Australian National University

Language matters in accurate measurement of the Sustainable Development Goals

Stephen Haslett
Professor of Statistics and Director
Statistical Consulting Centre
Australian National University
stephen.haslett@anu.edu.au

Abstract

Monitoring and evaluation, progress reports, assessment of benchmarks and milestones, and official statistics underlie and are integral to the Sustainable Development Goals. They provide required measures of progress, often represented through numerical summaries. The substantive aspects of such measurements may seem purely quantitative, but the qualitative/quantitative divide can be an artificial one.

Rather than considering the underlying statistical modelling, this paper will explore the crucial role of language in even highly technical statistical research such as fine-level food security and child undernutrition assessments and poverty mapping in aid projects undertaken for the UN World Food Programme in Cambodia, Nepal and Bangladesh. The focus will be on equivalence of meaning, a core issue which goes beyond choice of language (Khmer, Nepali, Bangla, etc.) and translation of terms from or into English in the various survey and census instruments. Equivalence of meaning extends further to the way local knowledge, encapsulated in local languages, is captured and represented in the various survey instruments and field manuals used for data gathering in multilingual communities. The human and financial resources needed for language-related tasks must be taken into account in the planning, implementation and analysis of such development aid projects, if maximum accuracy is to be achieved in collecting the measurements necessary to support effective interventions.

Sketch outline

1. Number of languages in Nepal, Cambodia, Bangladesh
2. Background to poverty mapping
3. Nutrition, food security, etc.
4. Use of small area estimation in aid allocation
5. “Equivalence” of meaning
6. Walls and roofs from Cambodia – definition / equivalence of meaning within one language
7. Equivalence table – roofs, floors or walls (Cambodia)
8. Questionnaires and field instructions in local languages, publicly available versions in English: alphabets and equivalence of meaning across languages? (Nepal)
9. Crucial role of local knowledge in how questionnaires are translated (in parallel to problem of the field manuals for roof and walls being in local language)
10. Conclusion: language and qualitative issues matter, even in substantially quantitative studies linked to the SDGs.

The word "Nepali" written in

Number of languages in Nepal, Cambodia, Bangladesh

Nepal – 123 languages

- 45% of population of 26.5 million speak Nepali as a first language, written in Devanagari: Official language and script with104 ethnic groups
- Nepali = Gorkhali = नेपाली भाषा Nepālī bhāṣā = खस कुरा Khas kurā
- Regional languages include: Limbu; Maithili; Newar; Angika; Tharu; Gurung; Tamang; Magar; Sherpa; Kiranti; Sunuwar; Bhojpuri; Rajbanshi

- *Source:* Nepal Central Bureau of Statistics, 2011 National Census

The word "Nepali" written in

Number of languages in Nepal, Cambodia, Bangladesh

Cambodia – 27 languages: 19 indigenous languages

- Most of population of 13.4 million speak Khmer as a first language, written in : Official language and script
- Khmer ภาสาเขฺมร or more formally ខេមរភាសា
- Indigenous languages include: Bahnaric, Cham, Chong, Jarai, Khaonh, Khmer, Kuy, Lao Nyo, Mel, Ra’ong, Sa'och, Stieng, Thmon
- *Sources:* https://www.ethnologue.com/country/KH

  Cambodia National Institute of Statistics, 2008 National Census

The word "Nepali" written in

Number of languages in Nepal, Cambodia, Bangladesh

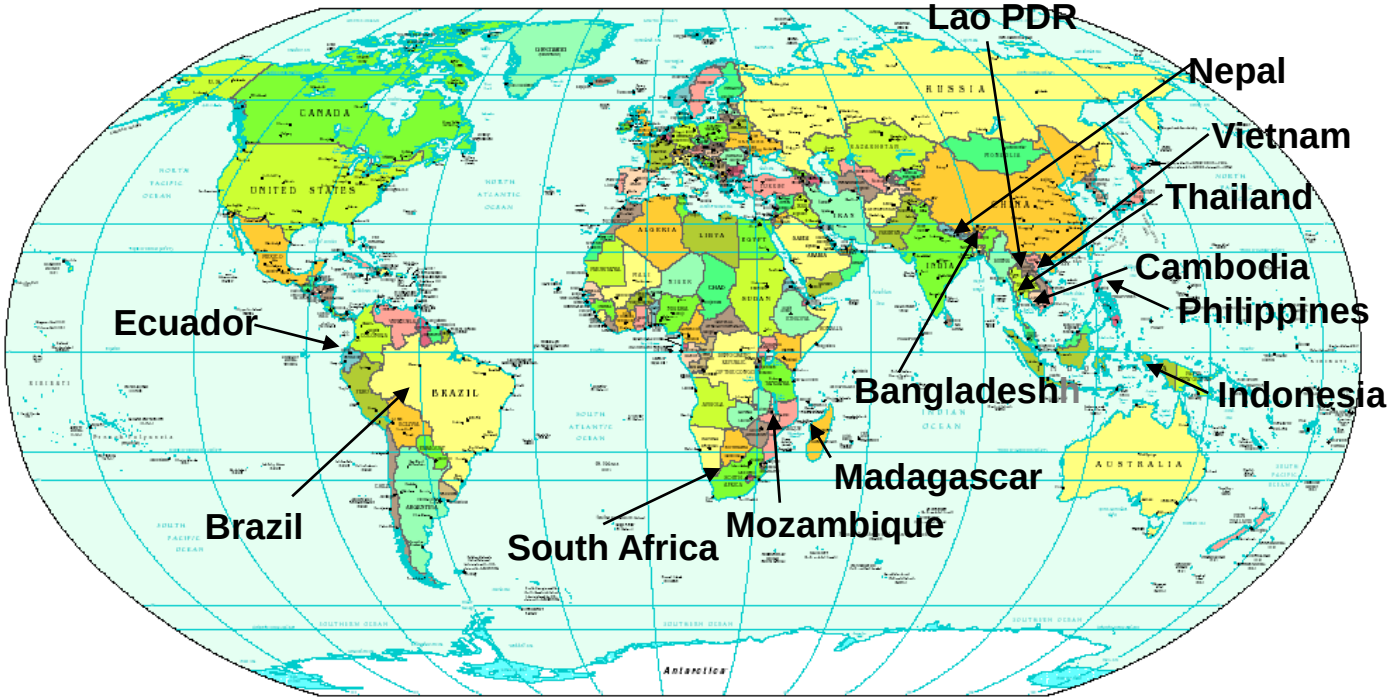
Bangladesh – 38 minority languages

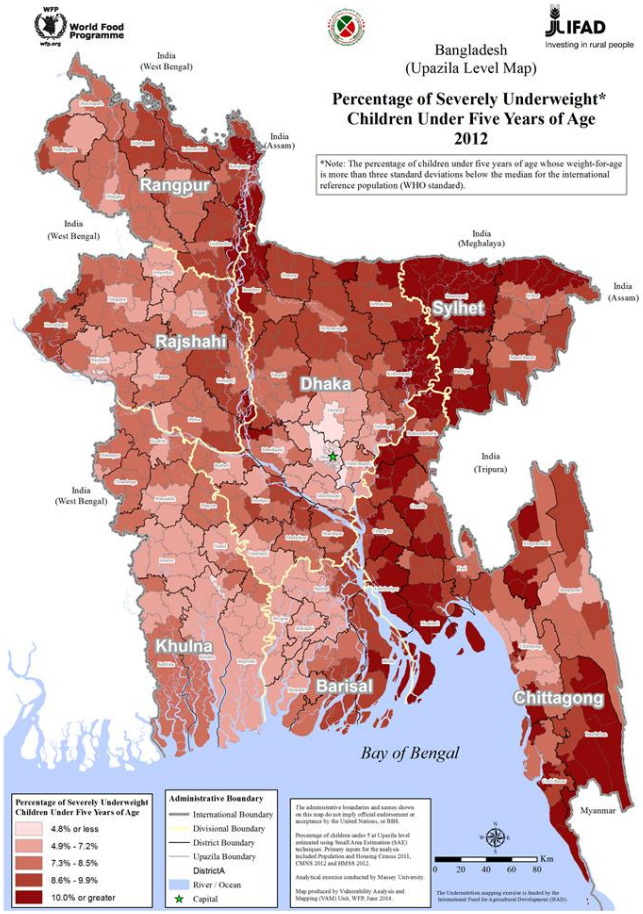
- 98% of population of 142.3 million speak Bangla as a first language, written in : Official language and script
- Bangla or Bengali script বাংলা লিপি *Bangla lipi*
- Indigenous languages include: Arakanese, Assamese, Bishnupriya Manipuri, Chakma, Hajong, Tangchangya, Oraon Sadri, Khasi, Koda, Mundari, Pnar, Santali, War-Jaintia, Kurukh, Sauria Paharia, A'Tong, Chak, Chin, Asho, Bawm, Falam, Haka, Khumi, Koch, Garo, Megam, Meitei Manipuri, Mizo, Mru, Pangkhua, Kok Borok, Riang, Tippera, Usoi
- *Source:* Bangladesh Bureau of Statistics, 2011 National Census

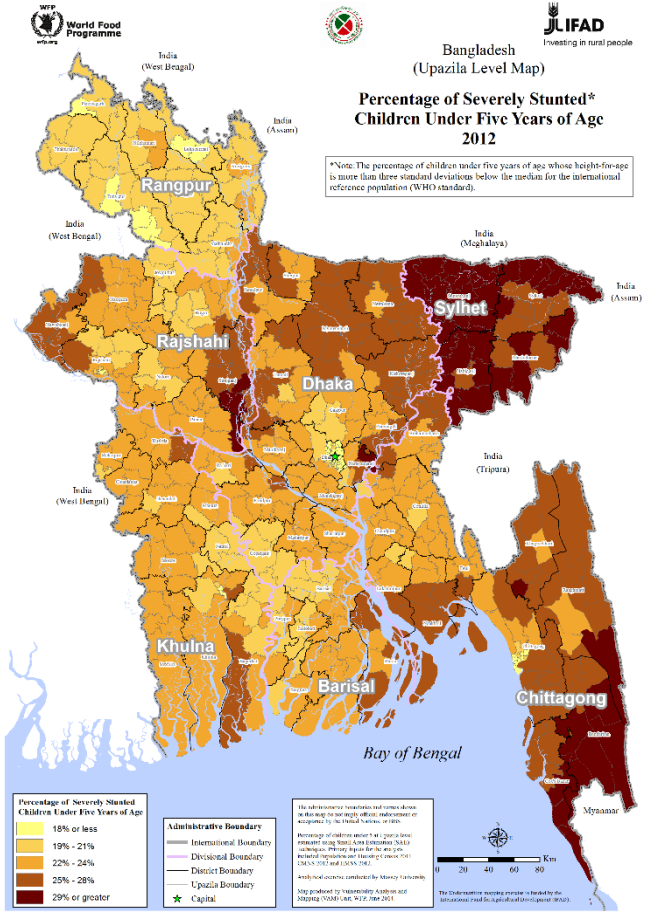
Background to poverty mapping

- Geographically-detailed information within a country is very expensive to obtain using national sample surveys because huge sample sizes are needed
- Poverty mapping is a statistical modelling technique, and one way of doing small area estimation (SAE), that models survey or survey and census data to produce fine-level estimates.
- SAE can be applied to provide fine-level detail for domains rather than areas (e.g. ethnic groups)
- It can also be applied for nutritional measures, eg kilocalorie intake, and stunting, underweight and wasting in young children
- The fine-level estimates are used as an important input into aid allocation.

Where?

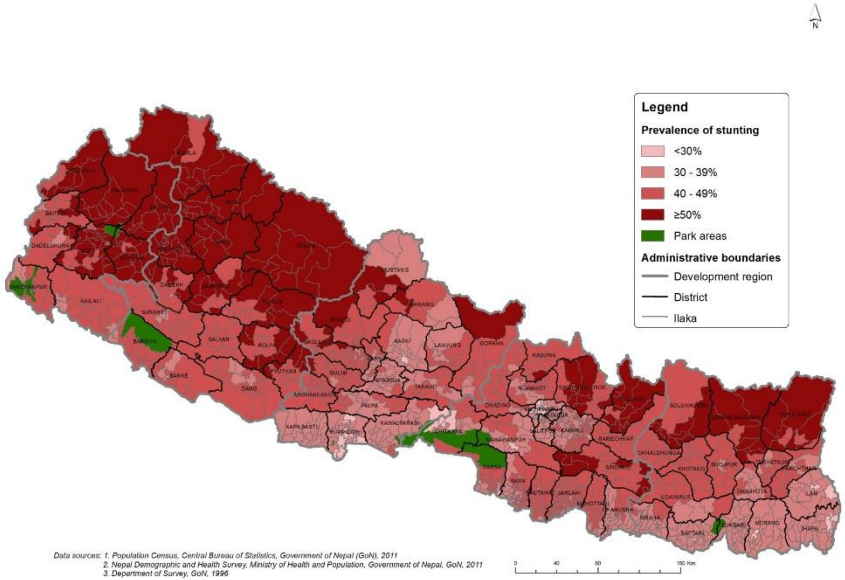

Lao PDR
Nepal
Vietnam
Thailand
Cambodia
Philippines
Indonesia
Ecuador
Bangladesh
Madagascar
Brazil
South Africa
Mozambique
CANADA
UNITED STATES
RUSSIA
BRAZIL
AUSTRALIA
Antarctica


World Food Programme
IFAD
Investing in rural people
Bangladesh
(Upazila Level Map)
Percentage of Severely Underweight*
Children Under Five Years of Age
2012
*Note: The percentage of children under five years of age whose weight-for-age is more than three standard deviations below the median for the international reference population (WHO standard).
India (West Bengal)
India (Assam)
India (Meghalaya)
India (Tripura)
Rangpur
Rajshahi
Dhaka
Sylhet
Khulna
Barisal
Chittagong
Bay of Bengal
Myanmar
Percentage of Severely Underweight Children Under Five Years of Age
4.8% or less
4.9% - 7.2%
7.3% - 8.5%
8.6% - 9.9%
10.0% or greater
Administrative Boundary
International Boundary
Divisional Boundary
District Boundary
Upazila Boundary
DistrictA
River / Ocean
Capital
Km
0 20 40 60 80


World Food Programme
IFAD
Investing in rural people
Bangladesh
(Upazila Level Map)
Percentage of Severely Stunted*
Children Under Five Years of Age
2012
*Note: The percentage of children under five years of age whose height-for-age is more than three standard deviations below the median for the international reference population (WHO standard).
India (West Bengal)
India (Assam)
India (Meghalaya)
India (Tripura)
Rangpur
Rajshahi
Dhaka
Sylhet
Khulna
Barisal
Chittagong
Bay of Bengal
Myanmar
Percentage of Severely Stunted Children Under Five Years of Age
18% or less
19% - 21%
22% - 24%
25% - 28%
29% or greater
Administrative Boundary
International Boundary
Divisional Boundary
District Boundary
Upazila Boundary
Capital
Km
0 20 40 60 80

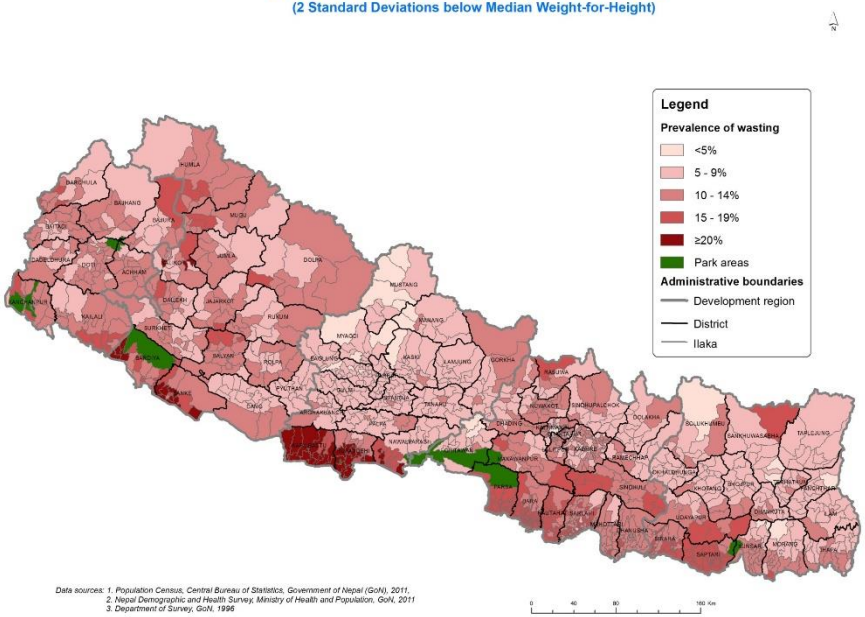
Prevalence of Stunting (S2) at Ilaka Level

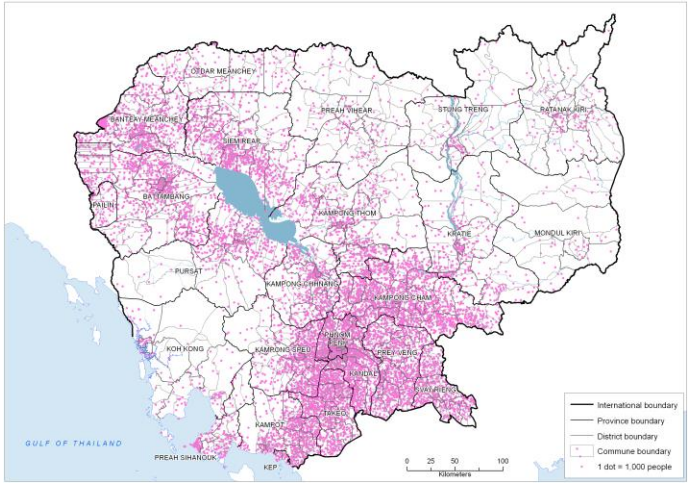
(2 Standard Deviations below Median Height-for-Age)

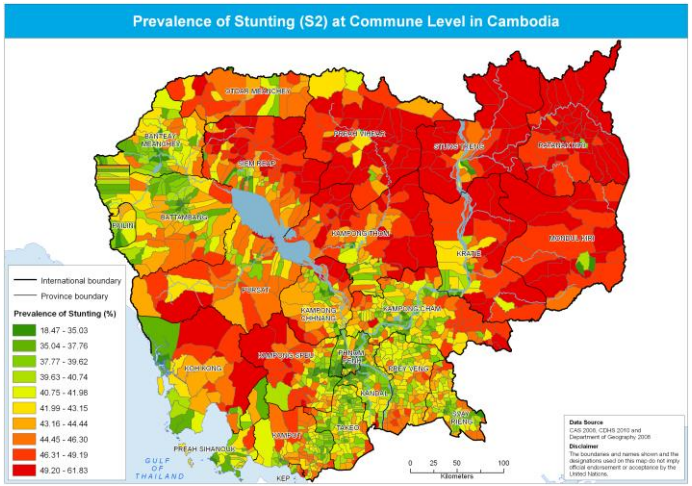


Prevalence of Wasting (W2) at Ilaka Level

(2 Standard Deviations below Median Weight-for-Height)




OTDAR MEANCHEY
BANTEAY MEANCHEY
PREAH VIHEAR
STUNG TRENG
RATANAK KIRI
SIEM REAP
PAILIN
BATTAMBANG
KAMPONG THOM
KRATIE
MONDUL KIRI
PURSAT
KAMPONG CHHNANG
KAMPONG CHAM
KOH KONG
KAMPONG SPEU
PHNOM PENH
PREY VENG
KANDAL
SVAY RIENG
TAKEO
KAMPOT
GULF OF THAILAND
PREAH SIHANOUK
KEP
International boundary
Province boundary
District boundary
Commune boundary
1 dot = 1,000 people
0 25 50 100
Kilometers


Prevalence of Stunting (S2) at Commune Level in Cambodia
International boundary
Province boundary
Prevalence of Stunting (%)
18.47 - 35.03
35.04 - 37.76
37.77 - 39.62
39.63 - 40.74
40.75 - 41.98
41.99 - 43.15
43.16 - 44.44
44.45 - 46.30
46.31 - 49.19
49.20 - 61.83
GULF OF THAILAND
PREAH SIHANOUK
KEP
0 25 50 100
Kilometers
Data Source
CAS 2008, CDHS 2010 and Department of Geography, 2006
Disclaimer
The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.

Use of small area estimation in aid allocation

- Because the poverty related estimates are at a much finer level than possible with a survey alone, better targeting of aid is possible
- Decisions on aid allocation can use other information, such as population density
- Actual allocation depends on the aims of the aid project

“Equivalence” of meaning

- But much depends on how well the statistical modelling is done
- This is turn depends on detailed data cleaning and comparison
- Part of this process is relatively technical, but an essential part is not, and depends instead on non-sampling complications
- Language, especially when questionnaires are conducted in local or national languages rather than English, and translation is involved, is one of these complications
- Even in a single language a question is how to interpret a single question or whether two questions in different questionnaires ask the same thing.

Walls and roofs from Cambodia: definition / equivalence of meaning within one language

- Questions that appear straightforward can be complicated, even in one language
- This is more obviously true of qualitative questions, but is also so for quantitative ones.
- So let’s look at a very simple situation where the initial impression is that there should be no problem at all

Roofs, floors or walls (Cambodia)

Final draft approved by CTC on 24 August 2006

FORM A HOUSELIST

STRICTLY CONFIDENTIAL

Royal Government of Cambodia
General Population Census of Cambodia, 2008

Page Number.....
Total Number of pages used for the EA.....

Identification Particulars

	Khet / Krong	Srok / Khand	Khum / Sangkat	Phum	Enumeration Area No.
Name					
Code					

Building / Structure and Household Particulars

Line No.	Building/ Structure Number	Predominant Construction Material of Building / Structure*			Purpose of Building/Structure (Enter Code)	Household No.	Particulars of Head of Household		Number of Persons Usually living in the Household			Remarks
		Wall	Roof	Floor	1. Residence 2. Residence & Shop 3. Residence & workshop 4. Residence & any other establishment (specify)		Name	Sex 1 = Male 2 = Female (Enter Code)	Males	Females	Persons	
1	2	3	4	5	6	7	8	9	10	11	12	13
1												
2												
3												
4												
5												
6												
7												
8												
9												
0												
	(**Count the number of entries and give total)				**Total			Total				

*KEY TO CODES

Wall Material (Column 3)
1. Bamboo / Thatch / Grass / Reeds
2. Earth
3. Wood / Plywood
4. Concrete / Brick / Stone
5. Galvanised Iron / Aluminium / Other metal sheets
6. Asbestos cement sheets
7. Salvaged / Improvised materials
8. Other (specify)

Roof Material (Column 4)
1. Bamboo / Thatch / Grass
2. Tiles
3. Wood / Plywood
4. Concrete / Brick / Stone
5. Galvanised Iron / Aluminium / Other metal sheets
6. Asbestos cement sheets
7. Plastic / Synthetic material sheets
8. Other (specify)

Floor Material (Column 5)
1. Earth / Clay
2. Wood / Bamboo planks
3. Concrete / Brick / Stone
4. Polished stone
5. Parquet / Polished wood
6. Mosaic / Ceramic tiles
7. Other (specify)

Name of Enumerator : .....

Signature Day / Month / Year

Name of Supervisor : .....

Signature Day / Month / Year

1

Roof type in Cambodia

Bamboo / thatch / grass
Tiles
Wood / plywood
Concrete / brick / stone
Galvanised iron / aluminium / other metal sheets
Asbestos cement sheets
Plastic / synthetic material sheets
Other (specify)

Wall type in Cambodia

Bamboo / thatch / grass / reeds
Earth, wood / plywood
Concrete / brick / stone
Galvanised iron / aluminium / other metal sheets
Asbestos cement sheets
Salvaged / improvised materials
Other (specify)


BOREY

Crucial role of local knowledge in how questionnaires are constructed and translated

- In parallel, there is the problem of the field manuals (e.g. for roof and walls – Bangladesh, Cambodia, Nepal) being in local language only
- When quantitative questions in two questionnaires (e.g. census and survey or two surveys – Bangladesh, Cambodia, Nepal) require comparison for “equivalence”, the problem compounds
- Even for an identical question in two questionnaires, context may alter sensitive responses (e.g. child mortality in Nepal)
- For qualitative questions (e.g. “Can the person [being interviewed] read and write” - Cambodia), issues are more subtle again.

Questionnaires and field instructions: alphabets and equivalence of meaning across languages? (Nepal)

१) मुख्य प्रशिक्षक प्रशिक्षण कार्यशाला

केन्द्रीय तथ्याङ्क विभाग
राष्ट्रिय जनगणना २०६८
मुख्य प्रशिक्षक प्रशिक्षण कार्यशाला

मिति २०६७	१०:०० – ११:३०	स्रोत व्यक्ति	११:३०–११:४०	११:४० – ०१:१०	स्रोत व्यक्ति	१:१०–१:५०	०१:५०– ०३:२०	स्रोत व्यक्ति	३:२०–३:३०	३:३०–०५:००	स्रोत व्यक्ति
१०/१३	उद्घाटन सत्र		विश्राम	नेपालमा जनगणना र रा.ज.ग. २०६८ को उद्देश्य, महत्व, विधि, सन्दर्भ समय, कार्यसंगठन र कार्यतालिका		खाजा समय	राष्ट्रिय जनगणना २०६८ मा सालैंगिकीकरण		विश्राम	राष्ट्रिय जनगणना २०६८ मा सामाजिक समावेशीकरण	
१०/१४	जनगणना अधिकृतको काम, कर्तव्य र अधिकार (जनगणना अधिकृत निर्देशिका)			इलाका सुपरिवेक्षकको काम, कर्तव्य र अधिकार (इलाका सुपरिवेक्षक निर्देशिका)			गणना सुपरिवेक्षकको काम, कर्तव्य र अधिकार (गणना सुपरिवेक्षक निर्देशिका)			गणकको काम, कर्तव्य र अधिकार (गणना निर्देशिका)	
१०/१७	तालिम निर्देशिकाको प्रयोगसम्वन्धी जानकारी			गणना क्षेत्र नक्साको प्रयोग - १			गणना क्षेत्र नक्साको प्रयोग-२			प्रयोग हुने लगतहरु तथा अवधारणा र परिभाषाहरू (गणना क्षेत्र, अक्सर बसोवास, घर, परिवार, परिवारमूली, व्यक्ति, आदि)	
१० /१८	घर/परिवार सूचीकरण लगत : परिचियात्मक विवरण र महल १-९			घर/परिवार सूचीकरण लगत : महल १०-२८			घर/परिवार सूचीकरण लगत: महल २९-३० र पृष्ठान्तका विवरण			छनौट विधि र परिवार छनौट गर्ने तरिका	
१०/१९	परिवार छनौट विधि (प्रयोगात्मक) र घर/परिवार सूचीकरण लगत उतार फाराम भर्ने तरिका			लगत १: परिचयात्मक विवरण र पारिवारिक विवरणको प्रश्न नं १-२ सम्म			लगत १: पारिवारिक विवरणको प्रश्न नं ३ देखि ७ सम्म			लगत १: पारिवारिक विवरणको प्रश्न नं ८ देखि १० सम्म	

क्रमशः

२४

Conclusion

- In quantitative studies, for example quantitative monitoring and evaluation, the final quality of the research is often determined by the “weakest link”
- Even highly technical studies with excellent methodology can flounder on small and apparently unimportant aspects
- Language, whether local or international or the connections between them, is one such aspect
- Language and qualitative issues really do matter, and should not be overlooked even in substantially quantitative studies linked to the SDGs.

References

Haslett, S. Isidro, M. and Jones, G. (2010) Comparison of survey regression techniques in the context of small area estimation of poverty, Survey Methodology, Vol 36, No 2, 157-170.

Haslett, S. and Jones, G. (2010) Small area estimation of poverty: the aid industry standard and its alternatives, Australian and New Zealand Journal of Statistics, Vol 52, No 4, 341-362.

Haslett, S., Jones, G., Isidro, M., and Sefton, A. (2014) *Small Area Estimation of Food Insecurity and Undernutrition in Nepal*, Central Bureau of Statistics, National Planning Commissions Secretariat, UN World Food Programme, UNICEF, and World Bank, Kathmandu, Nepal, December 2014, 154 pages, ISBN: 978-9937-3000-976. https://www.wfp.org/content/nepal-small-area-estimation-food-insecurity-and-undernutrition-december-2014

Haslett, S. Jones, G. and Isidro, M. (2014) *Small-Area Estimation of Child Undernutrition in Bangladesh*, Bangladesh Bureau of Statistics, United Nations World Food Programme and International Fund for Agricultural Development, November 2014, 76 pages, ISBN 978-984-33-9085-1.
https://www.wfp.org/sites/default/files/Undernutrition%20Maps%20of%20Bangladesh%202012_Technical%20Report_Massey%20WFP%20BBS%20IFAD.pdf

Haslett, S., Jones, G. and Sefton, A. (2013) *Small-area Estimation of Poverty and Malnutrition in Cambodia*, National Institute of Statistics, Ministry of Planning, Royal Government of Cambodia, and the United Nations World Food Programme, June 2013, 88 pages, ISBN 9789996375507.
http://www.wfp.org/content/cambodia-small-area-estimation-poverty-and-malnutrition-april-2013

Isidro, M., Haslett, S. & Jones, G. (2016) Extended Structure Preserving Estimation (ESPREE) for updating small area estimates of poverty, *The Annals of Applied Statistics*, 10: 1, 451–476, DOI: 10.1214/15-AOAS900.