Evaluating Data quality in Time Diary Surveys Using Paradata

Ana Lucía Córdova Cazar
Robert F. Belli

This material is based upon work supported by the National Science Foundation under Grant No. SES - 1132015.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Focus of this research

- How can we measure ‘interview complexity’?
- Is there a relationship between interview complexity and data quality in calendar and time use surveys?

Theoretical Background:
Data Quality in Calendar and Time Diary Surveys (1)

- Measurement error
  - Occurs when the respondent's answer deviates from a "true" value
  - Can be attributed to the respondent, interviewer, instrument, mode
  - Used in time use diaries to examine data quality

Theoretical Background: Data Quality in Calendar and Time Diary Surveys (2)

- Generally, no gold standard in time use diaries
  - Need to look at possible indicators of data quality that may reflect the concept itself (Lyberg, 2012)
    - Failure to remember an activity
    - Failure to provide sufficient information (where, with whom)
    - Refusal to provide an answer
- Interview complexity as a determinant of data quality in calendar and time diary surveys:
  - More complicated interviewing situations will lead to poorer data quality than less complicated ones (Belli et al., 2004)

Theoretical Background: Paradata and Measurement Error (1)

- Use of paradata based on the notion that measurement error occurs when there is a breakdown in the cognitive response process (Olson and Parkhurst, 2012)
  - Comprehension, Retrieval, Judgment, Editing
- Paradata
  - May reflect these cognitive response processes
  - May serve as proxies to identify breakdowns
- Several examples of previous empirical research
  - E.g. Longer response times may be a sign of question complexity and / or difficulties in the response process or potential problems with survey questions (Yan and Olson, 2013; Yan and Tourangeau, 2008; Bassili, 1996)

Theoretical Background: Paradata and Measurement Error (2)

- Types of paradata:
  - Response times
    - Measured in milliseconds (automatic process thanks to CAI software)
  - Mouse clicks
    - Back-ups and changes
  - Keystrokes or audit trails
    - Highest level of granularity
  - Call and Case history files
    - Number of call attempts before obtaining a complete interview
    - Number of completed interviews by interviewers

The *beauties* of Paradata: An example from the Survey of Income and Program Participation (SIPP)

"2/9/2013 11:23:40 AM","Enter EHC","Key:00000010"

"2/9/2013 11:23:50 AM","Mouse:108,119","Message:LeftDown","HitTest:Client"
"2/9/2013 11:23:50 AM","Mouse:108,119","Message:LeftDown","HitTest:Client"
"2/9/2013 11:23:50 AM","Mouse:108,119","Message:LeftUp","HitTest:Client"
"2/9/2013 11:23:50 AM","Mouse:108,119","Message:LeftUp","HitTest:Client"
"2/9/2013 11:23:50 AM","EHC Action Performed: Topic Selected: 1 Landmarks"

"2/9/2013 11:24:10 AM","Leave Field: BCore_Middle.TEHC[1].BLandMark_Screener2","Cause:Leave Text Field","Status:Normal","Value:1"

"2/9/2013 11:24:10 AM","EHC Action Performed: Radio button checked Screener2"

"2/9/2013 11:24:10 AM","Leave Field: BCore_Middle.TEHC[1].BLandMark_Screener2","Cause:Leave RadioButton click","Status:Normal","Value:1"

"2/9/2013 11:24:11 AM","Leave Field: BCore_Middle.TEHC[1].BLandMark_Screener2","Cause:Leave Screener2 TextBox","Status:Normal","Value:1"

"2/9/2013 11:24:56 AM","Leave Field: BCore_Middle.BLandMark[1].PeriodNum","Cause:Leave Text Box","Status:Normal","Value:1"

"2/9/2013 11:24:56 AM","Leave EHC","Key:00000010"

Data

- 2010 American Time Use Survey (ATUS)
  - Annual time diary survey
  - Respondents report each activity from previous day (duration, with whom, where)
  - CATI survey (every interviewer keystroke is captured automatically).
  - 13260 respondents (Final n= 13,144)

Measures (1)

Predictor Variables

- Interview complexity latent factor
  - Paradata indicators (14 observable paradata variables >> next slide)
  - Substantive variable (Number of reported activities)
- Respondent demographics (Age, gender, education, marital and employment status)

Outcome Variables (Data Quality)

- ATUS Error 1: Insufficient detail error (Reported activity could not be coded)
- ATUS Error 5: Memory gap error

Measures (1)

Predictor Variables

- Interview complexity latent factor
  - Paradata indicators (14 observable paradata variables >> next slide)
  - Substantive variable (Number of reported activities)
- Respondent demographics (Age, gender, education, marital and employment status)

RQ 1

Outcome Variables (Data Quality)

- ATUS Error 1: Insufficient detail error (Reported activity could not be coded)
- ATUS Error 5: Memory gap error

Measures (1)

Predictor Variables

RQ 1

- Interview complexity latent factor
  - Paradata indicators (14 observable paradata variables >> next slide)
  - Substantive variable (Number of reported activities)
- Respondent demographics (Age, gender, education, marital and employment status)

Outcome Variables (Data Quality)

RQ 2

- ATUS Error 1: Insufficient detail error (Reported activity could not be coded)
- ATUS Error 5: Memory gap error

Measures (2)

Indicators of Interview Complexity

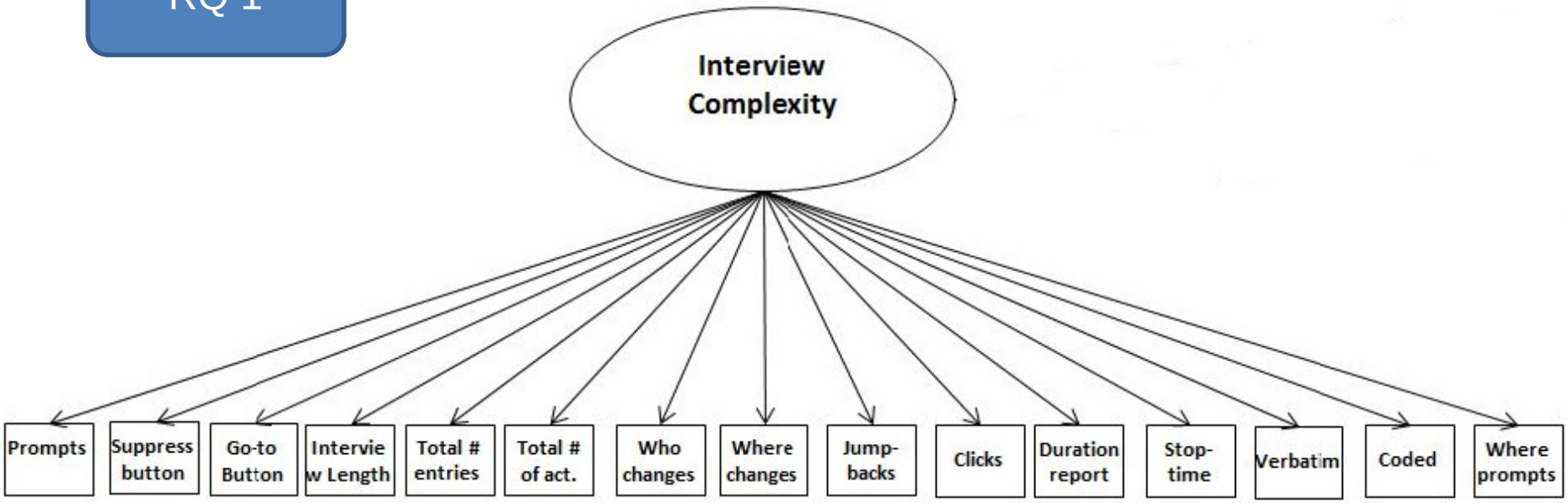
1. Total number of prompts per interview
2. Use of “suppress” button during interview
3. Use of “Go to” button during interview
4. Interview length in minutes
5. Total number of entries
6. Total number of “who” changes
7. Total number of “where” changes
8. Total number of entry edits (“jump-backs”)
9. Total number “clicks” without edits
10. Total number of times activity was reported as a duration
11. Total number of times activity was reported as a stop-time
12. Total number of times activity was reported as verbatim
13. Total number of activities that were directly coded
14. Total number of ‘where’ prompts
15. Total number of activities reported

Analytic Rationale (1)

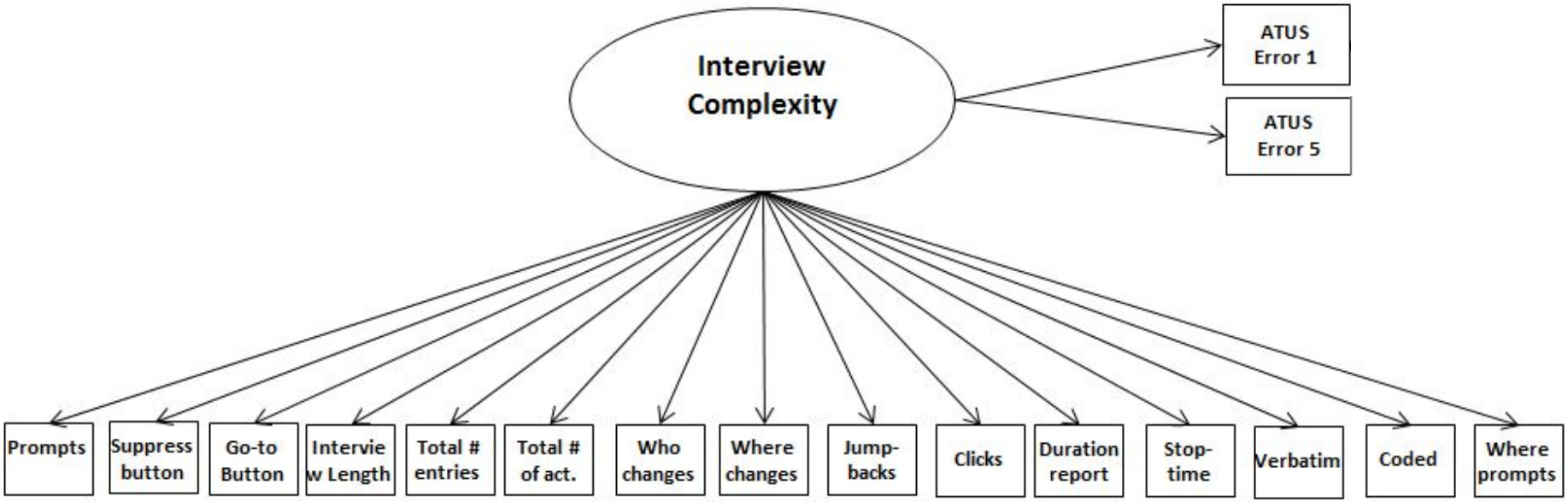
- A structural equation model (SEM) is estimated to predict the relationship between interview complexity and data quality.
  - A measurement model for interview complexity RQ 1
  - A structural model to predict both errors simultaneously RQ 2

Analytic Rationale (2)

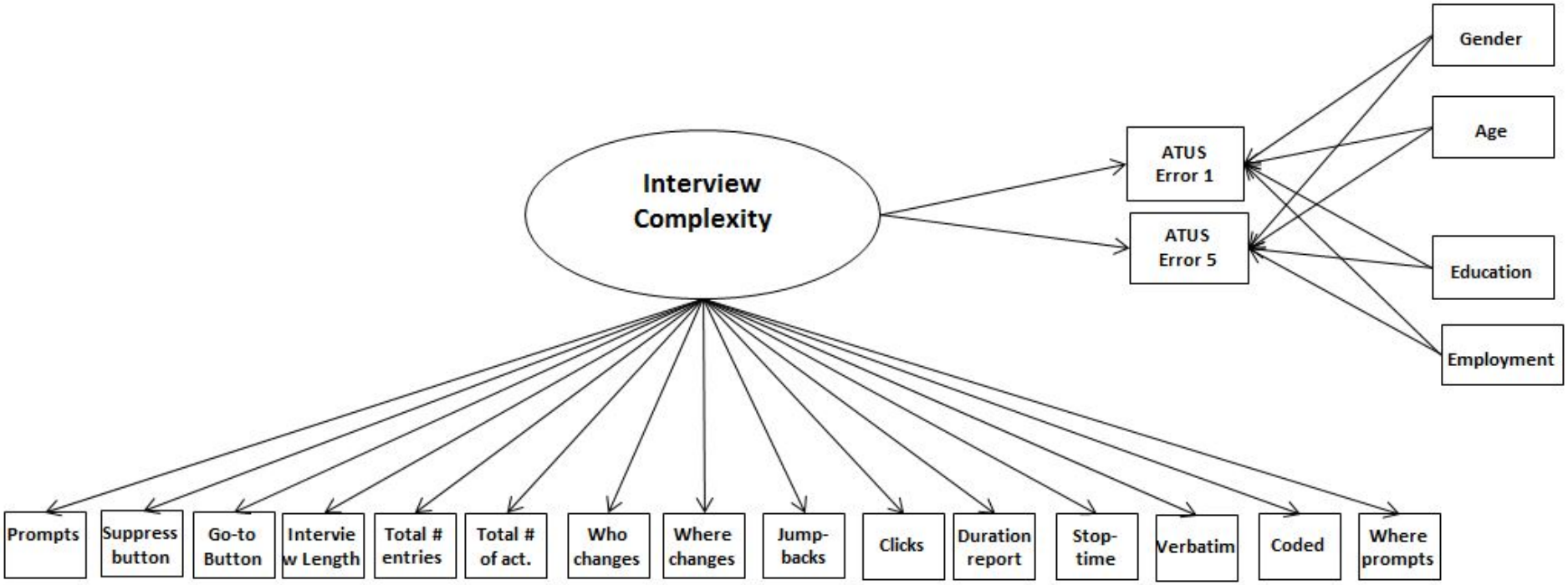
RQ 1



Analytic Rationale (3)

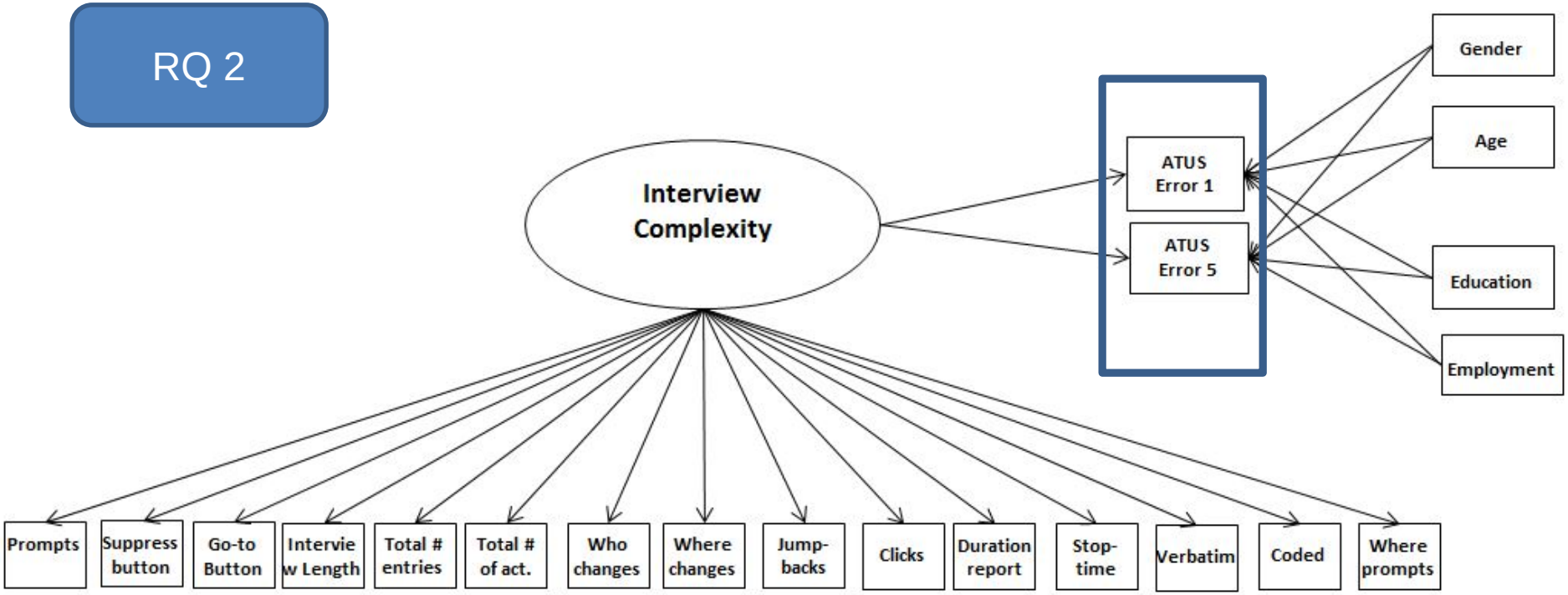


Analytic Rationale (4)



Analytic Rationale (4)

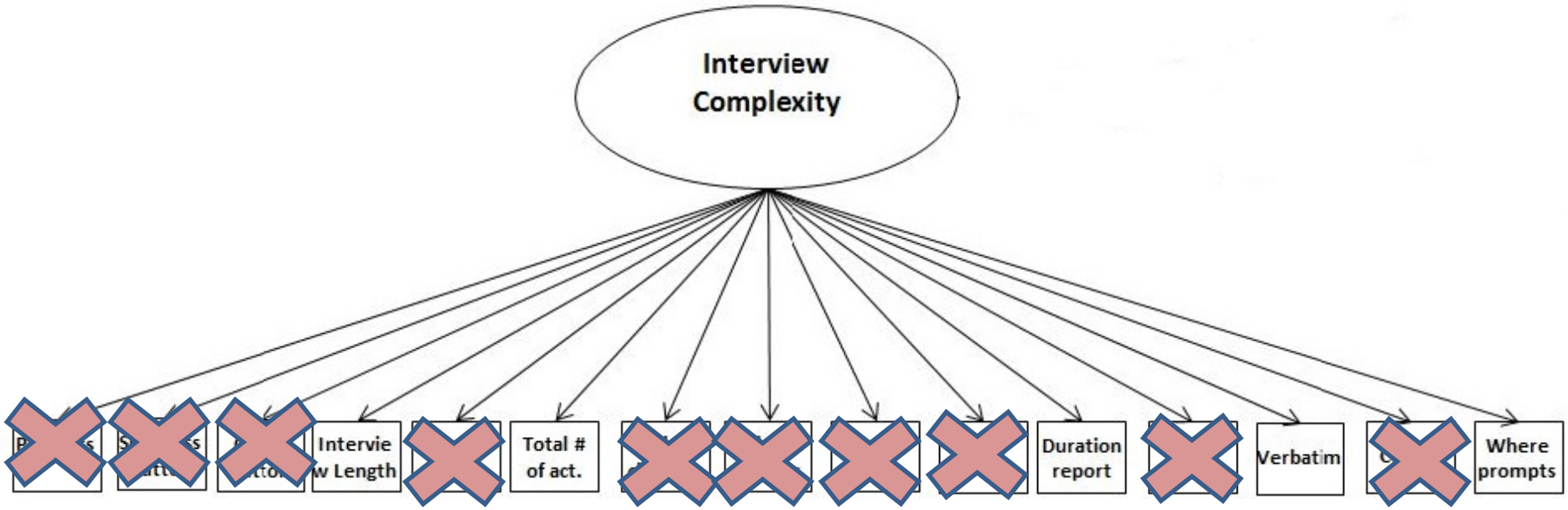
RQ 2



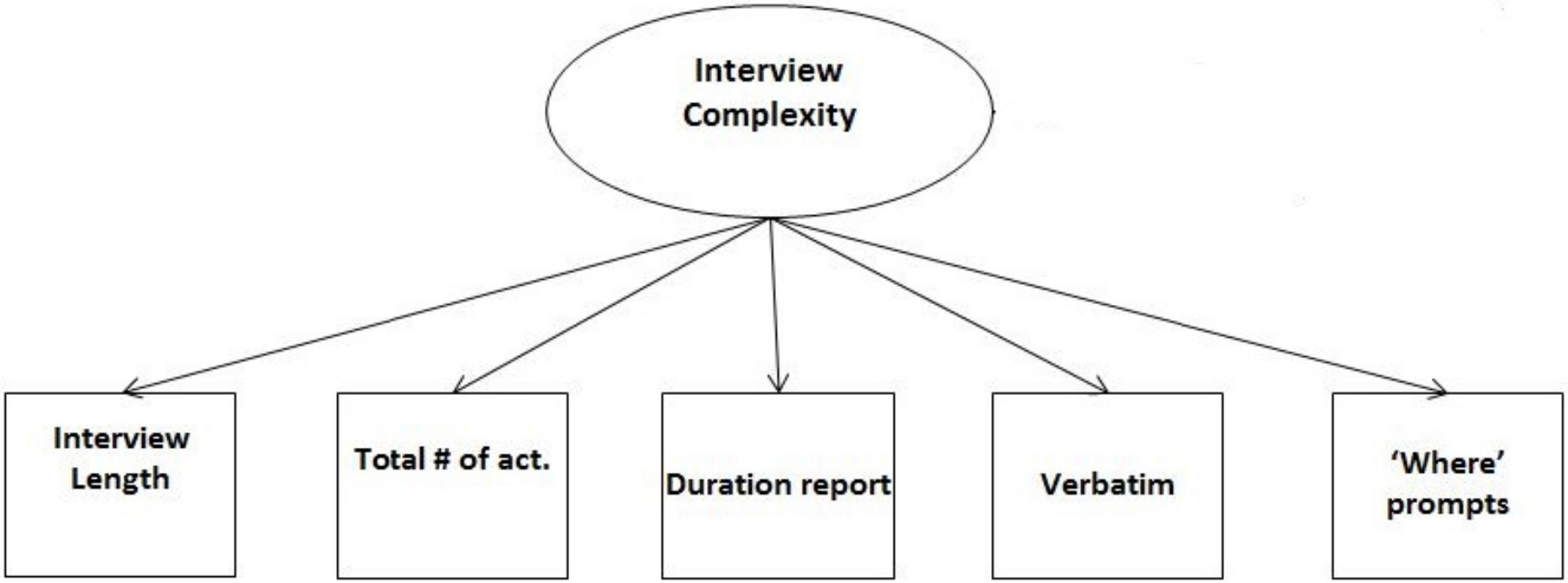
Results (1): Measurement Model (Measuring Interview Complexity)

- Variances and covariances of observed variables are analyzed
- Reliability and dimensionality of fifteen items were assessed using CFA
  - Robust maximum likelihood estimation (MLR)
  - CLUSTER option was used to indicate that respondents are nested within 69 interviewers in this dataset.
- This model was estimated such that higher values indicate greater levels of interview complexity for all items

Results (2)
Measurement Model



Results (3)
Measurement Model



Results (4)
Measurement Model: Model Fit Statistics

Model	# Items	# Estimated Parameters	Chi-Square Value	Chi-Square Scale Factor	Chi-Square DF	Chi-Square p-value	CFI	RMSEA Estimate
One-Factor	14	42	14566.890	8.132	77	<.0001	.140	.120
One-Factor (Error covariance between Coded and Verbatim)	14	40	11284.550	9.917	79	<.0001	.335	.104
One-Factor (All previous items set to 0 now removed)	11	34	8362.969	6.323	43	<.0001	.350	.121
One-Factor (Error covariance between clicks and jump-backs)	11	35	6622.109	6.539	42	<.0001	.486	.109
One-Factor (Remove Jump-backs, clicks and coded activity)	5	15	229.341	10.393	5	<.0001	.870	.058
One-Factor (Error covariance between where prompts and total number of activities)	5	16	27.390	14.150	4	<.0001	.986	.021

Results (5)
Measurement Model: Item factor loadings

Model Parameter	Unstandardized		Standardized	
	Estimate	Standard Error	Estimate	Standard Error
Factor loadings				
Duration of the interview	0.261	0.010	0.671	0.026
Total number of activities	0.759	0.016	0.951	0.004
Number of activities reported as a duration	0.105	0.003	0.943	0.005
Number of activities reported as verbatim	0.715	0.014	0.947	0.002
Prompts asking "where"	0.727	0.013	0.953	0.006
R2 for Item Variances				
Duration of the interview			0.450	0.035
Total number of activities			0.905	0.007
Number of activities reported as a duration			0.888	0.009
Number of activities reported as verbatim			0.896	0.003
Prompts asking "where"			0.908	0.011

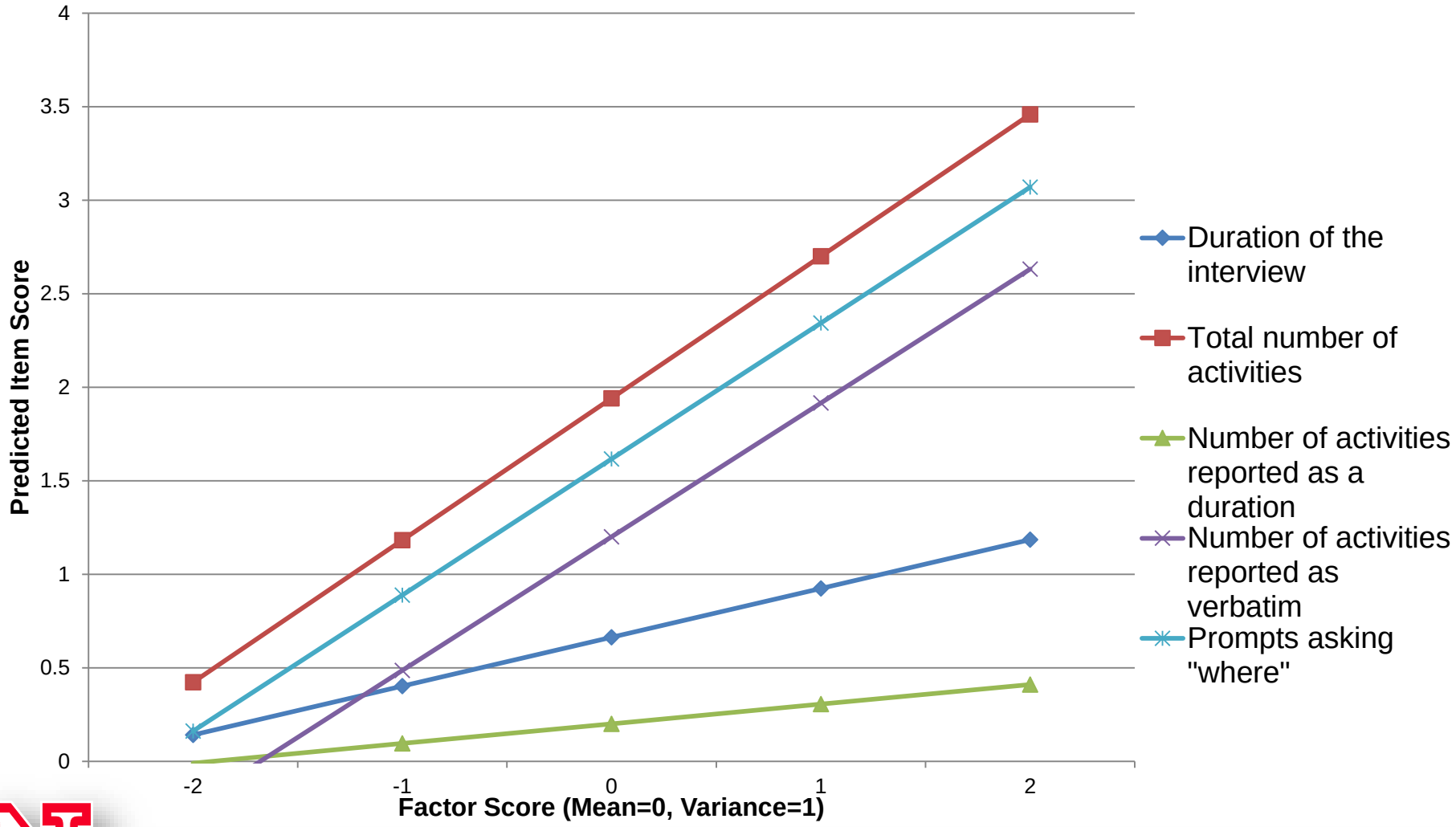
Results (5)
Measurement Model: Item factor loadings

Model Parameter	Unstandardized		Standardized	
	Estimate	Standard Error	Estimate	Standard Error
Factor loadings				
Duration of the interview	0.261	0.010	0.671	0.026
Total number of activities	0.759	0.016	0.951	0.004
Number of activities reported as a duration	0.105	0.003	0.943	0.005
Number of activities reported as verbatim	0.715	0.014	0.947	0.002
Prompts asking "where"	0.727	0.013	0.953	0.006
R2 for Item Variances				
Duration of the interview			0.450	0.035
Total number of activities			0.905	0.007
Number of activities reported as a duration			0.888	0.009
Number of activities reported as verbatim			0.896	0.003
Prompts asking "where"			0.908	0.011

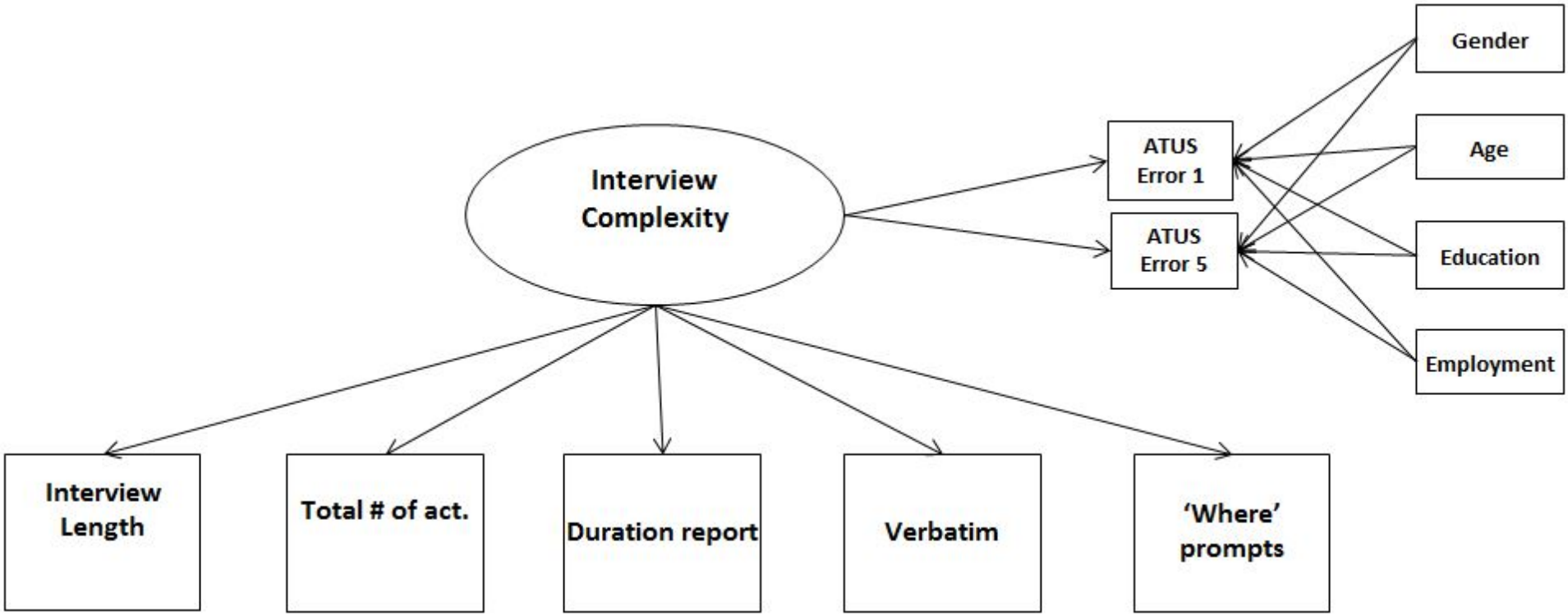
Results (5)
Measurement Model: Item factor loadings

Model Parameter	Unstandardized		Standardized	
	Estimate	Standard Error	Estimate	Standard Error
Factor loadings				
Duration of the interview	0.261	0.010	0.671	0.026
Total number of activities	0.759	0.016	0.951	0.004
Number of activities reported as a duration	0.105	0.003	0.943	0.005
Number of activities reported as verbatim	0.715	0.014	0.947	0.002
Prompts asking "where"	0.727	0.013	0.953	0.006
R2 for Item Variances				
Duration of the interview			0.450	0.035
Total number of activities			0.905	0.007
Number of activities reported as a duration			0.888	0.009
Number of activities reported as verbatim			0.896	0.003
Prompts asking "where"			0.908	0.011

Results (6)
Measurement Model: Discrimination of Items



Results (6)
Structural Model



Results (7)
SEM: Logistic regression odds ratio results

Insufficient detail error	
Interview Complexity	1.575*
Graduate degree	1.241*
Age	0.994*
Employment	0.785*
Memory gap error	
Interview Complexity	1.326*
Graduate degree	0.713*
Age	1.021*
Employment	0.746*

Conclusions (1)

- Interview complexity predicts both errors in a significant positive way
  - For an additional unit in the interview complexity trait, the odds of providing insufficient detail for the activity to be coded increase by 58% and the odds to produce a memory gap increase by 33% for every additional unit in the interview complexity factor

Conclusions (2)

- Both errors are significantly predicted by age, employment status, and graduate education, though in different directions.
  - Compared to people with less than a graduate degree, the odds of respondents with a graduate degree to provide insufficient detail increase by 24%.
  - Compared to respondents with less than a graduate degree, the odds of respondents with a graduate degree have a memory gap decrease by 29%.
  - For every additional year of age, the odds of providing insufficient detail decrease by .006%, whereas the odds of having a memory gap increase by 2%.

Discussion

- Interview complexity can be measured by observable variables, though variance explained is not very large (5.9% for insufficient detail error and 2.2% for memory gap error)

- Age, employment and education seem to play a significant role in the production of error.
  - Specifically, the older the respondent, the more memory gaps and the less insufficient detail
  - Only having a graduate degree makes any difference in whether either error is produced.

Future Research

- Identify additional indicators of interview complexity
  - From fifteen observed variables, only five were shown to have significant and practical loadings on an interview complexity factor
- Estimate a multi-level structural equation model
  - Possible more variance will be explained if the effects at the interviewer and the respondent level are disentangled
- Utilize the SIPP to cross-validate results
  - Possibility of examining the correspondence between survey responses and administrative records

Questions?
Comments?

Thank you!

Ana Lucía Córdova Cazar
SRAM – Gallup Research Center
University of Nebraska – Lincoln
al.cordovacazar@huskers.unl.edu