Mode Effect on Data Quality: Testing Measurement Invariance of Sensitive Measures among Minorities

Galit Gordoni
The Open University
Israel

Peter Schmidt
University of Giessen
Germany

Yiftach Gordoni
Haifa University
Israel

European Survey Research Association Conference _Warsaw 2009

Outline

- Introduction
- Mode effect on data quality – traditional approach
- Mode effect on data quality – SEM approach
- Research hypotheses
- Analytical procedure
- Research design
- Results
- Conclusions
- Implications and recommendations for mode selection and mix-mode

Introduction

Mode selection: data quality implications

- Accuracy of data produced in surveys depends on the amount of measurement error (random and systematic).

- Measurement error is especially crucial when measuring latent constructs with manifest indicators, which are more vulnerable to response errors.

(Groves et. al., 2004)

Understanding the characteristics underlying mode differences in data quality is essential for making informed decisions of which mode or modes to use in a specific survey.

(de Leeuw 2005; Groves et. al., 2004; Voogt & Saris 2005)

The **present study aims** to:

Compare data quality between face-to-face and telephone modes among a special population

by

Testing mode invariance in psychometric properties

of

Measurement models relating sensitive constructs with their indicators

within

Structural equation modeling approach.

Mode effect on data quality – traditional approach

- Data quality Indexes
- Variables mediating mode effect on data quality
- Mode effect on the mediating variables
  - In general
  - Among minority status groups
  - Among collectivistic-oriented minority groups
- Empirical trends

Traditional data quality indexes

Traditional mode comparison consecrate on observed satisficing response strategies as indicators of random and systematic measurement error.

The main satisficing strategies are:

- **Non-differentiation response style**
- **Acquiescence response style**
- **Extremeness response style**

In sensitive subjects mode comparison concentrate mainly on:

- **Socially desirable response style**

Psychological mediating variables of mode effect on data quality

In general, data quality is affected by:

- **Cognitive burden -** The cognitive requirements needed for an optimal response process generating the true value.
- **Respondent's Motivation -** The degree of effort and persistence involved in the execution of the response process.

These mediating variables might cause random and systematic error.

(Krosnick 1991; Tourangeau et al. 2000)

In sensitive scales, data quality is also affected by:

- **Threat of disclosure** – the concerns about disclosing the true value to an interviewer

that,

might disapprove the respondent socially undesirable response,

or

won't keep data confidentiality, so respondent might be harmed by a third party.

This factor might cause systematic error due to social desirability response style.

(Tourangeau & Yan, 2007)

Mode effect on the psychological mediating variables

Modes differ in two main characteristics:

1. **Media-related characteristics:** Social conventions and customs associated with the media utilized in the survey mode.

2. **Information transmission characteristics:** Availability of communication channels

(verbal – words; non verbal – gestures, expressions; paralinguistic – tone of voice).

(de Leeuw, 1992; 2005)

1. **Media-related characteristics** effect the psychological mediators through two main processes:

- Rules of behavior in telephone conversation such as
  - Avoidance of lapses
  - Regulation of communication by the interviewer

give

less time to perform optimal cognitive process.

Therefore,

higher cognitive burden in telephone mode is expected.

(Sloan et. al., 2006)

In addition,

- Interviewer higher social presence in f-t-f mode

increases

fear of embarrassment and social disapproval.

Therefore,

higher threat of disclosure in the f-t-f mode

is expected.

(Tourangeau et. al., 2000)

2. Information transmission characteristics effect the psychological mediators through two main processes:

- The limited communication channels in telephone mode

require

explicit verbal communication for performing an optimal cognitive process.

Therefore,

higher cognitive burden in the telephone mode is expected.

(Groves et. al., 2004)

In addition,

- Interviewer in telephone is less able to
  - Develop rapport with the respondent
  - Maintain respondent's interest
  - Increase sense of legitimacy, trust, and perceived confidentiality in survey institution.

Therefore,

lower respondent's motivation and higher threat of disclosure is expected in the telephone mode.

(de Leeuw, 2005; Groves et. al., 2004)

Mode effect among minority status groups

Due to their inferior position in society, minority status group members share strong:

- Beliefs about group exploitation
- Suspicious of the wider society and its institutions
- Need for social approval from the majority group.

(Johnson et. al., 1997; Johnson & Van de Vijver, 2003)

Therefore,

scales regarding majority-minority relations might be perceived as very sensitive,

and

mode effect on threat of disclosure might increase.

Mode effect among collectivistic-oriented cultures

In collectivistic oriented cultures norms, obligations and duties tend to guide social behavior more then personal interests.

(Triandis et. al., 1984)

Respondents from collectivistic culture share strong:

- Reliance on non-verbal cues for communication.
- Resistance to self-disclosure in general and to strangers, in particular.

Therefore,

mode effect on cognitive burden, motivation and threat of disclosure might increase.

(Johnson et. al., 1997, 2002; Johnson & Van de Vijver, 2003)

Empirical trends

1. In sensitive questions higher social desirability bias is evident in the telephone mode,
   (de Leeuw & van der Zouwen, 1988; Holbrook et. al., 2003; Tourangeau & Yan, 2007)
   especially among minority group members.
   (Aquilino & LoSciuto 1990; Aquilino 1994)

2. Lower non-differentiation, acquiescence and extremeness response styles are evident in the f-t-f mode (de Leeuw 2005;de Leeuw & van der Zouwen, 1988; Sloan et. al., 2006), especially in complex items (Sloan et. al., 2006).

3. Consistent, but insignificant, trend of lower reliabilities and mean item-total correlations is evident in the telephone mode (de Leeuw, 1992).

Mode effect on data quality – SEM approach

Structural equation modeling (SEM) approach – application for mode comparison in data quality

SEM approach use an explicit definition of a measurement model.

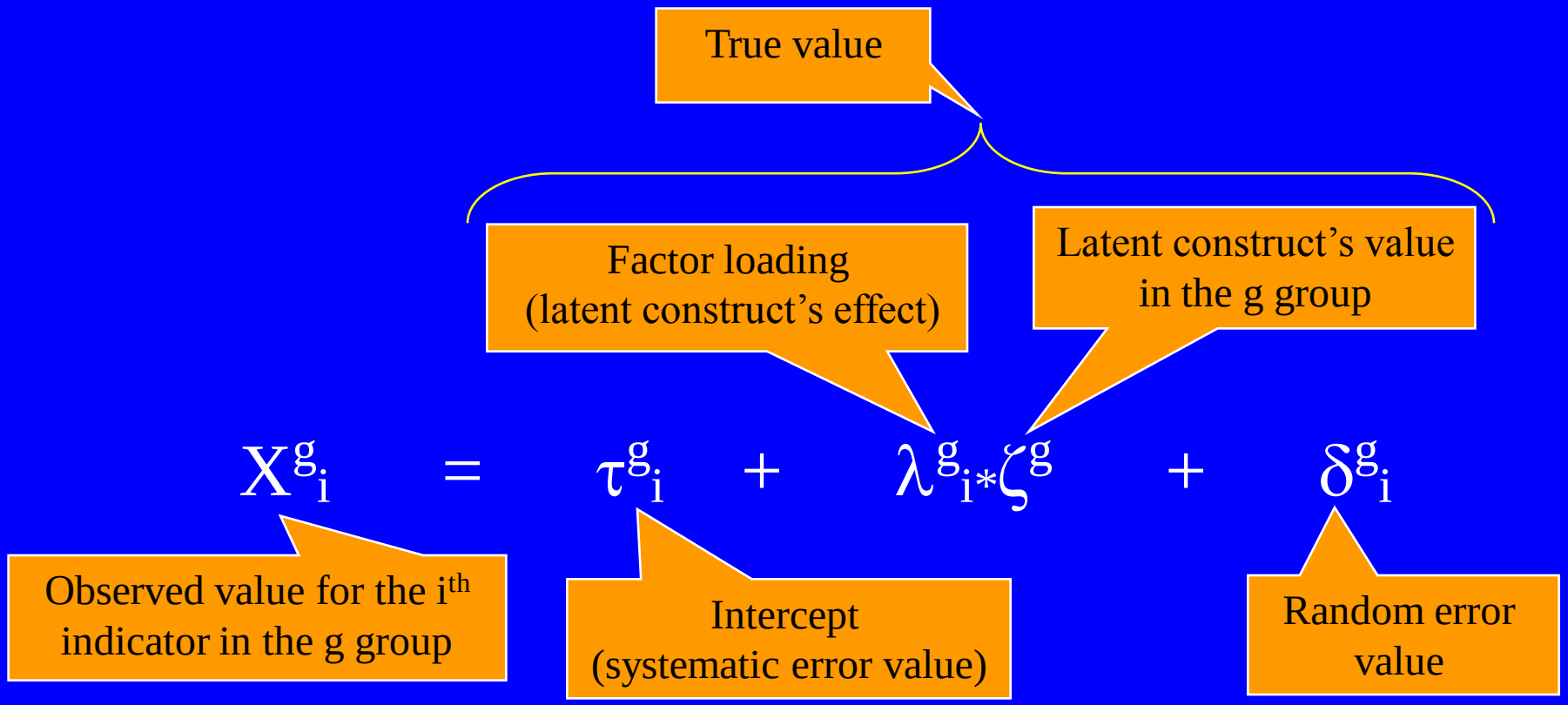
Therefore, SEM has several advantages for mode comparison in data quality:

- Psychometric properties comparison for each item.
- Separable diagnosis of measurement error components (random, systematic).
- Positive indication of data quality comparison (latent construct's effect on the indicators – validity coefficient).

(de Leeuw, 1992; de Leeuw, Mellenbergh & Hox, 1996)

Measurement model – basic equation

A measurement model is a set of equations relating manifest indicators to their latent construct.
The basic equation structure is:



measurement model – visual description

FTF equations (group A)

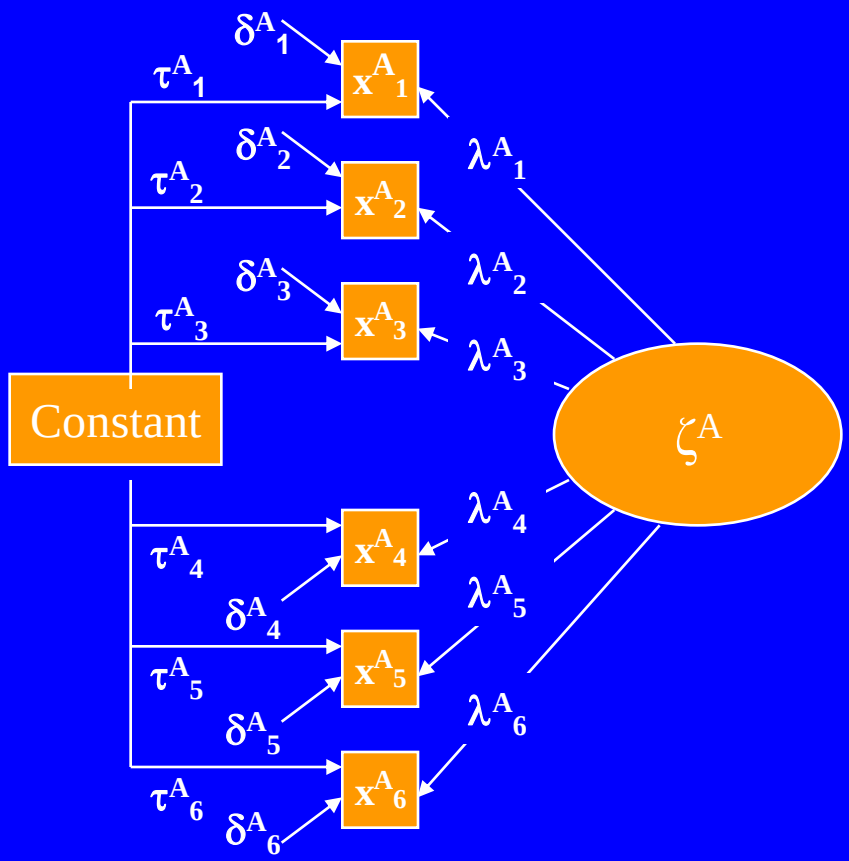
$$x^A_1 = \tau^A_1 + \lambda^A_1\zeta^A + \delta^A_1$$

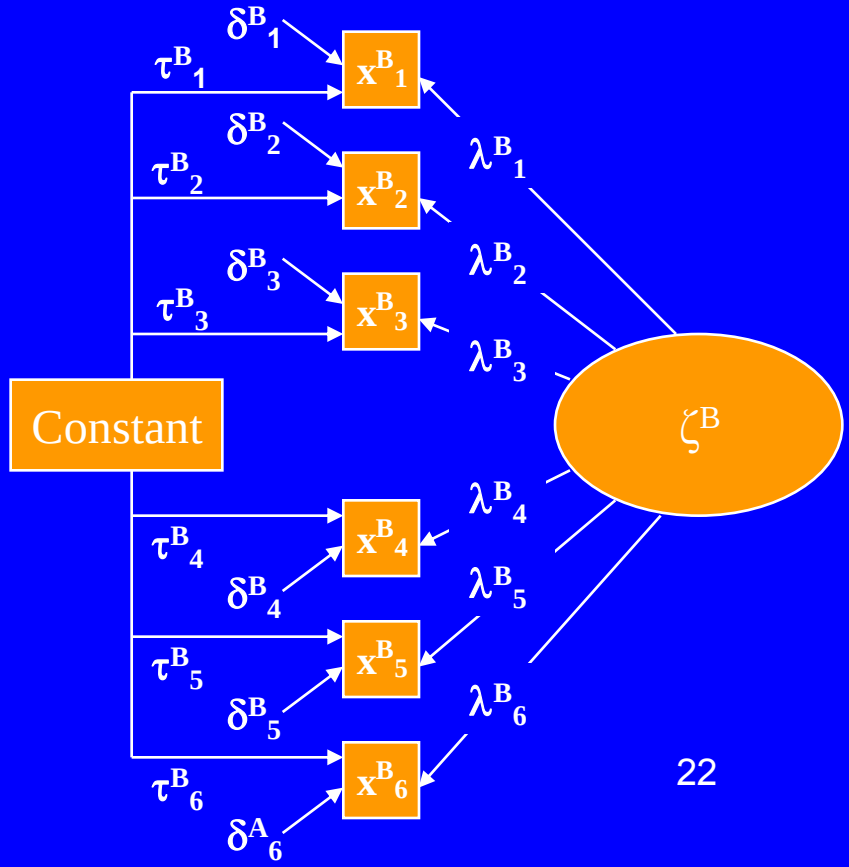
$$X^A_2 = \tau^A_2 + \lambda^A_2\zeta^A + \delta^A_2$$
$$x^A_3 = \tau^A_3 + \lambda^A_3\zeta^A + \delta^A_3$$
$$x^A_4 = \tau^A_4 + \lambda^A_4\zeta^A + \delta^A_4$$
$$x^A_5 = \tau^A_5 + \lambda^A_5\zeta^A + \delta^A_5$$
$$x^A_6 = \tau^A_6 + \lambda^A_6\zeta^A + \delta^A_6$$

Tel equations (group B)

$$x^B_1 = \tau^B_1 + \lambda^B_1\zeta^B + \delta^B_1$$

$$x^B_2 = \tau^B_2 + \lambda^B_2\zeta^B + \delta^B_2$$
$$x^B_3 = \tau^B_3 + \lambda^B_3\zeta^B + \delta^B_3$$
$$x^B_4 = \tau^B_4 + \lambda^B_4\zeta^B + \delta^B_4$$
$$x^B_5 = \tau^B_5 + \lambda^B_5\zeta^B + \delta^B_5$$
$$x^B_6 = \tau^B_6 + \lambda^B_6\zeta^B + \delta^B_6$$





Research hypotheses

Based on the research literature it is possible to formulate the following hypotheses regarding measurement model's parameters:

1. Lower factor loadings in the telephone mode.
2. Higher absolute intercepts in the telephone mode.
3. Higher random errors in the telephone mode.

(No differences, however, are assumed in latent constructs' means and variances).

- **Analytical procedure**

Procedure for elimination of confounding results due to total survey error:

- Excluding non-telephone household respondents, in order to eliminate confounding results due to under-coverage.
- Adding demographic variables as covariates, in order to eliminate confounding results due to non-response (age, gender, education, religion, income).

Procedure for analyzing mode effect using SEM measurement invariance tests:

- Latent factor scaling was achieved by using two constraints – average loading equal 1, and intercepts sum to zero.

  These constraints enable a meaningful interpretation of the parameters' estimates (Little et al., 2006).

- Partial invariance was used when full invariance could not been achieved (Byrne et al., 1989; Baumgartner & Steenkamp, 1998).

- **invariance tests include the following five steps:**

Step 0: A baseline model test

A separate pre-test of the hypothesized factor structure in each mode, which

- assures acceptable baseline fit of the hypothesized cognitive frame of reference, used to make items responses in each mode.

Step 1: A configural invariance test

A test of same factor structure across modes, which

- assures similar meaning of latent construct in each mode.

Step 2: A metric invariance test (weak invariance)

A test of same factor loadings across modes - $\lambda^{A} = \lambda^{B}$, which

- assures **same meaning** of latent construct in each mode. In other words, construct is manifested and calibrated by items in the same way.

Weak invariance is a prerequisite for meaningful modes comparisons.

Step 3: A scalar invariance test (strong invariance)

Test of same intercepts across modes - $\tau^A = \tau^B$.

This means a same systematic bias in items in each mode (same calibration of construct's zero value).

- Strong invariance assures same scaling for the latent construct.

Step 4: An error invariance test (strict invariance)

A test of same error variances across modes – $\delta^{A} = \delta^{B}$.

- This means that the amount of measurement error in items is the same in each mode.

- Error invariance also means invariance in the items and scale reliabilities across modes (assuming equal latent factor variance).

Research design

- Population
- Sample Selection
- Questionnaires
- Data collection

Population

- **Definition**

Israeli Arab citizens, 18 years old and older.

- **Characteristics**
  - Constitute 16.5% of the Israeli population.
  - Near Eastern, semi- traditional society.
  - Working class community, within a middle-class Jewish society.
  - A non-assimilating and dissident minority which loyalty is suspect by the Jewish state and population.
  - In sharp disagreements with the Jewish majority on ideological issues.
  - A discriminated Minority. (Smooha, 2005)

Sample Selection: frames and method

In the face to face sample:

- A multistage stratified sample was chosen from the Localities database of the 1995 census.
- A simple random sample from each of the localities, sampled in the first stage, was chosen from the official list of eligible voters of 2003.

In the telephone sample:

A multistage stratified sample of telephone numbers was chosen from the computerized white pages. Respondents were selected by the last birthday method.

Questionnaires

- **In the present study** we used three measures – social integration, threat, and prejudice - which are part of an Arab-Jewish Relations Index.

  The index was developed by Smooha (2005) at the Jewish-Arab Center in Haifa university (**http://soc.haifa.ac.il/~s.smooha**).

- Since 2003 the index is measured in ongoing surveys among the Israeli Arabs.

- The three measures were validated in two independent studies (2002, 2003) using confirmatory factor analysis (Gordoni, 2006).

Social Integration scale

To what extent do you agree or not agree to:	Do not agree			Agree
SI1. Friendship relationships between Jews and Arabs	1	2	3	4
SI2. Residence of Arabs in Jewish Neighborhoods	1	2	3	4
SI3. Arab attendance in Jewish high-Schools	1	2	3	4
SI4. Arab-Jewish youth encounters	1	2	3	4
SI5. Arabs spending time in parks and swimming pools in Jewish localities	1	2	3	4
SI6. Arab-Jewish joint organizations	1	2	3	4

Threat scale

To what extent do you fear or not fear that in the next few years there will be:	Do not fear			Fear
TH1. Severe violation of Arab civil rights	1	2	3	4
TH2. Massive land expropriation	1	2	3	4
TH3. State violence against Arab citizens	1	2	3	4
TH4. Jewish violence against Arab citizens	1	2	3	4
TH5. Annexation of the Arab Triangle to the Palestinian state despite its Arab residents' opposition	1	2	3	4
TH6. Population transfer of some Arab citizens outside Israel	1	2	3	4

Prejudice scale

To what extent do you agree or not agree that:	**Do not agree**			**Agree**
PR1. Most Jews in Israel cannot be trusted	1	2	3	4
PR2. Most Jews in Israel tend to be violent	1	2	3	4
PR3. Most Jews in Israel are willing to give up their self-dignity in order to achieve comfort, money and advancement	1	2	3	4
PR4. Most Jews in Israel are racist	1	2	3	4

Data collection

Interviews were conducted in Arabic by Arab interviewers during September-November 2003.

- **F-T-F mode:**

Most interviewers were from the sampled locality. Response rate was **83.7%** (701 completed interviews out of 837 eligible cases).

- **Telephone mode:**

Response rate was **27.7%** (511 completed interviews out of 1841 eligible cases).

Results

- Descriptive statistics for observed scores
- Measurement invariance tests
  - Fit indices
  - Results summary
  - Conclusions
  - Parameters estimates

Descriptive statistics for observed scores

- Some differences in observed means and standard deviations are evident in all three scales.

These results raise the following questions:

1. Shell we attribute the observed **mean differences** to mode differences in systematic errors, in factor loadings, or in both?
2. Shell we attribute the observed **standard deviation differences** to mode differences in random errors, in factor loadings, or in both?
3. Does observed **indifferences** indicate measurement invariance across modes?

Social Integration scale – descriptive statistics

Item	M		SD	
	FTF	Tel	FTF	Tel
SI1. Friendship relationships	3.75	3.83	.60	.64
SI2. Residence Jewish neighborhoods	**3.02**	**3.04**	1.09	1.26
SI3. Attendance Jewish high-schools	3.11	3.19	**1.09**	**1.26**
SI4. Youth encounters	3.62	3.74	.77	.80
SI5. Attendance Jewish recreations	3.39	3.31	.95	1.14
SI6. Joint organizations	**3.63**	**3.77**	.73	.72
Composite manifest score	3.42	3.48	.63	.60
Reliability (Cronbach's α)	.81	.69		
Mean inter-item correlations	.46	.32		

Note. All correlations are significant at $p<.001$; Face to face mode (FTF): n=474, Telephone mode (Tel): n=483; All statistics were corrected for covariates effects.

Threat scale – descriptive statistics

Item	M		SD	
	FTF	Tel	FTF	Tel
TH1. Violation of civil rights	3.34	3.29	.97	1.11
TH2. Land expropriation	3.30	3.50	1.05	.99
TH3. State violence	3.15	3.11	1.10	1.18
TH4. Jewish violence	3.09	3.05	1.11	1.22
TH5. Annexation of the Triangle	2.60	2.57	1.27	1.37
TH6. Population transfer	2.74	2.84	1.26	1.33
Composite manifest score	3.03	3.06	.90	.84
Reliability (Cronbach's α)	.89	.79		
Mean inter-item correlations	.58	.40		

Note. All correlations are significant at p<.001;
Face to face mode (FTF): n=472, Telephone mode (Tel): n=466;
All statistics were corrected for covariates effects.

Prejudice scale – descriptive statistics

Item	M		SD	
	FTF	Tel	FTF	Tel
PR1. Cannot be trusted	2.66	2.86	1.14	1.31
PR2. Tend to be violent	2.28	2.42	1.12	1.35
PR3. Give up their self-dignity to achieve comfort, money, advancement	2.60	2.58	1.12	1.33
PR4. Racist	2.59	2.65	1.17	1.34
Composite manifest score	2.53	2.63	.89	.92
Reliability (Cronbach's α)	.80	.65		
Mean inter-item correlations	.49	.31		

Note. All correlations are significant at $p<.001$;
Face to face mode (FTF): n=453, Telephone mode (Tel): n=438;
All statistics were corrected for covariates effects.

Measurement invariance tests results

Social Integration scale – fit indices

	Model	Compare	χ^2 (df)	$\Delta\chi^2$ (Δdf)	RMSEA	CFI	AIC
	Baseline						
	F-T-F		152.06 (54)*		.062	.934	44.06
	Tel		108.59 (54)*		.046	.952	.59
	Multi-group						
A	Configural		262.86 (108)*		.039	.941	46.86
B	Full metric	A	264.97 (113)*	2.11 (5)	.038	.942	38.97
C1	Full scalar	B	291.10 (118)*	26.13 (5)*	.039	.942	55.10
D	Full error	C2	363.86 (120)*	98.67 (6)*	.046	.905	123.86

*** p<.001**

Note. Fit indices and significant tests are based on Satorra-Bentler robust correction for non-normality (Satorra & Bentler, 1988). FTF: n=474, Tel: n=483

Social Integration scale – invariance tests results summary

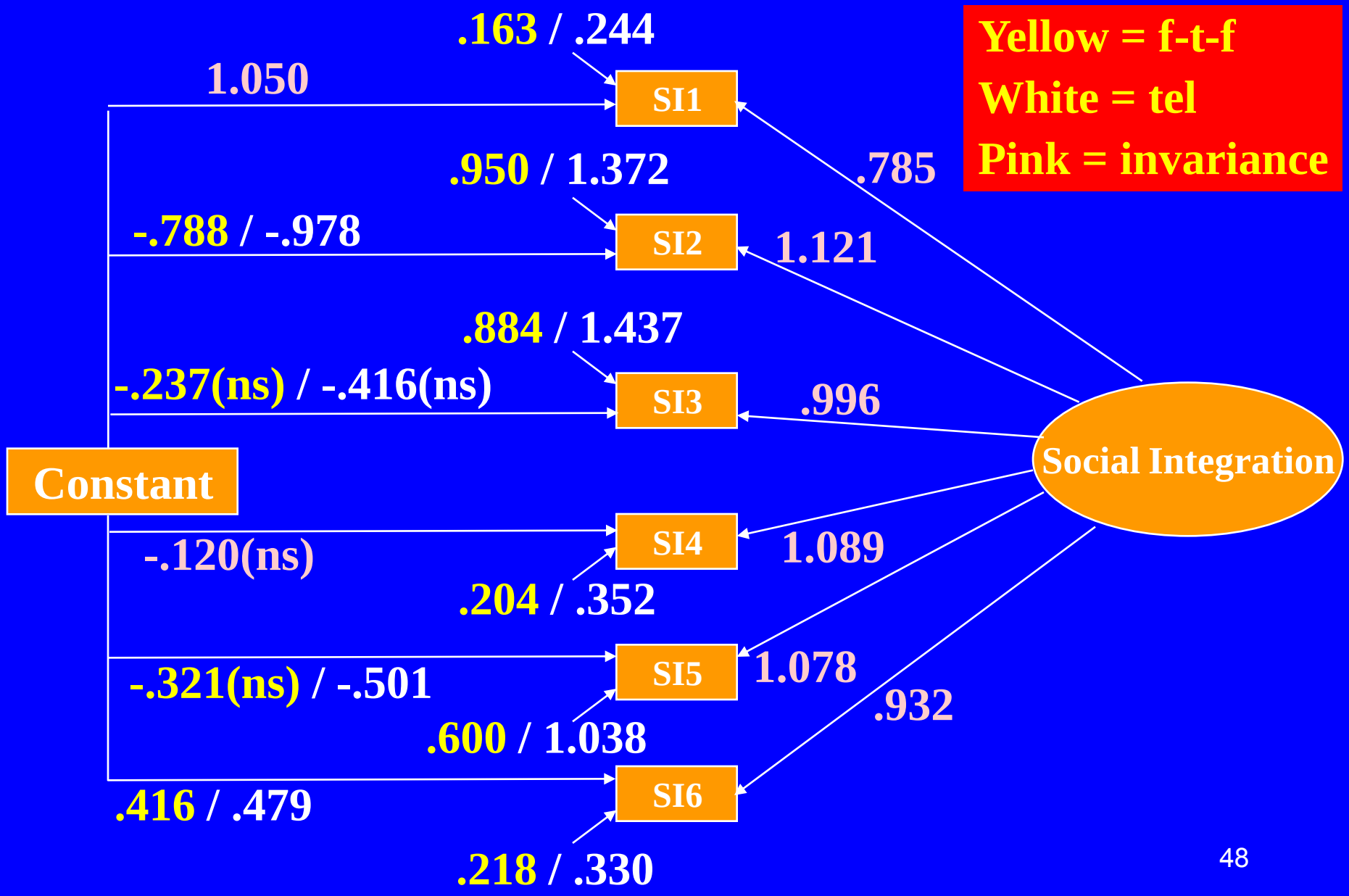
The results point to:

- Adequate fit of hypothesized factor structure in each mode.
- Full weak invariance – same structure and loadings.
- Scalar non-invariance - higher absolute intercept in telephone mode for 4 items.
- Error non-invariance - higher error variances in telephone mode for all items, and therefore lower reliabilities.

Social integration scale - Conclusions

1. The scale apparently measure the same construct in both modes.
2. The higher systematic error in telephone supports the claim that the indirect mode effect on item responses through confidentiality concerns is higher than its indirect effect through concerns of interviewer presence.
3. The higher random error evident in telephone is consistent with the higher cognitive burden and lower motivation assumed for this mode.

Social Integration scale – parameters estimates



Threat scale – fit indices

	Model	Compare	χ^2 (df)	$\Delta\chi^2$ (Δdf)	RMSEA	CFI	AIC
	Baseline						
	F-T-F		108.01 (53)***		.047	.977	2.01
	Tel		84.71 (53)**		.036	.976	-21.29
	Multi-group						
A	Configural		193.07 (106)***		.030	.976	-18.93
B1	Full metric	A	205.39 (111)***	12.32 (5)*	.030	.974	-16.61
B2	Partial metric	A	197.75 (110)***	4.68 (4)	.029	.976	-22.25
C1	Full scalar	B2	218.52 (115)***	20.77 (5)***	.031	.973	-11.48
D	Full error	C2	333.46 (119)*	135.11 (6)***	.044	.941	95.46

***p<.05 **p<.01 ***p<.001.** Fit indices and significant tests are based on Satorra-Bentler robust correction for non-normality (Satorra & Bentler, 1988).. One modification was added to the baseline models – correlated errors between items 5 and 6. FTF: n=472, TEL: n=466.

Threat scale – invariance tests results summary

The results point to:

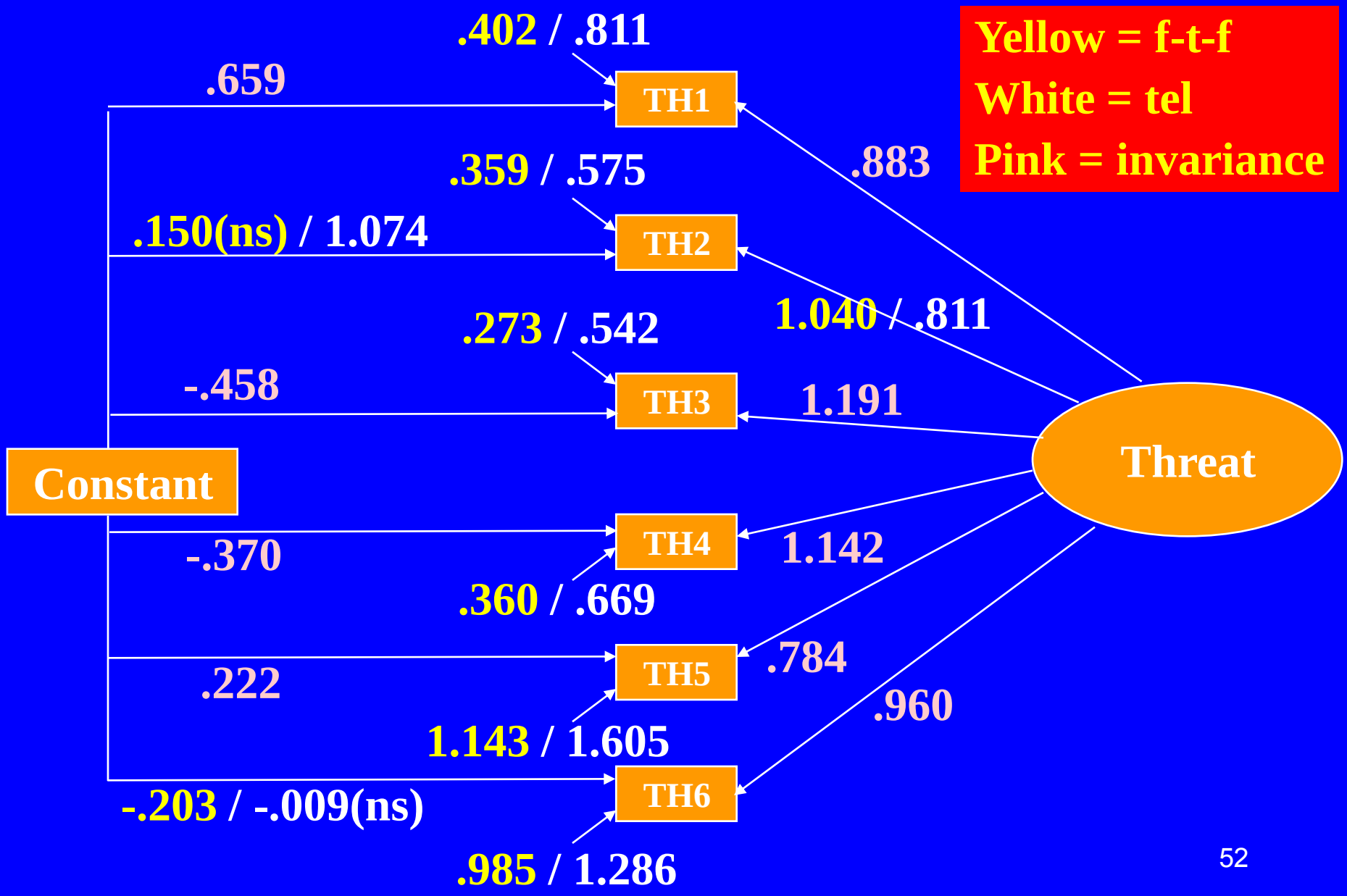
- Adequate fit of hypothesized factor structure in each mode.
- Metric non-invariance – lower factor loading in telephone mode for one item (partial weak invariance was achieved).
- Scalar non-invariance - intercepts of two items vary between modes, in opposite directions.
- Error non-invariance - higher error variances in telephone mode for all items, and therefore lower reliabilities.

Threat scale – conclusions

1. The major part of the scale (5 out of 6 items) measure the same construct in both modes.

   The lower loading in telephone is consistent with the hypothesized moderate effect of mode on the latent construct manifestation through the item.
2. The opposite directions found in systematic bias do not support the claim that telephone mode yield more socially desirable responses.
3. The higher random error evident in telephone is consistent with the higher cognitive burden and lower motivation assumed for this mode.

Threat scale – parameters estimates



Prejudice scale – fit indices

	Model	Compare	χ^2 (df)	$\Delta\chi^2$ (Δdf)	RMSEA	CFI	AIC
	Baseline						
	F-T-F		47.30 (29)*		.037	.985	-10.70
	Tel		37.81 (29)		.026	.990	-20.19
	Multi-group						
A	Configural		85.32 (58)*		.023	.987	-30.68
B1	Full metric	A	114.97 (61)***	29.65 (3)***	.032	.974	-7.03
B2	Partial metric	A	85.70 (59)*	.38 (1)	.023	.987	-32.30
C	Full scalar	B2	123.43 (62)***	37.73 (3)***	.033	.973	-.57
D1	Full error	B2	190.05 (63)***	104.35 (4)***	.048	.940	64.05

***p<.05 **p<.01 ***p<.001.** Fit indices and significant tests are based on Satorra-Bentler robust correction for non-normality (Satorra & Bentler, 1988). FTF: n=453, TEL: n=438.

Prejudice scale – invariance tests results summary

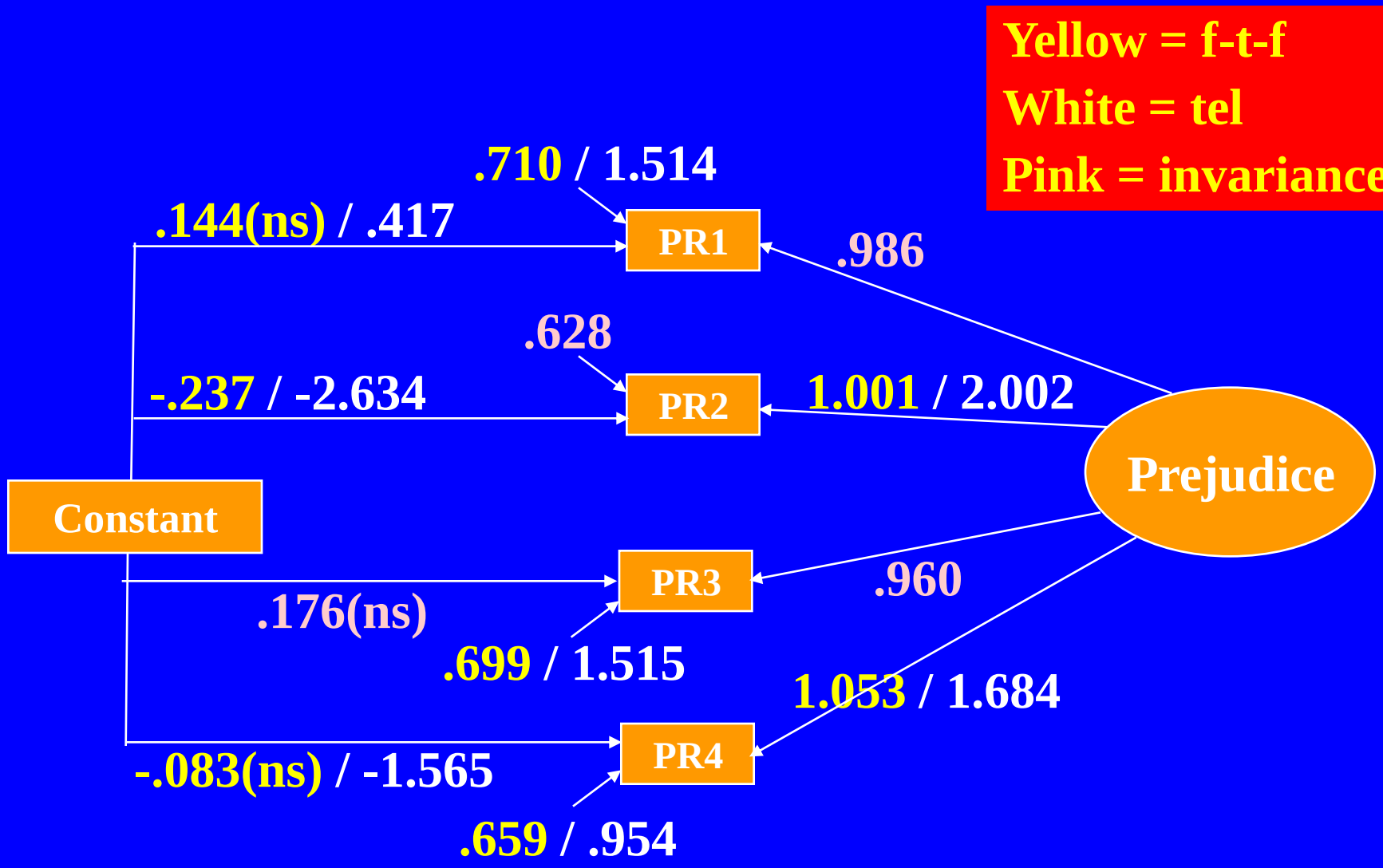
The results point to:

- Adequate fit of hypothesized factor structure in each mode.
- Metric non-invariance – higher factor loadings in telephone mode for two items (partial weak invariance was achieved).
- Scalar non-invariance - higher absolute intercepts in telephone mode for three items.
- Error non-invariance - higher error variances in telephone mode for three items.

Prejudice scale - conclusions

1. The scale apparently measure the same construct in both modes.
2. The higher systematic error in telephone supports the claim that the indirect mode effect on item responses through confidentiality concerns is higher than its indirect effect through concerns of interviewer presence.
3. The higher random error evident in telephone is consistent with the higher cognitive burden and lower motivation assumed for this mode.
4. The higher loadings in telephone is opposite to the hypothesized direction.

Prejudice scale – parameters estimates



- **Implications and recommendations for mode selection and mix-mode**

Implications for mode selection and mix mode

- F-T-F mode seems fully superior in regard to **random error** (reliabilities) and partially superior in regard to **systematic error**.

Therefore, using composite manifest score based on telephone or mix mode data will cause:

1. Attenuated correlations (due to higher random error).
2. Higher standard error of the mean and lower estimation precision (due to higher random error).
3. Attenuated mean (due to higher systematic error).

recommendations for surveys using telephone or mix modes for data collection

- Random error problems can be reduced by
  - Increasing the sample size, which decreases the standard error of the mean estimator.
  - Investing effort in techniques that enable to reduce cognitive burden and increase respondent motivation.
  - Excluding the problematic non-invariant items.
  - Using SEM, which corrects for random error.

- Systematic error problems can be reduced by
  - Investing effort in techniques that enable to increase trust and sense of confidentiality.
  - Excluding the problematic non-invariant items.