

**Data collection in South  
Africa**

**Jare Struwig**

**Benjamin Roberts**

*Social science that makes a difference*

HSRC RESEARCH OUTPUTS

4370



**HSRC**  
Human Sciences  
Research Council

# South Africa – some background

- ❖ The 2006 mid-year population is estimated at approximately 47,4 million.
- ❖ Fifty-one per cent (approximately 24,1 million) of the population is female.
- ❖ For 2006, life expectancy at birth is estimated at approximately 49 years for males and 53 years for females.
- ❖ The estimated overall HIV-prevalence rate is approximately 11%. The HIV positive population is estimated at approximately 5,2 million.
- ❖ People above 20 yrs of age who are totally illiterate-18%

# **South Africa-some background**

- ❖ **South Africa a young democracy-1994 first democratic elections n South Africa.**
- ❖ **South Africa: heterogeneous, multilingual society, with 11 official languages; commonly referred to as the 'rainbow nation'.**
- ❖ **History of Apartheid-separate development for white and black people**
- ❖ **Currently still very polarized – inequalities between essentially white and black – Gini coefficient highest in the world**
- ❖ **Urban rural divide very notable**

*Some national regular surveys in SA*

**General Household Survey (GHS) PI:STATS SA**

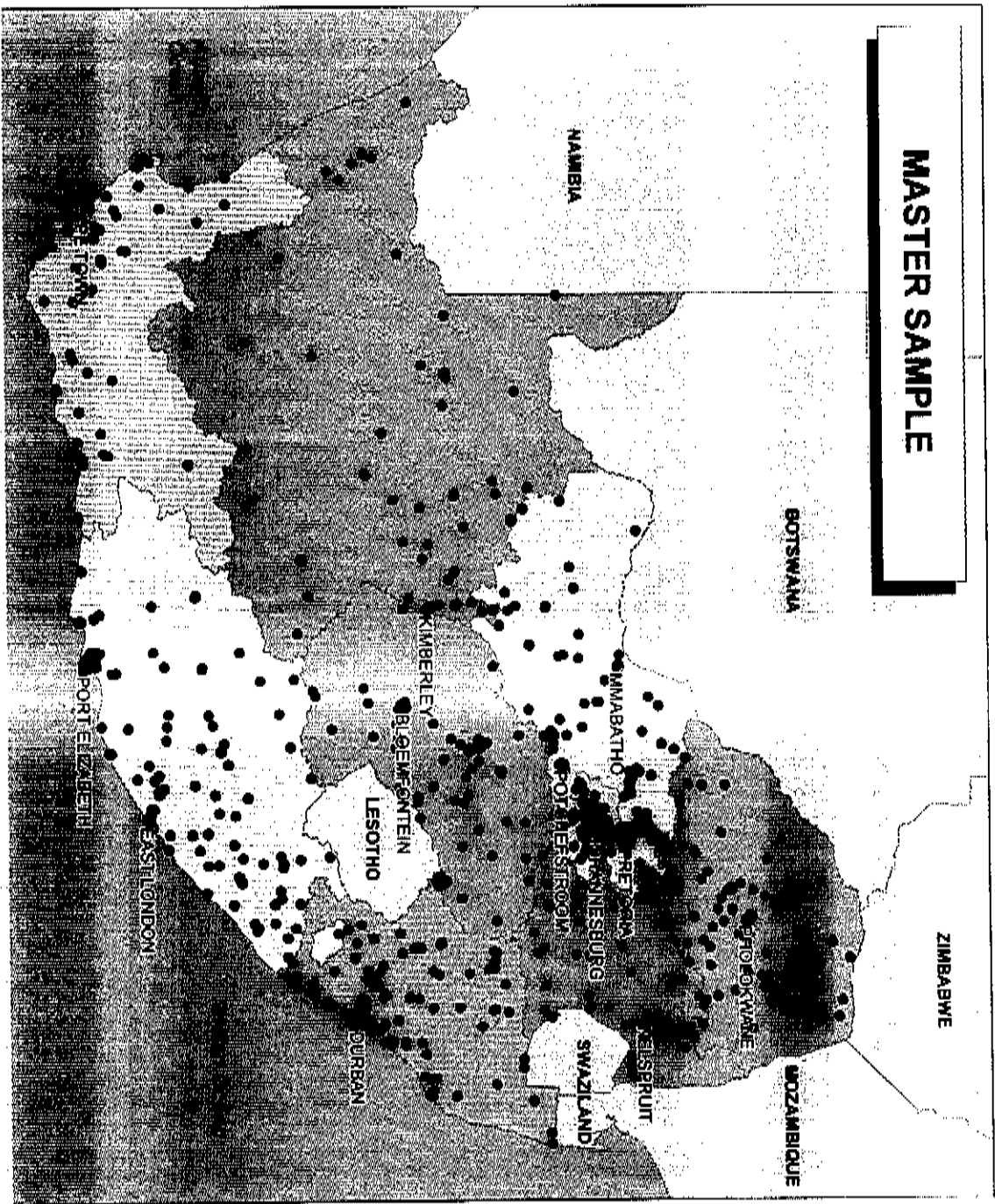
**Labour force survey (LFS) PI:STATS SA**

**Afrobarometer: IDASA**

**South African Social Attitude Survey (SASAS) HSRC**

# Sampling

MASTER SAMPLE



# ***Sampling***

A multi-stage stratified sample was drawn using probability proportional to size principles.

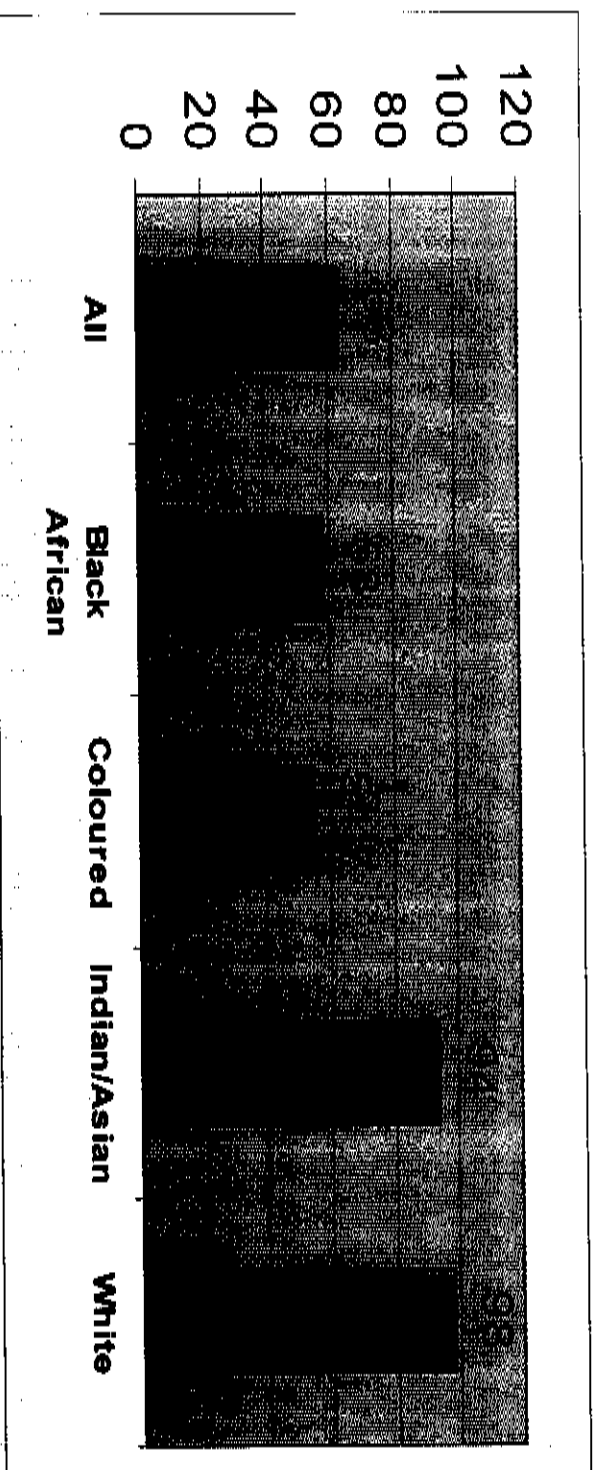
The sample is drawn from a master sample, The master sample is drawn from the database of enumeration areas (EAs) established during the demarcation phase of Census 2001. The sampling procedure for the master sample involves explicit stratification by province, and within each province, by urban and non-urban areas and race. Within each stratum, the sample is allocated disproportionately. A PPS sample of PSUs is drawn in each stratum, with the measure of size being the number of households in the PSU.

## ***Choice of Data Collection Mode***

- ❖ Most national surveys in South Africa uses face-to-face data collection..
- ❖ The preference for this mode of interviewing is influenced by a number of factors.
  - ❖ Relatively low levels of literacy
  - ❖ Positive influence that the presence of an interviewer has on response rates.
- ❖ Telephonic and mailed surveys: In the absence of telephones and recorded residential data, the only reliable approach to surveying a nationally representative sample is through face-to-face interviews.

## Choice of Data Collection Mode

- ❖ Household access to a landline and/or cellular telephone for regular use increased from 44% in 2002 to 55% in 2004 and 63% in 2005.
- ❖ BUT, inequality in access: Less than 60% of African and Coloured households have access to a telephone.
- ❖ In 2005, 77% of SASAS sample were black African respondents





## **SASAS Response Rate, 2003-2005**

	<b>2003</b>	<b>2004</b>	<b>2005</b>
Addresses issued	6917	6907	6930
Vacant, derelict and other out of scope	558	303	295
In scope	6359	6604	6635
In scope	100.0	100.0	100.0
Interview achieved	<b>78.3</b>	<b>84.5</b>	<b>86.4</b>
Interview not achieved	21.7	15.5	13.6
Refused <sup>1</sup>	11.7	9.3	8.1
Non-contacted <sup>2</sup>	10.0	6.2	5.5

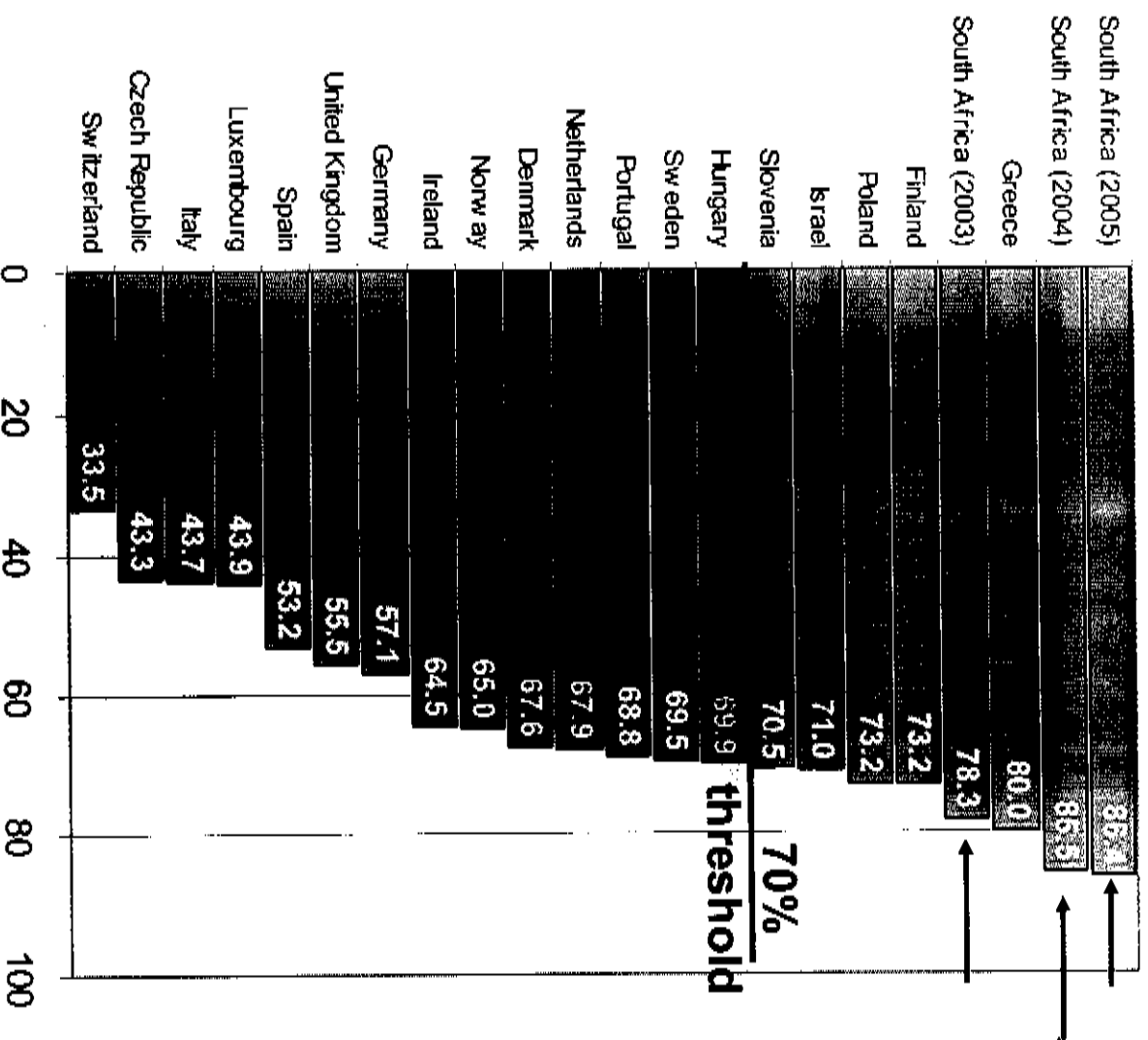
<sup>1</sup> 'Refused' comprises refusals before selection of an individual at the visiting point, refusal by the selected person, and 'proxy' refusals (on behalf of the selected respondent).

<sup>2</sup> 'Non-contacted' comprises households where no one was contacted and those where the selected respondent could not be contacted.

# Fieldwork outcomes

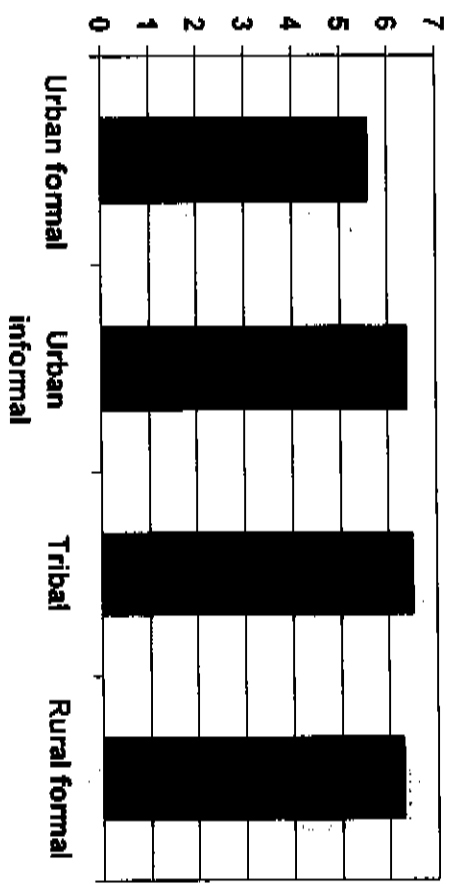
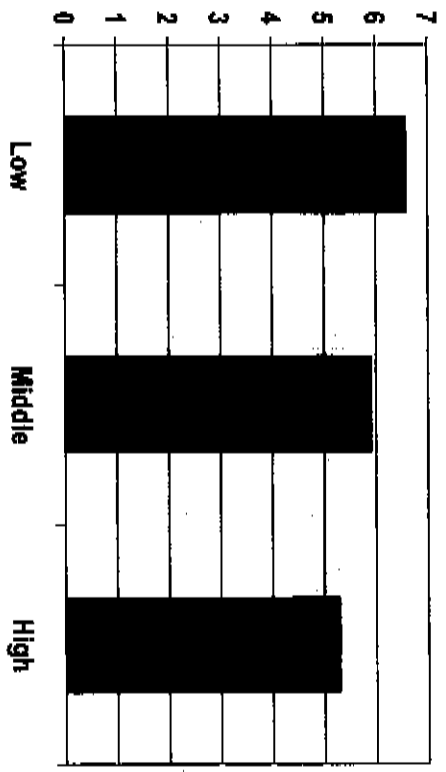
## Achieved response rates (in %): Comparing ESS and SASAS

❖ ESS 70% target response rate

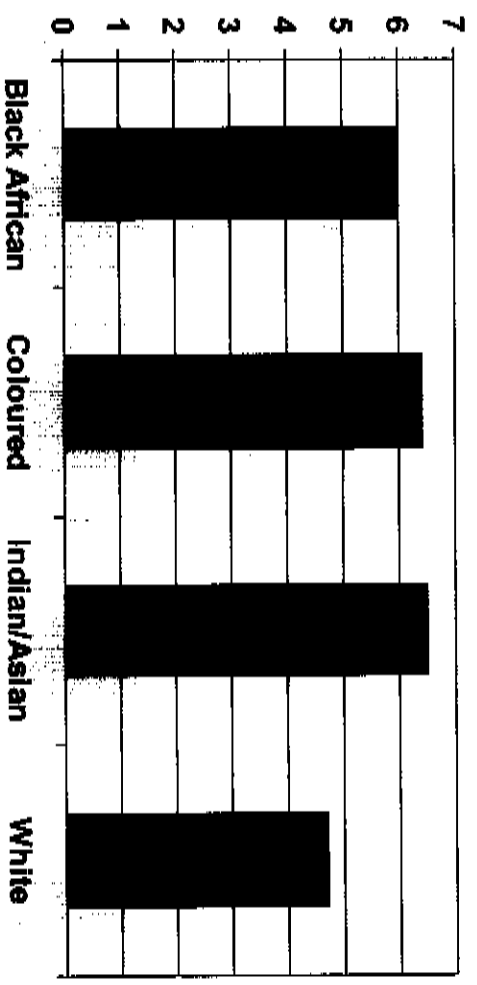


Source: Philipppens & Koch 2003

# Response rates by LSM, race, geotype



## Race



High response rate-  
 Young democracy  
 Unemployment  
 high  
 Oral culture

## ***Choice of Data Collection Mode***

- ❖ Although response rates in the examples have been good, a lingering problem is the ability to reach respondents in the upper end of the income distribution.
  - ❖ Escalating violent crime: led to the rising use of access control to restrict access to properties and suburbs over the last decade.
  - ❖ Increasing presence of high walls, ferocious dogs, security guards and gated communities poses a real challenge for gaining physical access to selected visiting points.
  - ❖ Compounded by the difficulty of finding people at home and a higher propensity for refusal due to questionnaire length and the sensitivity of certain topics.
- ❖ Recent SASAS planning meetings: begun to discuss experimenting with alternate or mixed mode approaches of data collection for difficult to reach localities in future rounds.

# ***Field structure***

**HSRC coordinator**

**Supervisor (each province)**

**Sub-supervisors (3)  
(Driver, selects Eas, houses, initial contact, police)**





# **Fieldworker training**

**Contracted companies use their fieldworkers**

Generally have a Grade 12 qualification

Ability to communicate

Computer literate

Mostly full time fieldworkers – main occupation

**Fieldworkers are trained for each project**

Two day training

Includes role play

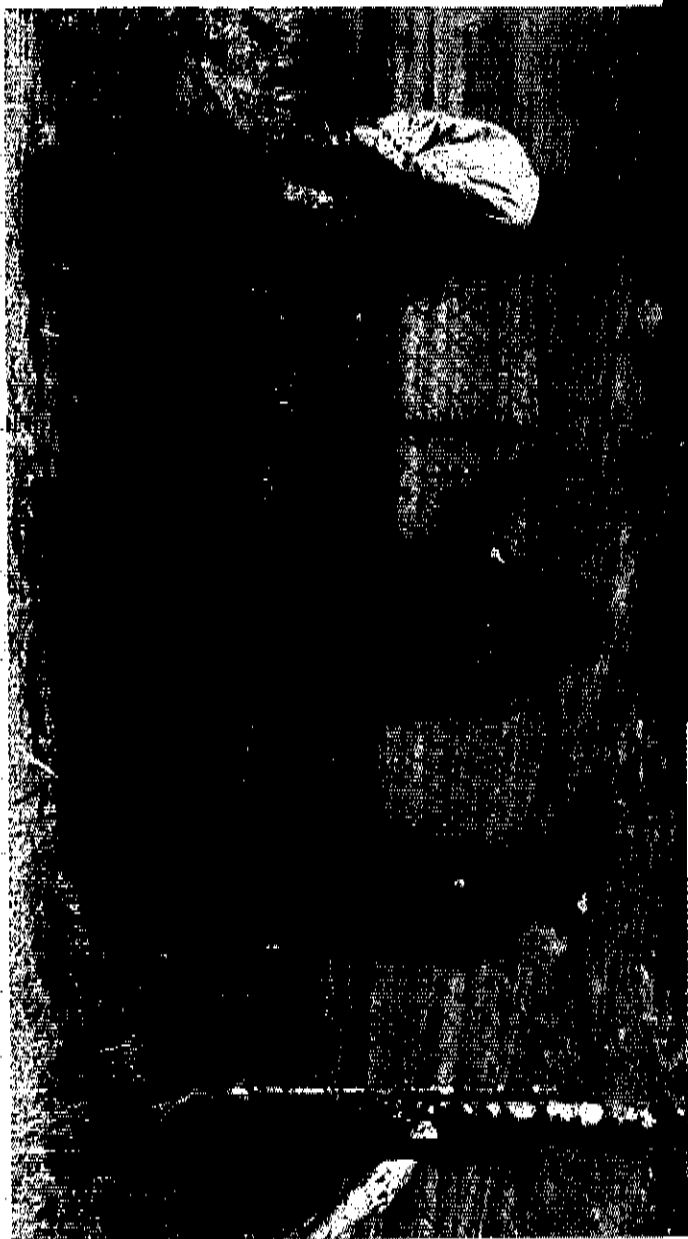
Ethics

Protocol

Approach

**Evaluated for each project (Graded A, B,C)**

**Database**





## Technical summary

<b>Pilot questionnaire</b>	Yes
<b>Pilot languages</b>	Yes
<b>Advanced information</b>	No
<b>First contact by</b>	Visit
<b>Minimum of required visits</b>	Three
<b>Ideal length</b>	45 min
<b>Respondent incentives</b>	None
<b>Back checks</b>	10-15%

# ***Protocol***

- **Whilst it is not the custom to seek permission to interview the public in established democracies, in Africa it is very often essential to seek permission from the relevant local authorities. inkosi / induna / local councillor Whilst clearances causes delays in fieldwork – ensure that field teams were in area**
- **Local police station**
- **Contact agriculture and mines for entrance to farms and houses**
- **Blocks of flats with security-need permission**
- **A letter of introduction to be given out to respondents.**
- **Name tags for identification**
- **Car identification stickers**

## ***Language issues***



- ❖ **South Africa: heterogeneous, multilingual society, with 11 official languages; commonly referred to as the 'rainbow nation'.**
- ❖ **It is important that respondents are interviewed and answer questions in a language they feel comfortable.**
- ❖ **Most local languages have many different dialects. Sometimes the same language is spoken differently in urban areas and rural areas- therefore crucial to get local fieldworkers.**
- ❖ **Ideally, this implies that questionnaires require translation into each of the major languages-not possible.**
- ❖ **SASAS translation context: fielded in 6 languages in R1 and 5 languages in R2. In R3, will be increased to 8 languages.**

## Lost in Translation?

No. Languages fielded per SASAS round and breakdown of South African population by home language (2001)

Home language	2003	2004	2005	South Africa
<b>West Germanic subgroup</b>				<b>21.5</b>
Afrikaans	X	X	X	13.3
English	X	X	X	8.2
<b>Nguni subgroup</b>				<b>45.7</b>
IsiZulu	X	X	X	23.8
IsiXhosa			X	17.6
SiSwati				2.7
IsiNdebele				1.6
<b>Sotho subgroup</b>				<b>25.5</b>
Sepedi			X	9.4
Setswana	X	X	X	8.2
Sesotho				7.9
<b>No subgroup</b>				<b>6.7</b>
Tshivenda	X	X	X	2.3
Xitsonga	X		X	4.4
Other				0.5
<b>Total</b>	<b>6</b>	<b>5</b>	<b>8</b>	<b>100.0</b>



## *Lost in Translation?* Salient Challenges

- ❖ Securing linguistic equivalence in a multilingual environment is extremely challenging.
- ❖ The diversity of cultures means that:
  - ❖ Attempts at finding shared meanings for concepts such as 'family', 'household', 'married', 'employed' are fraught with difficulty.
  - ❖ This difficulty does not only emerge along linguistic lines; also notable locational determinants, e.g. 'armed security force', 'burglar alarm system' are common in urban but not rural localities.
  - ❖ Job security
  - ❖ Translations may not always easily able to match English scale properties, e.g. distinction between 'strongly disagree' and 'disagree' gets blurred and fieldworkers may need to probe.
  - ❖ There exists a high risk for misrepresentation due to misunderstanding.

# *Lost in Translation?*

## Scope for Improvement



- ❖ Consequently, much time and resources need to be devoted to translation procedures.
- ❖ While the human and fiscal resources devoted to translation have been improving with successive rounds...
- ❖ Need for a more rigorous translation process.
- ❖ Translation framework for SASAS:
  - ❖ Interrelated processes to produce final translated version
  - ❖ Translation; Review; Adjudication; Pre-testing (source and translated questionnaires);
  - ❖ Documentation is essential, but being crowded out by competing SASAS and other survey demands.
  - ❖ Procedures and guidelines need to be more systematic
  - ❖ Need to improve data quality testing to estimate reliability and validity of questions before used in practice.
  - ❖ Ultimately: questions of planning, timing and procedure
  - ❖ Trade-offs between timeliness and accuracy:

## ***Poles Apart? Concluding thoughts***

- ❖ Despite the demands that prevailing social conditions in developing countries like South Africa pose for the design of social surveys...
- ❖ There remain similarities with the approach adopted in developed countries and cross-national surveys.
- ❖ There are parallels in the problems faced, though it is recognised that the frequency and severity of these vary considerably.
- ❖ This presents an important opportunity for enhanced knowledge-sharing and lesson-learning in the continuing process of improving the quality and accuracy with which social attitudes are measured in South Africa.