by wards within municipalities, stratification within wards by SALs (and at times kinds of SALs) to ensure spread, means of identifying sites to be visited, choosing of dwellings and then households within them, and protocols for revisits before substitution. A step forward in one respect, notably e.g. GIS-based sample-site selection, seems often to have been accompanied by a step backwards in another.

The sequence of bullets below draws on the exhaustive details, arranged historically by QoL, of Appendix A and its Table 1. But the argument will be structured instead by considering the key steps in defining and implementing a sound sample, with allusions as necessary to this or that QoL.

2.1 Some relevant numbers in one convenient place

In South Africa, wards are geopolitical subdivisions of municipalities used for electoral purposes. They are delimited by the Municipal Demarcation Board, and each ward receives two seats on the relevant municipal council. In 2016 the number of wards in Gauteng was adjusted upwards from 508 to 529. The numbers of wards in each of the nine municipalities that comprise Gauteng are shown in Table 1 below. One sees that the "big three" municipalities of Johannesburg, Ekurhuleni and Tshwane have upwards of 100 wards each, whereas at the Lesedi and Midvaal have only 13 and 15 wards respectively.

Also shown in the table are the number of small area layers (SALs) per municipality, to a total of 17,840 in Gauteng. Unlike wards, the SALs are a census geography, and not a political geography. From Census 2001 onwards, to assist the protection of respondents' anonymity, Statistics South Africa (Stats SA) no longer released information down to the level of census enumeration areas (EAs; n=80,787 nationwide), but where necessary combined two or three small and adjacent EAs within a particular sub-place, to create an SAL. This is now the lowest level at which census data is released.

Table 1: Wards and SALs within municipalities

Municipality	Wards	SALs
Johannesburg	135	5 800
Ekurhuleni	112	4 610
Tshwane	107	4 524
Sedibeng District		
Emfuleni	45	1 136
Midvaal	15	184
Lesedi	13	187
West Rand District		
Mogale	39	594
Rand West	35	500
Merafong	28	305
Total	529	17 840

These polygons are now colloquially called "SALs". Their distribution by municipality is shown in Table 1. The average is 52 SALs per ward. Excluding 58 SALs that were empty, there were 220 households per SAL on average.[2] SAL boundaries do not align perfectly with ward or municipal boundaries, and there are currently some areas which are not covered by an EA or an SAL, as these were not populated when the current statistical geography was developed.

It's worth noting that the population of Gauteng was 13.4m according to Stats SA's 2016 Community Survey, of whom 9.5m were adults of 18 years old and above: 78.5% Black African, 3.4% Coloured, 2.7% Indian or Asian, and 15.4% White. According to Stats SA's 2019 mid-year population estimates, the total population of Gauteng had increased to 15.2m.

2.2 Initial parameters for QoL VI and perhaps beyond

Sub-section 2.3, following this one, will attempt to argue a coherent choice among the quite technical options for the various sequential steps in devising and administering sampling, such as the number of interviews per ward, whether or not to cluster interviews, the choice between SALs and EAs, and different criteria and methods for substitution of respondents. The argument will draw on GCRO experience - what turned out to be good and bad choices in running past QoLs; the practices in sampling similar large surveys as offered by experts in the sampling workshop; and insights from a couple of handy review articles of the massive literature, as well as methodology documentation from Stats SA. It will avoid equations in favour of what they convey. In some instances, there are options, either course of which would work. The conclusions will then be summarised in six integrated recommendations in section 2.4.

But before tangling with the specifics, some broad parameters can be quite summarily recommended. A parameter is a major consideration that, in the modern jargon, "frames" the argument. Some may endure beyond the next QoL, but others may well be affected by changes as dramatic as those experienced in the last decade, e.g. in GIS information for sample specification in the field, and the use of tablets for questionnaire administration as well as monitoring of

[2] SALs per ward range from 11 to 149; households per SAL range from 3 to 3,696!).

fieldworkers. This should be kept in mind, though I shall for convenience speak of "future QoLs", for which some aspects will surely endure.

2.2.1 The QoL boundaries should continue to be those of Gauteng province

The Gauteng City Region, which is GCRO's formal remit, extends beyond the boundaries of the Gauteng province. QoL I accordingly planned to conduct fieldwork in six hundred wards, a quarter of them in Northwest, Mpumalanga and Free State. However, the subsequent four QoLs have been confined to Gauteng, for sound practical reasons.

Firstly, the core grant of the GCRO, from which the majority of each QoL is funded over two years, is provided by GPG, the Gauteng Provincial Government, not by its neighbours. The funds allocated to QoL are already extremely constrained, despite supplementation by two or three of the larger Gauteng municipalities. Efforts to implement the survey with a budget that is below the market rate for the desired sample size has had various seriously undesirable consequences – even when the survey has been confined to Gauteng. Consequently, coverage of a broader area is financially inadvisable. Secondly, the boundaries of the province are clearly known, which is invaluable for sampling and weighting; whereas the GCR is continually expanding into the neighbouring provinces without specified boundaries. This would vex broad comparability over time, and when it was attempted in QoL I it already required different sample-weighting approaches for within Gauteng and beyond.

2.2.2 QoLs should continue to be conducted by face-to-face interviews

The difficulties of securing access for interviews through the security barriers of blocks of flats and gated communities,[3] as well as through the psychological barriers of respondents in all milieux who are anxious about admitting strangers, have been experienced since QoL I. They are daunting, and are increasing. They were underlined in the field-management workshop that informs Section 3: there is a "hit rate" of successful first approaches in more affluent areas of only 7-14%. Moreover, Stats SA reports its highest refusal rates in the Gauteng province. In more evenly developed countries, cold-call survey interviews are now widely conducted by telephone, and in some countries households may volunteer to complete their census returns on the internet. However, at least three factors make these approaches unworkable for QoL.

Firstly, fully 87% of South African households only use cellphones[4] for which there is no available listing to be used as a sample frame, still less one confined to a specified province. The same applies to internet addresses, with the additional challenge that only 15% of households had in-home access in 2016. Secondly, the questionnaire is much too long to be handled by telephone, unless it were to be ruthlessly confined to a greatly reduced set of QoL index items (identified in Section 4) and essential supporting demographic items. Yet additional modules on topics like transport or health have been requested by the Province or other users and have proved invaluable, and some are worth sustaining, or better rotating – as we shall discuss. Thirdly, the acceptance rate for telephonic cold-calls, and especially response to requests to self-administer

[3] This also has implications for the motivation and management of fieldworkers, noted in section 3.
[4] Statistics South Africa, *General Household Survey*, 2016.

questionnaires, is poor even in developed countries, and is also being experienced by Stats SA in its current experiments.

Possible innovations such as partial telephone or web interviewing can better be tested and assessed on smaller samples as QoL VI unfolds. Where telephone, and the internet via social media, may have more immediate application to QoL is in seeking appointments assisting access to respondents in secured dwellings. This will be mentioned in Section 3.

2.2.3 QoLs should retain the approach of including all Gauteng wards in the sample[5]

Retaining sampling by ward, which has been used in one way or another since QoL II, is a good idea for two main reasons: user-driven and analytical. The most prominent users of the QoLs are the Gauteng province and its constituent municipalities, for whom wards make intuitive sense as being associated with ward councillors. The historical continuity that enables trends to be drawn – allowing perhaps for some backwards reanalysis, following the findings of this Review – is a rare attribute. Finally, analytically, GCRO and other researchers report having found the ward-level data amazingly fruitful, not for inter-ward comparisons (for reasons covered below) but because wards can generally be aggregated in various ways: into relatively distinctive communities as well as municipalities, which can then be characterised and analysed demographically.[6]

Specifically, ward-averaged results provide the data for the highly regarded GCRO analytic maps, in which each of the 529 wards is a "pixel", yielding illuminating geographical insights into spatial distributions of poverty, access to different services, etc. The proviso is that the spread of intervals being visually represented, such as quintiles, is sufficiently broad compared to the margins of error (see below).

It is worth noting that abandoning ward-based sampling would not of itself lead to a substantially reduced overall sample, but merely to a re-allocation of the sample. To get an minimally acceptable margin of error of, say, 2.9% – as shown later in Table 3 – for within-municipality trends and inter-municipality comparisons, one would still need 9 samples of 1,120 interviews each, approximately 10,000 in all. And a 2.0% margin of error for all municipalities would require samples of 2,200 each, a total of nearly 20,000 – which is of the order contemplated for by-ward sampling in any event, subject to funding.

[5] This should more be put more precisely as "…so long as overall sample size permits". Failing this, one would possibly adopt – in the terms to follow – a disproportionate allocation of the affordable number of wards to municipalities (to retain near-comparable precision for the smallest municipalities), followed within each municipality by a PPS sampling of wards, to each of which an equal number interviews would be assigned.
[6] Of course the historical continuity, spatial congruence and homogeneity of wards are less than perfect, but the variations are minor. Some ward boundaries were redrawn in 2016; and a few align imperfectly with census boundaries, especially because the respective authorities may make adjustments at different times. GCRO makes the necessary pragmatic adjustments in its successive datasets, e.g. offering earlier and later sample weights. And some wards span quite diverse SALs (as when an informal settlement may develop adjacent to formal housing).

2.2.4 QoLs should continue to be based upon multistage stratified cluster sampling by area

Within wards, "the crucial advantage of multi-stage sampling is that it results in a concentration of fieldwork and consequently a saving of time, labour and money";[7] and, I would add, the viability of support and oversight from team supervisors in what are taxing circumstances for fieldworkers.

GCRO initially sought to circumvent this approach in QoL V, with predictably disastrous effect. It would be in the company of other large-sample surveys in South Africa and beyond in retaining the multi-stage approach for several QoLs ahead: Stats SA's General Household and Community Surveys, the HSRC's SABSSMs, and large international comparative surveys like the Demographic and Health Surveys. Consequently, the GCRO should make financial and planning allowance for the time and expense of the sampling and weighting challenges, until sound alternatives have been piloted and implemented to great scale by experienced operators.

2.3 A critical selection of the key sequential elements of QoL sampling

Within this broad framework, and informed by current good practice, the sub-section will review what has been learned for sampling from previous QoLs, and cognate efforts, that should be maintained, or improved... or certainly not attempted again! Each consideration quickly becomes technically intricate the further one probes, and indeed may become irrelevant as technology changes. This will be noticed in a couple of instances below. Resolving what to do next at that level will have to be the call of the GCRO QoL leader, after consultation with the provider concerned.

2.3.1 Number of interviews per ward

In QoL V, *an equal number of interviews was assigned to wards* for the first time, on the guidance of Professor Paul Fatti. More precisely, after some initial over-ambition, 50 interviews per ward were conducted for metros that had assigned funds to the QoL, and 30 for all other wards. The final weighting corrected for this difference, so we may speak of a fixed sample per ward.

By contrast, in QoLs II to IV, interviews were assigned to each ward in proportion to its share of the overall Gauteng population,[8] so the number of interviews per ward initially varied greatly.[9] Some *ad hoc* reassignments of interviews adjustments were accordingly made, from the more to the less populous wards: for example, in QoLs III and IV to yield a floor of 60 interviews per metro ward and 30 otherwise. The logic of this was political and financial rather than statistical: the large municipalities erroneously thought they needed larger samples per ward, and chipped in extra money.

---

[7] Moser and Kalton, p. 108. The further benefit, "that it obviates the necessity of having a sampling frame covering the entire population" is not applicable, in that there are now the GTI and Stats SA spatial sampling frames under continuous improvement. The former – GTI's Building Based Land Use layer – has been used already in QoLs IV and V.

[8] Or, "proportionately to population". This is *not* the same as the PPS method of selecting SALs as in QoL IV, and referenced below.

[9] In Census 2011, the average population of a ward in Gauteng was 24,154 with a standard deviation of 13,451. The latter means that just more than two-thirds of wards had populations between 11,000 and 37,000, and the remaining third varied even more widely. The unadjusted samples per ward could thus vary greatly.

However, the arrangement of a single fixed sample size per ward is greatly preferable, whether metro or district. This is because the margin of error[10] of the sample of a ward is *dependent on the chosen size of the sample, not on the size of the parent ward population* – subject to the proviso that the sample is truly "random", plus some other specifications, of which more below.

This remarkable feature is widely misunderstood because it is so counter-intuitive. It seems common sense that a larger population somehow needs a larger sample to achieve a specified margin of error. But it is wrong. The matter deserves some elaboration.

The QoLs work (mainly) with populations of households. But since each is represented by the responses of a single adult, we may speak for the moment of a population of individuals. A socio-political poll of the opinions of individuals, as reported in the newspapers, may typically have a more-or-less random sample of 2,000 individuals. The relevant sampling formula[11] then implies that the responses in the sample to a typical question (e.g. a majority of 60% approve abortion on demand, and the balance disapprove) will have a margin of error of ±2%; *and this is so whether the sample of 2,000 was randomly drawn from among the individuals in a single ward, or Johannesburg municipality, or the Gauteng, or the whole of South Africa!*[12]

Unfortunately, the formula also implies that as the sample size diminishes, the margin of error increases. If the random sample above were halved to 1,000 households, the margin of error rises only modestly to ±3%. But the margin of error rises more rapidly for smaller samples.

But in the last three QoLs, although around 25,000 individuals were interviewed, when spread across the 529 wards of Gauteng this allowed an average of only 50 interviews per ward, approximately; for which the margin of error is dismayingly large: ±14%. Table 2 shows the margin of error for other per-ward samples of this order. Let us unpack further what this means. If in some ward the pro-abortion proportion were measured at 60%, among a sample of 50 respondents, the *actual level* among all the people in the ward may lie somewhere – anywhere – between 46% and 74%. There may be minority rather than majority support for abortion on demand in that ward! This also means that, for samples of 50, the scores between two wards – or for a given ward at two points in time – have to differ very widely before one can say that they are actually significantly different.[13]

[10] The more technical, less expressive term is "confidence interval".
[11] There are handy calculators on the web that furnish the results for different specifications: for instance, at https://www.surveysystem.com/sscalc.htm.
[12] The margin of error for a given sample size also depends on the "split" of the answers to the given survey question. When it is very uneven the margin of error is slightly better. The statistical theory also requires a specified "confidence level", namely, how sure one may be that the true level in the population will fall within the margin of error indicated by the sample. The confidence level is typically set at 95%, i.e. one may be confident that the true level in the populations falls within the sample's margin of error for 95 such efforts in 100. Finally, the sample is assumed to be much smaller than the parent population, which is clearly the case here.
[13] It is sometimes thought that for differences of this kind to be significant the margins of error "must not overlap". This is not quite true: they must not overlap a lot! See Geoff Cumming and Sue Finch, "Inference by eye: Confidence intervals and how to read pictures of data", *American Psychologist* 60:2 (2005), pp. 170-80.

Table 2: Margins of error for small sample sizes

Ward sample size	Margin of error
10	31%
20	22%
30	18%
40	15%
50	14%
60	13%
70	12%
80	11%
90	10%
100	10%

The bad news, then, is that it would always have been a vain hope for municipalities to use ward-by-ward results from QoL to identify reported respondents' differences among particular wards, or to identify ward-level changes over time… except where the differences being considered were very large.

There is good news, however, of two kinds. Firstly, in practice, users require *aggregations* of wards: into the province as a whole, or municipalities, or occasionally communities. The respective samples are then much larger, and the margins of error shrink. For example, retaining the example of doing 50 interviews per ward, the largest municipality, Johannesburg, with its 135 wards, would have 6,750 interviews, and a margin of error of only 1% - allowing quite discerning comparisons over time. However, Lesedi with only 14 wards would have only 700 interviews, with a margin of error of 4%, so it would have to show appreciably larger changes before they could be said to be significant. We take this up again below.

Secondly, a fixed number of households per ward allows convenient *scaling* of the overall sample if funding is tight. One might drop from 50 to 40 interviews per ward, i.e. from 26,450 down to 21,160 across the 529 wards of Gauteng, noticing from Table 2 that the per ward sampling error deteriorates negligibly from 14% to 15%. If one drops further to 30 per ward, i.e. a total of 15,870 interviews, the per-ward sampling error deteriorates somewhat more steeply, to 18%. Since one cannot plausibly analyse by ward anyway, this could be acceptable; but the implication for the ward "pixels" on GCRO's maps is that the intervals being displayed should be consonant with these margins of error, e.g. quintiles, not to be misleadingly fine-grained.

In what follows, on the understanding that for QoL VI the GCRO budget and the supplementation from the province and municipalities may be constrained, we shall take forward the illustration of 40 rather than 50 interviews per ward.

2.3.2 Number of interviews per cluster, and clusters per ward

It is quite possible in principle to choose, and specify the geographic location of, visiting points for separate individual interviews spread across a ward; and this does have the advantage of

optimising randomness. Indeed, this was done for QoL V. A commercial company, GeoTerraImage (GTI), maintains a regularly updated building-based land-use (BBLU) database, providing the coordinates and number of dwelling units for all residential buildings in the province. From this list, the QoL V fieldwork service provider sampled at random the appropriate number of visiting points from each ward (together with "substitution points" for each sample point, discussed below).

However, as GCRO reports, this spread "substantially increased the difficulty of completing data collection".[14] A more usual approach is for the interview visiting points to be planned in "clusters". QoLs II and IV used clusters of 4 and 5 interviews respectively. The arrangement has considerable practical advantages, and is accordingly taken for granted by Stats SA and major fieldwork houses: transport is generally arranged with four or five fieldworkers, one of whom is also the supervisor, travelling by car. Aligning clusters to teams reduces travel costs and time, and enables closer supervision. As Professor Fatti notes, if the selected visiting points of a cluster "are not too close together, then the interviews could hopefully be considered as being 'reasonably' random."[15]

Given our illustrative 40 interviews per ward, one might plan for 4 fieldworkers per car, corresponding to 10 clusters of 4 interviews. It might be tempting to plan for 8 clusters of 5 interviews, corresponding to 5 fieldworkers per car, but this would offer less variability across the ward. Therefore, if 40 interviews per ward cannot be afforded, one would preferably drop to 9 clusters of 4 interviews, so 36 per ward, total of 19,044 interviews; or at a push to 8 clusters of 4, so 32 per ward, total 16,926.[16]

Table 3 shows the number of wards per municipality in column A, and the numbers of interviews per municipality in column B, for an indicative 10 clusters of 4 interviews per ward.

The approximate[17] margin of error for these municipality samples is in column C, which shows that the margin of error for tiny Lesedi at 4.3% or Midvaal at 4.0% is about three times larger than that of Johannesburg at 1.3%, obviously because of their much smaller samples. This is relevant to whether variations in results over time can be reported as actual changes. As we have noted, if either of these small municipalities go up or down by less than 4% (on an evenly split answer) from one QoL to the next, strictly all that can be said is that the result is unchanged within sampling error.

[14] Appendix A, p. 12.
[15] Appendix D, p. 4
[16] For even further reductions in ward sample size, smaller clusters would be desirable to retain variability, e.g. rather than dropping to only 7 clusters of 4, do 9 clusters of 3 interviews, so 27 per ward, total 14,283. However, the savings on total sample size would then be offset by less efficient transport arrangements.
[17] "Approximate" is because the margins of error indicated here are not adjusted for "design effect", by which – compared to a simple random sample (SRS) – the margins of error of a multistage stage sample may differ. Some aspects of sampling (e.g. stratification, discussed below) may improve the multi-stage margins of error compared to SRS sampling, while others (e.g. clustering of interviews) may worsen them. See C.A. Moser and G. Kalton, *Survey Methods in Social Investigation, 2nd edn* (London: Heinemann, 1971), pp. 89 and 105. This relatively accessible text is the Old Testament of sampling. Note that the authors' use of "cluster" is more general, and more orthodox.

Table 3: Wards per municipality and margins of error for 10 clusters per ward, and alternative model

		10 clusters of 4 per ward			Varying number of clusters per ward, and "floor" of 1120 interviews, per ward		
	A	B	C	D	E	F	G
	No. of wards	No. of interviews at 40/ward	Margin of error of municipal sample	No. of clusters of 4, per ward	No. of interviews with an 1120 "floor"	Margin of error of municipal sample with a 1120 "floor"	No. of clusters of 4, per ward, with "floor"
Johannesburg	135	**5400**	1.3	10	**4973**	1.4	9
Ekurhuleni	112	**4480**	1.5	10	**4126**	1.5	9
Tshwane	107	**4280**	1.5	10	**3941**	1.6	9
Sedibeng District							
Emfuleni	45	**1800**	2.3	10	**1800**	2.3	10
Midvaal	15	**600**	4.0	10	**1120**	2.9	22
Lesedi	13	**520**	4.3	10	**1120**	2.9	19
West Rand District							
Mogale	39	**1560**	2.5	10	**1560**	2.5	10
Rand West	35	**1400**	2.6	10	**1400**	2.6	10
Merafong	28	**1120**	2.9	10	**1120**	2.9	10
Total	529	21160			21160		

To address this, the samples in Lesedi and Midvaal could be increased, as shown in column D, to match Merafong's 1120 interviews. This move could be "fed" by reducing the sample sizes of the three biggest municipalities, as shown at the top of column E and then doing 9 rather than 10 clusters of 4 per ward in those three municipalities, as shown in column G. The three middle-sized municipalities are unaffected. In the three smallest municipalities, there would be many more clusters per ward, again seen in column G, because they have so few wards across which to draw the 1120 interviews!

This manoeuvre improves the margin of error of Lesedi and Midvaal to 2.9%, in column F, the same as Merafong; whereas the samples in the three largest municipalities are still so large that their overall precision is negligibly affected. However, where previously the same weighting was applied for across-the-province and by-municipality analysis, to achieve this benefit would entail that nine separate municipality-specific weights be calculated for use in by-municipality analysis, adjusting for the disproportional allocations of interviews.

2.3.3 Invoking SALs (or EAs), and assigning clusters to them, optionally after stratification

Given our illustrative example of 10 clusters of 4 interviews per ward, how may they best be drawn? A particular challenge arises at this juncture, strongly emphasised on past experience by Graeme Gotz: namely that some wards may be "lumpily" heterogeneous. For example, an informal settlement may be adjacent to small-holdings in the ward, or to formal suburbs. How may it be ensured that sampling correctly represents these components?

QoL II adopted a *sui generis* approach, of overlaying a grid of perpendicular lines on each ward, with the lines closer together over urban than rural areas. Then the requisite number of

intersections[18] on the grid was chosen at random, with this point providing both the location for the first attempted interview of each cluster of four, and the starting point for "random walks" to select three more interviews. The critical problem with this approach was that the variation in grid density was not precisely specified, so that it could not be known whether each household across the ward had an equal chance of selection.

The SALs provided by Stats SA, mentioned in sub-section 2.1, come to the rescue. There are between 11 and 149 SALs per ward, an average of 52. Out of this plethora, in each ward 10 need to be selected, to host a cluster of 4 interviews. But the *size* of SALs differs: households per SAL vary very widely around an average of 220. So the selection technique of probability proportional to size (PPS) is required.[19] Under PPS, SALs that are larger (according to an available measure of size, notably population) have a greater probability of being selected than smaller ones. When coupled with subsequently visiting a specified number of households per selected SAL (we are working with 4 households), this technique meets the key requirement for the sample of households to be truly random, i.e. across the visited SALs every household will have had an equal chance of being selected into the sample.

QoL III took heed of SALs, but then ambitiously sought to assign single interviews per SAL, rather than clusters of interviews in fewer SALs. Of course in small wards, even with the floor of 30 interviews, this meant more than one interview per SAL; or, in large wards, some empty SALs. Assignment to SALs was done by PPS. This provided, obviously, a better spread than clusters of interviews. The supplier felt that the extra travel and time costs were not material, given the manageable size of a typical SAL. But this view was not shared by other providers or by the literature.

QoL IV also took advantage of some sub-ward geography , but with two variations.[20] The sampling was done by a sampling consultant, Dr Ariane Neethling, working closely with GTI. GTI's lowest level of geography consists of Census EAs from 2001, rather than the SALs subsequently issued publicly by Stats SA. The per-EA attributes are updated by modelling that is informed by the periodic population estimates, and other on-the-ground information gathered by GTI. Dr Neethling applied PPS to select the requisite number of EAs in each ward that would furnish the clusters of interviews.

Secondly, Dr Neethling invoked "stratification" of their chosen small areas, EAs, to ensure that coverage of the "patchy" heterogeneous areas of wards was not left to chance. Before doing the PPS selection, within all the wards the EAs /SALs are divided into "strata" according to one (or more than one) *predominant* attribute, e.g. by the four race groups, or formal/informal.[21] The number of clusters per stratum is then assigned proportionately. For example, if 30% of the EAs/SALs in a given ward are majority-informal, 3 of the 10 clusters would need to be drawn from the informal stratum, and 7 from the formal stratum. Thereafter the PPSs are done

[18] One remembers again that the number of households to be sampled per ward varied for QoL II to QoL IV.
[19] A welcome worked example is given in Moser and Kalton, *op. cit.*, 115-6.

[21] Stats SA has moved from urban-formal/urban-informal/farm/rural to urban/farm/rural classification of SALs. GTI still retains the formal/informal difference in characterising its modelled EA-level data.

separately for each stratum, to identify which 3 of the majority-informal SALs and 7 of the majority-formal SALs will have interviews.

Assigning interviews to EAs/SALs by PPS-within-strata offsets some of the loss of variation caused by clustering interviews.[22] And, happily, "a proportionate stratified sample can never be less precise than a simple random sample".[23] However, the more heterogeneous the strata, the less the gain.

And there are practical disadvantages to stratifying SALs. For the manoeuvre to be useful, the stratifying attribute needs to correlate strongly with the outcome variable(s). In the case of our multivariable indexes, one would choose strata "because they seem relevant to most variables".[24] Given the enduring effects of spatial apartheid, confirmed periodically by GCRO's own maps, race would be a prime candidate. Secondly, if therefore four strata are used ,the PPS process has to be repeated 529*4=2116 times. Dr Neethling in QoL IV reported using stratification by race.[25] Some deft programming would have been required. Finally, stratification may have unknown consequences, because of the requirement of "painting" wards as predominantly as this or that, in order to assign them to strata before PPS-selecting the SALs/EAs to be visited within the strata.[26]

In principle, with a clear and salient stratifying variable to hand, stratifying seems an obvious choice. In practice, given possible anomalies and especially given the extra effort, Professor Fatti doubts whether it is worth it. GCRO may wish to apply the well-known empirical precept, "try it and see", using the QoL V data for a small municipality like Lesedi: for each of its 13 wards, doing PPS sampling of SALs within proportionately allocated SALs in each of the four race strata, or across all the SALs in the ward. Then the resulting un-weighted distributions of races for the two attempts could be compared, as well as the breakdown by race of selected variables that contribute strongly to the the QoL sub-indices – *all done at municipal level.*

This last phrase is important. GCRO has agonised by comparing, from different QoLs, the spatial distribution of sample points within and across SAL "patches" *of a single ward* (see for such a ward from QoLs III and IV). This is a misplaced undertaking. *Of course* these distributions will vary, and *of course* there will be more clumping when clusters of interviews are employed, as in QoL IV, rather than mainly single interviews per SAL as in QoL III: that is the nature of randomness. (This is seen in how often numbers in the weekly Lotto may offer two or three nearly adjacent numbers.) And therefore, even after cumulating across the SALs in a ward, one must expect, as we have seen, variation in some outcome measure in the given ward, even assuming no changes of time: typically ±14!

We have therefore stressed, in sub-section 2.2, that wards can never be separately analyse-able units, except for supplying the 529 pixels on a vivid GCRO map. A hundred or many more wards, spanning a thousand or many more SALs, are summed to the nine municipality levels in Gauteng (and then, moreover, weighted by raking to match known Stats SA marginal values by ward!). As

[22] Valliant et al. quoted in Appendix E by Dr Ariane Neethling, p. 2.
[23] Moser and Kalton, p. 91.
[24] *op. cit.*, p. 92.
[25] Appendix E, p. 2.
[26] Suppose most houses in Alexandra had two or more backyard shacks. The every SAL in Alexandra would be "painted" as informal, and there would be no formal stratum.

UWC's Professor Daniels reminded, "Let statistics do its work!" It is thus, therefore, only by comparing results at municipal level that one may want to test whether stratification is worth it.

The remaining issue is this: does GCRO use EAs or SALs, for the sub-ward geography, i.e. for selecting the 10 entities within which interviews are to be done? Choosing between the use of GTI's 2001-based, extensively modelled EAs and Stats SA's post-2001, public SALs is not obvious. Moreover, the decision carries forward into the geo-data lists of the areas, from which the points are specified at which at interviews are to be done (plus additional substitution points). This is considered further in the next sub-section. GTI has a longer spatial provenance; its building information is currently more detailed, even according to Stats SA; and their data was successfully used for QoL IV's sampling (therefore using EAs), with stratification thrown in. Stats SA's building information is steadily accumulating, from its vast field-force engaged in its various surveys; it has the advantage of being official, and free (like the GCRO QoL data itself), rather than at risk of future supply arrangements; and its SALs were successfully applied in QoL III, if not for clusters.

A hybrid solution for QoL VI may be this (remembering that SALs either coincide one-on-one with EAs, or else are selective combinations of 2 small EAs). GCRO (i) overlays GTI's EA-layer with Stats SA's SAL boundaries, (ii) uses the GTI-based lists of building co-ordinates now within the SAL boundaries to give the measure of size of the SALs for use in the PPS selection process; (iii) applies PPS to select the 10 SALs per ward; (iv) samples visiting points and substitutes within each selected SAL. The issue can be revisited two years hence, to see whether to move to using Stats SA's own building lists and co-ordinates, in conjunction with their SALs, for QoL VII.

2.3.4 Identification of visiting points/buildings in wards, or in their SALs/EAs

A sampling frame involves a list, from which each entity – in our case households – has an equal opportunity of being selected. In the past, lacking a list, face-to-face fieldworkers had to start from some randomly selected point, and make a "random walk" according some rule[27] to identify households to attempt until the cluster of interviews was fulfilled, perhaps with some substitutions en route. This method was used in QoL III, although it sought one interview per SAL rather than a cluster; with the location of the interview being set at the centroid of the applicable SAL on a GIS map. If a replacement or more than one interview within an SAL was required, a random-walk protocol was set. But "human beings are notoriously bad at randomizing, and recent research shows that enumerators' instructions on conducting a random walk…systematically affect both the selection probabilities and the survey results".[28]

As the last decade unfolded, improving GIS technology has provided lists of the spatial co-ordinates of all buildings in an area, whether SAL or EA, as described at the end of sub-section 2.3.3. After scrambling the entries on these lists, a specified number of visiting points (and substitution points) for interviewers can be selected and mapped. (Thereafter, with some additional protocol described in the next sub-section, households and respondents can in turn be

[27] E.g., count off twenty visiting points, i.e. distinct sites, on the current side of the street, if necessary moving round the block, and then select the house across the street. This covers some ground, for smalloldings and farms.

[28] Noam Lupu and Kristin Michelich, "Advances in survey methods for the developing world", *Annual Review of Political Science,* 21 (2018), 195-204, p. 199.

selected from the buildings with an equal possibility of being included.) This approach was accordingly used by Dr Neethling for QoL IV, for each of the EAs that had been selected by PPS from within the race strata of every ward. And it was used in QoL V, but the visiting points were selected, and located, from ward-wide randomised lists of the buildings, i.e. without the intermediate use of EAs.

The general approach is thus well-precedented, and is recommended for use in QoL VI. Whether EAs or SALs are used as the basis for clustering, a random selection of building coordinates can be selected from either the GTI or StatsSA listings of building co-ordinates. The *practical* complications are not trivial, however: the buildings on the GIS map may be less-than-clearly "dotted", requiring a specified "geo-fence"; interview points may represent blocks of flats, or gated estates, or - at a smaller scale - formal houses in a township with backyard shacks, or suburban homes with a resident domestic servant on the premises.[29]

2.3.5 Selecting among dwellings and households at visiting points, and respondents

At the "visiting points" or buildings, within a specified "geo-fence", in practice, there is often more than one dwelling per pre-sampled site, such as backyard shacks behind a formal township house, or a main house plus workers' dwellings on a smallholding, or a block of flats. And for a gated community the visiting point on the map among scores of houses may well be only the security gate. In that case the fieldworker needs to follow some procedure for random selection of one of the several *dwellings*.[30] This was by the throw of a dice in QoL III. But in QoL IV the randomising procedure called a "Kish grid" was incorporated in the fieldworker's tablet. The stalwarts Moser and Kalton[31]give a worked example.

In a block of flats, the grid may select first the floor to be attempted, and then the dwelling on that floor. This is unworkable for sequentially numbered gated communities. For this, the tablet would need to have an app(lication) to throw up random numbers, and the first such number less than the total of dwellings is selected. (To repeat, in practice the final selection may have to be much more haphazard, and negotiated with the entrance security, even if an attempt has been made by the provider to find and then contact and then secure permission from the body corporate in advance.)

Next, there may be more than one household in the dwelling, in which case the tablet application or the automated Kish grid is again applied to randomly select the household among them. Separate households tend to be defined by variants of the traditional definition, "people who eat together". On this basis, a resident domestic servant may well be counted as a separate household, and be included for possible selection in the Kish grid.

---

[29] GCROs Christian Hamman conducted an eyeball comparison of the respective maps for this Review: "Look out for differences in the rural areas; the absence of backyard dwellings in the StatsSA Dwelling Frame [and] reduced coverage in the GTI BBLU with regard to cluster housing slides; but similarities between the datasets in suburbs, as well as differences in informal settlements where GTI BBLU often 'guess' that there must be dwellings, but there are also similarities [there] between the datasets."

[30] In the GTI data these various points are typically all listed – either as separate points (backyard shacks), or as multiple units at a point. So in theory one could select the exact dwelling or unit upfront, but in practice the difficulties of access made this refinement too challenging.

[31] Moser and Kalton, *op. cit.*, pp. 198-9.

And finally, there is the selection of the interviewee. Strictly speaking, for a household survey this choice might best *not* be random, but rather the adult in the household – often the mother, if applicable – who is most knowledgeable about household arrangements and its services! However, the QoLs have evidently been intended as *not only* as surveys about household-level matters: once on the spot, they have also canvassed the views of the selected adult respondents on a variety of notably individual opinions, such as their assessment of the state of democracy, their mental wellbeing, and their feelings about foreigners. The hope evidently has been that these tens of thousands of respondents, suitably weighted,[32] will represent the views of adult individuals across Gauteng and its municipalities. And certainly, these individual-respondent answers then feature alongside household descriptions and services in the indexes.

Therefore, the remaining requirement is that, after the listing of all the adults in the household, including their ages and genders, the fieldworker randomly selects which person s/he is to interview. This was done in QoL III by selecting whomever had the nearest next birthday to the interview date, from the initial listing. With a Kish grid, as used in QoLs II, IV and V, only the ranked ages need be asked, for women and men. With the Kish grid computerised on tablets, this is preferable for the future.

2.3.6 Rules for substituting interviews

Critically, the randomly selected person may be absent at the time the interviewer arrives. Accessing him or her may well necessitate a specified number of revisits, if that person is for example away at work. It is widely recognised as unacceptable to take a substitute interviewee from that household, on the first occasion or later. This would bias the sample towards at-home people, e.g. elderly people, or unemployed of any age, or young adults.) Revisits may thus most often need to be scheduled by the fieldworker for an evening or weekend.

QoL III expected a revisit to be undertaken the same evening, failing which a substitute visiting point could be taken up by a random walk. QoLs II, IV and V demanded up to two re-visits before taking up a substitute visiting point shown on the GIS-based map, and there repeating the dwelling/household/interviewee selection.[33] One notes that, if a return visit that evening fails, the second such revisit attempt would most usually require the fieldworker to travel back to the household again the next day, possibly a costly undertaking. And it could well occur that the respondent, when reached, refused! (As is being repeatedly remarked, the implementation of a sample in the field is an always exhausting and often demoralising undertaking.)

A recent overview of survey practice in developing countries (which included a survey of established survey practitioners as well as a literature review) is worth quoting, because it may seem heretical:[34]

> *Among the surveys conducted by our expert respondents, just over two-thirds revisited a selected household at least once…In principle, revisits—or callbacks in phone surveys—should increase response rates and reduce the potential for nonresponse bias.*

[32] See sub-section 2.3.6 below.
[33] In QoL V's initial phase, when solo fieldworker were deployed, s/he had to request "base" for the co-ordinates of a replacement visiting point.
[34] Lupu and Michelich, "Advances…",*op. cit.*, p. 199.

> *But revisits can also be costly…We know of no studies focused on the effects of revisits in developing contexts, but recent research suggests* ***there is little gain from repeated costly attempts to contact potential respondents.***[35]

QoL III required only one revisit, whereas QoLs II, IV and V demanded two. Given the massive overruns of time and, reportedly, cost by the providers, the QoL III substitution policy may be worth following; but subject to careful monitoring by the provider from the outset of the project, using the information transmitted from the tablets, of the implementation of the one-evening revisit. And, unlike QoL III, the ensuing substitutions would be at geo-pre-specified substitute visiting points, rather than by random walk.

2.3.7 Weighting of cases

There are many kinds of practical exigencies "on the ground", touched upon above and in Section 3, because of which the "realised sample" will not correspond to the appropriate benchmarks from Stats SA. To adjust the former to respect the latter, one is reassured that "post-stratification weights, with raking being the most widely accepted method"[36] are the uncontroversial modern technique. The realised sample of respondents is summarised into a matrix of, say, eight race-by-gender columns shown for 529 rows of wards. Then:

> The actual algorithm involves repeatedly estimating weights across each set of [column and row] variables in turn until the weights converge and stop changing. Essentially, raking forces the survey totals [by race and gender for the columns, and by ward for the rows] to match the known population totals by assigning a weight to each respondent.[37]

The basis on which GCRO has done this in QoLs III and V is to weight the distribution of the interviewed individuals to individual Stats SA benchmarks, using race and gender (which are known to be subject to fieldwork distortion). Since the interviewed individual represents the household, the result is then also construed as a household-weighted sample: a "one person, one household" vote. This is a compromise. Not surprisingly, the QoLs have been described at different junctures as individual surveys and household surveys! Perhaps the compromise is for the best, given that variables dealing with individual-level opinions and household-level circumstances are – not unusually among quality of life surveys – combined in the indexes, which are the central purpose of the QoLs.

However, if one were strictly to follow Stats SA, there would be two distinct weights: an *individual* as against a *household* weight. The former is easier to comprehend. The respondent represents the voices of all the adults in the household – be they 1, or 2, or 7 – captured in the initial listing, and so her/his responses are weighted upward accordingly before extracting the matrix used for raking.[38] As for the *household weight*, it turns out that Stats SA draws on UN

[35] Legleye S, Charrance G, Razafindratsima N, Bohet A, Bajos N,Moreau C. 2013. Improving survey participation: cost effectiveness of callbacks to refusals and increased call attempts in a national telephone survey in France. *Public Opin. Q.* 77:666–95.
[36] Lupu and Michelich,*op. cit.*
[37] A clear worked example is given in L. Anderson and R.D Fricker Jr., "Raking: An Important and Often Overlooked Survey Analysis Tool", at http://faculty.nps.edu/rdfricke/docs/RakingArticleV2.2.pdf.
[38] This approach, using a "design weight", appeared at first blush to have been applied in QoL IV. Had this been so, this QoL might have needed reweighting to be consonant with the others, or vice versa. But it turns

methodology to model an intricate "household" construct from individuals' statistics![39] Given the combination of household and individual items in the indexes, noted above, it may be well for GCRO to continue consistently, for comparability. If future empirical research were to suggest otherwise, i.e. applying the intricately constructed household, all the previous surveys could be back-weighted and inspected for any relative differences in the QoL index scores.

2.4 Summary of the key sequential elements of QoL sampling

This section presents a summary of the approaches recommended above, for the sequence of steps involved in defining a QoL sample. (The lettering of the items below corresponds to that of the sub-sections above.) Ironically, one hopes that – after the chequered history of the last three QoLs – the recommendations would be seen as rather unsurprising by any of the large -scale survey providers in South Africa. Given that QoL are a very costly venture of around 20-30 thousand arduously achieved face-to-face interviews of individuals in their households, that will be a merit.

The extent to which these steps carry forward may well, of course, be adjusted in later QoLs the light of further developments, much as the implementation of QoLs hitherto has been profoundly affected by technological innovations although not always, as we have seen, to beneficial effect. But, broadly, something much like the following sequence should be manageable by the GCRO QoL team, in concert with a reputable fieldwork provider, providing that it has proven experience in managing the sound implementation of very large samples. (This requirement will be stressed in Section 3 on fieldwork.)

A final caution, that will be expanded in section 3.3: the recommendations below seek to be practical and affordable as well as sound. Among a large team of fieldworkers working for months in often arduous circumstances, some mis-identified venues, too-easy substitutions, or outright cheating will occur. But with a reputable provider, a seasoned and well-supervised field-force, a large probabilistic sample, and weekly GCRO monitoring of prompt and appropriate corrections – to be discussed in the next section – most of these aberrations will be promptly repaired, and the remainder "come out in the wash". As Professor Daniels emphasised, learn to "Let the statistics do the work!"

2.4.1 Number of interviews per ward

In the light of the approved tender, specify the affordable number of interviews to be sought in every one of the 529 wards, e.g. 50 or 40 per ward. Recall that the fixed sample per ward is because the margin of error of the sampled results depends only on the sample size, not on the size of the ward

out that the "average is used, instead of the observed number of persons 18+ in a household, to obtain more smooth design weights with less variation."[38] If the author and Professor Fatti are correct, this means that the QoL IV realized sample, taken into the raking, was merely a uniformly scaled-up version of that applied in the "hybrid" approach for QoLs III and V! So no harm has been done to inter-QoL comparability.

[39] "Household estimates that were developed using the UN headship ratio methodology were used to weight household files. The databases of Census 1996, Census 2001, Community Survey 2007 Census 2011 were used to analyse trends and develop *models to predict the number of households* [our emphasis] for each year. The weighting system was based on tables for the expected distribution of household heads for specific age categories, per population group and province." Statistics South Africa, *General Household Survey, 2017*, pp. 66-7. cf. Statistics South Africa, *Victims of Crime Survey, 2017/18,* p. 100.

population. At 40 per ward, a total of 529 x 10 x 4=21,160 interviews would be involved. This assumption will be carried forward for the next steps.

2.4.2 Interviews per clusters

Determine the number of interviews to be included in each "cluster" of interviews. Assuming 4 fieldworkers per car is specified (one of whom is also the supervisor), then for 40 interviews per ward one may assume 10 clusters of 4. We carry this example forward below.

> **Option:** Increase the number of SALs per ward from which samples will be drawn in Lesedi and Midvaal to match the total number of clusters in Merafong, so these three municipalities would each have 1120 interviews. This could be "fed" by doing 9 rather than 10 clusters of 4 per ward in the three largest municipalities. The three middle-sized municipalities are unaffected. This manoeuvre improves the margin of error of Lesedi and Midvaal to 2.9%, the same as Merafong; whereas the samples in the three largest municipalities are still so large that their overall precision is negligibly affected. Appropriate weighting would be provided.

2.4.3 Selection of SALs in the wards

Each ward comprises a number of SALs (an average of 52 per ward) as demarcated by Stats SA. *Use the probability proportional to size (PPS) procedure to randomly choose which of the SALs will receive a cluster of interviews*, viz., 10 SALs of 4 interviews in our running example. PPS is desirable because SALs vary appreciably in the number of households they contain.

> **Option:** In every ward, divide the SALs in all the wards into strata by race, before selecting the appropriate number of SALs within each stratum.

2.4.4 Reaching the respondent

Identification of buildings/visiting points: GCRO will have purchased from GTI their latest listing of the geo-references of all the buildings in Gauteng. GCRO can overlay a SAL layer. Then *for each of the 10 selected SALs in a ward, randomise the ordering of all georeferenced residential dwelling points; and then do a systematic sample*[40] *of, say, 10 units from the list: 4 dwelling points to be visited initially, and the further 6 to be available for possible substitutions.* Then the fieldwork provider needs to generate A3 paper maps to a workable scale, per ward and where necessary per SAL, showing the selected dwelling points to be visited for the fieldworkers.

Selection of households: *The fieldworker is to use the automated tablet-based Kish grid: to randomly select a dwelling in the building/site, and the household in the dwelling.*[41] *Ad hoc* arrangements may, of course, be necessary to secure any interview at all!

Selection of the respondent: After listing the adults (over 18 years) in order of age, and recording the gender, *the fieldworker is to use the automated tablet-based Kish grid provided in the tablet to select the interviewee.* S/he is to note via the tablet the beginning and the end of the

[40] If, say, there are 230 buildings in a SAL (or in a stratum within a SAL), for 10 interviews (4 plus 6 spares), choose every 23rd building from the randomised list, after drawing a random number less than 23 for the starting point.

[41] In blocks of flats, the Kish grid can select a floor and then a flat. In gated communities, the houses may be numbered sequentially, and the tablet will need to provide a random number.

interview, the location, and if possible the cell number of the interviewee for possible telephonic back-check.

2.4.5 Replacement

If the household or selected respondent firmly refuses to participate, the fieldworker will need to move on to a substitute building/site (not a substitute respondent in the same household), after documenting the event on the tablet. However, if the selected respondent is absent at first call, the fieldworker must seek an appointment for the evening of the same day, or else for a suggested time the next day. If this one *revisit fails, the reasons must be documented, and the fieldworker is to move on to another of the pre-selected substitute buildings/sites.*

2.4.6 Weighting the "realised" sample

The sample is to be weighted by "raking", as described in the previous section to compensate for over- or under-representation of respondents from the chosen reference categories (presently males and females within African, Coloured, Indian and White) in each ward, while taking advantage of the marginal estimates in Stats SA's Community Survey that will be more recent than the Census until 2023.

Figure 1a: A ward and its SALs, showing visiting points by random-walk, in QoL III

Figure 1b: The same ward and SALs, showing pre-defined visiting points in clusters, in QoL IV

3 QoL survey-management requirements in field and in GCRO

The real purpose of this section, as it has emerged from the briefs, expert workshops, and interviews, is not primarily about the management of sound fieldwork by the private-sector provider on the one hand, or the necessary management structure on the GCRO side on the other. It is about securing and sustaining a constructive and trusting working relationship between them. And this in turn demands a balance between what we may call demand and supply.

On the one hand GCRO – having chosen a provider with an effective managerial structure, sound finances, and a known track record of delivering massive-sample high-profile surveys – may expect to receive what current GIS and cellphone technology have made possible and tempting, namely real-time access to the information being relayed from interviewers tablets, and daily emails of surveys completed. But, a critical but, GCRO must not expect the engage the provider daily on deficiencies it may have uncovered as they arise. This is micro-management that no reputable and well-managed provider should be expected to tolerate. Rather, the provider must have the opportunity (i) to identify such difficulties; (ii) correct them (which probably will occur only the next day for piecemeal deficiencies, and later if a pattern of more serious misconduct has to be accumulated, e.g. for dismissal of a supervisor); and (iii) assemble a summary of such corrections for its weekly meeting with GCRO. From its side GCRO, by manageably sampling the work of different field teams at different times, can assemble its own summary over the week of what it has identified, and of what it expects the provider to have identified, in order check up on monitor the adequacy of the provider's self-corrections, and if necessary demand and follow up improvement.

In other words, GCRO's task is not to monitor the fieldwork, but to monitor the adequacy of the provider's doing so, and the adequacy of their corrective actions.

It follows that the fieldwork provider's project management team must be equipped: (i) not only to canvass the GCRO counterpart daily on major contingencies arising whose solution may require GCRO input; but also (ii) supply GCRO weekly with up-to-date, comprehensive, quantified reports and analyses ☒ of how interviews and back-checks are proceeding, and how the problems that have arisen the previous week were diagnosed and have been or are being remedied; as well as (iii) after some days' delay for tidying and correction, provide weekly batches of the proportion of survey data undertaken. This will replace the daily upload to GCRO that the provider has not scrutinised.

The rest of this section slightly amplifies this account, in three sections. Firstly, as was done in the previous section, assumptions or parameters are noted, in sub-section 3.1. Sub-section 3.2 summarises some of the GCRO choices and experiences of fieldwork provision for past QoLs that led to the "perfect storm" described in Section 1, This is in order to underline what especially needs to be retained in future QoLs – or strenuously avoided! Sub-section 3.3 recaps, as recommendations in point form, the main good practices that may be gleaned from sub-section 3.2 (with some supplementation from the GCRO briefs prepared for the Review). In the main, these recommendations are what should be routinely arranged by an experienced large-scale provider; and they are presented in any of the canonical texts[42] on sample-survey conduct and monitoring. The points may be useable as a broad checklist to be covered in future procurement investigations. The last part, sub-section, 3.4, contemplates a suitable management structure for GCRO: for its engagements with the provider, tailored to an appropriate level of concurrent monitoring activities, and for keeping GCRO executive management apprised.

3.1 Initial assumptions for managing QoLs

3.1.1 External private-sector provider

Firstly, it is assumed that – given the proper arrangements adumbrated above and expanded below – GCRO will continue to commission the QoL fieldwork from an external private sector provider. Managing fieldwork, whether only for the QoLs or possibly for additional projects, is simply not the core business of a co-university based agency oriented to policy relevant research. Far better that an experienced commercial entity – with the several necessary managerial, supervisory, call-centre, HR and financial staff in place on the basis of a sustainable through-flow of surveys – handle the unending nightmares of human resource disputes, subsistence payments, fieldworker injuries on the job, vehicle accidents, providing and programming and replacing tablets, large-scale database maintenance, and data-cleaning.... and make a fair profit in doing so. The stresses and overtime visited on several GCRO staff, in having to help to rescue QoL V as it painfully transitioned to orthodox field practice and data-management, should be proof enough.

[42] For example, Survey Research Center, *Guidelines for Best Practice in Cross-Cultural Surveys* (Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan, 2010).

3.1.2 Six-month duration for field-work

It was assumed in Section 2 that QoL VI, and perhaps beyond, will be conducted with a face-to-face interviews of respondents in their homes For such a survey, the consensus across the previous QoL service providers is that approximately 125 fieldworkers are the maximum that can be reliably managed and supported in-field at a time. For a sample of 18-20,000 this yields a realistic field-work time of 6 months for which to allow (with fieldworkers having a five-day week, but that includes evenings and weekends). This contrasts with the repeated previous QoLs' hope of 3 months – which led to unpredictable and extensive over-runs, with alarming numbers of interviews having to be repeated because of over-large and under-trained field-forces. (By way of a reassuring comparison, Stats SA has its 30,000 General Household Survey in field for a year.) This duration is quite manageable within a two-year cycle.

Planning for a longer time in field may increase the chances that respondents' answers are affected by prominent events or seasonal variation – e.g. the effect on trust in Eskom of a sudden resumption of heavy load shedding, or the effect on respondents' assessments of their health in winter as against summer. This is true. But these effects will be mitigated if all municipalities are tackled concurrently and interviewing is randomly timed across their selected wards and their SALs, as far as the cost-efficient organisation of transport allows. Indeed, some averaging across seasonal effects leads to a sounder answer, which is why Stats SA took to spreading the calibration survey for the CPI across a year! Besides, for reasons that will be evident below, previous attempts to work faster have failed, with an unpredictable impact of appreciable proportions replacement interviews. And, of course, if only a smaller sample is affordable, the six months in field can be safely reduced accordingly.

3.1.3 Standard provision for fieldworker teams

Finally, there is a harsh reality (rather than an assumption) that has to be borne in mind in any arrangement of fieldwork: good fieldworkers face a personally daunting and potentially demoralising task, day after day. As specified in Section 2, they have to personally reach an identified visiting point in milieux as various as informal settlements, apartment blocks and small-holdings; there secure entry to a selected a household and randomly pick a respondent; and – often after one or more necessary re-visits in the evening or on the weekend to find them at home – secure their assent to be questioned for up to an hour. The reported experiences of fieldworkers were in common in all the QoLs. The briefing document vividly records the experiences at the various junctures:[43]

> *Key challenges experienced during implementation related to limited community preparation, and included difficulties in accessing affluent areas with high walls… as well as complexes and mining hostels, high levels of household refusals in affluent areas and plots, clashes over access with community policing forums despite clearance from local police, [and] delays from tribal authorities and apartment block landlords…Fieldworkers and supervisors encountered … language barriers when respondents weren't able to speak a 'common vernacular', no resident of the appropriate age, or no responsible person available to interview …[as well as] widespread abuse while conducting data collection, much of it racial, and harassment by security personnel.*

[43] "GCRO's Quality of Life Survey: Sampling workshop briefing document", *passim.*

Given these typical experiences, which tend to be out of the sight of the principal, fieldworkers' employment conditions are important in sustaining their motivation and minimising attrition: such as decent remuneration,[44] a *per diem,* and working in a team with support as well as scrutiny provided on the spot by the supervisor. It is not surprising, then, that the model of 4 or 5 interviewers travelling in a small car, one of them a supervisor, is an industry standard based on long experience. Even in the best QoL case, 15% attrition was still reported. But departing from this standard, as will be seen below for QoL V, is disastrous: more than 500 interviewers passed through training and deployment before the job was done.[45]

3.2 Major experiences in previous QoL fieldwork arrangements

Broadly speaking, a different provider was recruited for each QoL to try to avoid the difficulties experienced with the previous one. So there has been no opportunity for providers to learn from their previous experience, nor for GCRO to build a relationship of trust and familiarity as a resource for handling the inevitable and numerous difficulties that arise. The effect has been that the experience has been increasingly traumatic as the QoL sequence unfolded, and especially as the sample size rose to the intended 25-30,000 in QoLs III to V. The thumbnail sketches below are offered for what may learnt from good or bad experiences. (They omit the choice and implementation of sampling, which was covered in Section 2.)

The inaugural **QoL I** had the advantage of a sample of only approximately 6 600 respondents. It was conducted by a research house with experience in working for academic projects, which was able to assist with questionnaire design, undertake modest in-field piloting, provide detailed fieldworker training with manuals, deploy experienced supervisors, and undertake spot-checks as well as telephone callback of 15% of the sample. Interviews were on paper, and so were reviewed in field and before capture. GCRO only received a dataset at the conclusion. The fieldwork was completed without delays in some 2 months. The chief difficulty was in the use of separate GPS devices to record location, which went awry on some 10% of questionnaires.

Therefore a principal attraction of the provider selected for **QoL II** was that, although paper-based questionnaires were still used, a digital pen was employed, intended to capture the geo-location as well as the responses to questions, which were then transmitted by cell-phone.[46] However, in practice "the accurate recording of GIS coordinates proved extremely challenging... As a result, coordinates were not available for over 5 000 of the 16 729 surveys completed... Where feasible, survey coordinates were calculated using available information, such as street address and fieldworker whereabouts."[47] Eventually, wards were unknown for about 6% of the questionnaires. The field work was spread over 7 months. A final wrinkle was that after GCRO received the dataset at the conclusion, errors were detected in the merging and had to be corrected.

---

[44] Since payment is generally per questionnaire, a cap on questionnaires completed per week may be helpful in discouraging hasty preparation and careless interviewing. Survey Research Center, *Guidelines, op. cit.,* p. 473.

[45] "Sampling workshop briefing document", *op. cit.,* p. 17.

[46] Similar arrangements to QoL I were made for training, piloting, in- and post-field checking of questionnaires, plus an external consultant for back-checks. Back-office computerised quality control was added.

[47] "GCRO's Quality of Life Survey: Survey management workshop briefing document", p.10.

So, in turn, the provider was selected for **QoL III** for offering three relevant attractions: for the first time, data-gathering was on tablets, which supplied reliable geo-positioning of the interview and also made live upload of the data available to GCRO. In addition to the usual controls,[48] the geo-referencing allowed "viewing of survey locations on an online interface, and cross-checking survey locations with coordinates and field maps."[49] The provider was responsible for digitising the questionnaire – an extra step that manifested some problems during fieldwork. Nevertheless, having had experience with GIS and with large-scale surveys for government, the provider was able to complete the first full-scale QoL, a sample of 26 387, in just over the specified 3 months. But then, as with QoL II, GCRO checks on the ostensibly final product uncovered widespread duplicate and mis-located questionnaires that, according to the provider, arose from database merging problems, and the provider had to return to field for a further month.

With hindsight after the next two QoLs, both of which ultimately required 10 months in field, GCRO might have recognised QoL III as a near-creditable performance. But lacking that resource, GCRO accepted the tender for **QoL IV** from a fourth new provider, a large market-research agency.[50] This turned out to have two large, related problems. Firstly, it was not apparently made clear in the tender that the provider would subcontract responsibility for fieldwork to a range of small providers. Finally more than 500 interviewers were involved. Although it was reported that the sub-contracted team leaders and supervisors conducted in-field quality control, and there were check-backs and office scans on incoming data, GCRO's independent quality assurer unsurprisingly "identified major problems with in-field implementation of the survey".

Secondly, the provider had not previously done much GPS work. Once co-ordinates were ultimately provided, "serious problems quickly became apparent, with clear clustering of surveys in areas such as fast food outlets, malls, and taxi ranks."[51] As many as 6 000 interviews had to be redone, stretching fieldwork over 10 months to reach the 30 000 target. While batch uploads of survey data were successful, for concurrent GCRO scrutiny, in practice the full scale of the problems could only become evident upon initial handover of a full dataset with co-ordinates.

Perhaps understandably in view of this last experience, but certainly rashly for such a large and expensive face-to-face survey, GCRO departed altogether from orthodox fieldwork practice for **QoL V**. Previously inexperienced fieldworkers drawn from a youth employment programme were given extensive training and then – on the initial plan – expected to make their way alone by public transport to geo-specified visiting points (with nearby substitution points) that had been randomly defined across the ward, without any convenient clustering. This arrangement was proposed by a University of Johannesburg partner organisation that had developed a platform for "live tracking of fieldworker whereabouts, and integrated mapping of multiple coordinates for every survey on a web-based viewer which appeared to be an ideal fit for GCRO's

[48] Similar arrangements to QoL II were made for training, piloting, in-field checking of questionnaires and back-office computerised quality controls, some 6,000 telephonic call-backs, plus an external consultant for back-checks. The provider suggested week-long training and piloting in future because of the tablet technology.
[49] "GCRO's Quality of Life Survey", p. 12.
[50] This is the first QoL which reported translation of the questionnaire into four vernacular languages.
[51] *"GCRO's Quality of Life Survey: Survey management",* p. 15.

needs".[52] And there were indeed two positives. Firstly, "the application also collected extensive paradata together with survey responses and coordinates… This meant that the implementation of sampling protocols could (and did) receive very close scrutiny." Secondly, the daily survey-data uploads onto the web application could receive automated and manual checks.[53]

However, there were three crippling deficiencies. Firstly, and most importantly, the model that fieldworkers could work largely independently and use public transport to reach interview locations, and that supervision and support could all be remote, unsurprisingly failed. "Fieldworker attrition and cheating were widespread." This initial plan was quickly abandoned, and arrangements were gradually adapted towards the usual 4-5 interviewers per car. But the result was that the target of 38 000 had to be reduced to 24 800, for which fieldwork stretched over ten months, as for QoL IV. Secondly, the software proved ill-suited, compared to the packages used in QoL III and IV, for digitisation of the questionnaire. Problems with skips etc. persisted throughout the project. Finally, although data could be viewed in real time on the web server, it proved extremely challenging to extract for concurrent scrutiny, and GCRO eventually took on responsibility for conversion to SPSS.

3.3 Critical learnings from QoL (and advisers') experience

The thumbnail summaries above indicate fieldwork-related problems that QoLs have had in common, such as reliable geo-referencing (with the exception of QoL V); and particular problems, such as the lack of provider accountability for fieldworker quality in QoL IV. Others major issues, of both kinds, have been mentioned in the various inputs to this Review. They have informed the headings and the items mentioned under each. Neither is exhaustive: rather, they seek to be informed by past QoLs towards relevantly improving future ones. Exhaustive, and exhausting, prescriptions covering the entire survey sequence in general, and thus spanning all the sections of this Review and much more, are available is in the famous Michigan University Survey Research Center's 700-page *Guidelines for Best Practice*, footnoted earlier.

For painfully accumulated, QoL-specific, fieldwork-specific requirements, one might look at the specifications in the tender document for QoL V. However, the fact that the ensuing QoL V implementation was nevertheless so deficient almost from the outset, and was such a horrific experience for GCRO, is a reminder that the real requirements and capabilities may be not so easily documented, appraised in the procurement process, or delivered. The points below try to capture some of these aspects. Many may seem obvious: but they are included precisely because they specifically came up in the fieldwork briefing document, inputs at the three workshops, and in interviews with providers and counterpart GCRO researchers.

Before moving to the particular measures, intended to help avert a next manifestation of the "perfect storm", it may be well to identify two overall precepts or outlooks that "frame" such particulars; and that are necessary for such a vast and intense project to succeed:

- A massive survey has an unavoidable degree of imperfection. Fabricated interviews and/or patently inappropriate locations, uncovered in the course of periodic quality

[52] *Ib.,* p. 17.
[53] As in QoL IV, the questionnaire was translated; and 25% of respondents were reached in telephonic back-checks.

controls, are entirely unacceptable and need to be rigorously re-done at nearby surplus visiting points. But that is different from GCRO's using available real-time information to try to ensure day-by-day that each and every interview has occurred as prescribed in Section 2. Rather, it must be expected and countenanced that a proportion – to be monitored, and addressed if it escalates – of interviews will occur that are sound, but with mis-selected respondents, or at slightly mis-located visiting points. Such interviews are absorbed, if not in the multiple interviews per SAL, then in multiple SALs per ward in multiple wards per municipality. And if there nevertheless turn out to be cumulating biases – such as the well-known over-representation of at-home people like retirees or unemployed youngsters, or better-educated respondents – they will assuredly be mild, and will be brought into broad population alignment at ward level by the choice of relevant "raking" weights. In sum, "Let the statistics do the work", as UWC's Professor Daniel has reminded.

- In the daily trouble-shooting phone calls (not emails) between the respective operational managers on unexpected crises , and the weekly meetings in which GCRO monitors the provider's diagnosis-and-repair and data uploads, a transparent and trusting working partnership between GCRO and the provider will be built and strengthened on the understanding that dealing effectively with the inevitable contingencies is finally *both* parties' challenge... because if the provider fails, GCRO fails too. (That is not, however, a warrant for an ill-chosen, ill-equipped and inexperienced provider then to be bailed out by the heroic efforts of GCRO, as in QoL V!)

3.3.1 Tendering

The key requirement for the tender terms of reference (ToR) is to signal in the large, at the outset of the document, before all the details, that GCRO is seeking a well capacitated, well experienced private-sector fieldwork provider: namely, one with (i) several years' documented track record of (ii) very large-sample, (iii) GIS-informed, (iv) tablet-based, (v) pre-notified, (vi) well-databased and web-accessible, (vii) rigorously spatio-sampled, and (viii) transparently managed, probability-sample surveys in South Africa. The ToR could add that an additional advantage would be that be such work included (ix) public-sector themes like the QoLs, for (x) clients with the above-industry-standard academic and research-oriented expectations, like GCRO.

This has some corollaries: The provider will, *inter alia, be expected to* have:

- a substantial permanent-staff contingent of survey-related management staff (as required by its other continuing survey commissions), including a senior executive specifically accountable for field;
- in-house, or sound previous experience with outsourcing, specialist functions like the printing of the 6 000 A3 SAL maps for fieldworkers from the files displaying visiting points, provided by GCRO, and undertaking and classifying telephonic follow-ups of some 25% of sample;
- a roster of experienced in-house and/or contracting supervisors and field staff, well in excess of 25 and 125 respectively to allow for reasonable attrition, all with demonstrable previous record of satisfactory work for the provider;
- the organisational scale to sustain such a complement in field for six months undistracted by concurrent or new obligations;

- satisfactory work and remuneration conditions for this complement, including allowance for regular evening and weekend work, and associated transport;
- capacity for providing quantified, cumulating, weekly, written management reports on progress to GCRO, including problems and remedial measures; plus – after a reasonable preparation time – fortnightly uploads of clean geo-paradata and the substantive survey data itself.

And GCRO needs, *inter alia,* to:

- expect to have a three-hour working seminar (rather than a fifteen minute presentation) with each of the, say, two finalists to interrogate their proposals regarding the foregoing, as well as the many other detailed clauses in the tender document;
- provide a suitably sized, counterpart management and research contingent in GCRO dedicated for the duration of the QoL cycle, with appropriate delegations for prompt response to the provider, and capacity for upward reporting in GCRO – this is discussed in sub-section 2.4 below;
- anticipate the timetable and elaborate requirements for procurement through the GCRO board and the University;
- be familiar with what the PFMA and the Treasury Regulations actually demand and enable, to appeal effectively against the University procurement prescriptions that may, for example, obstruct the arranging of continuity with a satisfactory specialist provider in future;
- allow that a private-sector tenderer will quite justifiably plan to ring-fence a reasonable profit on the entire project, subject to actually delivering the project on budget and to deadline.

3.3.2 Questionnaire

Issues of questionnaire format, length, etc. are considered in sub-section 4.1. Some additional measures are:

- Known capability of well-established tablet software, and programming facility of the provider, to handle agreed geo-marking, Kish selections, and the follow-ups and skips in the QoL questionnaire;
- Translation of the questionnaire into four additional languages with sensible colloquial validation (rather than "back-translating" by academics), plus coverage of the languages in piloting.

3.3.3 Interviewer selection and training

Additional to the specifications in (3.3.1):

- Interviewer training for at least 5 days rather than 2-3 days has been recommended by a past provider, given the demands of using the tablet for dwelling, household and respondent selection and for the substantial interview;
- Interviewer testing on the specific training, to check whether all candidates meet the requirements;

- Inclusion of "soft skills" in training, with role plays etc. for hints on patience and empathy, sensitive topics, the presence of bystanders etc.
- Possibly arranging for what the literature calls "co-ethnic" interviewers, known to improve access.

3.3.4 Piloting

Although the QoL index items are established (as discussed in sub-section 4.2), if rotating modules are adopted piloting will still be an important feature of all QoLs. Some considerations:

- Three levels of piloting are optimal: (i) observation from "behind the glass" of how a variety of respondents cope with new questions (ideally in translation); (ii) if possible piloting of, say, 5 questionnaires each in field by highly experienced interviewers, with an occasional for detailed group feedback; (iii) a 1-2% sample after programming into the tablet, dispersed across wards, and data-analysed to pick up problems.
- Sufficient time has to be allowed for improvements from piloting stages to be incorporated, before the deadline for going into field!

3.3.5 Sampling design

This is covered at length in Section 2. It should be noticed that should stratification be adopted in the PPS selection of the particular SALS to visit per ward there may be appreciably more PPS exercises to conduct, to yield the SALs, than the 529 required for if stratification is not used. For example, if formal/informal stratification is applied in a given ward, then a PPS has to be conducted across the formal SALs in the ward to identify those to be visited; and another PPS, likewise, for the informal SALs.

3.3.6 Fieldwork

In addition to, or underlining, the requirements and capabilities mentioned elsewhere, the project requires:

- diverse prior publicity attempts by an inventive and persuasive provider's "desk", including mention if possible of the imminent survey in municipal rates bills or residents' association newsletters, telephone calls to bodies corporate, WhatsApp groups, etc.;
- supervisor experience and transport for evening and weekend follow-ups;
- an extended-hours telephonic support desk at the provider for difficulties arising;
- rigorous but not punitive criteria for substitutions, discussed in Section 2 (multiple call-backs may cause the respondent to be resentful and the interviewer to be demoralised);
- fieldworker flexibility, but arranged with supervisors and captured on the tablets for monitoring, to address situations of impossible access.

3.3.7 Field back-checks

- The purpose of back-checks is that interviewers and supervisors maintain good practice, because they are aware that any of their selections and interviews might be scrutinised. So they must also be aware that there is a reasonable chance of this occurring. Providers

have reported various rates of telephone call-backs, up to 15% within a week after the interview.

- More immediate are random call-backs by supervisors within the day. Occasional spot checks by the providers or GCRO's managers, and a concurrent exercise reviewing overall fieldwork roll-out by an independent consultancy will keep the provider itself similarly under scrutiny.
- The foregoing may be assisted, where addresses are not clear, that the practice continue of fieldworkers attempting to take a cellphone picture of the particular dwelling at the visiting point that was attempted.

3.3.8 Back-office quality control

Different forms and levels of automated and manual checks have been reported by providers, covering the geo-positioning of interviews relative to the pre-defined visiting points, the random selection of respondents, and the quality of questionnaire completion. GCRO may wish to define a suitable repertoire as part of the tender. This data, furnished by the provider, will inform the weekly fieldwork monitoring meetings and may be spot-checked by GCRO. GCRO may also wish, on occasion, to undertake highly selective, sampled computerised "dips" into the fortnightly survey-data uploads, and report back.

3.4 GCRO-side management arrangements

The GCRO briefing on fieldwork perhaps rather wistfully notes that New Zealand's Quality of Life Survey confines itself, for reasons of limited staff capacity, to "sampling guidance, questionnaire design, weighting and the final reports."[54] Everything in between is handled by the fieldwork provider. What have been the arrangements for the QoLs in GCRO?

3.4.1 QoL-specific management arrangements

The New Zealand approach also applied in QoLs I and II, where the then Executive Director was responsible, but was able to be "very hands off" with "minimal involvement" in day-to-day implementation.[55] He enjoyed some internal management support, and an external consultant helped with instrument design and fieldwork monitoring.

By contrast, as the briefing document thoughtfully traces, the extent of GCRO involvement has escalated through the next three QoLs, in two different respects. The one has been voluntary: as real-time access to geo- and survey data has become increasingly available, GCRO has actively sought ever more detailed day-by-day scrutiny and follow-up. This has allowed major issues in implementation to be picked up and resolved quite quickly, and has improved information about survey location and implementation. But the "perfect storm" described in Section 1 started to gather when GCRO felt that, because they *could* attempt to ensure the correct implementation of almost every interview, they *should* – which may not have been to the detriment of the overall end-product, as one provider surmised, but certainly affected GCRO's working relationship with the later providers, and the final timing.

[54] *"GCRO's Quality of Life Survey: Survey management",* p. 26.
[55] *Ib.,* pp. 8 and 10.

The other respect was unintended but perhaps inevitable, as the sample size moved upwards of 25,000, with attendant pressures of larger field management in an unaltered target time-frame. With each survey GCRO has had to intervene ever sooner, and ever more intensively, to avert serious deficiencies in the survey implementation process. As has been seen in sub-section 3.1, in QoL III it picked up widespread imperfections in the "final" dataset, requiring an extra 1 month in field. In QoL IV it had to micromanage subcontracted fieldwork via a recalcitrant provider, leading to a discarding of 6000 interviews before and after the "final" dataset, and an extra 7 months in field. And it pretty much had to steer in QoL V from an unprecedented and unworkable fieldwork model towards orthodox field arrangements, plus finally provide missing database skills. This lead to a one-third reduction in intended interviews down to 25,000, and again an extra 7 months in field.

These successive developments accumulated into the "perfect storm". As a consequence, a researcher had to be assigned full-time (in fact, much more than full-time) to QoL IV and QoL V, accompanied by the increasing and crisis-driven draw-down of the time of Graeme Gotz and the Executive Director Professor Rob Moore.

Now one hopes that various strategic recommendations of this document will alter this situation. Firstly, it has been recommended that GCRO refrain from detailed daily monitoring and intervention; reverts to regular weekly meetings with thorough quantified reporting of progress, diagnosis and repair; receives and supplies feedback from its weekly scrutiny of batch uploads; and thereby lets both the provider and the statistics "do the work". Secondly, it is hoped that this management arrangement is enabled by the conservative selection of a provider with the necessary track record of large scale orthodox delivery, with substantial on-board management and field staff as well as well-tried field- and back-office technology, and an ethos of transparency and collaboration.

However, even this more delimited supervisory and quality-monitoring role by GCRO warrants a less *ad hoc* contingent on its side, not only as a hands-on counterpart to the service provider's project manager and colleagues during fieldwork, but fore and aft. Preceding the fieldwork phase, which has been recommended to last six months, there is the budgeting, tendering, planning, user consultations, questionnaire modular design, and piloting (and follow-up dissemination from the previous QoL!). Following the delivery of the final database and tabulations, there is the initial analysis for and production of the main report, the launch, immediate follow-up briefings to municipalities and their interested departments, and the important and time-consuming processing and presentation of bespoke requests and original research publications.

This is clearly a two year cycle with demanding and interdependent flows of activity. The lead researcher would be expected to be on GCRO's complement, assigned to the project for the two-year cycle; and would need to be senior to handle the engagements with users, and numerate to handle or supervise the analysis, especially the indexing, and bespoke requests. But for the project to be sustainably handled into the future, it warrants – it is estimated – two additional posts.

One is an operations manager reporting to the lead researcher, especially to be the first point of contact in the day-to-day liaison with the provider, and also to handle the administrative aspects of the events through in the two-year cycle. The other is an additional post of an assistant to

support the researcher with routine calculations and administration arising from the weekly meetings with the provider. The latter could perhaps be filled by a research intern. The respective roles and competencies would evolve on the job under the oversight of the lead researcher.

It is essential for the GCRO Board to know who "carries the can". In the language of project management, the Executive Director Professor Moore is the "project sponsor", to whom reports the "project leader", the lead researcher, to whom report the operational manager and the assistant.

3.4.2 Wider GCRO management implications

Much as the adoption of a more delegated and timetabled approach to the QoL project requires a structured relationship of the QoL team with the provider, so it warrants a structured relationship within GCRO, especially to avoid the *ad hoc* and crisis-driven intrusions upon senior management that were experienced during the recent QoLs.

The following is presented as a suggestion, since it was not a topic at the workshops, and may already be in place in a similar form. Firstly, there would need to be a compact QoL-survey advisory group of three to four people supporting the lead researcher, convening in brisk regular meetings, to whom the lead-researcher would present digests and matters arising for discussion, covering the weekly fieldwork reports from, and quality feedbacks to, the provider. To offer support to the major facets of QoL, members might include the Executive Director qua project sponsor, Graeme Gotz, Melinda Swift and another. Importantly, the advisory group would *advise*: the lead researcher finally "carries the can" for what is decided and to be done.

Secondly, for any more far-reaching decisions arising as an emergency, the lead researcher would refer to the Executive Director.

Lastly, the Executive Director could present brief indications of progress and new developments, for noting, to the regular meetings of the management committee and the Board.

4 Questionnaire design and index construction

4.1 Questionnaire design

An apt description of the QoL questionnaire, as it has both mutated and expanded over the last decade, was provided at the third workshop by Dr Ben Roberts, who has run the HSRC's annual SA Social Attitudes Surveys (SASAS) for the last fifteen-odd years:

> *QoL is a big broad sprawling questionnaire. This is horrible for fieldwork but amazing as a dataset. It has meant the creation of whole new sets of information and analysis (marginalisation, accessibility, etc.)*

The "horrible" side was evident even in the earliest and briefest incarnation of the instrument. As the GCRO first Brief summarises it, about QoL I:[56]

[56] Providers' opinion was divided, in interviews. Cf. Appendix A, *op. cit.,* p.5.

> *Average questionnaire duration has been over 45 minutes, and poses a significant challenge to fieldworkers in the field. Further, many surveys are completed implausibly quickly, suggesting that fieldworkers take shortcuts in implementation.*

Although it is argued below that somewhat longer questionnaires don't necessarily imply longer overall time in field,[57] one must add the time, and cost, of quality control – which is multiplied across the very large sample. And one needs also to add, indeed prioritise, "respondent fatigue", as the textbooks call it. For unemployed or retired at-home respondents, the interview might provide the day's entertainment. But for a working respondent being interviewed after s/he has travelled home in the evening, "respondent exhaustion" might be more accurate, with likely diminishing accuracy of answers by the end of the questionnaire.

These considerations need to be held against the undoubtedly great gain to many users of the extensive range of core, added, retained, and temporary questions. Dr Roberts again: "This is of huge value - we need to maintain the way we think about QoL in terms of the diversity of uses." Can this plea be honoured while solving the problem of a distended questionnaire?

4.1.1 Background: the content and evolution of the QoL questionnaires

The problem has its origin in GCRO's laudable user orientation:

> *Content which has been retained across all or most survey iterations includes those questions thought from the outset to be essential to generating the QoL index, and those essential for government purposes.*

That is the easier decision. GCRO correctly consults, or is approached by, government users at each new QoL about their shorter or longer term perceived needs: examples include a battery of questions on health for the provincial Department that was once off; but also on transport for City of Johannesburg, that was repeated. Of course GCRO also entertains requests from its own and other researchers, such as a battery on mental health that was once off, but one on employment that has been sustained. But then there is the difficult decision:

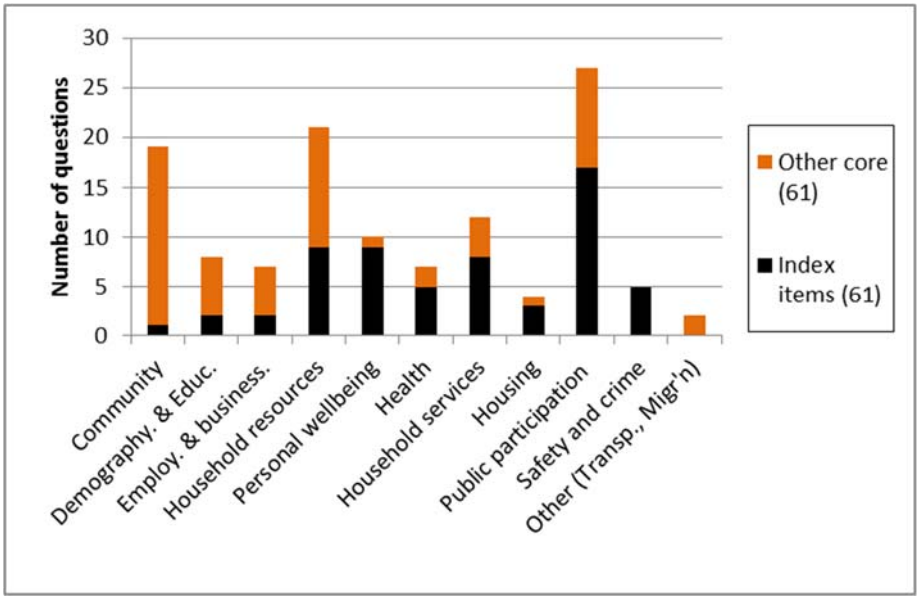
> *Once questions have been included in the questionnaire a few times, their retention is increasingly appealing because they allow for longitudinal comparability.*

A telling example is the battery of 15 items on service delivery, which unsurprisingly have been highly valued by different municipalities' users, and continued ever since QoL II.

The questionnaire includes a core of approximately 120 items that have broadly been sustained over the five QoLs: half being the original, intuitively selected, set of items for the index, approximately quarter being demographic items for analysis, especially of the QoL index, and a quarter being items deemed central from the outset for provincial and government reporting, including the 15 service-delivery items. GCRO has grouped questions into ten categories plus "other". Figure 2 shows the breakdown of QoL index and other core items:

[57] *Loc. cit.,* p. 7 ff.

Figure 2: Breakdown of core content between index items and other items



The expansion has occurred in the sets of additional items, some once-off and replaced, but others retained for two or more QoLs. The outcome is shown in Table 4.

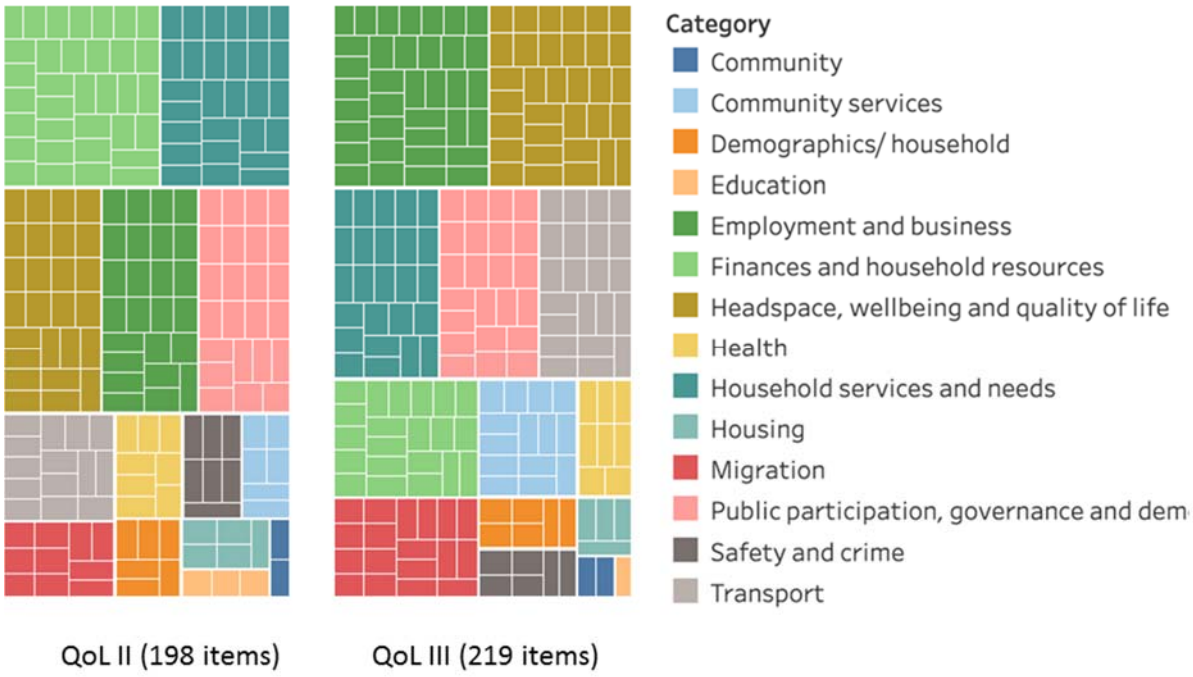
Table 4: Core and additional questions in QoLs I-V

	QOL I	QOL II	QOL III	QOL IV	QOL V
Core content	123	123	123	123	123
Additional questions	64	75	96	108	125
Total questions	187	198	219	231	248

One sees that the non-core component has doubled, from 64 items to 125, up to equalling the core. Depending on what questions were swapped out or retained, the overall balance of topics covers has of course altered. GCRO has illustrated this in an interactive graphic covering all the QoLs. Figure 3 below compares QoLs II and III as an example. The changes are seen to be quite substantial.

Thus, a set of extra Transport questions (grey) was introduced into QoL III, increasing coverage from 14 to 22 items; and likewise Migration (red), from 10 to 18, and Community services from 6 to 14. Other coverage was decreased compared to QoL II, but only slightly, e.g. Finance (pale green) from 24 to 21. So, as Table 4 showed, the number of non-core questions with which respondents had to deal rose from 75 to 96 between QoLs II and III (nearly by a third!).

Figure 3: Example of the changing proportions of themes covered in QoLs II and III



As Stats SA's Niël Roux put it at the third workshop:

> *Everyone wants to add questions... but no one wants to take any away. [After you] provide the client and stakeholder with what they want and are looking for, attempts to move or change questions... have resulted in outcry from external sources.*

4.1.2 Estimating a suitable questionnaire length

How long, then, is it acceptable for a questionnaire to be, and thus, allowing for an average number of "skips" and other contingencies, for an interview to be? Interestingly, as providers themselves have mentioned, the interview length is *not* really a factor for the ultimate duration of fieldwork. As emerged in Section 3, six months in field is more reasonable than the three months on which GCRO had always planned, given how many reliable fieldworkers any likely provider can properly manage per day. And then it turns out that the fieldworkers will average between 1.2 and 3 interviews per day over the duration, as mentioned by providers for QoLs III and II.[58] Either way, a longer or shorter interview could in principle be accommodated.

Rather, the main determining factor is "respondent fatigue": how long, on average, a respondent is prepared to agree to be engaged, and then to proceed without getting impatient and giving hasty or ill-considered answers, i.e. compromising data quality. One should add interviewer fatigue: how long an interviewer can be expected to be patient in each interview, explaining

[58] 120 interviewers per day in field, for five days a week (noting they should work weekends and evenings) over six months, at 1.2 interviews per day yields approximately 19,000 interviews, if no interviews are discarded.

issues and being fastidious in following even intricate questions, often after a sapping effort of securing access, day after day for several months. A longer questionnaire increases the temptation for an interviewer to take short cuts, or mis-record a particular answer so as to avoid a follow-up section.

The actual evidence from the previous QoLs is rather impressionistic. The provider in QoL IV outsourced the fieldwork with scant supervision, and provided little information. The interviewers in QoL V were, notwithstanding their training, too inexperienced to furnish an indication. We may note that QoL I reported a minimum of 45 minutes an interview, and an average somewhat longer, and QoL II reported an average of an hour. And both strongly recommended that it be shorter! This was for 187 and 198 questions respectively, i.e. 20 to 23 pages of typical layout. Indeed, in both these cases the questionnaires were still basically paper-based. Whether transferring to a tablet is faster, or just more accurate, would need to be researched. For QoL V, with as many as 248 questions on tablets, the administration time dropped to just 30 minutes by the end of fieldwork – whether this reflects interviewer familiarity or cheating, or respondent impatience, is not known. Other experience suggests such a drop-off is about 30%, suggesting again that a time at start of 45 minutes was experienced.

These reported times take account of the varying number of actual questions asked to widely differing respondents on average, for the given "skips". And of course there are finer considerations affecting the relation between the number of questions and the length of the interview that were noted at the third workshop: how well the questions were piloted to be well phrased, sequenced, and, ideally, pre-translated; how well the tablet has been programmed for the interviewer, and how it may be used to assist the respondent with multiple answers, and so on. These will all be improved by careful qualitative "behind the glass" testing of questions, and piloting of a modest number of questionnaires in field by supervisors who then give detailed feedback. (Piloting is discussed in Section 3.)

But the prime factor remains the number of questions; and on the evidence above, a maximum of 200 questions (of the broad mix of kinds hitherto used in the QoLs) would seem to be a target that is desirable for both respondents and providers, and affordable to GCRO.[59] If money were tight, even fewer would be welcome, say 180 or even 160.

4.1.3 Options for adapting to a shorter questionnaire

The key manoeuvre to fit within this suggested limit, suggested at the third workshop and based on the extensive experience of both the HSRC and Stats SA, is to use *rotating modules of questions* on top of the core content. But before contemplating how this might be arranged, the GCRO has a prior option, that will allow a more generous use of modules: that is, slimming the number of questions in the core content, enabled by a properly quantified revision of the QoL indexing that is elaborated is sub-section 4.2 below.

There it will be seen that, rather than the intuitively defined, original 10 dimensions, an empirically-based QoL index comprises 7 dimensions. And it requires appreciably fewer than the original set of somewhere between 58 and 61 variables, yet with a gain in statistical coherence: in an early version it used 41 variables, and this number has safely been further reduced to 28, less

[59] Stats SA for its GHS, and HSRC for its SASAS, both report about 250 questions. They also report much higher fieldwork charges per head than GCRO would be able to afford.

than half the original! However, some of the variables that no longer warrant inclusion in the reworked index may well be worth sustaining alongside the non-index core items for, say, their policy interest, and the trends they will continue to evidence. Suppose 15 such ex-index variables are continued.

What about the other 60 non-index core variables: demographics, and items defined and retained as core for governmental use: could they be reviewed and similarly slimmed? Two criteria come to mind. One is to identify those that have been less, or less revealingly, used in the analyses in previous QoL reports. Another is to check, among those that *have* actually been used, which are not in fact statistically noteworthy in the distinctions being made.[60] Together this might save, say, a further 15 core variables.

In that case, the core drops to 90 variables, freeing fully 70-90 variables for non-core use on a modular basis. The approach adopted by the HSRC in SASAS (itself following that of the ISSP, the International Social Survey Programme), reported to the third workshop by Dr Roberts, is this:

> *The rotating modules range from a minimum of 10-20 items to a maximum of 40. Each of these are strongly conceptually grounded, with clear core indicators of their own. The use of these modules is various - some of them appear every other year [recalling that SASAS is annual], while some only appear once every 5 or 10 years*

In addition, the HSRC avoids the inclusion of once off modules in SASAS, by carrying these in a parallel user-pays annual survey primarily for government users, akin to the "omnibus" surveys run for the private sector by market research companies. However, the QoLs, especially in that they occur only two-yearly, would probably want to retain the option of occasional one-off user-pay modules as they have done in the past.

In addition, the HSRC avoids the inclusion of once off modules in SASAS, by carrying these in a parallel user-pays annual survey primarily for government users, akin to the "omnibus" surveys run for the private sector by market research companies. However, the QoLs, especially in that they occur only two-yearly, would probably want to retain the option of occasional one-off user-pay modules as they have done in the past.

On this basis, with the 70-90 variables at play, a future QoL may, for example, carry the following:

(i) Two substantive government modules of, say, 10-15 items each that have assessed as worthy of repetition annually or (for slower-moving phenomena) every other year, to track new developments or priorities such as such housing densification, youth employment opportunities, trust in institutions, perceived municipal responsiveness, or xenophobic attitudes (the previous QoLs embody many examples);

(ii) Two explicitly one-off modules of, say, 10 items each (or one of 20 items) on a user-pay basis, such the health and other modules previously;

[60] Given the very large sample sizes in QoLs, distinctions being reported may well be statistically significant – in the sense of licensing inference to the parent population – while being substantively small.

(iii) Perhaps a suite of up to 10 occasional questions of topical interest established by GCRO;

(iv) Equivalent numbers of (i) and (ii) to be taken up by GCRO researchers or other researchers in SA or abroad.

For any of this to be possible, the existing non-core suites of questions need to revisited, which will involve some tough decision-making. This is taken up now.

4.1.4 Regulating systematic use of the non-core questions

Some important criteria should apply to regulating the use of these non-core options, and quality-assuring their content. This in turn requires a decision-making body to function. The following are only tentative indications to be discussed, rather than recommendations, since they were not considered at the workshop. They would be expected to evolve over time.

Firstly, it really must be remembered by GCRO that QoL is a vast and vastly expensive sample. Candidate modules for inclusion and retention in the questionnaire – including (iii), the occasional topical questions – should warrant the smaller-area detail that QoL enables. If not, they could be executed on *sub-samples*. On the cluster design proposed in Section 2, a module could be included as an extra page in every other or every fourth questionnaire, yielding samples that are still very large, notionally 10,000 or 5,000 respectively, and still with controlled ward-based spatial distribution. There would then be room to have two or four such sub-sampled modules in a given QoL instead of one of the repeated modules in (i) or one of the once-off modules in (ii).

Secondly, guidance on appropriate real-cost pricing of user-pay modules should be established with the help of a financial-cum-project consultant, mindful of commercial rates. (The "omnibus" surveys obviously have done this... and they aren't cheap!)

Thirdly, GCRO's existing processes could continue to apply for selecting, indeed soliciting, provincial and municipal take-up for (i) and (ii). Some existing suites may need discontinuing, after discussion with their principals, to allow for new modules.

However, for (iv), the modules (or sub-modules) available to GCRO's and other researchers, *competitive selection processes* to run well in advance will need to be designed and implemented, for the opportunities to be fairly and peaceably made available. Again, the following are tentative suggestions for consideration, not recommendations, since they were not discussed either:

- Selection might centre around a GCRO panel chaired by the Executive Director, with one representative each nominated by the Research DVCs of Wits and UJ, the GCRO lead QoL researcher, another GCRO senior researcher assigned by the Executive Director, and a prominent external researcher familiar with GCRO matter.
- A sub-committee of the panel (e.g. the GCRO lead QoL researcher and their GCRO colleague) could short-list applications to be sent for an external reviewer report; and then assess, perhaps with the assistance of the external researcher, which finalists are to be put to the panel for final award. The applications would be accompanied by a short motivation, and a copy of a prior publication if relevant. If the GCRO people are themselves applicants, they would recuse themselves as necessary.

- Continuation of existing non-core suites of researcher questions, including repetitions, should be similarly revisited; otherwise there is no room for fresh modules. Their being repeated could be made subject to published research outputs based on the first insertion of the questions
- The GCRO applicant researchers may perhaps not have to face a cost. The external researchers could well expect to have had to raise grant money to pay GCRO, perhaps for a reduced price.
- Proposals accepted that include repeat implementations of modules could be made subject to review by the panel or its sub-committee after the first cycle.

The competitions for modules need to be advertised and processed adequately in advance, so that GCRO can require the successful ones to undergo the qualitative and quantitative piloting phases. There will not be time for this for QoL VI. But requests or solicitations for modules, with motivations, can still be adjudicated by the panel described above.

Moreover, a draft slimming of the additional modules or suites currently comprising the 125 additional questions, and indeed of the 60 non-index core items, can certainly still be undertaken by the QoL lead researcher, and put to a mini-version of the wider panel, in time for QoL VI.

4.2 Index construction

The QoLs are the flagship project of GCRO, and their most prominent analysis and report has been through the QoL (Quality of Life) index. Unlike the design of sampling covered in Section 2 and the management of fieldwork covered in section 3, the four key aspects of producing the index have remained unchanged from QoL 1 onwards with negligible variation. These are: the selected core set of 58 questions; their intuitively-based arrangement into 10 dimensions; the coding of responses to the questions into simplified form, i.e. numerical variables; and a simple, equi-weighted additive approach to aggregating the value of each dimension and the final composite score. This indexing has won the attention and enthusiastic support of provincial and municipal governments and other users, and now is widely disseminated in the media by the Premier of Gauteng Provincial when it is launched every second year.

The purpose of this part of the review has been to assess the soundness of this four-step indexing approach. However, the enquiry, as outlined in this section, actually unfolded in a different order. It began by interrogating the intuitive conception of the dimensions, using the more orthodox and statistically founded method of index composition by factor analysis.[61] This suggested that most of the dimensions had deficient internal correlations. A corollary was that statistical model fit could not be achieved for the overall 10 dimensional model. Instead, the factor analysis yielded a coherent new index configuration that differed in six important respects. An evidence-based, stable and evocative composite model comprising *7 dimensions* emerged. This involved a *reduced number* of contributing variables, down from 58 to 33 significant items, *assigned differently* in notable instances to the revised dimensions (and allowing 25 weaker and unneeded items to be culled).

Moreover, in conducting the empirical investigation, it was found that the *coding of the variables* needed correction, and that their original *richer formats* could preferably be retained. Finally, the

[61] OECD, *Handbook of Indicator Construction* (Paris: OECD, 2010), p. 21.

empirical factor-analytic approach offers more nuanced options than the previous equi-weighting for aggregating *the index results*, respecting the differing strengths of the variables contributing to each dimension, and then on the considerably differing contributions of the 7 revised dimensions to the single score.

The exploration has been conducted on the QoL V data, and extended to QoLs IV and III, by a work-team of Dr de Kadt and Samy Katumba of GCRO, Professor Fatti, and the author. At the urging of Graeme Gotz at the third review workshop, other potentially relevant variables available in the wider QoL selection, that were not included the original intuitive choice of 58, were brought in for consideration. In fact 4 of them proved to be relevant and interesting. In particular, some of them empirically combined with other variables into a strengthened factor of the socio-political kind that had proved statistically insignificant in the initial transition to an empirical solution.

This has now allowed the calculation for the first time of scores for the 7 data-driven dimensions that are weighted by their importance in the views of the respondents; and likewise their weighted combination into the single-score QoL composite index measure. This new 7 -factor evidence-based solution has been tested backward into QoLs III and IV, and performs well across these datasets too. This provides confidence that it may be used as the basis for a revised QoL index into the future.

The presentation below will unpack the original intuitive 10-factor index, and show the empirical assessments thereof, in sub-section 4.2.1. Then in sub-section 4.2.2 it will outline the tidying of the variables; the model choice and slimming of the empirical 7-factor solution, in its updated version, and how that product relates to the original intuitive solution. Finally, sub-section 4.2.3 will set out the corresponding procedures of aggregating the dimensions and the single score, in ways that take advantage of the refined empirical data. Particularly relevant from the literature will be the OECD's *Handbook of Indicator Construction,* and a very recent review article of the literature.[62]

4.2.1 The existing, intuitively-based QoL Index definition of dimensions and approach to aggregation

The genesis of the extensive range of questions that was included in QoL I is described in section 4.1. The then-Director of GCRO and a consultant drew upon a mix of 58 subjective and objective items to create a QoL index, while grouping them into 10 dimensions, as shown in Appendix F. The dimensions were "developed and refined" from two previous South African studies.[63] They overlap with those in the literature on "well-being", as overviewed, for example, in the report of the Stiglitz Commission.[64] Most of the dimensions comprise 4 or 5 items; Infrastructure and

[62] OECD *Handbook,* op. cit.; and S. Greco *et al.*, "On the methodological framework of composite indices: A review of the Issues of weighting, aggregation, and robustness" *Social Indicators Research* 141 (2019), pp. 61–94.
[63] David Everatt, "Quality of life in the Gauteng City-Region, South Africa", *Social Indicators Research* 130 (2017), pp. 71-86, p. 81.
[64] The Stiglitz Commission list is: health, education, their everyday activities (which include jobs and housing), their participation in the political process, the social and natural environment, and the factors that shape their
personal and economic security.

Work have 7 items; and the Socio-economic dimension has 12. This structure was been retained all the way through to QoL V, to allow comparability.

An alternative, empirical approach to grouping candidate items into dimensions was subsequently commissioned from Talita Greyling,[65] now a professor and an authority in the field. It used principal component analysis (PCA), a cognate (in its outcome) of factor analysis. Based on the responses of interviewees PCA (i) groups items into independent dimensions, (ii) shows relative strengths of items within in each dimension, and (iii) shows the relative strengths of the dimensions, in accounting for the myriad variations in the survey data. The weighted aggregation of items within dimensions is yielded by (ii), and the weighted aggregation of the dimension totals into the single QoL score is yielded by (iii). In this way, the successive aggregations are based on the scored answers of the numerous respondents, rather than the developer's intuitions. After some consideration of whether any particular dimensions, e.g. employment, should indeed be weighted up because of their importance in the South African context, it was reportedly tested for QoL I "whether or not PCA... generated results that were significantly different from the unweighted approach. The simple answer was 'no'."[66]

Instead, the decision taken at the time was to do a simple equi-weighted sum of the items in each dimension. An immediate problem for this method was that the scoring of the items was mixed: most were five-point Likert scales, from "strongly disagree" to "strongly agree" or equivalent, and most of the balance were "no or yes" dichotomies. For comparability, therefore, all the items were dichotomised as 0-1. The number of resulting dichotomies varied among dimensions. So, after they were summed, the totals were scaled to run from 0 to1. Then the 10 separate dimension totals could be summed to yield a single-score QoL index running from 0 to 10.[67]

The 10 dimension totals, between 0 and 1, are compared for the 5 QoLs in Table 5 below, as gathered from Appendix F. The bottom line of the table shows that, from QoL II onwards, there was a steady increase in the single-figure QoL index from 6.0 to 6.3, albeit by only 0.1 in a scale of 0 to 10. Comparing QoL IV to QoL V, for example, one sees that the increase represented a mix: deterioration on 5 dimensions and improvement on 4, with 1 unchanged.

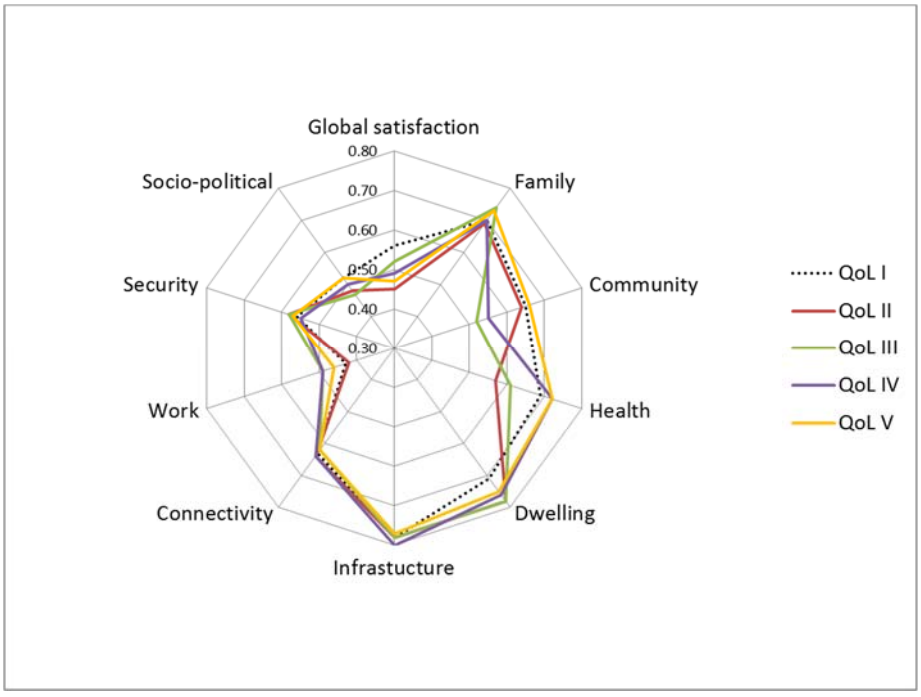
[65] T. Greyling, "A composite index of quality of life for the Gauteng city-region: a principal component analysis approach", GCRO Occasional Paper No. 7, 2010.
[66] David Everatt, "Quality of life in the Gauteng City-Region", *op. cit.,* p. 82.
[67] The approach in the text was offered as an alternative from QoL IV. The initial approach from QoL I was simply to add all variables together and then scale the total to a score out of 10. There are scales in psychology with dozens of items. But they deal with a particular construct, e.g. PTSD. In this case, given the substantive variety of the dimensions, first to score each dimension is sensible and revealing.

Table 5: The ten summed dimensions, running from 0 to 1, for the five QoLs

	QoL I	QoL II	QoL III	QoL IV	QoL V
Global satisfaction	0.56	0.45	0.52	0.49	0.47
Family	0.70	0.69	0.74	0.70	0.73
Community	0.65	0.64	0.52	0.55	0.66
Health	0.69	0.57	0.61	0.72	0.72
Dwelling	0.71	0.78	0.78	0.76	0.75
Infrastucture	0.78	0.78	0.78	0.80	0.77
Connectivity	0.63	0.64	0.62	0.64	0.62
Work	0.43	0.42	0.49	0.49	0.46
Security	0.56	0.58	0.58	0.55	0.57
Socio-political	0.52	0.48	0.47	0.50	0.52
Average	**0.62**	**0.60**	**0.61**	**0.62**	**0.63**

The body of the table is rendered more revealing by a polar graph, Figure 4 (if it is reproduced in colour!). If we continue to set aside QoL I (shown by a dotted line), as having preceded the use of by-ward sampling, one sees that the nearly all the dimensions of the four QoLs (which run from 0 to 1) coincide within 0.1, or even less. The most notable exceptions are that Qols III and IV are substantially lower than the other two QoLs on the dimension of Community, as are QoLs III and II on Health. But on inspecting the item-level detail of the respective questionnaires, one sees that this may be due to changes of wording, or non-inclusion of items.

Figure 4: Mean values of intuitive 10 dimensions for QoLs I-V



4.2.2 Limitations of the existing, intuitive QoL indexing

This overall approach to QoL indexing, as was noted, has been in place since QoL I. But some reservations of increasing, and finally decisive, seriousness needed to be addressed.

Firstly, if the variables, dimension and weights are "chosen by the developer of an index, naturally this means that it is conceived as 'subjective' since it relies purely on the developer's perceptions."[68] The attempt, secondly, to avoid this complication by equal weighting of items and dimensions is not a defensible remedy. According to the literature, "This is obviously convenient but also universally considered to be wrong."[69] Indeed, it is argued that making this choice, as against using interviewees' responses to assign weights – weights which may *or may not* be equal – is itself subjective.[70]

Thirdly, collapsing the many 5-point Likert scales into dichotomies, as required by the equi-weighting within dimensions, is a sub-optimal use of the information so carefully solicited from respondents. In particular, "Individuals close to but on opposite sides of the cutpoint are characterised as being very different rather than very similar", and this made lead to "spuriously significant results."[71]

More serious, fourthly, is that missing responses – an unavoidable feature of surveys – have been inappropriately coded in the preparation of the data from QoL I onwards. They were assigned the value of zero, but not identified as "missing" in the calculations. So they have functioned as an extreme bottom score in the Likert scales, and as the lower of the two valid answers in dichotomies, biasing these scores downwards. The result will have been erroneous outcomes of unknown extent within variables, the dimension-indexes, and probably the single-score. This has to be corrected.

The most serious problem, finally, is that the majority of the ten intuitively defined dimensions do not meet a critical requirement: adequate inter-correlation among the component items of each dimension. To constitute an acceptable summative scale, the component items must be meaningfully correlated with each other (albeit not too strongly, in which case certain of them are redundant). To check for this, the OECD *Handbook* advice is "to use the Cronbach coefficient alpha, which is the most common estimate of internal consistency".[72] Additionally, all candidate dimensions can then be tested concurrently in an overall model, using confirmatory factor analysis (CFA), to see if the chosen assignment of the items to the respective dimensions empirically "fits".

Below, in Figure 5, are the results for the Cronbach's alpha on the intuitive 10-dimension solution, compared with an empirical 7-dimension solution that has a rearranged subset of the same items. Alpha is expected to exceed 0.7 for a sound scale in psychology, but since we are dealing with the gamut of socio-economic items, we have allowed for a marginal zone down to

---

[68] Chowdhury, S. and Squire, L. "Setting weights for aggregate indices: An application to the Commitment to Development Index and Human Development Index", *Journal of Development Studies*, 42-5 (2006), 761-771.
[69] *loc. cit.*
[70] Greco et al. *op. cit.*
[71] Douglas G Altman and Patrick Royston, "The cost of dichotomising continuous variables" *BMJ* 332 (2006), p. 1080.
[72] OECD *Handbook*, op. cit., *p.26*

0.6. One sees that only 1 of the 10 intuitive dimensions makes even the marginal cut. This deficiency was carried forward by the retention of the intuitive 10-dimension model in the ensuing four QoLs. In the 7-dimension solution, only one dimension does *not* make the cut; three are good, and three are in the marginal zone.

Figure 5: Cronbach's alpha test of internal consistencies for the intuitive 10-dimension and the empirical 7-dimension models

CRONBACH'S ALPHA* PER FACTOR TEST OF INTERNAL CONSISTENCY FOR INTUITIVE 10-FACTOR AND EMPIRICAL 7-FACTOR MODELS

Less than 0.600	
0.600 to 0.699	
Greater than 0.700	

	Intuitive 10-factor model		Empirical 7-factor model		
	Factor	alpha	alpha	Factor	
Global life satisfaction	g1-g4	0.267			
Family	f1-f5	0.538	0.643	f2-f4, c2, w3	Family
Community	c1-c4	0.133			
Health	h1-h5	0.535	0.755	h1-h3	Health
Dwelling	d1-d5	0.531			
Infrastructure	i1-17	0.543	0.779	d3, i3-i6, e4	Infrastructure
Connectivity	e1-e5	0.380	0.661	e2, e5, w5, w6, h5	Socio-economic status
Work	w1-w7*	0.104			
Security	s1-s5	0.625	0.627	s1-s4, c1	Security
Socio-political attitudes	p1-p12	0.53	0.759	p6-p9, r5	Satisfaction with government
			0.402	c4, p10, p11, r4	Participation
	58 items			33 items	

* M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha", *International Journal of Medical Education 2:2011, 53-55.*

As for the additional, whole-model test, the 10-dimension version, with its assignment of items to the dimensions, could not be induced to fit, indicating that the model is "mis-specified".[73] When, in the next sub-section, a model that does comfortably fit is generated empirically (with the 7 dimensions, as it transpires), two evidently coherent dimensions do largely carry over from the intuitive model - health (in part) and security (as a whole). But the technique signals that some in the intuitive dimensions actually belong in other dimensions; and a good number are found not to contribute significantly anywhere in the new model, i.e. to vary randomly with respect to the dimensions in respondents' minds.

[73] In a CFA, a model is said to be mis-specified if more than a handful of items present more significant correlations with dimensions outside the one to which they have been assigned. R.B. Kline, *Principles and Practice of Structural Equation Modelling*, 3rd edn (New York: Guilford, 2011), p. 240.

4.2.3 The sequence of establishing a defensible empirically-based model

As indicated above, there are five necessary steps in the sequence towards defensible QoL dimensions and their summation: [74]

(i) treat missing values correctly;

(ii) reinstate into the analysis the richer 5-point Likert scales that were widely used on the actual questionnaire;

(iii) demonstrate in QoL V how an empirically driven, well fitting, model that turned out to have 7 dimensions is extracted by exploratory factor analysis (EFA), in comparison with 8, 9, and 10-dimension options;

(iv) following the Greyling approach, establish and apply the weights that follow from the factor analysis for aggregation of items into dimensions, and aggregation of dimensions into the single QoL score, ;

(v) validate the QoL V model in the datasets of QoLs III-IV (with some back-and-forward adjustments if necessary), then repeat step (iv), and display the historical trends.

Before setting forth, time was spent on resolving a suitable QoL dataset. The hope of simply pooling of the thousands of records across QoLs was dashed by Professor Fatti. He pointed out that what would properly be required was a cell-by-cell weighted pooling of the QoLs' vast correlation matrices. This would have been an intricate and error-fraught exercise. Instead, it was decided to work on QoL V first, and then check whether the the resulting model adequately fitted the data from earlier QoLs.

Steps (i) and (ii) above, are interlinked. As noted in the previous sub-section, from QoL I onwards the many Likert scales were recast as dichotomies (usually the two positive Likert options versus the rest), in order that they could be simply summed with actual dichotomous "no/yes" questions of each respondent, towards yielding a score on the particular dimension. Missing values were uniformly coded as 0 – equivalent to a negative response. This will have erroneously boosted the proportion of negative responses.

In correcting this, firstly the five-point Likert scales were retained in the calculations.[75] Then two kinds of missing values were distinguished. *Structurally missing* values (arising from "inapplicable" questions, such as a question on marriage satisfaction for single people) were set to the neutral mid-point for Likert scales, "where they do least harm". Likewise, any dichotomies with missing values were converted to trichotomies with the missing value set at the centre. For values which could be classed as *missing at random* (i.e. where respondents said "Don't know" or declined to answer, or the fieldworker simply overlooked a question), general practice in

[74] The OECD *Handbook* distinguishes ten steps. The last three concern the application of an index. More relevant are the first seven, which are subsumed into the sequence followed here: Framework, data selection, imputation of missing data, multivariate analysis, normalization, weighting, and sensitivity analysis.
[75] At the same time, the Likert responses were recoded to run from 0 to 4, instead of the initial 1 to 5, for processing alongside dichotomies that were coded 0 and 1.

statistical agencies was followed:[76] the non-responses were imputed using non-parametric multiple imputation, in this case using a routine in the open-source programming language R.[77]

Step (iii) is next, the empirically based sorting of questionnaire items into dimensions to which they significantly attach (and the setting aside of items that in fact aren't significantly differentiated among dimensions). As noted in sub-section 4.2.1 above, at the time of QoL I Greyling's statistically based method was canvassed but was abjured in favour of the developer's intuition. The method that was actually subsequently demonstrated on the data by Greyling[78] would have used the technique of principal components analysis (PCA). The current approach is similar, but slightly more ambitious in three respects.

Firstly, the cognate method of exploratory factor analysis (EFA) has been used instead of PCA. The dimensions extracted in PCA are by definition "orthogonal", or uncorrelated – which is rather artificial given the way, for example, socio-economic status is in fact correlated with health. In EFA the dimensions in the model may sensibly be recognised to be empirically correlated, by invoking what is called an "oblique" rather than orthogonal relation among the dimensions.[79]

Secondly, instead of a prior selection among potentially significant variables in the light of the literature, as Greyling recommended before running the PCA,[80] the current exploration has brought into reckoning the wider range of variables of potential local relevance that were usefully included in the QoL I questionnaire.

Thirdly, the PCA was initially demonstrated on the fully dichotomised dataset.[81] The current approach has, as noted above, retained to the richer, mainly Likert-scaled form of the raw data, alongside the actual dichotomies, and has also used the appropriate variant of EFA.[82]In conducting an EFA, one stipulates models with different numbers of dimensions for inspection, in order to adjudge the most satisfactory model according to statistical criteria and conceptual clarity. If too few factors are specified for the model being considered, items will be thrust together that are not conceptually coherent; if too many factors, the last few factors will be seen to share too few items, that moreover only "load" weakly onto the factor. This is less subjective than it sounds: the obvious factors, with their strongly loading items, tend to emerge very insistently among the models, with the most strongly loading variables indicating the factors'

[76] For instance, Statistics South Africa, *General Household Survey* (Pretoria: Statistic South Africa, 2017), p. 68.
[77] This was, reassuringly, found to yield the same results as when the imputation was carried out by default in Mplus in the course of checking the factor analysis.
[78] Greyling, "A composite index", at n. 65.
[79] The choice is debated in Greyling, "A composite index", *op. cit.,* pp. 13-14. EFA has the additional explanatory connotation that the dimensions are effectively "underlying constructs" that dispose the observed items, as against the mere linear combinations of the observed variables to account for maximum variance, in PCA. See A. Afifi *et al.*, *Practical Multivariate Analysis 5th Edn* (Boca Raton, CRCPress, 2012), pp. 379-80.
[80] *Ib.*, p. 9: "Selection of the measuring items to be included in a composite index of quality of life was guided by the reviewed theoretical approaches and the reviewed current quality of life measures."
[81] In a later paper, Greyling revisits the QoL I data with a categorical variant of PCA more attuned to dichotomies: Talita Greyling, "An analysis of the quality of life of migrants in Gauteng, a province of South Africa", *Journal of Economic and Financial Sciences*, 8:2 (2015), pp. 495-517.
[82] This is "polychoric" EFA, which is designed for the dichotomies and the ordinal Likert scales.

conceptual content. There are also the statistics of how strongly the respective dimensions contribute to the overall model, and how well the overall model fits; though these too are only guidelines.[83]

Table 6: Comparison of 7- to 10-dimension exploratory factor analyses (EFAs)

	Label	Indicator	7 factor	8 factor	9 factor	10 factor	Reduced 7 factor		Legend
Infrastructure and Services	d3	Dwelling structure made of bricks or concrete	0.822	0.815	0.806	0.813	0.800		Factor 1
	i3	1.7 - Flush toilet facility	0.934	0.938	0.940	0.920	0.951		Factor 2
	i4	1.4 - Piped water source	0.655	0.657	0.662	0.660	0.675		Factor 3
	i5	1.9_Electricity supply	0.836	0.823	0.828	0.820	0.827		Factor 4
	i6	1.8 - Rubbish disposal	0.780	0.782	0.781	0.789	0.793		Factor 5
	e4	5.7_3 - Household assets: Television	0.648	0.642	0.640	0.645	0.657		Factor 6
	d1	1.17 - Satisfaction: Dwelling	0.377	0.374	0.365	0.373			Factor 7
Socio-Economic Status	h5	11.6 - Personally covered by medical aid	0.842	0.838	0.829	0.836	0.895		Factor 8
	e2	12.1 - Highest level of education completed	0.670	0.665	0.656	0.662	0.662		Factor 9
	e5	5.7_5 - Household assets: Internet connection	0.583	0.580	0.553	0.570	0.547		Factor 10
	w5	Employment Status	0.493	0.506	0.522	0.507	0.491		
	w6	12.10 - Total monthly income all household members	0.802	0.799	0.799	0.782	0.850		
	f5	5.9 - Children skipped meal past year?	0.486	0.494	0.498	0.486			
	e3	5.7_1 - Household assets: Telephone or landline	0.390	0.381	0.376	0.379			
	w4	8.25 - Satisfaction: Working conditions in your job	0.422	0.420	0.424	0.425			
	p12	6.24 - Asked to pay bribe to govt. official?	-0.361			-0.351			
Government Satisfaction	p6	6.6 - Satisfaction: National government	0.815	0.811	0.824	0.825	0.848		
	p7	6.7 - Satisfaction: Provincial government	0.850	0.869	0.880	0.874	0.902		
	p8	6.8 - Satisfaction: Local municipality	0.664	0.671	0.675	0.669	0.676		
	r5	8.14 -Government that improved quality of life	0.581	0.582	0.582	0.578	0.553		
	p9	6.23 - Most government officials are doing their best?	0.451	0.449	0.425	0.433	0.429		
Life Satisfaction	f2	7.4 - Satisfaction: Family	0.678	0.675	0.669	0.687	0.658		
	f3	7.2 - Satisfaction: Time to do things you want to do?	0.603	0.602	0.592	0.605	0.615		
	f4	7.7 - Satisfaction: Leisure time	0.645	0.644	0.634	0.638	0.664		
	c2	7.5 - Satisfaction: Friends	0.530	0.535	0.527	0.527	0.547		
	w3	7.6 - Satisfaction: Standard of living	0.398	0.394	0.415	0.415	0.360		
Safety	s1	9.3 - How safe walking in your area during the day?	0.735	0.783	0.707	0.763	0.764		
	s2	9.4 - How safe walking in your area after dark?	0.555	0.524	0.552	0.547	0.556		
	s3	9.5 - How safe do you feel at home?	0.623	0.650	0.613	0.636	0.620		
	s4	9.1 - Crime improved or got worse?	0.559	0.533	0.558	0.552	0.534		
	c1	6.43 - People in your community can be trusted	0.467	0.449	0.469	0.464	0.461		
	s5	9.2 - Have you been a victim of crime in the past year?	0.368		0.366				
Health Status	h2	11.16 - Health status prevent from doing daily work?	0.910	0.944	0.778	0.787	0.899		
	h3	11.17 - Health prevent from usual social activities?	0.908	0.928	0.780	0.785	0.898		
	h1	11.15 – Health status in the past 4 weeks	0.440	0.409	0.472	0.474	0.464		
Political Engage	c4	10.1 - Participated in the activities of any clubs	0.412	0.367	0.384	0.400	0.443		
	p10	6.4_3 - Attended Community development forum	0.663	0.678	0.633	0.656	0.711		
	r4	8.22 - Communication with municipality	0.521	0.544	0.541	0.550	0.548		
	d4	1.3 - Dwelling owned		0.464					
	p11	6.2 – Planning to vote in the 2016 local election	0.499	0.569	0.543	0.522	0.459		
	p4	6.40 - Blacks and whites never really trust each other			0.919				
	r1	5.05 - Travel time to most frequent destination			0.791	0.918			
	e1	6.39 - The press is free to say or write what it likes				-0.693			
	p5	6.42 - Positive attitudes to foreigners in Gauteng				-0.737			

[83] One guideline is that dimensions with an "eigenvalue" less than 1 are candidates to be discarded: Afifi *et al., op. cit.,* p.363. These may be inspected on a "Scree plot". And it is hoped that models will have satisfy overall measures of fit called RMSEA, to be < 0.05, and CFI and TLI > 0.95. But there is the deceptive feature that including more dimensions will yield marginally better fit statistics, even as they have eigenvalues less than 1 and erratic mixes of items.

So it was with the EFA on the QoL V dataset. The analysis included for empirical consideration the original intuitive QoL I selection of 58 variables and a half dozen others of potential relevance, such as transport and political engagement. The models for 7, 8, 9 and 10 dimensions are displayed in Table 6.

Several things are evident. First, the first seven dimensions are very robust in all four solutions (the columns), with variations among them of only one or two isolated items (such as feeling that elections will be free and fair, and being asked for a bribe). Secondly, while an additional dimension or two seems stable at the bottom of both the 8- and 9-dimension columns, these dimensions have only three items, some of the loadings are below the conventional cutoff of 0.35, and they are not obviously conceptually coherent. So they do not provide reason for moving above 7 dimensions. And in the 10-dimension solution, the last dimension is a melange. Taken together, this reasoning would suggest the 7-dimension solution is best.

Moreover, the dimensions of this model are sound when tested separately for Cronbach's alpha, as indicated to the right in Figure 5 in sub-section 4.2.1 earlier. And the model as a whole has good fit according to the three customary measures for EFA,[84] whereas it will be recalled that the intuitive 10-dimension model could not be brought to fit at all. This 7-dimension solution thus accords well with the OECD *Handbook* demands.

It is also seen that it involves 41 variables. The further 17 that were intuitively included in the original 58-variable, 10-dimensional model turn out to be too weakly related to any dimension to come into contention. Moreover, if questionnaire length is an issue, the 7-dimensional model can be trimmed further without loss of insight. One may retain items that have loadings of strength greater than 0.5, and then add back in variables with loadings of between 0.35 and 0.5 to a minimum of five variables for factor (which is desirable for the stability of factors). This takes the 7-dimensional model down to a much handier 33 variables. This was welcomed in the section 4.1.3, on addressing bloating of successive QoL questionnaires.

There is, however, a countervailing consideration. Some of the trimmed variables concern issues of high salience in Gauteng communities. Items such as water, poverty, crime and uncaring authorities have been prominent in community, labour, and service-delivery protests. They could well be retained by GCRO as questions for separate reporting.

The 7-dimensions of the new solution, printed vertically in the leftmost column of Table 6, are: Infrastructure and services, Satisfaction with government, Political engagement, Life satisfaction, Health status, Socio-economic status and Safety. A defender of the intuitive model might ask: since four or five of these dimensions seem broadly similar to dimensions in the intuitive model, how importantly different is the new model?

The firm answer is this: prompted statistically by the patterning in the million-odd discreet interviewee responses, it is the changes in assignment of items by the factor analysis into the three newly conceived dimensions, but also in alterations of the four similar-seeming dimensions, plus the dropping of many redundant items that give the evidence-based 7-dimension model a defensible empirical fit, in contrast to the empirically unfittable intuitive 10-dimension model.

[84] RMSEA = 0.045 (robust & standard), CFI = 0.950 (robust); 0.969 (standard), TLI =0.945 (robust); 0.966 (standard).

Figure 6 below shows these crucial re-assignments and removals of variables. For example, we see in the figure that – in the experience of respondents – having access to the internet is associated with secure middle-class status, alongside high education and income and being employed, rather than with an abstract dimension of "Connectivity" in the old schema; and likewise, having a dwelling of bricks is associated with good municipal services, rather with an abstract dimension of "Dwelling" in the old schema. The re-assignments, reflecting respondents' overall thinking, make startling good sense when one notices them.

The factor analysis also supplies the numerical information for **step (iv).** This has two stages, namely a nuanced, empirically based method of aggregating each respondent's scores within each of the seven dimensions, and then compounding these scores for the QoL single-score. Less numerically enthusiastic readers may prefer to skip this paragraph and the next. There are various options of arriving at a dimension score for each respondent. Greyling's has the advantage of being precedented in OECD work, and tried and published on GCRO QoL data.[85] Reverting back to Table 6 one sees, say for the Health dimension, the three respective "loads" of the separate items. In the reduced 7-factor model, on a scale of 0 to 1, the item loading of 0.899 betokens that the answers across all respondents to "Health status prevents me from doing my daily work" (whether agreeing or disagreeing) are manifested as being more strongly correlated than the answers to "Health status in the past 4 weeks" (whether agreeing or disagreeing), which has an item loading of only 0.464.

After some necessary arithmetical transformation, these loadings provide the correct weights to apply to each respondent's distinctive answers, that are then summed to yield or her/his score between 0 and 1 on that dimension.

The second stage does something similar, to weight the dimension scores in aggregating them to the single-value QoL score. The dimensions are shown by the factor analysis to have differing strengths in securing the fit of the overall model: for instance, the Infrastructure and Services is much more influential in explaining the patterns in the data than the Safety dimension. These strengths are expressed in what is called the "eigenvalue" of each factor. Again with some necessary arithmetic transformation,[86] these are applied as weights to each respondent's dimension scores, that are then summed for her/his single-score QoL index value.

The two stages yield dimension scores and single-score index values that may be averaged across all respondents in Gauteng; or across residents of a chosen municipality, or community; or with different demographic breakdowns thereof. The remaining **step (v)** also has two stages. Firstly, it has to be checked whether the overall 7-dimensional model that was so clearly established in QoL V also fits the data from QoLs III-IV, after the errant coding of missing values has been corrected! This provides a demanding check of the robustness of the model, on similarly gathered but independent datasets.[87]

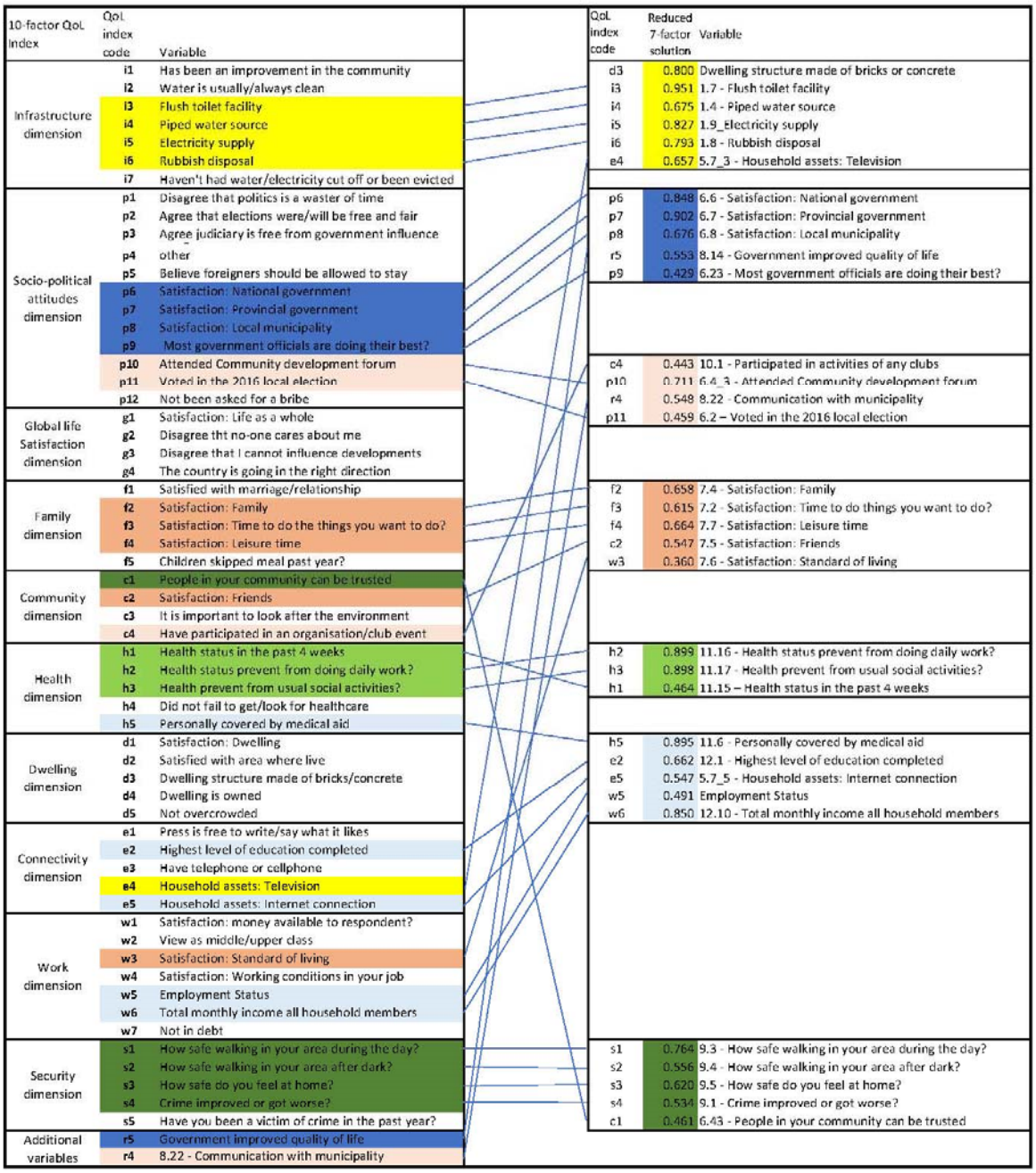
---

[85] Greyling, "A composite index", *op. cit.* See also Talita Greyling, "An analysis of the quality of life of migrants", *op. cit.* Nicoletti, G., Scarpetta, S. & Boylaud, O., 2000. *Summary indicators of product and market regulation with an extension to employment protection legislation.* Available at: ww.oecd.org/eco/eco.

[86] The weight applied to the 0-1 score for each dimension is its eigenvalues as a proportion of the sum of all the eigenvalues of the retained factors.

[87] Attempting CFAs with QoLs II and I was not yet attempted by the time of this Report, because of the demands of first correcting their erroneous coding and conducting the necessary missing-value imputations.

Figure 6: Comparison of intuitive 10-dimension and empirical 7-dimension models and their items



This was done by Confirmatory Factor Analyses (CFAs). It was expected that some to-and-fro tweaking in respect of this or that variable might be necessary, which may, have been used different phrasing. However, the CFAs were successful. This means that the within- and across-dimension weights established for QoL V can validly be applied to the records of every respondent in the earlier QoLs, and trends over time calculated and graphically displayed for Gauteng, municipalities etc.

This business of analysing questionnaire items into dimensions, and the creating and summing of weighted scores within and across dimensions, is taken seriously for good reason. In the OECD, certain budgets are allocated in relation to index scores to member countries. So, even

more intricately adjusted methods have been investigated and applied. These could be a topic for advanced further research on the existing QoL datasets in the future. For the moment, it is recommended that the move from intuitively assigned items, with its equi-weighting within and across dimensions, to the empirically based selection and assignments of items and weights (and the corresponding recalculation of QoL V, followed by Qols IV and III, as a foundations for the indexing of subsequent surveys) is already one giant leap for GCRO.

4.2.4 Recapitulation, and recommendations on index construction

This sub-section on indexing has differed from those on sampling and fieldwork in that, historically, there was only one approach that had been used in all five QoLs. This involved an intuitive selection of 58 questionnaire items to be included and the 10 dimensions into which they were assigned; dichotomizing of items from Likert scales; and then summing within and across dimensions. This was shown in sub-section 4.2.1 to be sub-optimal: the 10 intuited dimensions did not satisfy the internal consistency demands for any scale, and the overall model could not be brought to fit a confirmatory factor analysis, suggesting a substantial misallocation and redundancy among the items.

The alternative here recommended is, in the large, similar to one that was canvassed at the time of QoL I, but not adopted, apparently because it was tried and found to make no difference,[88] and was more intricate. The latter is undoubtedly true. But given that the analysis is handled by computer in any case, once the within- and across-dimension weights are established and added to each respondent's record, the process proceeds in the background, and analyses can proceed as usual of producing dimension and one-score results for Gauteng and the separate municipalities, with breakdowns by demographic and related other variables.

One great advantage of the method undertaken – tested in the OECD and in successively more refined analyses of QoL I data by Professor Greyling in South Africa – is that the selection, assignment and weighting of indicators is based on orthodox statistical analysis of the patterning in the responses of thousands of respondents to dozens of questions, rather than on the intuition of one or two index developers. The method is significantly less intricate than other statistical approaches reported in the OECD literature. Thus, in lieu of a theoretical review of these other possibilities, this OECD/Greyling method was re-tested in practice as part of this Review, but using the much larger and more recent QoL V dataset.

En route, it was uncovered that the problem – common to all surveys – of missing values in the data had been incorrectly handled from QoL I onwards. For the test, this was first remedied for QoL V. Thereafter, the analysis was conducted largely following the chosen OECD/Greyling precedents, with some minor preferred variations. Unsurprisingly, the outcome of the technique – a robust 7-dimensional model – did satisfy the internal consistency demands for its separate dimensions; and the overall model did meet the requirements of empirical fit.

The other great advantage of this method is the effective halving of the number of indicator variables required, by the shedding – from the indexing - of those that did not bear strongly or at all on the empirically uncovered number of dimensions. Even if some of the shed variables are retained for their descriptive interest, this has been a substantial contribution to reducing the

[88] Everatt, *op. cit.,* at n. 63.

questionnaire length, minimising respondent and interview fatigue, and thereby enhancing quality.

The decisive confirmation of the new and leaner index structure from QoL V in QoLs III and IV (after correcting the recoding of their data) has meant that analytic approach is ready for QoL VI. The recommendations of this sub-section can thus comprise the steps to be followed:

(i) treat missing values correctly ;

(ii) retain in the data analysis the richer 5-point Likert scales that were widely used on the actual questionnaire;

(iii) test by CFA whether the existing empirically driven, 7-dimension model fits the QoL VI data, with minor adjustments if necessary

(iv) following the OECD/Greyling approach illustrated above, apply to respondents' records the new weights for aggregation of items into dimensions, and of dimensions into the single QoL score;

(v) compute the within-dimension scores and the across-dimensions single QoL score, and thereafter draw the trends with previous QoLs for Gauteng, municipalities etc.

A final recommendation regarding revised indexing is actually beyond the scope of this review. The communicating of the new index from QoL VI onwards will need to be considered. However, the familiar titles and spread of the revised seven dimensions, in relation to the original ten, will be reassuring. As previously, the main interest will be in the by-municipality scoring on the different dimensions, and overall. Comparisons with QoL V, to the extent necessary, will also be made on the revised dimensions. The need for "rebasing" time series is fairly familiar, for example in Stats SA's changing the ingredients and weighting of the Consumer Price Index every five years to reflect changing goods and patterns of consumption. Their advice can be sought on how this is conveyed.

5 Appendices

A. GCRO Sampling brief

B. GCRO Field management brief

C. GCRO Questionnaire and indexing brief

D. Dr Paul Fatti's sampling review, towards QoL V

E. Dr Ariane Neethling's sampling and weighting report, for QoL IV

F. Original intuitive QoL dimensions and items vs factor-analysed model.

G. Workshop participants and interviewees.