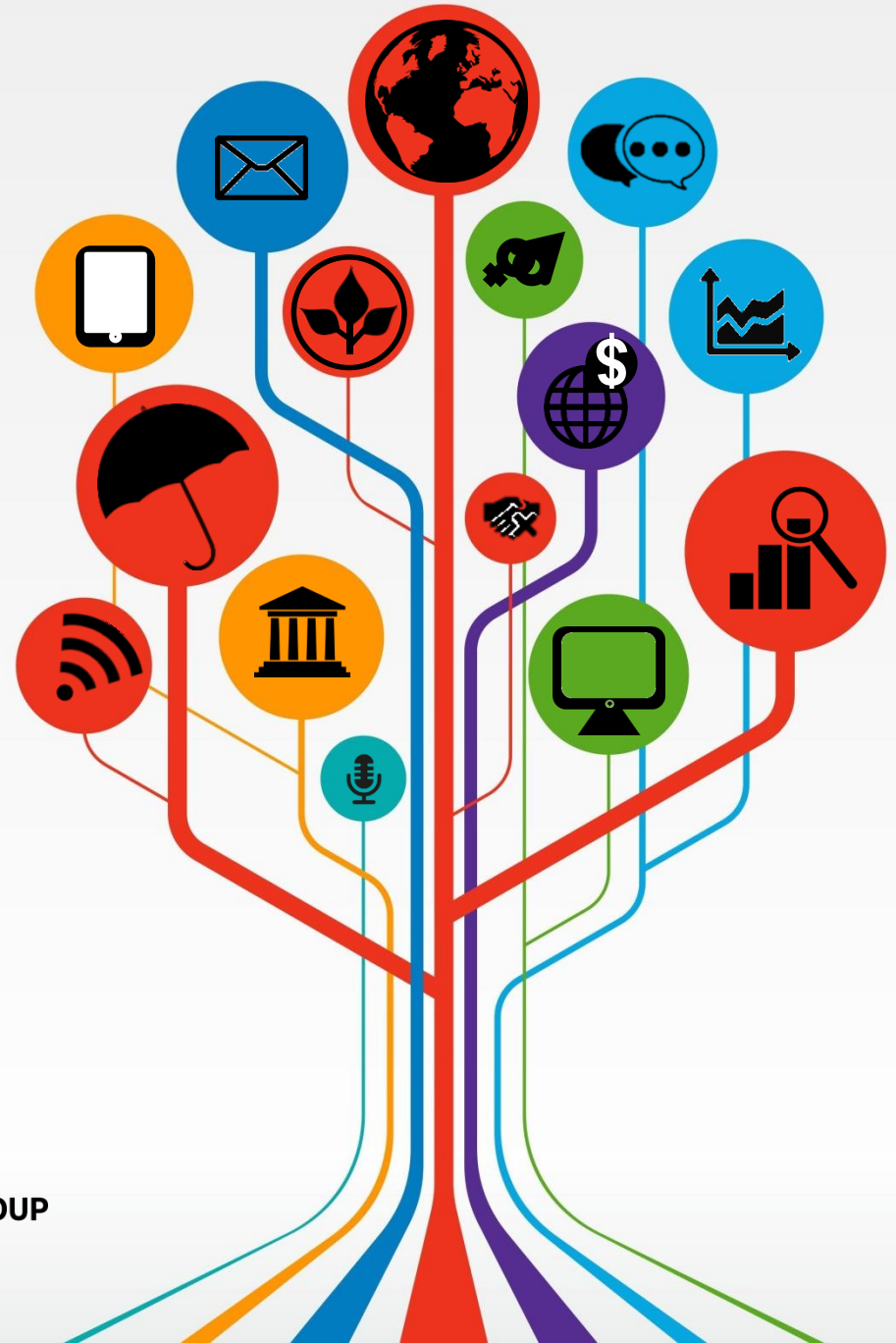# Survey Instrument Design and Pilot
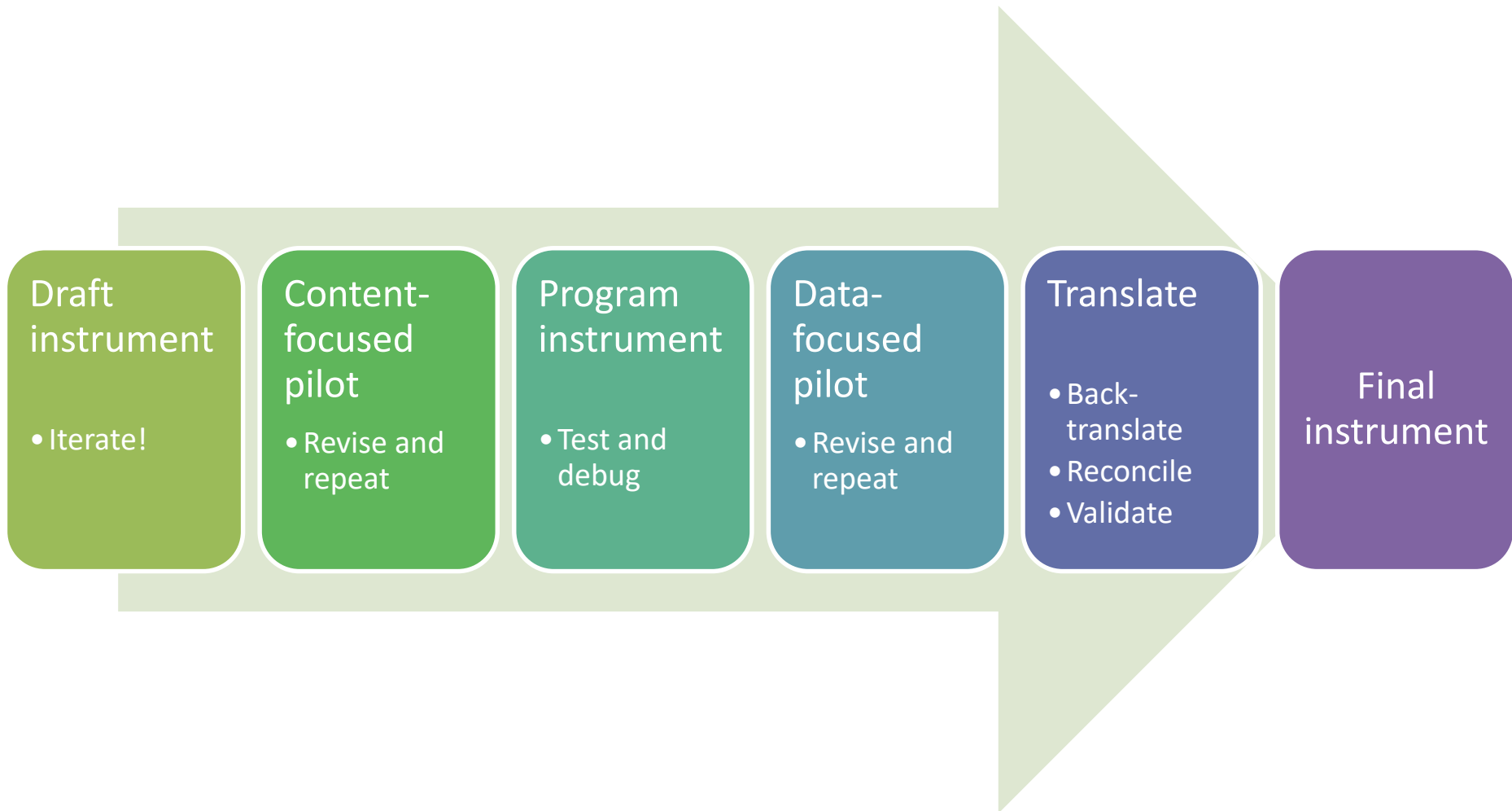
Steven Glover

Sékou Kone

10 June 2019
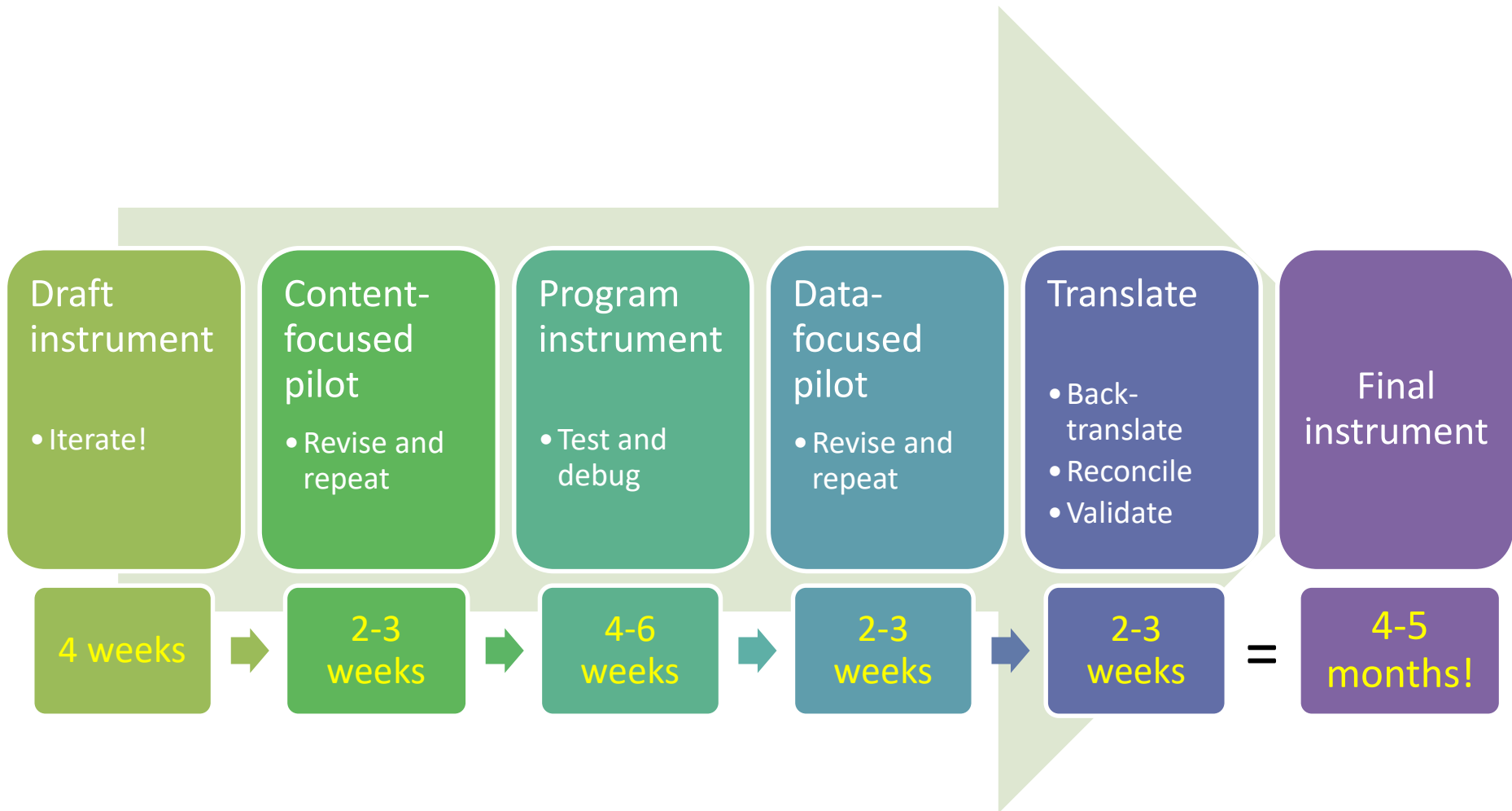
i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Process

Process

Draft instrument	Content-focused pilot	Program instrument	Data-focused pilot	Translate	Final instrument
• Iterate!	• Revise and repeat	• Test and debug	• Revise and repeat	• Back-translate • Reconcile • Validate	
4 weeks	2-3 weeks	4-6 weeks	2-3 weeks	2-3 weeks	= 4-5 months!

Process

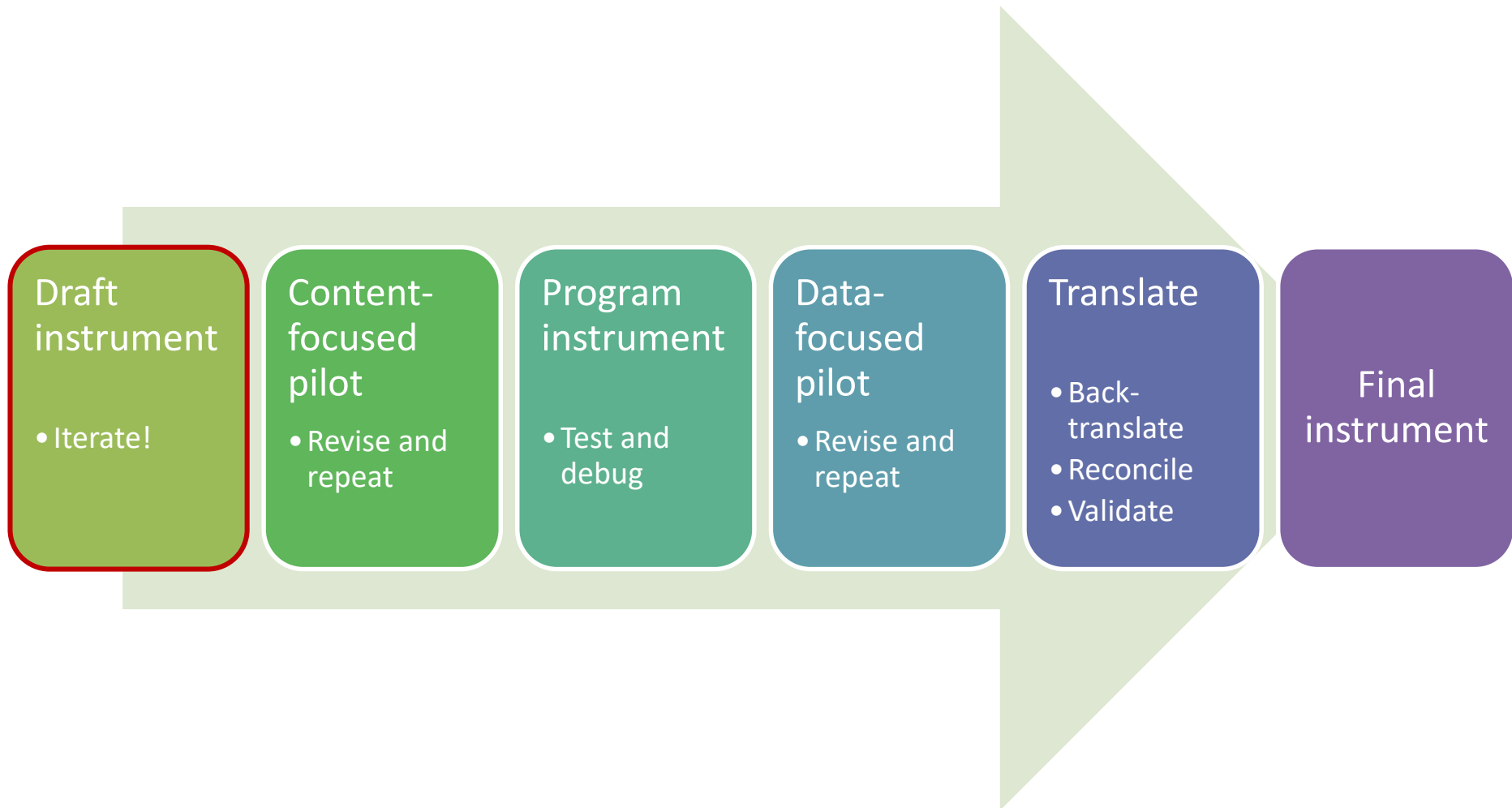| Draft instrument	Content-focused pilot	Program instrument	Data-focused pilot	Translate	Final instrument
• Iterate!	• Revise and repeat	• Test and debug	• Revise and repeat	• Back-translate • Reconcile • Validate	

Draft survey instrument

1. **Start by outlining modules**
   - Work from theory of change and pre-analysis plan
   - Input from research team (RT) critical

2. **For each module:**
   - Draft bullet list of key indicators
   - Discuss and agree on relevance and looping
     - Does the section apply to all HHs?
     - Repeat questions for all HH members?
     - Repeat for all plots a HH has or set a maximum?

3. **Develop survey questions for each module**
   - Do not start from scratch! (see next slide)

→ If a follow-up survey, work from the previous round, and only modify questions if absolutely necessary. Better to add or subtract than change.

Draft survey instrument

- Do not create questions from scratch!
- Start with a literature review of existing, reliable, well-tested surveys
  - Surveys in same country (regardless of sector)
  - Surveys in same sector (regardless of country)
  - A good source is microdata.worldbank.org; all catalogued studies include questionnaires and related documentation
- Compile relevant questions from existing surveys for each module
  - Add a column noting source for each question (e.g. LSMS, DHS, original)
- New tools from DIME Analytics forthcoming…
  - Questionnaire library on the new World Bank SurveyCTO server
  - Gold Standard modules, reflecting best practices in content and programming

Measurement challenges

- Think carefully about which indicators will be hard to measure
  - Things people do not know very well
  - Things people do not want to talk about
  - Abstract concepts

- Best to directly observe indicators whenever possible
  - Please read this sentence to me "…" is a better question than "Can you read a simple sentence?"

Challenge 1: things people don't know well

Anything respondent has to estimate, particularly over time.

- **E.g.** distance to grocery store, profit, consumption, income, plot size
- Prone to error due to long recall, math errors, low levels of numeracy

Strategies:

- Avoid asking the question, take another measure
  - E.g. through frequent follow-ups or diaries, or direct observation
- Build consistency checks into the instrument
  - Error! Amount harvested < amount consumed.
  - Enumerator verifies with respondent in real time
- Multiple measurements of same indicator
  - How many minutes does it take to walk to the grocery store?
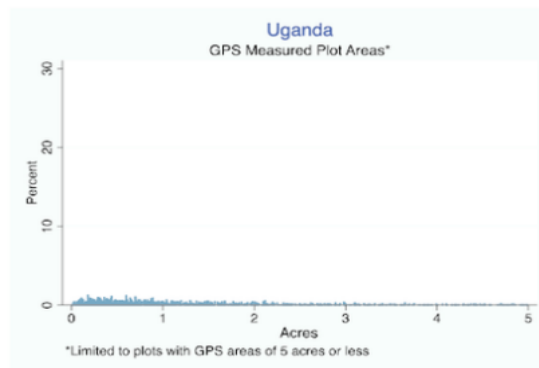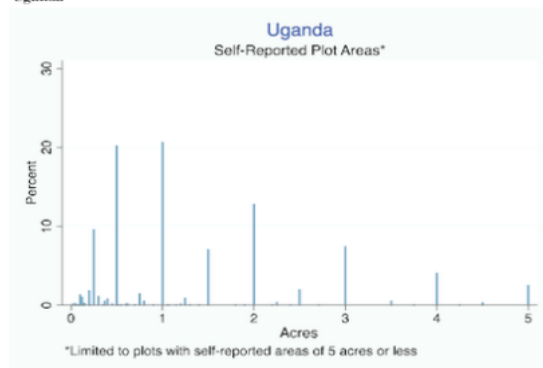 - How many miles away is the grocery store?

Challenge 1: things people don't know well
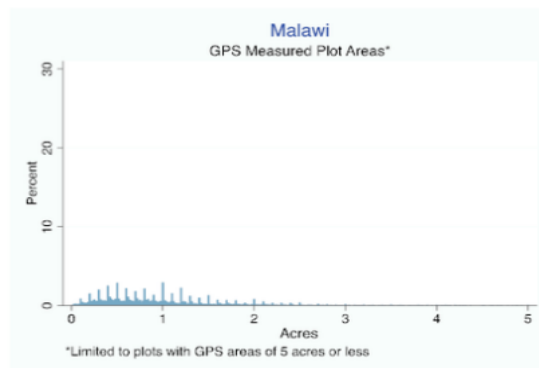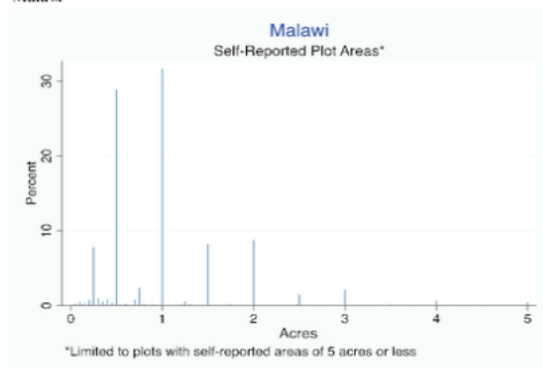
Carletto, Calogero; Gourlay, Sydney; Winters, Paul. (2013)
From guesstimates to GPStimates : land area measurement and implications for agricultural analysis (English).
Policy Research working paper ; no. WPS 6550. Washington, DC: World Bank.

Challenge 2: sensitive questions

Anything socially "risky" or something painful
- E.g. sexual activity, alcohol and drug use, domestic violence, etc

Strategies:
- Don't start with the hard stuff!
- Always ensure the comfort and privacy of respondent
- Consider asking the question in third person
- Possibility of self-administration of certain modules?
- Frame questions to avoid social desirability bias
- Methods: list randomization

Challenge 2: sensitive questions

Question:	Estimated Prevalence per 100 Males in Population	
Have you ever . . .	*Paper Version*	*Audio-Computer Version*
Had sex with a prostitute	0.7	2.5
Had intercourse with a female	68.1	63.9
Made a girl pregnant	7.9	6.5
Had sex with another male	1.5	5.5
Taken street drugs using a needle	1.4	5.2

Data: 1995 Survey of Adolescent Males (aged 15-19)

F. Turner, Charles & Ku, Lydia & M Rogers, S & Lindberg, Laura & Pleck, Joseph & Sonenstein, Freya. (1998).
Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology.
Science 280: 867-873. Science (New York, N.Y.). 280. 867-73. 10.1126/science.280.5365.867.

Challenge 3: abstract concepts

Hard to measure empowerment, bargaining power, social cohesion, risk aversion…

- E.g. "I feel more empowered now than last year"

Strategies:

1. Define what you mean

2. Choose the outcome

3. Design a good measure

- E.g. "I decide together with my partner whether to send my child to private vs. public school"

Challenge 3: abstract concepts

SECTION M: RISK AND AMBIGUITY AVERSION

| M1.1 | You are going to play a game where you draw a ball out of a bag without looking. If the ball you choose is the "right" color, then you win. You get to decide which bag to choose the ball from.

Bag One: In Bag One there are 4 RED balls and 6 YELLOW balls. You must pick a RED ball in order to win.

Bag Two: In Bag Two there are 10 balls – some are RED and some are YELLOW. You decide what color ball wins. You must then pick this color ball to win.

Which bag would you like to choose from? | 1 = Bag 1
2 = Bag 2
88 = Don't know | \|___\| |

Bag 1   Bag 2
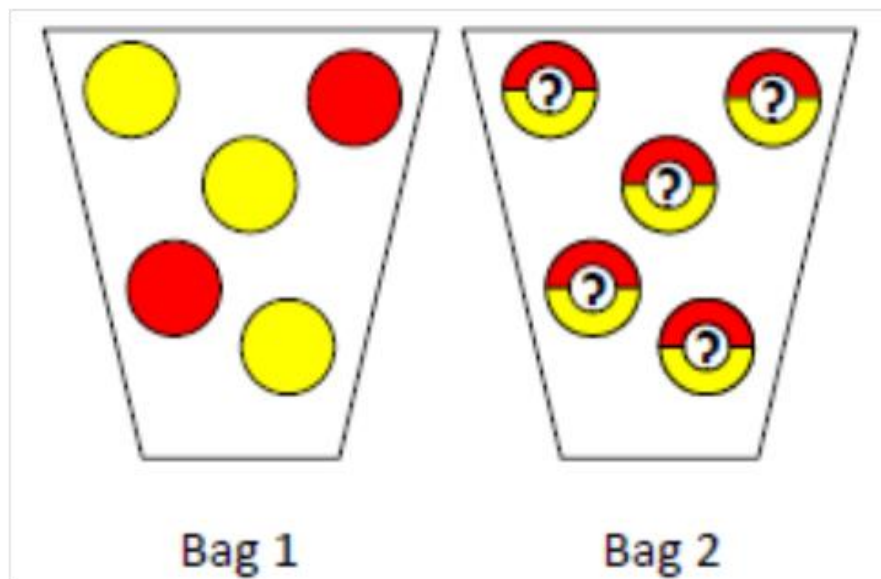
Process
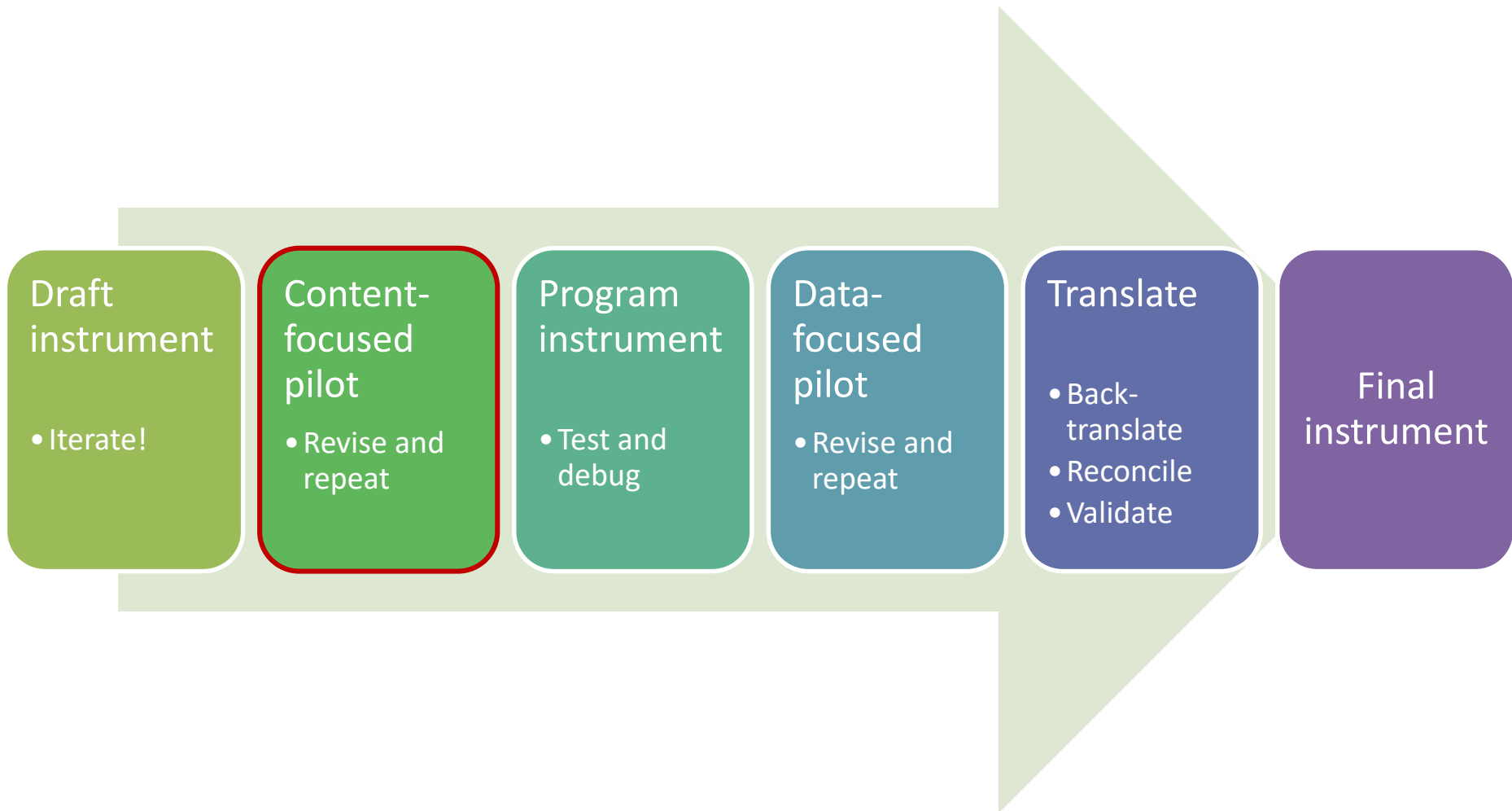
Content pilot

- "Pre-pilot"

  – **Objective**: Answer broad questions about survey design and context

  – **Mode**: Semi-structured interviews and focus groups

  – **Applies**: Necessary if designing survey from scratch

- Content-focused pilot

  – **Objective**: Missing questions? Missing answer options? Survey flow? Question wording? Survey length? Response variance?

  – **Mode**: Structured interviews

  – **Applies**: All new surveys. May omit for follow-up surveys *with few changes.*

Why pilot on paper?

- Get **better information**
  - Record open-ended responses quicker
  - Draw lines and arrows between questions to suggest restructuring
  - Record observations and feedback in the margins
  - Make edits to wording or note translation problems directly in the text

- **Pilot likely to result in significant changes** to the survey instrument
  - If already programmed, may hesitate to implement positive changes because of work required

- Revising programming can take longer than doing it from scratch

Who is involved?

Small team (2-4)

≥ 5 days

Field Coordinator

PI(s)

Experienced interviewers

Staff from gov./ implementing partner

Local consultants, students, project staff
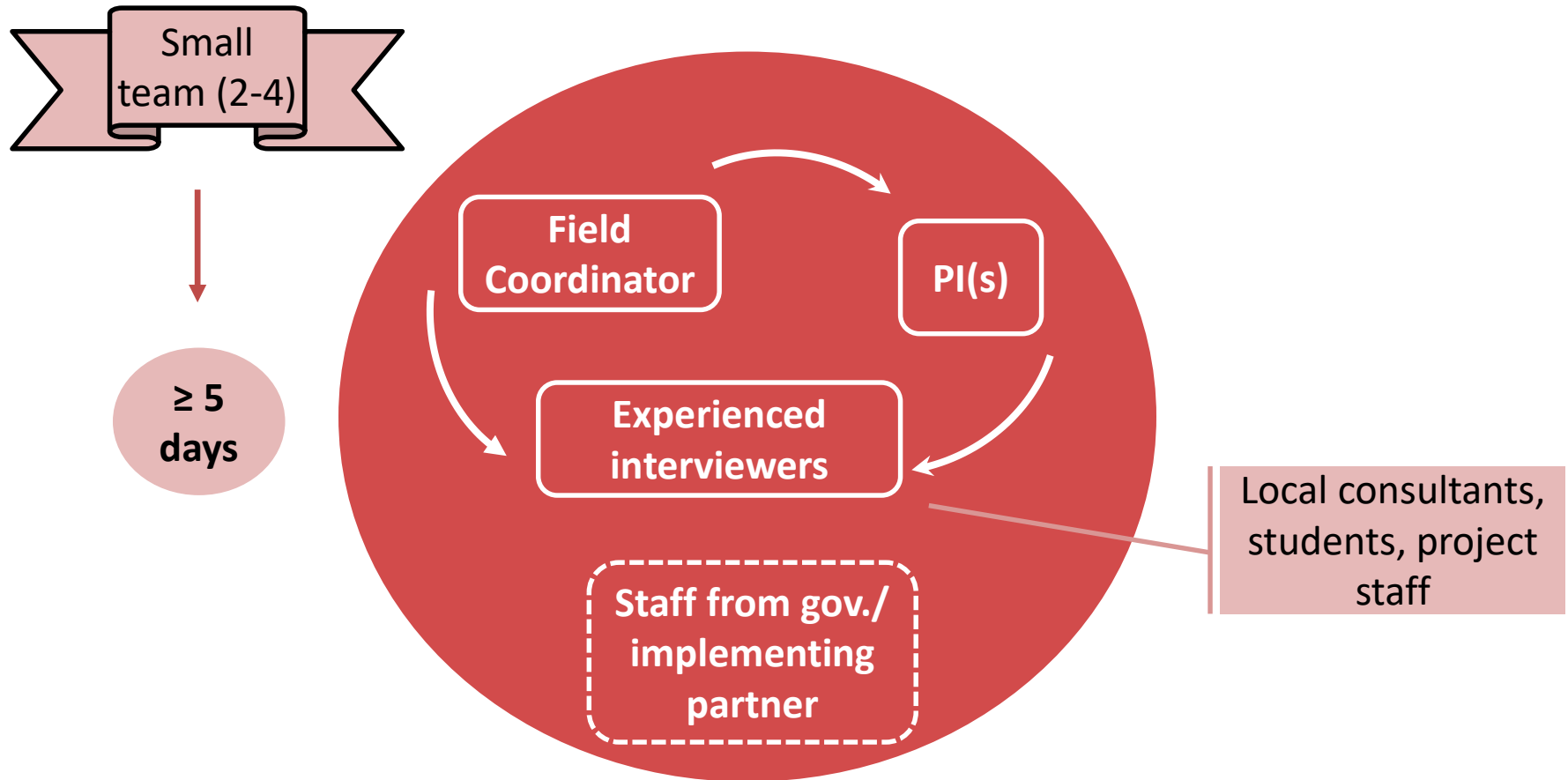
Planning an effective pilot

- Plan sufficient time for:

  - **Training** interviewers

  - Group **feedback**

  - **Revisions** based on feedback

- **Goal**: pilot until there are no more changes, min 1 week

  - PI present: debrief and edit survey at the end of each day

  - On your own: alternate pilot days with feedback days in which you debrief with PIs

- Take lots of **notes**! It will help focus enumerator training.

What to test?

Survey Design

- Do the questions make sense to the respondent?

  - Require explanations? Reaction time?

  - Follow-up with the enumerator (and possibly the respondent) on questions that seemed problematic: is the issue translation? Phrasing? Conceptual? Cultural?

- Are answer options comprehensive?

  - Ensure that all 'other' responses are specified and recorded

- Is the enumerator following the scripted translations?

  - If not, ask the enumerator to note any translation issues & discuss later

  - If you do not speak the language, you can still note if interviewer's questions were noticeably longer/shorter than the written question

What to test?

Interview flow and timing

- Interview flow?

  – Pauses may be areas where interviewers need more instruction

- Times when the respondent looks bored?

  – Uncomfortable? Losing interest?

- Could the order of modules be improved? The order of questions?

- How long is the interview?

  – For each module, note start and stop time

  – Pilot interviews will take **at least twice as long** as actual interviews: extra probing, notes, flow not great yet

Idea: 2-stage approach

- **LAB – first couple of days:**
  - Invite respondents to the office and be honest about the reason they're there -- to help you refine the survey instrument
  - Encourage them to tell you if something is unclear or if they believe someone could interpret it differently
  - Ask them if they would interpret it any other way
  - Ask for their opinion on the answer options

- **FIELD – when you have a survey that you're happy with:**
  - Test it in a scenario closer to the reality during fieldwork

Don't forget to test survey protocols!

- **Interview scheduling:**
  - Good/bad times?
  - Possible to book appointments?

- **Infrastructure**
  - Electricity? If blackouts are frequent, are there generators (and fuel)?
  - Coverage of mobile phone networks? Does mobile data work?

- **Sampling**
  - Use sampling protocol to select pilot respondents, to test
  - Is sample frame up-to-date?
  - If doing own listing, pilot listing survey

- **Geo-data**
  - If using tablet, how long to lock in signal? How accurate?
  - If using GPS unit, test and refine protocol for saving data

Process

Draft instrument
- Iterate!

Content-focused pilot
- Revise and repeat

Program instrument
- Test and debug

Data-focused pilot
- Revise and repeat

Translate
- Back-translate
- Reconcile
- Validate

Final instrument
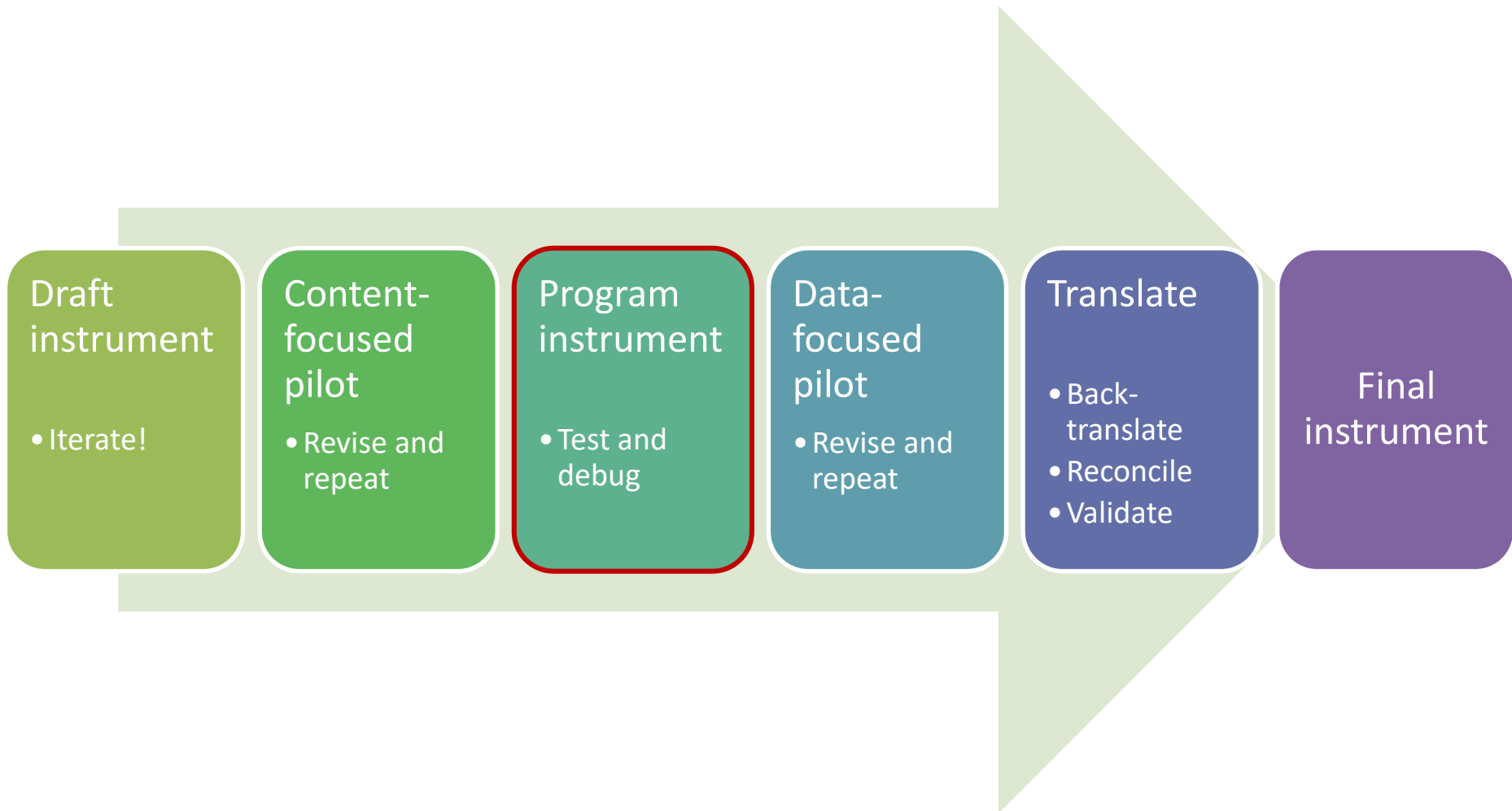
Timeline

- After questionnaire content and design is finalized
  - Changing questionnaire design after programming is done is a very common cause of data quality problems

- Build in sufficient time for programming!
  - Time needed depends on:
    - Length and complexity of instrument
    - Experience of primary programmer (and competing demands on their time)
    - Whether programming will start from scratch

- Typical timeline
  - 2-3 weeks for programming
  - 2 weeks for testing, de-bugging, and revising

Programming

- Many software options available; programming principles are consistent across platforms

- DIME surveys typically use *SurveyCTO,* where we have developed technical support and general protocols

- Data-quality assurance presentation will explore fundamentals

Start with pseudo code

- From list of modules, for each one describe at a high level what the code will do
  - On paper, or white board
  - Whether module will apply to all, how it flows with other modules, whether it is repeated
- After creating overall structure, add details
  - As details are added, language changes from plain English to something that closer to code
- Pseudo code example:

- Household Roster
  - Ask how many people in the household then repeat over that number to ask about name, age etc.
- Employment
  - repeat. One repeat for each household member over the age of 15
- Savings
  - Group. Ask if house hold have savings, if so ask savings question
- Access to Maternal Health
  - Group. If there is a female between the age of 16 and 45, then ask this module to the household head.

Start with pseudo code

- The more time you spend on pseudo coding the easier the actual coding

- When you have a final draft, discuss with someone else in the project.
  - Proofreading pseudo code is just as useful as asking someone to proofread text documents

- Think about data quality checks you will implement (especially built-in range and consistency checks) as part of this process

Process
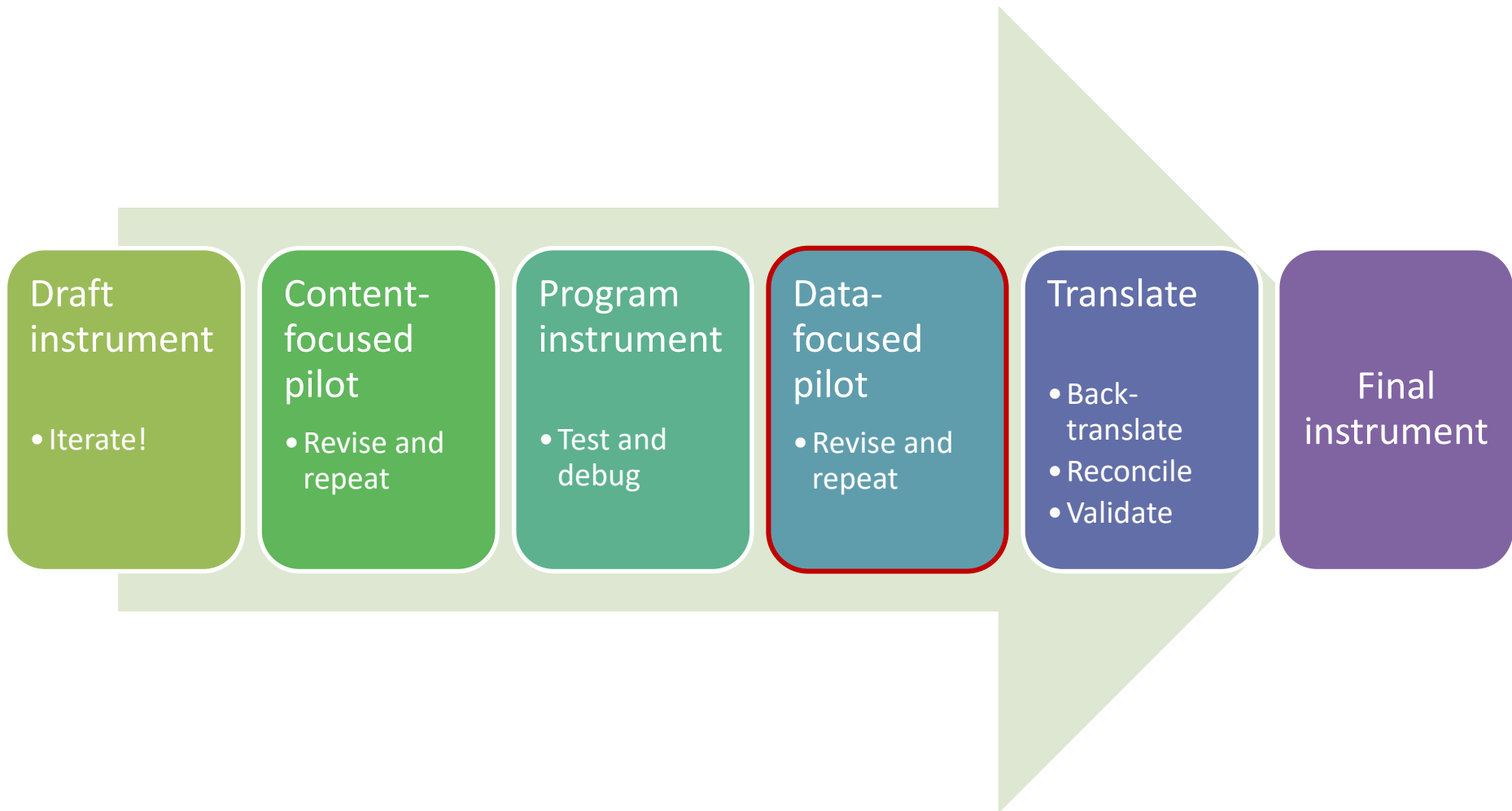
Data pilot - timeline

- After all content-pilot revisions made

- After programming complete and de-bugged

- After contract with survey firm signed

  - Ideally involve supervisors as data-pilot enumerators

  - Data-focused pilot has to be included in TORs

    - Must specify it is different than field testing during enumerator training

- To be finished at least 1 week before training

  - Allow for final revisions and debugging

Who is involved?
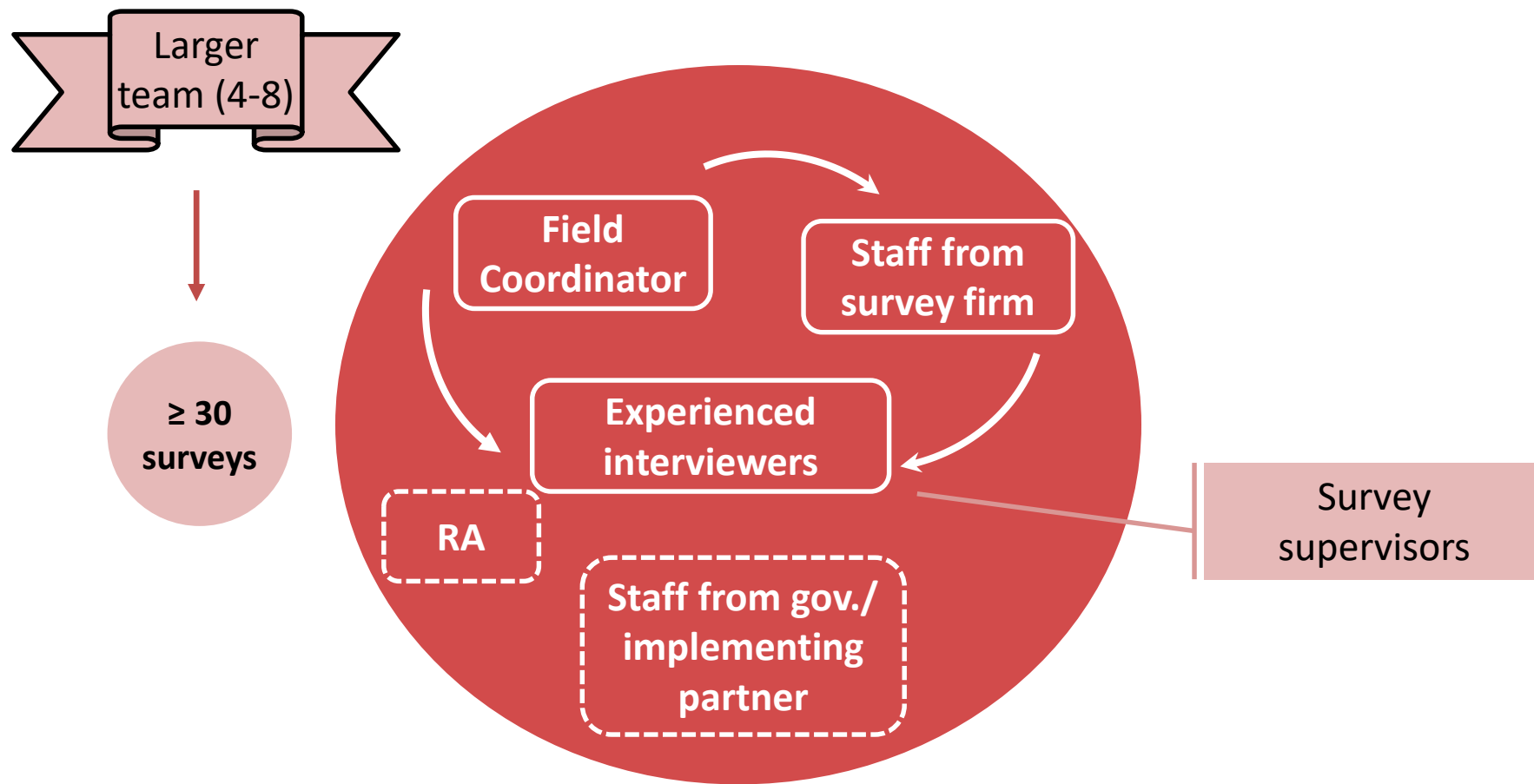
What to test?

Survey design and interview flow and timing

– Pay close attention to any survey design issues that came up in content pilot (and revisions made)

Programming

– Are all **skip patterns** working as expected?

– Are questions **displaying** properly on the screen?

– Are there any questions that should be **grouped / ungrouped**?

– Did all **modules appear**?

– Are **built-in data checks** working correctly (for outliers or inconsistent responses)?

What to test?

Data

- Download pilot data from server and import to Stata:
  - Use Stata template provided by SurveyCTO

- Check that:
  - All variables appear and are correctly labeled (and not too long)
  - Value labels in research team language (and not too long)
  - All 'pre-loaded' data appears as expected

- Test all skip patterns, check for (unexpected) missing data

- Check variance: both high and low
  - Low → if all respondents give same answer, data point may not be informative
  - High → question may need to be more precise or checks built in to alert enumerator of extreme values in real time

What to test?

High frequency checks

- Use the dataset to program a **do-file for high-frequency checks**
    - More on this in later session and hands-on session

- **Test the do-file** on pilot data and de-bug as needed

- Export **results of checks**
    - Use as example to discuss and agree with the survey firm on a final format for communicating and resolving issues
    - Test the survey firm responses of the HFC flags – often these are misunderstood
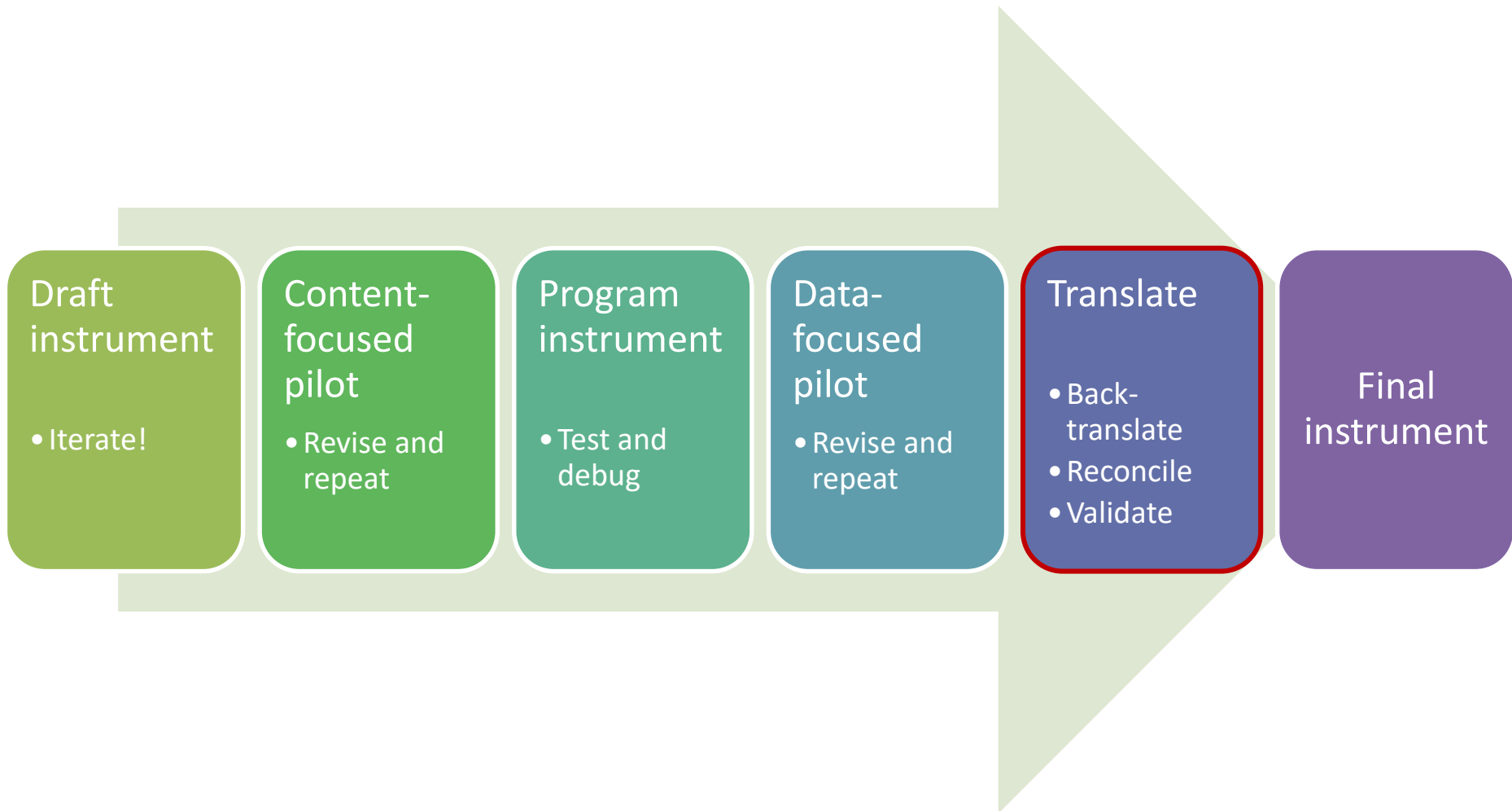
Data-focused pilot using SurveyCTO

- Pilot SurveyCTO form should be uploaded to project server (not development server)

- Indicate *pilot* in questionnaire name and id
  - keep pilot in name until finished collecting pilot data
  - reduces risk that someone uses pilot form to collect final data and ensures data is saved separately

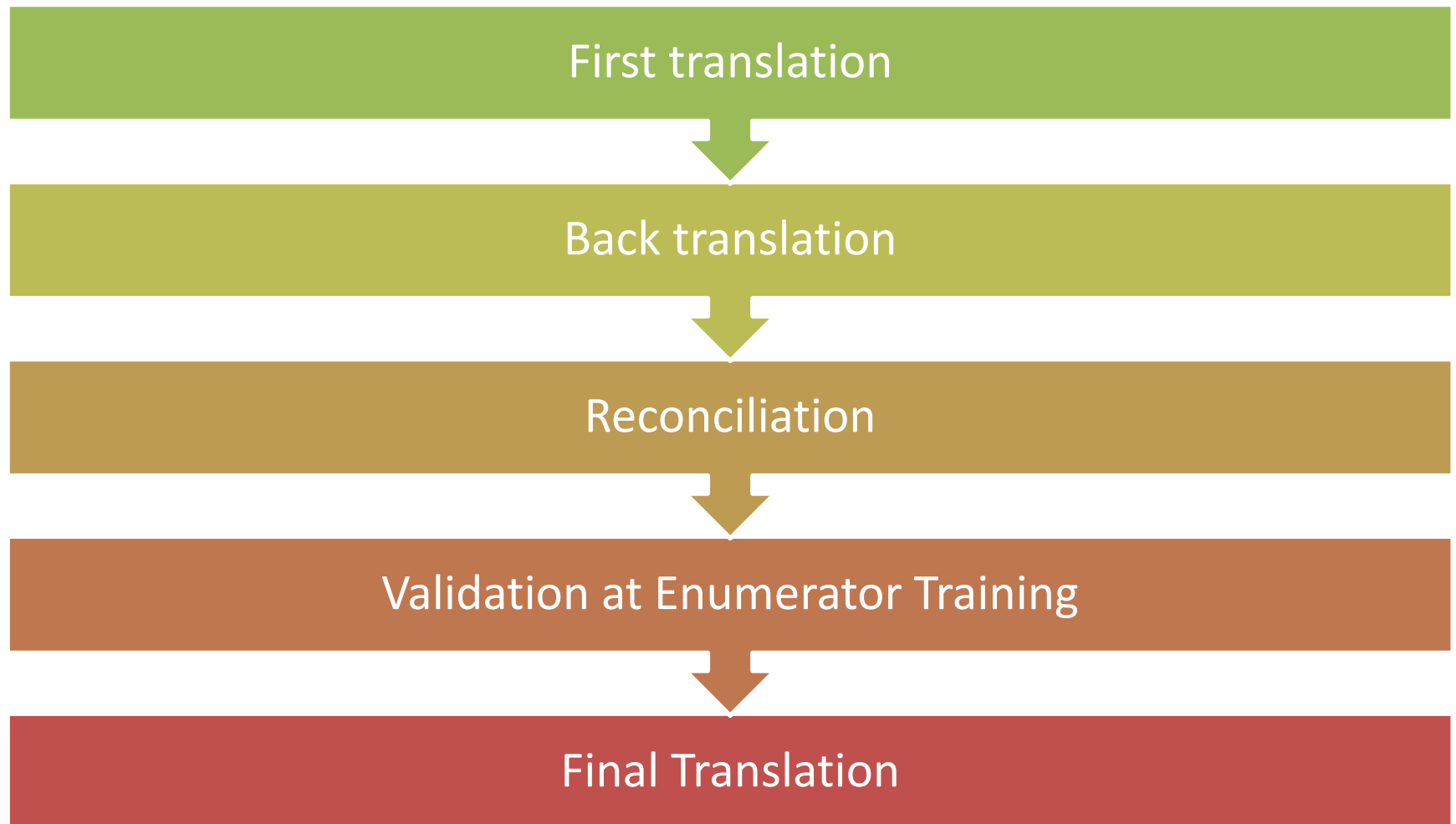# Process

Translation

- Enumerators should **NEVER** have to translate on the fly

- Goal: all enumerators and respondents must have exact same understanding of each question

- Translate to each language in population of interest
  - Always start first translation with version written in research team language

- Be very careful of version control between printable and programmed surveys!!

Translation

- Beware!!
  - Bad translation → field errors, confusion, bad data, etc.

- Who does the first translation matters
  - Good: professional translator; sector-specific knowledge; survey experience
  - Bad: survey firm management, government counterparts with other tasks

- Translation takes significant time and skill
  - People who do not have translation experience often underestimate time
  - Fluency in both languages is necessary but not sufficient

Translation process

First translation

↓

Back translation

↓

Reconciliation

↓

Validation at Enumerator Training

↓

Final Translation

Translation process

- Back translation
  - Translate local language to English (without seeing original English version)
  - Done by skilled translator with no connection to first translator
- Reconciliation
  - FC or experienced local consultant compares and reconciles discrepancies
- Validation
  - Expect to make **lots** of corrections and refinements in pilot and training
  - Discuss each translation during training as a triple-check (ensure you have bilingual staff dedicated to correcting translation)
- Survey form should include all relevant languages
  - Use multiple language columns in SurveyCTO
  - This helps with version control as changes can be made in both languages simultaneously
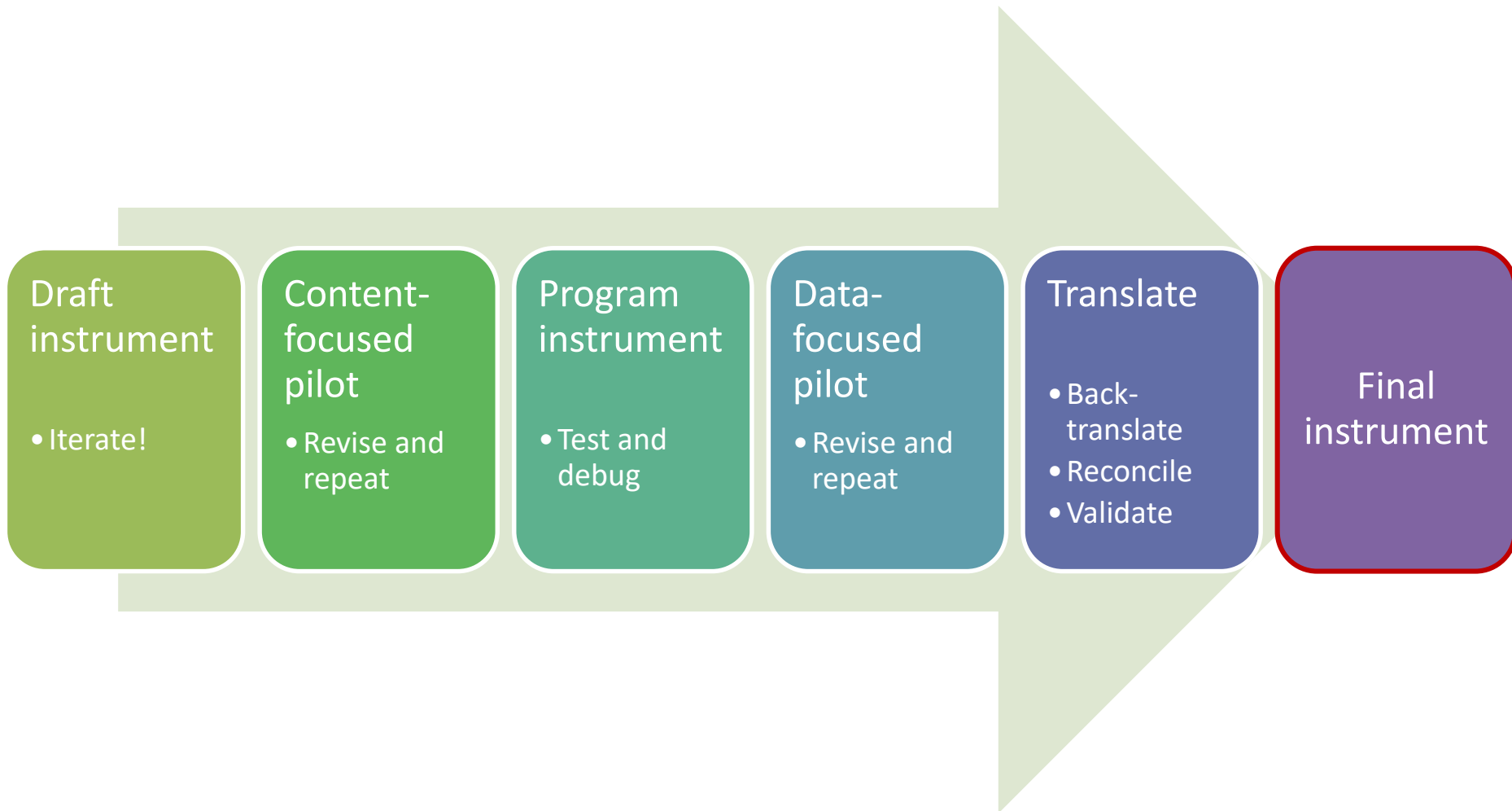
What about oral languages?

- During training, practice every question in the local language(s)

- Get several people to say out loud how they would translate it and agree on a common version

- Do demonstrations in front of the class in the local language

- Have a printed survey with space for writing the transliterated version of the questions

- Include transliterated key-terms as prompts in the programmed survey

Process

Draft instrument	Content-focused pilot	Program instrument	Data-focused pilot	Translate	Final instrument
• Iterate!	• Revise and repeat	• Test and debug	• Revise and repeat	• Back-translate • Reconcile • Validate	

Final instrument

- Need to have a readable, printable version of final instrument
  - This is NOT the programmed instrument!

- Important for:
  - Enumerator training
  - Sharing with government counterparts
  - {Possibly} IRB or local ethics approvals
  - Submission to donors
 - Archival in micro-data catalogue

Final instrument

- No perfect technical solution

- **Best practice** → excel workbook with multiple tabs and linked content
  - One tab per module (formatted), one tab for all programming
  - Cells with question text linked to programming tab

- **Other options**

  - Export printable version from SurveyCTO server
    - Export can be messy, especially if long list of response choices
    - Example: an irrigation survey, initially 900+ pages. In a couple hours cleaned up to <200 pages. But still unwieldy

  - Import 'survey' tab into stata and clean up with a do-file
    - Drop unnecessary rows and columns
    - Will need to add formatting to help distinguish repeat groups etc

Thank you! Questions?

Steven Glover

Sékou Kone

i2i DIME
TRANSFORM DEVELOPMENT
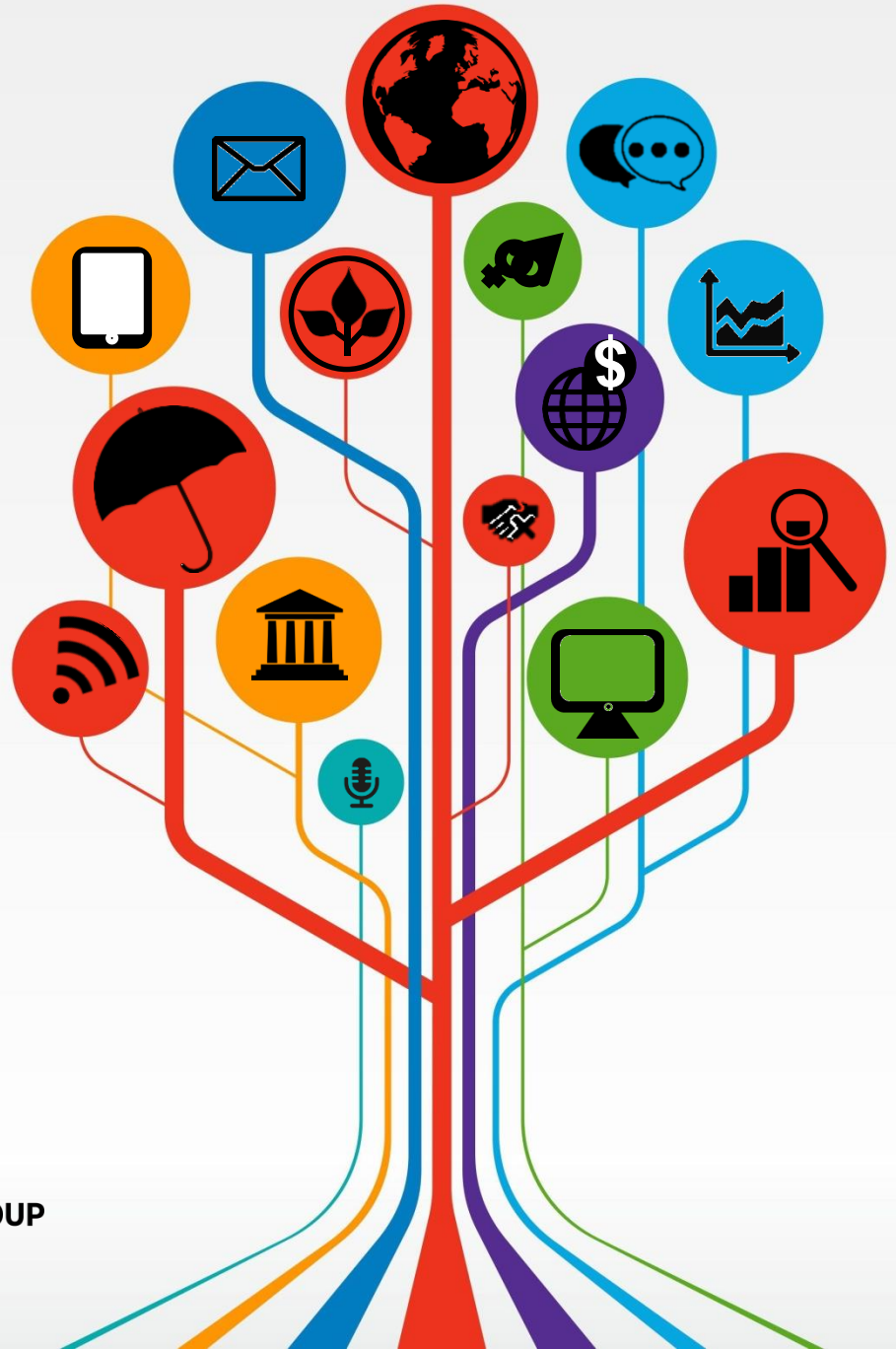
WORLD BANK GROUP

Appendix slides

A. List experiments

B. *SurveyCTO* printable forms

C. Programming

D. Further resources

A. List experiments

- Randomly divide sample into two groups
  - *Direct response:* report how many of N items are true, where items are neutral and non-sensitive
  - *Veiled response:* report how many of N+1 items are true
    - N items are identical to control group's items
    - the N+1st item is a sensitive item

- Estimate the population mean for the sensitive item (N+1st) by differencing out mean of sum of N items from the control

A. List experiments

- Example from Coffman et al, 2013

Panel A: Comparison of Direct Report and Veiled Report Treatments

Direct Report	**Veiled Report**
• I remember where I was the day of the Challenger space shuttle disaster.	• I remember where I was the day of the Challenger space shuttle disaster.
• I spent a lot of time playing video games as a kid.	• I spent a lot of time playing video games as a kid.
• I would vote to legalize marijuana if there was a ballot question in my state.	• I would vote to legalize marijuana if there was a ballot question in my state.
• I have voted for a political candidate who is pro-life.	• I have voted for a political candidate who is pro-life.
	• I consider myself to be heterosexual.
Please fill in the bubble that corresponds to the total number of statements above that apply to you. 0 1 2 3 4	**Please fill in the bubble that corresponds to the total number of statements above that apply to you.** 0 1 2 3 4 5

Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson. "The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated." *Management Science* (2016).

A. Potential issues with list experiments

- require people to count/add, possibly introducing noise to the data
  - especially if the list is long.
- unless the "innocent" questions are completely unrelated and have a known distribution, then there is a chance that the treatment in your RCT might have an effect on its distribution.
  - But designing your common questions that way makes your sensitive ones stand out even more.

B. Programming

- Never start with an empty excel sheet.
  - SurveyCTO provides template forms with basic settings in place

- Form ID controls how data saved on server and name for export
  - When ID updated, all new data collected saved in new dataset on server with reference to new id.
  - Should update ID if have to change instrument while survey is ongoing
    - Will result in 2 datasets and 2 import do-files.
    - Merge in Stata, adding 'version' variable first (likely useful for analysis)

- Good practice to set both form name and id to same value
  - Some functionality in questionnaire name
    - If start name with word *test*, form will not be visible to enumerators

- Naming convention: PROJECT_SurveyRound_Unit_Version
  - Example: KBMAP_BL_HH_v1

C. Printable version from *SurveyCTO*

- On the [Design](#) tab, choose *Download*, then *Printable version* from within the *Your forms* list
- The printable version is in HTML (web) format. To edit before printing, open the .html file in Microsoft Word, and *Save As* .docx
- May include notes, to explain when groups or fields will appear (when *relevant*), what restrictions there are on entries (*constraints*), etc
  - In the *survey* sheet, simply enter explanations into the *note* column
  - All notes will be included in the printable version (and will not appear anywhere else)
- May also enter text to appear in the response area to the right of questions, by including that text in the *response_note* column of your *survey* sheet
  - Put|____|____| if you are looking for two letters or numbers
  - For a checkbox, put a hollow square like □ (a special HTML character: enter "□", without quotes, into *response_note* column)
  - For a radio button, simply enter a capital O.

D. Further Resources

- WB microdata catalogue:

  http://microdata.worldbank.org/index.php/catalog/impact_evaluation

- JPAL microdata catalogue: https://dataverse.harvard.edu/dataverse/jpal

- DHS microdata catalogue: http://microdata.worldbank.org/index.php/catalog/dhs

- IFPRI microdata catalogue: https://dataverse.harvard.edu/dataverse/IFPRI

- International Household Survey Network - survey catalogue:

  http://catalog.ihsn.org/index.php/catalog