Submission Title
(What would be the problems of interpreting and comparing scores of psychological measures
in applied settings across different countries)

What would cause inflation of self-ratings on personality aptitude scales?

─ The effects of impression management and self-enhancing tendency ─

○Shiho Imashiro, Naoya Todo, Cheng Wang, Shino Sako, Yutaro Sakamoto

Focus of this research

- To apply Thurstonian IRT model (Brown & Maydeu-Olivares, 2011) to personality test responses of office workers in Japan
- To study the difference between forced-choice scales and Likert scales under a natural test-taking situation
  (no faking instruction)
- To examine whether a forced-choice format can reduce faking under a natural test-taking situation

Personality test use in organization

- Previous research shows that personality tests have an **acceptable levels of validity**. (Barrick & Mount, 2005; Hough & Ones, 2001; Hurtz & Donovan, 2000; Ones, Viswesvaran, & Dilchert, 2005; Salgado, 1997)
- Still, in order to use personality tests in organizational settings, **"faking" is probably the biggest obstacle**. (Hogen, Hogen, Roberts, 1996)

Inconsistent research findings on faking

- Faking is not a big threat to the validity. (Barrick & Mount, 1996; Hogan, 1991; Ones, Viswesvaran, & Reiss, 1996) ○
- Test scores do differ between applicants and non-applicants. (Birkeland, Kisamore, Brannick, & Smith, 2006) ✕
- Instructed faking ≒ naturally occurring faking
- Need for theoretical research (Griffth & Peterson, 2011)

Forced-choice format

- A suggested remedy to faking (Bass, 1957)
- Are forced-choice scales more resistant to faking than Likert scales?
  Yes (Jackson, Wrobleski, & Ashton, 2000) or
  No (Heggestad, Morrison, Reeve, & McCloy, 2006)
- Are forced-choice scales more valid in predicting performance than Likert scales?
  Yes (Cristiansen, Burns, & Montgomery, 2005) or
  No (Converse, Oswald, Imus, Hendroicks, Roy, & Butera, 2008)
- Using Thurstonian IRT model (Brown & Maydeu-Olivares, 2011), normative information can be properly extracted from forced-choice item responses.
- How would the introduction of IRT model change the research results shown above?

Data collection & Analysis

- **Sample**; 644 Japanese office workers
  all male, age 25-55, various occupations, working for companies with more than 500 employees from various industries
- **Scales;** Managerial aptitude test in English (Recruit Management Solutions)
- **Procedure**; Internet survey after being screened by English ability
- **Analysis**; Force-choice items are analyzed and scored using Thurstonian IRT modeling. Likert items are analyzed by multi-graded IRT.
- **Item examples；**
< Forced-choice items >
  "Choose the one that you consider <u>most like you</u>, and the one that you consider <u>least like you</u>"
    I am able to make friends with others quickly.
    I am good at building consensus in my group.
    I like taking charge when I work in a group.
    Even when facing a problem, I can be very optimistic.

< Likert items > (1 strongly disagree – 4 strongly agree)
  ※ Items are basically same with the forced choice items.
    Each sentence is rated separately.
< Impression Management > (1 strongly agree – 7 strongly disagree)
    I always obey laws, even if I'm unlikely to get caught.   (Paulhus & Reid, 1991)
    I don't gossip about other people's business.
< High self-regard > (1 not satisfied at all – 7 highly satisfied)
  "How satisfied are you with your performance regarding the behavior described below"
    to proactively build a relationship with many people
    to take the initiative and lead others as a leader

Result 1   (correlations among test scores)

Correlations (N=644)

	L_Extra.	L_Influ.	L_Innov.	L_Tough.	L_Deci.	F_Extra.	F_Influ.	F_Innov.	F_Tough.	F_Deci.	Mean	SD
Likert_Extraverted	1										−.010	.912
Likert_Influential	.553**	1									−.006	.868
Likert_Innovative	.417**	.626**	1								−.008	.887
Likert_Tough-minded	.449**	.451**	.443**	1							−.006	.865
Likert_Decisive	.409**	.594**	.502**	.524**							−.001	.885
Forced-choice_Extraverted	.640**	.377**	.252**	.282**	.230**	1					−.075	.987
Forced-choice_Influential	.403**	.597**	.428**	.267**	.356**	.663**	1				−.006	1.031
Forced-choice_Innovative	.317**	.475**	.561**	.286**	.369**	.523**	.795**	1			−.034	.995
Forced-choice_Tough-minded	.289**	.280**	.307**	.544**	.415**	.471**	.455**	.566**	1		−.041	.958
Forced-choice_Decisive	.225**	.391**	.362**	.423**	.553**	.364**	.573**	.685**	.815**	1	−.018	.942

** p < .01

- <u>Correlations between Likert and forced-choice scores of the corresponding scales are moderate, but higher than those of non-corresponding scales</u>.
- Correlations among the scales tend to be higher with the forced-choice scores than the Likert scores.
- Some inter-scale correlations of forced-choice scores are too high to claim measuring independent constructs.

Result 2   (Structural Equation Modeling)

$X^2 = .78$ (df=176)
CFI= .91
RMSEA= .07
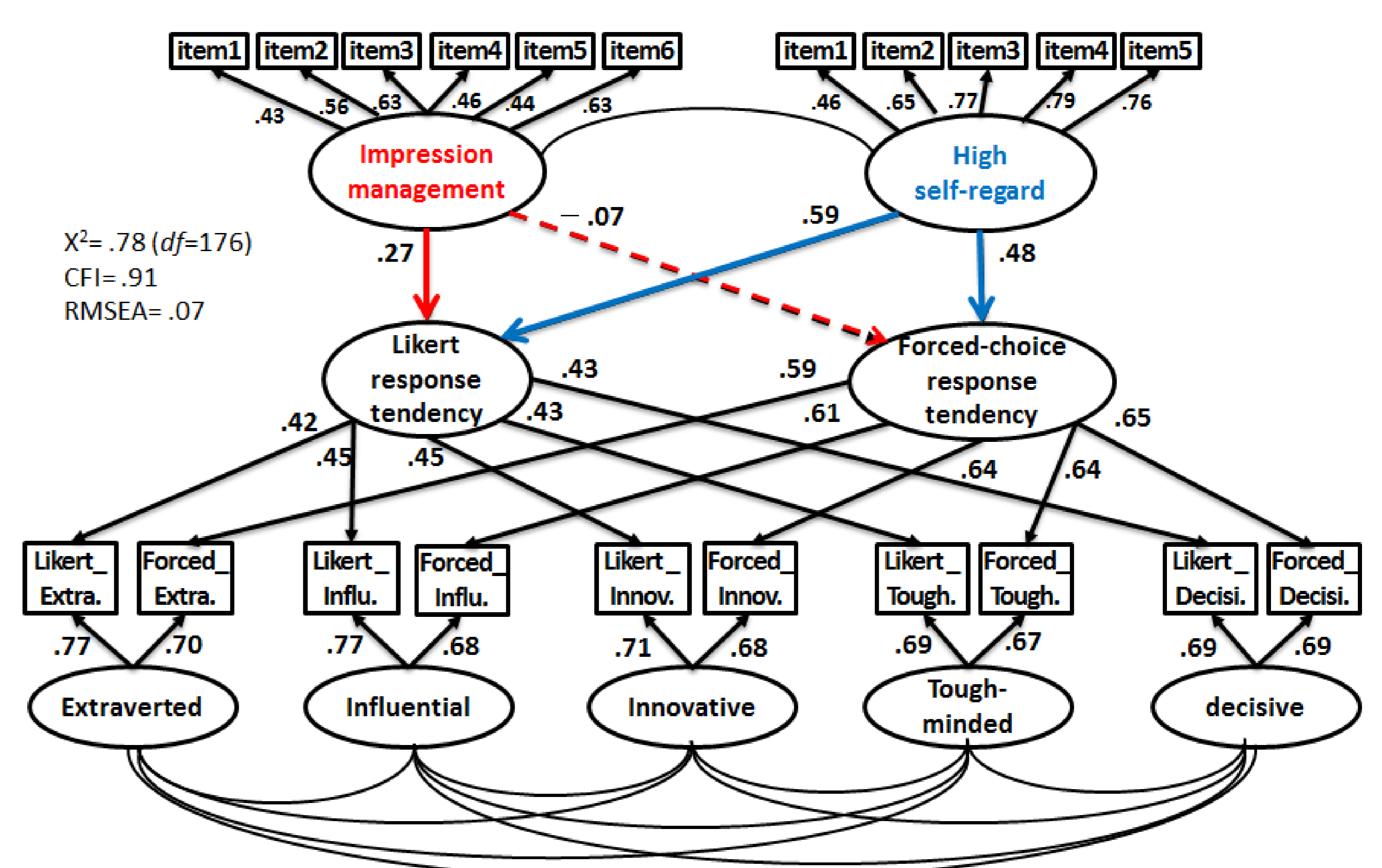
- <u>Impression management did affect Likert but not forced-choice scores.</u>
- <u>High self-regard affected both Likert and forced-choice scores at the same extent.</u>
- A limitation of self-reporting personality tests.
- What would be included in the forced-choice response?
- Previous studies showed forced-choice scores were affected by general intelligence (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006; Cristiansen, Burns, & Montgomery, 2005).

Future studies

- To find ways to reduce high inter-scale correlations of the forced-choice scales.
- To examine the extent and reason why general intelligence affect the forced choice scores.
- To conduct a similar study with Japanese female office workers.
- To conduct a similar study in other countries.
- To conduct a validation study with performance criteria (i.e., supervisor's ratings).