

Understanding the Determinants of Reliability

Ting Yan, Roger Tourangeau, Hanyu Sun, Douglas Williams, Rosalynn Yang

ITSEW 2018
June 4, 2018

Acknowledgements

- The work reported here was funded by a grant from the National Institute on Drug Abuse (5R01DA040736-02 to RT).
- We are grateful to Tammy Cook, Vanessa Meldener, and Victoria Vignare for their oversight of the data collection effort, to Martin Baer and Ed Dolbow for the work in programming the instrument, to Andrew Caporaso for help with double coding, and to Mike Jones for selecting the sample and creating the sample weights.

Background

- Reliability is central goal of questionnaire design process
- Relatively little is known about respondent characteristics and item characteristics affecting reliability of answers
 - Saris and Gallhofer (2007)
 - Most ambitious effort, but conclusions built on
 - Particular statistical technique and data source
 - Attitudinal items mostly

Question characteristics affecting reliability

- **Number of response categories** (Alwin, Baumgartner, and Beattie 2018; Revilla, Saris, and Krosnick 2014)
 - More categories, low reliability
- **Type of questions**
 - Factual questions > subjective questions (Alwin, 2007)
- **Type of response formats**
 - Factual questions with numeric open-ended responses > factual questions with vague response categories (Alwin, 2007)
- **Length**
 - Long questions, low reliability (Alwin, 2007)
- **Middle option**
 - Presence of middle categories => low reliability (Alwin et al., 2018)

Question characteristics affecting reliability (2)

- Polarity
 - Unipolar rating scales > bipolar rating scales (Alwin et al., 2018)
- Labeling
 - Fully labeled scales > partially labeled scales (Alwin 2007; Alwin and Krosnick 1991)

Respondent characteristics affecting reliability

- Age

- Rs over age 60 < Rs under 60 (Alwin, 1989; Alwin and Krosnick, 1991; Rodgers, Andrews, and Herzog, 1992)

- Education

- Higher level of education, higher reliability (Alwin, 1989, 2007; Alwin and Krosnick, 1991)

Our study

- Identifies R-level and Q-level characteristics informed by survey response model (Tourangeau, Rips, and Rasinski, 2000)
- Re-interview data
- A wide-range of question types and topics
- Research questions:
 - What respondent characteristics affect reliability?
 - What question characteristics affect reliability?

Data

PATH Reliability and Validity Study (PATH RV Study)

- Replicated the systems and procedures of the main PATH Study
 - Same instruments and software
 - Same pool of interviewers
- ABS sample of 9,782 addresses in 39 PSUs
- Mail screener to sampled addresses
 - \$5 incentive (later reduced to \$2)
 - Overall response rate: 25.1%
- Two in-person visits about 6 -24 days apart
 - ACASI interviews
 - \$35 (adults) and \$25 (youths)
 - Saliva sample

Sample Sizes

	PATH RV Study
# of PSUs	37 + 2 (delay in Houston; Miami)
Addresses sampled/screeners sent out	9,782
Households selected	862
Persons selected	862 adults 266 youths
Completed two interviews	176 adult users 266 adult non-user 129 youths

GDR's and Kappas

- Gross discrepancy rate: proportion of answers that differ in first interview and second (1 – agreement rate)
- Kappa: Chance corrected agreement rate

$$\kappa = \frac{p_a - p_e}{1 - p_e}$$

- p_e is the expected rate of agreement (assuming independence)

Exact Agreement

	Time 2 Answers		
Time 1 Answers			

Question level models

- DVs
 - GDR
 - Kappa
- Question type
 - Factual, attitudinal, demographic
- Length of questions
 - Number of sentences
 - Number of words per sentence
- Number of response options
- Type of response options
 - Not a scale, frequency scale, other types of scale
- Position of item in questionnaire
- Presence of social desirability concerns
- Centrality of question topic
- Reference period

	Adults		Youths	
	DV=GDR	DV=KAPPA	DV=GDR	DV=KAPPA
Attitudinal (vs. factual)	0.05	-0.09	0.01	-0.01
Demographic (vs. other)	-0.04	0.10	0.00	0.07
Number of sentences	0.00	-0.01	0.00	-0.02
Number of words per sentence	0.00	0.00	0.003	-0.01
Number of response options	0.06	-0.04	0.04	-0.02
Position	0.00	0.00	0.01	0.01
Response scale (vs. not a scale)	0.00	0.00	0.04	-0.07
Frequency scale (vs. other scales)	0.01	-0.05	-0.07	0.02
Social desirability concerns present	0.01	0.07	-0.08	0.06
Central topic	0.02	0.03	-0.03	-0.06
Present/Future reference period (vs. past)	-0.01	0.08	-0.09	0.09
N	426	419	224	213
R-square	0.70	0.31	0.67	0.51

	Adults		Youths	
	DV=GDR	DV=KAPPA	DV=GDR	DV=KAPPA
Attitudinal (vs. factual)	0.05	-0.09	0.01	-0.01
Demographic (vs. other)	-0.04	0.10	0.00	0.07
Number of sentences	0.00	-0.01	0.00	-0.02
Number of words per sentence	0.00	0.00	0.003	-0.01
Number of response options	0.06	-0.04	0.04	-0.02
Position	0.00	0.00	0.01	0.01
Response scale (vs. not a scale)	0.00	0.00	0.04	-0.07
Frequency scale (vs. other scales)	0.01	-0.05	-0.07	0.02
Social desirability concerns present	0.01	0.07	-0.08	<i>0.06</i>
Central topic	0.02	0.03	-0.03	<i>-0.06</i>
Present/Future reference period (vs. past)	-0.01	0.08	-0.09	0.09
N	426	419	224	213
R-square	0.70	0.31	0.67	0.51

	Adults		Youths	
	DV=GDR	DV=KAPPA	DV=GDR	DV=KAPPA
Attitudinal (vs. factual)	0.05	-0.09	0.01	-0.01
Demographic (vs. other)	-0.04	0.10	0.00	0.07
Number of sentences	0.00	-0.01	0.00	-0.02
Number of words per sentence	0.00	0.00	0.003	-0.01
Number of response options	0.06	-0.04	0.04	-0.02
Position	0.00	0.00	0.01	0.01
Response scale (vs. not a scale)	0.00	0.00	0.04	-0.07
Frequency scale (vs. other scales)	0.01	-0.05	-0.07	0.02
Social desirability concerns present	0.01	0.07	-0.08	0.06
Central topic	0.02	0.03	-0.03	-0.06
Present/Future reference period (vs. past)	-0.01	0.08	-0.09	0.09
N	426	419	224	213
R-square	0.70	0.31	0.67	0.51

Relaxed agreement

- Relaxed agreement=answers within one category of earlier answer

	Time 2 Answers		
Time 1 Answers	Dark Blue	Green	White
	Green	Dark Blue	Green
	White	Green	Dark Blue

- Approximate GDR's and Generalized Kappas

	Adults		Youths	
	DV=AGDR	DV=GKAPPA	DV=AGDR	DV=GKAPPA
Attitudinal (vs. factual)	0.02	-0.05	-0.01	0.02
Demographic (vs. other)	-0.03	0.08	-0.01	0.09
Number of sentences	0.00	-0.01	0.004	-0.04
Number of words per sentence	0.001	-0.01	0.00	0.00
Number of response options	0.03	-0.03	0.02	-0.03
Position	-0.01	0.00	0.00	0.02
Response scale (vs. not a scale)	-0.03	0.06	-0.01	-0.01
Frequency scale (vs. other scales)	0.02	-0.07	-0.04	0.02
Social desirability concerns present	-0.02	0.09	-0.03	0.07
Central topic	0.00	0.04	0.02	-0.11
Present/Future reference period (vs. past)	0.00	0.08	-0.06	0.08
N	426	419	224	213
R-square	0.18	0.21	0.27	0.21

	Adults		Youths	
	DV=AGDR	DV=GKAPPA	DV=AGDR	DV=RKAPPA
Attitudinal (vs. factual)	0.02	-0.05	-0.01	0.02
Demographic (vs. other)	-0.03	0.08	-0.01	0.09
Number of sentences	0.00	-0.01	0.004	-0.04
Number of words per sentence	0.001	-0.01	0.00	0.00
Number of response options	0.03	-0.03	0.02	-0.03
Position	-0.01	0.00	0.00	0.02
Response scale (vs. not a scale)	-0.03	0.06	-0.01	-0.01
Frequency scale (vs. other scales)	0.02	-0.07	-0.04	0.02
Social desirability concerns present	-0.02	0.09	-0.03	0.07
Central topic	0.00	0.04	0.02	-0.11
Present/Future reference period (vs. past)	0.00	0.08	-0.06	0.08
N	426	419	224	213
R-square	0.18	0.21	0.27	0.21

	Adults		Youths	
	DV=AGDR	DV=RKAPPA	DV=AGDR	DV=RKAPPA
Attitudinal (vs. factual)	0.02	-0.05	-0.01	0.02
Demographic (vs. other)	-0.03	0.08	-0.01	0.09
Number of sentences	0.00	-0.01	0.004	-0.04
Number of words per sentence	0.001	-0.01	0.00	0.00
Number of response options	0.03	-0.03	0.02	-0.03
Position	-0.01	0.00	0.00	0.02
Response scale (vs. not a scale)	-0.03	0.06	-0.01	-0.01
Frequency scale (vs. other scales)	0.02	-0.07	-0.04	0.02
Social desirability concerns present	-0.02	0.09	-0.03	0.07
Central topic	0.00	0.04	0.02	-0.11
Present/Future reference period (vs. past)	0.00	0.08	-0.06	0.08
N	426	419	224	213
R-square	0.18	0.21	0.27	0.21

Summary of item-level models

	Effect	Significance
Attitudinal (vs. factual)	-	3/8
Demographic (vs. other)	+	6/8
Number of sentences	no effect	
Number of words per sentence	-	3/8
Number of response options	-	6/8
Position	+	1/8
Response scale (vs. not a scale)	+/-	4/8
Frequency scale (vs. other scales)	+	2/8
Social desirability concerns present	+	5/8
Central topic	-	1/8
Present/Future reference period (vs. past)	+	5/8

Cross-classified models: Adults

- DV
 - Whether or not same answer is given at both times
- Age
 - Over 60 (vs. under)
- Education
 - High school or less (vs. more)
- Need for cognition
- Conscientiousness

Cross-classified models: Adults

- Random effects

Unconditional Model	Estimates	ICC
Respondent	0.19	3%
Item	2.75	44%
Conditional Model	Estimates	% Reduction
Respondent	0.17	13%
Item	0.85	69%

■ Fixed Effects

	Estimates in log
Over age of 60 (vs 60 or under)	0.01
High school or less (vs. more than HS)	-0.29
Conscientiousness	0.09
Need for cognition	-0.01
Attitudinal (vs. factual)	-0.52
Demographic (vs. other)	0.95
Number of sentences	-0.03
Number of words per sentence	-0.02
Number of response options	-0.51
Position	0.04
Response scale (vs. not a scale)	-0.02
Frequency scale (vs. other scales)	-0.26
Social desirability concerns present	-0.16
Central topic	-0.37
Present/Future reference period (vs. past)	0.01

Cross-classified models: Youth

- DV
 - Whether or not same answer is given at both times
- Age
 - 12-19
- Education
 - 9th grade or more (vs. less)
- Conscientiousness

Cross-classified models: Youth

- Random effects

Unconditional Model	Estimates	ICC
Respondent	0.27	4%
Item	2.60	42%
Conditional Model	Estimates	% Reduction
Respondent	0.27	0%
Item	0.70	73%

- Fixed Effects

	Estimates in log scale
Age	-0.06
9 th grade or more (vs. less)	-0.14
Conscientiousness	0.03
Attitudinal (vs. factual)	-0.12
Demographic (vs. other)	0.47
Number of sentences	-0.04
Number of words per sentence	-0.04
Number of response options	-0.23
Position	0.06
Response scale (vs. not a scale)	-0.41
Frequency scale (vs. other scales)	0.70
Social desirability concerns present	0.71
Central topic	0.06
Present/Future reference period (vs. past)	1.03

Conclusions and Next Steps

Conclusions

- Large variation by survey items
- Adults R-level characteristics affecting reliability
 - Education and conscientiousness
- Q-level characteristics affecting reliability
 - Type of questions
 - Length of response options
 - Type of response options
 - Presence of social desirability concerns
 - Reference period

Next steps

- Relationship between reliability and
 - Item nonresponse
 - Time taken to answer the item
- Convergence of different measures of reliability

Thanks!!!

tingyan@westat.com

HAP Health Assessment
of the
Population