DEVELOPMENT RESEARCH | **AFRICA**
www.dra.co.za

Final data report v2
for the Integrated Quality of Life/ Customer Satisfaction Survey in the Gauteng City Region (GCR)

25 January 2010

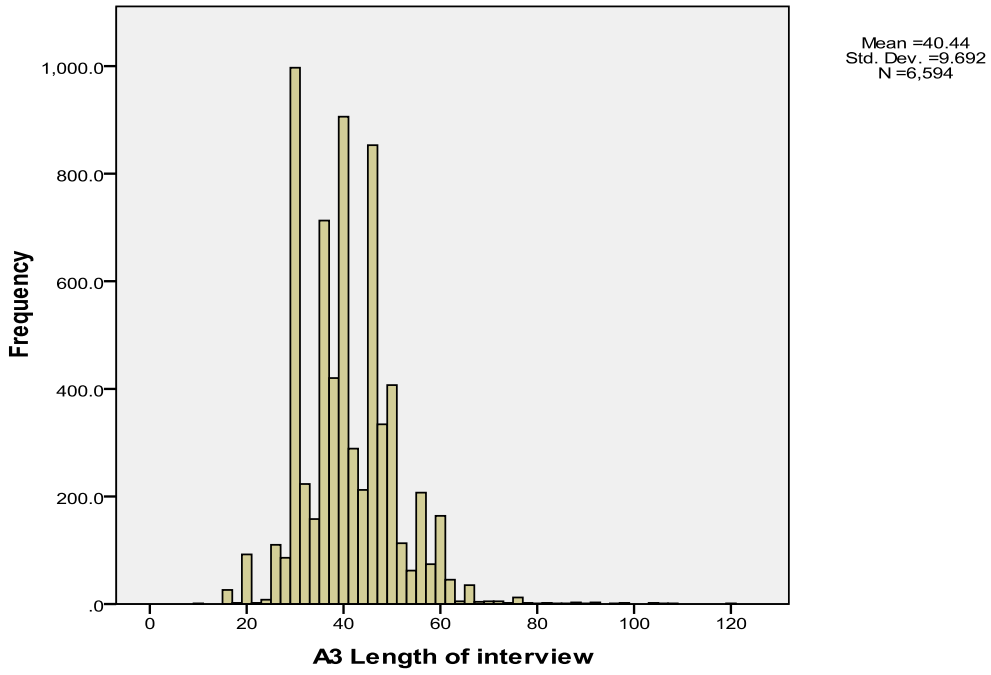
By DEVELOPMENT RESEARCH AFRICA Pty Ltd

General comments:

This data report should be read in conjunction with the completed dataset in SPSS, and serves to explain some of the inconsistencies that were found in the data during validation of the data in SPSS.

The general comments from fieldworkers were that this questionnaire was easy to follow, and there were no complicated skip patterns in the questionnaire. Interviews were conducted between 31 July and 8 October, which included 2 public holidays (9 August and 24 September). The weather during this period was mostly fine, sunny with no rain, except for 1 August (extremely rainy and cold).

The average duration of the interview was 40 minutes (see graph below).



For most questions, the number of missing data for individual questions was exceptionally low for a sample of this size. Missing data in ALL variables were checked against the questionnaire, and a number of variables were cross-checked for inconsistencies. Wherever possible, the respondent was phoned to verify their responses or to see if there was an explanation for the inconsistencies.

Universal CODES

-4 = Missing data

This code is used when the fieldworkers have left out information.

-3 = Refused to answer

This code is used when the respondent has told the fieldworker that they refuse to answer a specific question.

-1 = Do not know

This code is used when the respondent has told the fieldworker that they do not know the information for a specific question.

-2 = Not applicable

This code is used when a question is not applicable, for example when a skip pattern exists.

The following notes refer to information about specific problems in specific questions in relation to the questionnaire database.Please note that the databases attached have been validated for any inconsistencies in the data. The information listed below only refers to cases where the answers were inconsistent, however, many more checks have been carried out on the data.

Q1.4, Q1.5 and 1.6 (Satisfaction with dwelling, and reasons for satisfaction/dissatisfaction)

A number of respondents had not answered 1.4, but had answered 1.5 or 1.6. The questionnaires were checked, and based on the answers provided in 1.5 or 1.6, a response was entered for 1.4 (either code 2=satisfied) or code 4=dissatisfied) (DRAID 4013, 2731, 2708, 1687, 3052, 3854, 3156, 348, 1615, 5560, 5564, 5967, 5836).

Q1.7 and 1.8 (respondent lived in this dwelling 5 years ago, if not, when they moved into the dwelling)

One respondent (DRAID 195) said that they weren't in this dwelling 5 years ago, and then said they moved in in 2003. This was checked on the questionnaire, and left as is. They should probably have answered "Yes" in Q1.7 and skipped to 1.11.

Q1.10 (If the respondent came from Gauteng, North West, Free State or Mpumalanga, which municipality did they come from)

A response code 0 was added to this question for people who came from one of the 4 provinces, but outside the study area. For example, if someone came from Bloemfontein, they couldn't give one of the municipality codes even

though they came from the Free State. They would have answered code 0 (outside study area).

Respondents were asked to look at the showcard and tell us which municipality they came from. On the showcard and on the map itself, the codes were correct (Johannesburg=3 and West Rand=4). However, on the back of the map, the codes for municipality code 3 and 4 were swapped around (Johannesburg=4 and West Rand municipality = 3).Most fieldworkers would have used the codes from the showcard or the actual map, rather than the list on the back of the map, but answers with code 4 in Q1.10, Q1.13, 1.14, 1.15 and 1.17 could mean either 4- West Rand, or 4- Johannesburg. All the fieldworkers were contacted to confirm which codes they had used, and most indicate that they had used the map, and the list on the back of the map.

However, because West Rand municipality consists mainly of nonresidential areas, we can probably assume that most of the responses in 1.13 to 1.17 actually mean “Johannesburg”.

In the municipality list on the showcard, however, one could also think that people say they moved “from the West Rand” which would include Mogale City, Roodepoort etc, instead of West Rand municipality, which could explain the relatively high number of responses coded 4.

Q2.3 (Refuse disposal)

There were 84 respondents (1.3%) who hadn’t answered this question. The questionnaires were all checked for explanations, but none were found.

Q2.4 (household recycles waste)

There were 53 respondents (0.8%) who didn’t answer this question. Maybe they didn’t understand the question.

Q2.5 and 2.6 (main water source, and if not in the dwelling, where people get water from)

Some people answered that they have a borehole in Q2.5, so they had to answer Q2.6. In Question 2.6, they answer that they have a borehole in yard, which supplies water into their dwelling, so they answered code 1 (n/a water in dwelling) in 2.6. Some respondents had other sources of water in 2.5, and still said that they had water in their dwelling in 2.6. These questionnaires were all checked.

Q2.13 (Type of electricity supply)

There were 40 respondents (0.6%) who didn’t answer this question, and most of them were using electricity for cooking, lighting and heating water. Maybe their electricity was through an illegal connection and they were not comfortable divulging this.

Q2.14 (Total amount spent on water, electricity and other fuels)

Two DRAIDS (4888 and 6013) had very high amounts. The questionnaires were checked for explanations but none were found and the information was left as is. DRAID 4888 has a high income, and runs a business from home, which may explain the high usage. No explanation was found for the other questionnaire.

Q4.13 (In the last year has there ever been a time when you did not have enough money to feed the children in the household?)

A code 0 was added for those respondents who said that there are "no children in household", as they felt they couldn't answer this question. Some people reported to not have children in Q12.7, but still answered Q4.13, maybe they referred to adult children as this was not specified in Q4.13.

Q4.1 vs Q4.2 ("Do you owe money to anyone", and "are you paying back…")

There are a number of respondents who say "no" to all the options in 4.2, despite saying they owe money to someone. Maybe they felt that their option was not listed in Q4.2. The option of Q4.2.7 is confusing to them, because both answers can be used if they don' pay anything back ("no, I don't pay anything back", or "yes, I don't pay anything back").

Q5.14.12 (How do you generally find out about what your municipality is doing?)

A number of respondents said "none of the above" and then specified this with "loudspeaker", indicating that their way of finding out about the municipality was not listed in the options above, and they felt the need to specify.

Q6.3 (How satisfied are you with your marriage or relationship with your partner) and Q6.5 (How satisfied are you with your friends)

A response code 0 was added in 6.3 for those who are not in a relationship, and in 6.5 for those who don't have friends.

Section 7 (employment)

DRAID 4917 self employed, answered all questions in section 7, maybe they felt that all applied to them because business fluctuates.

Q7.2 (What is the main reason that you did not work in the past 7 days?)

A code "other" was added, as DRAID 2545 is not allowed to work, not and SA citizen, and DRAID 2118 was retrenched.

Q10.6 (What is the biggest health problem facing your community)

A code 0 was added for "none".

Q11.3 (Amount spent every month on public transport)

Some respondents said “0” without explaining why. Maybe they use public transport but don’t pay? Alternatively, maybe they didn’t know, or couldn’t answer.

Q11.7 (How long after leaving home, does it take you to reach your place of work or study or the place where you look for work?)

A code was added (9=Work from home), as a number of respondents work from home and don’t need transport.

Q11.10 (Number of cars owned by this household)

One respondent (DRAID 1485) had 10 cars, this was confirmed by QC by phoning the respondent to verify this.

Q11.11.4 (Average mileage per year)

There is a relatively high number of missing data, probably because people couldn’t answer this question. However, the number of missing is still well below 2% of the expected number of responses.

Q12.2 (Age)

One respondent (DRAID 5857) was only 17 years old. Looking at the KISH grid and the rest of the questionnaire, they said they were 18 years old. The person was selected for interview, and only right at the end of the questionnaire said that their age was 17 instead of 18. They also didn’t include themselves in the total number of children in the household. ***As this interview was done in an oversampled ward, we could decide to delete this DRAID, but this would have an effect on the calculation of the weights.***

Q12.2 vs Q7.2 (Age vs Reason for not working)

A small number of respondents said they are retired in Q7.2, but they are too young to be pensioners. Maybe they mean that they had stopped working, or someone else in the household is a breadwinner and they don’t need . All the questionnaires were checked.

Q12.3 vs Q7.1 (Employment status vs “have you worked in the past 7 days”)

A small number of respondents (DRAIDS 5573, 2153, 5358, 6138 and 4863) said they didn’t work in the past 7 days, but they are employed full time. The questionnaires were checked for explanations, but none were found. Maybe these respondents were on leave – 4 of them are females who could be on

maternity leave. All the questionnaires were checked for explanations, none were found, and the information was left as is.

Q12.4 vs Q7.2 (Reason for not working vs disability)

Some people said that the reason they hadn't worked in the past 7 days was because they are "disabled" , but in Q12.4 they said they had no disabilities (DRAIDs 6336, 781, 6410, 5296). All the questionnaires were checked and no explanation was found.

Q12.5 (Total number of people in the household)

Two questionnaires have a very high number of people in the household (DRAID 2171: 26 people, and DRAID 4546: 23 people). These questionnaires were checked and the household sizes verified.

Q12.5 vs Q12.7 (Total number of people in the household, vs total number of children in the household)

A number of questionnaires (about 20) have the same numbers for the total household size and the total number of children here. The only time this should happen is if all the people in the household are under 18 (child headed households), but these questionnaires are not child headed households. The questionnaires were all checked and no explanation was found for this. Maybe they are referring to adult children instead of children under 18.

Q12.5 vs 4.8.6 (Total number of children under 18 in the household vs the total number of people receiving a Child Support Grant)

Two respondents (DRAID 4034 and DRAID 5764) had a higher number of CSG recipients than total household size. There may be someone in the household who receives a CSG for a child who doesn't stay in the household. DRAID 4034 also had a higher number of children than the total household size, so maybe they didn't include the children in Q12.5.

Q12.10 (Main home language of respondent)

A code was added: 28= "other" for a couple of respondents who had a home language that wasn't listed in the questionnaire.

Q12.11 (Main language spoken by the children in this household)

A number of people answered this question despite there not being children in the household. This could be because they have adult children in the household, and there was no limitation put on the ages of the children for this question.

Problems encountered with GIS coordinates

During the fieldwork, GIS coordinates were taken at every household where interviews took place. Data validation filtered out the most obvious coordinate problems, for example where 23 degrees had been captured instead of 28 degrees. During the validation process, it was found that two different coordinate settings had been used in field (dd.mm.ss and dd.ddddd). Because the spaces where this information was filled in on the questionnaire did not allow for a second dot to be put in, all the coordinates were captured with in the same way (2 digits for the degrees, followed by a dot and then 5 digits), which made it almost impossible to see what format was used for which questionnaire. To correct this, the following procedure was used:

1) if the second two digits (being the first 2 digits after the dot) were 60 or higher, it was assumed that the format was dd.ddddd. In the dd.mm.sss format, the minutes in the dd.mm.sss format cannot be higher than 59
2) we checked the formats used by the different field managers throughout the study. If coordinates were consistently in the same format for a field manager, the assumption was made that they had always used the same setting (they don't know how to change the setting) and that they had used the same GPS device throughout the study
3) in those cases where the format changed, we looked at the dates that the questionnaires were administered. There were only a few dates on which the GPS devices could have been handed in and possibly changed to another device, for example when the batteries were running low.
4) In the remaining unresolved cases, the coordinates were checked against the spatial data to determine which coordinates were the correct ones.

For each of the cases, the coordinate format was determined in this way. The cases that had been captured in dd.mm.sss format were then converted to dd.ddddd.

When the dataset was completed, and checked against the spatial information, problems were encountered with a large number of GIS coordinates. A total of **1691** out of **6639** questionnaires had coordinates that did not fall within the ward that was captured on the questionnaire.

A number of possible causes for these discrepancies were investigated, and we met with the fieldmanagers to determine the possible causes for these discrepancies. All the questionnaires were then looked at on a ward by ward basis. The following explanations were found:

1) the coordinates were written down wrong on the questionnaire or captured incorrectly (some numbers could be either a 1 or a 7, or a 2 and a 3 were confused because of the way it was written down). The fieldmanagers explained that they would write down the

coordinates in their field books during the time that the fieldworker was doing the interview. The coordinates and ward numbers were then copied over onto the questionnaires, resulting in a number of mistakes which happened during the copying. Most of these cases were identified by looking at each ward individually. If all the coordinates for that ward had 25.71, and only one had 25.11, the incorrect coordinates were corrected. This resulted in **246** cases being fixed. *(Ward number accepted, coordinates corrected)*

2) In a large number of cases, the ward numbes were mixed up. This could be clearly seen when we looked at the location of the questionnaires: clustered around the intended starting point in a neighbouring ward. The fieldmanagers explained that they would often split the team and work on two neighbouring wards at the same time – to increase the amount of time spent in each ward and allow for returning to the same households. When entering the ward number, they could have been swapped around. This could be verified by checking the dates of interview and the field manager code. A total of **678** cases had the ward numbers swapped around in field. *(coordinates accepted, ward number corrected)*
3) When we looked at the spatial information, it became clear that in some cases the fieldteams had crossed into a neighbouring ward. They would have started around the intended starting point and following the correct procedure, identified households at the required intervals. In some cases, the starting points were located almost on the ward boundary. Some wards had changed, with new areas being built up, making it hard to identify the correct location on the map. The google maps that were used were the most up to date aerial maps available, but in some places the information was more than a year old. A total of **657** cases had crossed into a neighbouring ward. *(coordinates accepted, ward number corrected)*
4) In some cases, the interviews took place in the ward, but the person taking the GPS coordinates was standing on the other side of the road, just across the ward boundary, or the ward boundary was very close to the property involved. On smallholdings, the driveway was long, and the property would fall within the correct ward, but the fieldmanager took the coordinates on the driveway, resulting in incorrect coordinates. A decision was made to allow these cases if they would fall within 50m of the ward boundary, unless there was evidence that they crossed into the neighbouring ward as described above. This happened in **80** of the cases. *(left as is, not problematic despite showing as discrepancies)*
5) In a few cases, the boundary of the study area was crossed as per point 3 above, but the neighbouring ward was not part of the study area. Where we were confident that the actual interviews had taken place and the coordinates were making sense in terms of their location, the points were accepted and added to the study area, as their location was part of a neighbourhood that was included in the study area, despite the ward boundary being crossed. *(ward number corrected, coordinates accepted)*

At the end of this exercise there were **3** cases for which no solution could be found. These cases were removed from the sample, resulting in a total sample of 6636 cases. Because of the errors made in field, some 36 wards remained unsampled, and 5 new wards were added on. In total, 572 wards were sampled instead of 603.

After all these corrections were made, the weights were corrected. Because of the number of empty wards, a decision was made to use municipal level weights.