Extreme Response Style and Faking: Two Sides of the Same Coin?

Matthias Ziegler and Christoph J. Kemper

Abstract

Using Mixed Rasch Models, several studies identified two distinct classes present in personality test data collected under lowstakes (i.e. the test person has nothing to gain or lose based on the test results) and anonymous conditions. Based on threshold diagrams the classes were interpreted as two response styles – extreme response style and midpoint responding. Within psychological research concerned with faking, there are some studies in which Mixed Rasch Models were also applied. However, in those studies data were derived from high stakes (i.e. the test person can gain or lose something based on the test results, e.g., a job application situation) and non-anonymous conditions. Again, two response bias classes emerged. This time they were interpreted as representing slight and extreme faking. Whereas slight fakers deviate only little from their actual honest answer in a personality questionnaire, extreme fakers deviate strongly. It is unclear, though, in how far these findings actually represent distinct response biases. Nevertheless, Mixed Rasch Models might offer the chance to correct survey data for these response biases. The present chapter shortly summarizes prior findings and then reports a study in which both conditions, i.e. low and high stakes settings, were realized within one sample, allowing to investigate the relationship between the different biases found in different conditions. The findings suggest that the previously reported response styles for different high and low stakes conditions might be less distinct than assumed. This raises the question whether the Mixed Rasch Model can really capture faking. Practical implications are considered.

Introduction

Questionnaires are an efficient and widely applied source of research data in the social sciences. Behavior, attitudes, or personality are usually measured by having respondents indicate their level of agreement to a series of statements on a multi-point rating scale. However, variance due to a substantive construct is not the only source of systematic variance in these measurements. Observed responses and scores may also contain variance due to response bias. The present book deals with these error sources mainly from a perspective in which an interviewer is thought to have manipulated the data. While this is extremely important, it should also be kept in mind that the interviewee might distort answers to the questions. This perspective is taken in the present chapter. From the perspective of the interviewee different response biases can be distinguished. Paulhus (1991) refers to response bias as a systematic tendency to respond to questionnaire items on some basis other than the specific content. This systematic tendency might be a temporary reaction to a situational demand (response set) or a stable and consistent behavior (response style). For example, in high stakes situations, i.e., situations in which the test taker can gain or lose something based on the test results, such as job applications (Zickar, Gibby, & Robie, 2004) persons may deliberately distort their answers in order to create a positive impression. This response bias is referred to as faking (MacCann, Ziegler, & Roberts, 2011). A more general term is social desirability. Yet, social desirability can also include unconscious distortions like self-enhancement. Therefore, we use the term faking to clearly indicate that a willful distortion is meant. Two distinct classes of fakers were identified in previous studies – slight fakers (SF) and extreme fakers (EF) (Robie, Brown, & Beaty, 2007; Ziegler, 2011). Slight fakers are described as people who distort their responses only little. Extreme fakers, however, strongly deviate from what would actually be their honest response. In low stakes situations, i.e., there is nothing to gain or lose for the test taker, response biases may occur due to stable and consistent preferences of persons to answer rating scale items in a specific way. For example, some persons prefer extreme response categories of a rating scale while others prefer moderate categories to describe themselves (extreme responding, ERS; midpoint responding, MPR). Numerous studies have shown that these biases occur when persons respond to rating scale items and that they may contaminate the measurement of substantive constructs (Austin, Deary, & Egan, 2006; Gollwitzer, Eid, & Jürgensen, 2005; Meiser & Machunsky, 2008; Rost, Carstensen, & Von Davier, 1997). The existence of SF and EF as well as ERS and MPR raises at least two interesting questions for the survey context. On the one hand, if such individual response styles could be modeled, survey data could be corrected for such distortions. On the other hand, especially ERS is also of interest within survey research concerned with identifying interview protocols faked by the interviewer (Bredl, Winker & Kötschau, 2012). Thus, if the same response bias can be found for the interviewee, more care is needed when using ERS to identify dishonest interviewers.

Thus, response biases threaten the data quality of surveys and the validity of conclusions based on survey data. Moreover, such response biases can be found on the level of the interviewee as well as on the level of the interviewer. In the research presented in this article, we focus on two interviewee response biases that have been identified applying Mixed Rasch Models (MRM), i.e. ERS/MPR and SF/EF in previous research. Our research questions are (1) whether people responding to a questionnaire with rating scale items in a low and a high stake situation can be classified as belonging to a specific response bias class, and (2) how these classifications of ERS/MPR and SF/EF are related. The findings bear practical implications for survey research. If it is possible to identify such response biases, correction methods could be applied to improve data quality in surveys.

On the following pages we will first shortly summarize findings with regard to ERS and MPR. Afterwards we will introduce the basic ideas of the Mixed Rasch Model followed by a short overview of psychological faking research applying this method. Finally, we will present results from an empirical study where both response styles, i.e. ERS/MPR and SF/EF, were studied on the level of the test taker. The chapter closes with a discussion of these results as well as some practical implications.

Extreme Response Style and Midpoint Responding

ERS refers to the tendency to disproportionately favor the endpoints or "extreme" categories of rating scales, irrespective of particular item content (Naemi, Beal, & Payne, 2009). In contrast, MPR refers to the tendency to prefer middle or neutral categories of rating scales. Both response styles are an issue of concern for researchers using questionnaires as a source of research data. Distinct groups of persons who prefer to describe themselves with either extreme or moderate ratings were found for diverse constructs, e.g. the Big Five personality dimensions (Austin et al., 2006; Hinz, Brahler, Geyer, & Korner, 2003; Rost, Carstensen, & von Davier, 1999b), personal need for structure (Meiser & Machunsky, 2008), anger expression (Gollwitzer et al., 2005), anxiety sensitivity (Kemper, 2010), and in organizational surveys (Eid & Zickar, 2007). These results suggest that ERS and MPR might affect the measurement of constructs in general, especially when multi-point rating scales are used. ERS and MPR may introduce additional construct-irrelevant variance to the measurement leading to biases on the individual and on the group level. On the individual level, a person's tendency to prefer extreme categories may result in a higher score and thus in a higher level of the intended-to-be-measured trait (Bolt & Johnson, 2009). On the group level, another bias might occur. When items function differently in subgroups of the population, observed effects can be a result of group specific response styles. In this case, mean differences between groups cannot be validly compared (Bolt & Johnson, 2009; Eid & Rauber, 2000). As underlying dispositional factors of ERS/MPR Naemi et al. (2009) propose so-called epistemic constructs (Kruglanski, 1989), e.g. judgment complexity, rigidity, intolerance of ambiguity, and dogmatism. Such dispositions describe the manner in which people process and respond to information. Besides possible dispositional antecedents of these response styles, researchers have explored their relations to various individual difference variables. Substantial correlations were found between ERS/MPR and sex, age, cognitive ability, education, ethnicity, and personality, especially extraversion and anxiety (Austin et al., 2006; Berg & Collier, 1953; Eid & Rauber, 2000; Greenleaf, 1992; Meisenberg & Williams, 2008; Naemi et al., 2009). However, results are considerably less consistent and conclusive regarding the nature and direction of the observed relations.

The present study will not focus on possible relations between ERS/MPR and personality characteristics. Instead, the relationship with SF/EF will be investigated to determine the uniqueness of these supposedly distinct response biases.

The Mixed Rasch Model

The most sophisticated method to identify ERS and MPR is the so-called Mixed Rasch Model (MRM; Rost et al., 1997). The MRM combines a Latent Class Analysis (LCA) with a Rasch analysis. In the present analyses a LCA was combined with an ordinal Rasch model. The basic equation is:

$$p(X_{vi} = x) = \sum_{g=1}^{G} \pi_g \cdot \frac{e^{(x \cdot \theta_{vg} - \sigma_{ixg})}}{\sum_{s=0}^{m} e^{(m \cdot \theta_{vg} - \sigma_{img})}}$$

$p(X_{vi} = x)$ = Probability of person v to chose category x for item i

$\sum_{g=1}^{G}$ = Sum from class g = 1 to class G

π_g = Class size parameter

σ_{ixg} = Sum of all threshold parameters of item i up to chosen category x in class g

θ_{vg} = Person parameter for person v in class g

σ_{img} = Sum of all threshold parameters of item i in class g

m = Number of thresholds

s = threshold

Consequently, in MRM different qualitative classes can be assumed in which person parameters may vary quantitatively. Results of the MRM analysis are indicative of whether there are subgroups in the sample data for which the items are differentially difficult to answer or subgroups that exhibit different response behaviour. To determine the number of classes representing the data best, information criteria, e.g. the Consistent Akaikes Information Criterion (CAIC) are usually compared (Rost et al., 1997; Zickar et al., 2004). Thus, this combination of a Latent Class Analysis and a Rasch model allows for qualitatively distinct classes in which people can differ quantitatively on the trait being measured. As could be shown, under certain conditions, the results of an MRM analysis can be interpreted as evidence of a response bias measured along with a trait.

Using MRM, two types of information can be obtained for each individual. Each person is assigned to one class and within each class a person parameter representing the individual's trait level is estimated. Person parameters theoretically range from minus to plus infinity but scores between minus and plus 3 are most common. A score of zero indicates an average trait level. The higher the score, the higher the trait level. On the item level, threshold parameters are estimated for each item. The item thresholds represent the boundary values between neighboring rating categories of an item. Thresholds are scaled on the same dimension as the person parameters and represent the level a person parameter must exceed to achieve a higher probability for the person to pass from one rating category to the next, e.g. from "very likely" to "extremely likely". Because the classes differ qualitatively, threshold values cannot be compared between the classes. However, it is possible to find out whether the classes found in the data emerge due to differences in traits or differences in response bias by inspecting the item threshold diagrams. If the thresholds run parallel in both classes, this indicates a response bias. Most studies that found ERS/MPR could show that the estimated threshold parameters run parallel in these two classes. This is interpreted as indicating two classes that only differ in their response bias but not the trait actually measured. This method was also used in the research presented here.

Extreme and Slight Faking

MacCann et al. (2011) define faking as: "… a deliberate set of behaviors motivated by a desire to present a deceptive impression to the world. Like most other behavior, faking is caused by an interaction between person and situation characteristics. (p.311)". Faking is further seen as a conscious effort to manipulate responses to self-descriptive questionnaire items (Ziegler, 2011; Ziegler, MacCann, & Roberts, 2011). The ultimate goal of faking is to create a specific impression depending on the demands faced by a person in everyday situations. For example, the deliberate promotion of competence, fearlessness, and physical prowess is often seen in job applicants or males trying to impress a dating partner. Another example is the deliberate minimization of faults involving excuse-making and damage control. This behavior might be seen in religious settings or legal defendants trying to avoid punishment (Paulhus, 2002). According to Paulhus, the former is termed agency management and the latter communion management. Both forms of impression management or faking are considered a problem for researchers working with questionnaire data as they may introduce a systematic error to the measurement of substantive constructs (Heggestad, George, & Reeve, 2006). When faking questionnaires, persons may distort item scores and aggregated scale scores slightly or extremely (slight vs. extreme fakers) impacting on the criterion (Ziegler, Danay, Schölmerich, & Bühner, 2010) and construct validity (Ziegler & Bühner, 2009) of the measurement. Intensive research focused on the underlying factors and processes of faking behavior. Concerning situation attributes, a psychological demand is necessary motivating and pressuring the person to create a certain impression, e.g. looking more favorable or impressive than one actually is (faking good) or looking more problematic or psychologically troubled than one actually is (faking bad) (Ellingson, 2011; Mueller-Hanson, Heggestad, & Thornton, 2006). Concerning person attributes, faking seems positively related to self-efficacy of positive self-presentation (Pauls & Crost, 2005; Ziegler, 2007), self-monitoring, and cognitive ability, and negatively related to integrity (McFarland & Ryan, 2000). As far as Big Five traits are concerned, results are less consistent. However, all Big Five traits were found to be related to faking, i.e. affected by it (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Grubb & McDaniel, 2007).

As was the case with ERS and MPR MRMs are often used to identify slight and extreme faking in questionnaire data. In both cases, resulting classes are interpreted based on the fact that the item thresholds for the found classes run parallel, indicating the same trait is measured but with the impact of a response bias. For the data acquired under low stakes conditions the classes are called ERS and MPR. For data derived from high stakes settings SF and EF. However, the differences between the respective classes oftentimes look pretty comparable. Despite the fact that the methods used and the way results are interpreted are the same, the relationship between ERS/MPR and SF/EF is unknown. If the two response biases were indeed distinct, it would be possible to distinguish between a mere response style and actual faking. However, if the biases were related, MRMs could still be used with survey data to correct the scores if either one is thought to be likely to have occurred.

Method

Participants and Procedure

326 undergraduate psychology students of a German University participated in the study. Due to technical problems during the administration of the measures (see below) used, 14 data sets were lost. The final sample consisted of N = 312 participants between 21 and 53 years of age (M = 25, SD = 5.5). 79% were females. Participants were invited to a lab on campus and randomly assigned to either an experimental group (n_E = 157, 121 females) or a control group (n_C = 155, 126 females). After providing informed consent, participants were seated in front of a computer and received a briefing. A test battery containing several tests was administered on the computer. Part of this battery was the Revised NEO Personality Inventory (NEO-PI-R; Ostendorf & Angleitner, 2004). Several other scales which are not the focus of the present research were also administered (e.g., the Intelligence Structure Test 2000-R (IST 2000-R; Amthauer, Brocke, Liepmann, & Beauducel, 2001)). For the NEO-PI-R, participants received different instructions. In the control group, participants were instructed to fill out the NEO-PI-R honestly on the first measurement occasion as well as on the second. In the experimental group, participants were instructed to respond to the personality test honestly on the first occasion (low stakes situation) and fake-good on the second occasion (simulated high stakes situation). Thus, data from the low stake situation can be used to model ERS and MPR. Data from the simulated high stakes situation can be used to model SF and EF. The faking instruction used in the simulated high stakes condition contained a realistic scenario for all participants with a warning to discourage obvious faking as proposed by Rogers (1997). As participants were university students, we adapted the instruction to this context: “Universities have to select their students. To achieve a good selection, a number of instruments like the following are usually used. Please imagine that you are participating in a student selection procedure. Of course, it is your goal to get an admission as a psychology student. Therefore, you have to fill out the following questionnaire in a way that assures your admission. However, you have to be careful since a test expert will check the results for obvious faking and you do not want to be identified.” After completing the test battery, participants received course credit for their participation in the study and were debriefed.

Measures

Personality was assessed with a German adaptation (Ostendorf&Angleitner, 2004) of the NEO-PI-R (Costa & McCrae, 1992). The NEO-PI-R allows for a comprehensive assessment of personality based on the Five-Factor Model (FFM). Containing 240 items, the NEO-PI-R measures the Big Five personality dimensions of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness with 48 items per dimension. Each dimension consists of several facets, e.g. Neuroticism consists of Anxiety (N1), Hostility (N2), Depression (N3), Self-conscientiousness (N4), Impulsiveness (N5), and Vulnerability (N6). Each of these facets is measured by 8 items. Items contain statements and a five-point rating scale for respondents to indicate their endorsement ranging from "strongly disagree" to "strongly agree". The NEO-PI-R is a comprehensive, reliable and valid measure of the Big Five personality dimensions and its facets (for details on the psychometric quality see Costa & McCrae, 1992; Ostendorf & Angleitner, 2004). Internal consistencies here were comparable to those reported in the German manual and ranged from $\alpha = .46$ to $\alpha = .88$ for the personality facets.

Statistical Analysis

Data was analyzed with SPSS 20. Furthermore, the computer program WINMIRA (von Davier, 2001) was used for the MRM and the LCA. To avoid response bias associated with the middle category of the five-point rating scale other than MPR/ERS or SF/EF, e.g., its correlation with item clarity (Kulas & Stachowski, 2009), the scale was transformed to a four-point scale for the MRM analyses as proposed by Rost, Carstensen and von Davier (1999a) as well as Austin et al. (2006). This was accomplished by collapsing category 2 with 3. Models with one to four classes were estimated for the data and the best fitting solution selected based on information criteria – Akaike's Information Criterion (AIC), Schwartz Information Criterion (BIC) and Bozdogan's Consistent AIC (CAIC) (for details see Bozdogan, (1987); Read & Cressie, (1988)). Threshold parameter profiles were compared between classes to specify the nature of the found classes. These analyses were conducted for all participants using the data from measurement occasion 1. Here no specific instruction other than to respond honest to NEO-PI-R items was given. Thus, we were looking for two class solutions resembling ERS and MPR. These analyses were run for all 30 facets of the NEO-PI-R separately. As a consequence, each person was classified as either applying ERS or MPR for those facets where these biases could be found. We then used a Latent Class Analysis with these classifications to determine whether people consistently used a response style across facets or switched between styles. Using only data from the experimental group (fake-good condition), the same analyses were conducted for data from occasion 2. This time, the two class solution was supposed to represent SF and EF. Thus, after the analyses two classifications were made for each participant from the experimental group. Each participant was classified as either ERS or MPR based on the data from occasion 1 (low stake situation, i.e. honest responses can be assumed) and as SF or EF based on data from occasion 2 (simulated high stakes situation, i.e. faked responses can be assumed). These classifications of ERS/MPR and SF/EF were then compared using a McNemar test.

Items yielding estimation problems in the MRM analyses or causing class solutions based on content (i.e. the thresholds did not run parallel in the classes) were excluded from the analyses. If a facet had less than five items remaining, the facet was also excluded from further analyses because scale length is an important factor in the accurate identification of classes with WINMIRA (Zickar & Burnfield, 2003).

Results

Using data from occasion 1 ERS/MPR could be shown to occur in 12 facets. For eight out of the 30 facets, 2-class solutions could be found without any further changes. However, for some facets estimation problems occurred for some items which were then dropped from the analyses. As a consequence four facets were no longer analyzed because less than 5 items remained. Having eliminated items with estimation problems 2-class solutions that could be interpreted as ERS/MPR emerged for four more facets. The remaining facets either had 1-class solutions as best fitting result or were not clearly interpretable as ERS/MPR. The 1-class solutions could be due to the small number of items and persons (for a detailed description see Gerber-Braun, 2010). Thus, classifications in 12 facets were the basis for the LCA. LCA results can be found in Table 13.1.

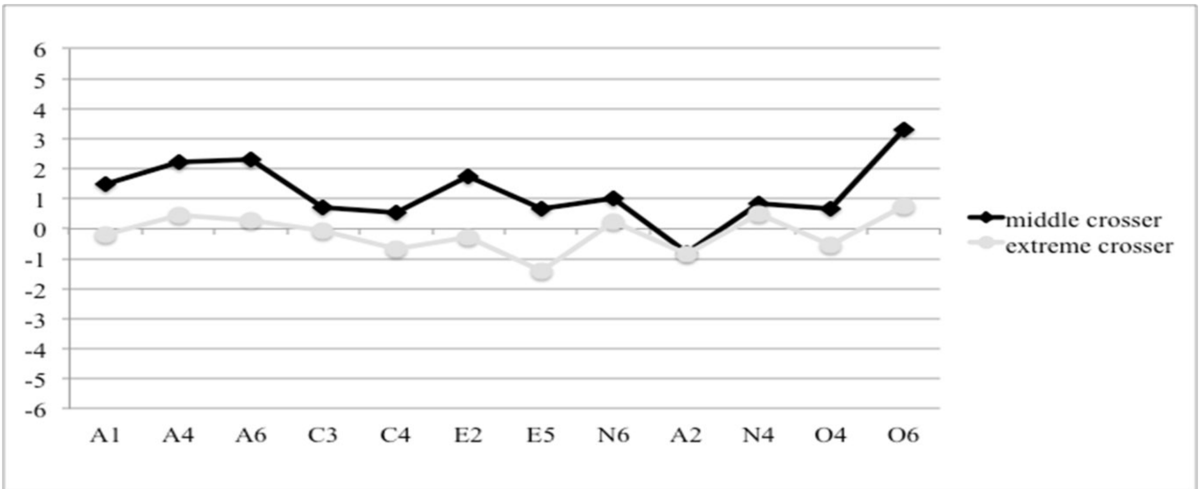

Fig. 13.1: Item locations for the two-class solution of the LCA for the 12 facet classifications; A1 = Trust; E2 = Gregariousness; A2 = Straightforwardness; C3 = Dutifulness; N4 = Self-conscientiousness; O4 = Actions; A4 = Compliance; C4 = Achievement striving; E5 = Excitement seeking; N6 = Vulnerability; O6 = Values; A6 = Tender-mindedness

Table 13.1: Latent Class Analysis with 12 facets of the NEO-PI-R for Occasion 1

Number of classes	class nr.	class size	sum score	Information Criteria		
				AIC	BIC	CAIC
1	1	1	4.27	4537.13	4582.04	4594.04
2	1	.67	3.21	4426.02	**4519.60**	**4544.60**
	2	.33	6.40			
3	1	.40	3.47	**4424.70**	4566.93	4604.93
	2	.31	6.54			
	3	.29	2.95			
4	1	.35	3.39	4425.18	4616.07	4667.07
	2	.23	5.23			
	3	.23	2.32			
	4	.19	6.99			

Notes: sumscore = sum of the means of the facets included in the LCA; bold class solutions indicate the best solution according to the particular criteria

Table 13.1 shows that the two-class solution fitted the data best according to the commonly chosen CAIC. Moreover, Figure 13.1 shows that the item locations for the two classes run almost parallel. This is usually seen as representing a response style. Thus, the analyses support a two class solution with about one third of participants being classified as applying ERS.

Before conducting the analyses with the data from occasion 2 (simulated high stakes), ANCOVAs for all facets were conducted to identify facets with substantial differences in personality scores between the experimental and control group. Data from Occasion 2 served as dependent variables and data from Occasion 1 as covariate. Experimental vs. control group was used as independent variable. Only those facets with significant and substantial differences between both groups after controlling for baseline differences were considered to be affected by faking and included in the subsequent MRM analyses. Only two of 30 facets were not faked (Openness to aesthetics and Openness to values, $p > .05$, partial $eta^2 < .013$). Thus, there were substantial and significant differences between the experimental and the control group in 28 of the 30 facets indicating faking. After the following analyses mirroring the procedure described above, thirteen facets yielded two class solutions consistent with an interpretation of classes as SF and EF. Classifications from these thirteen facets were used for the LCA. Results can be found in Table 13.2.

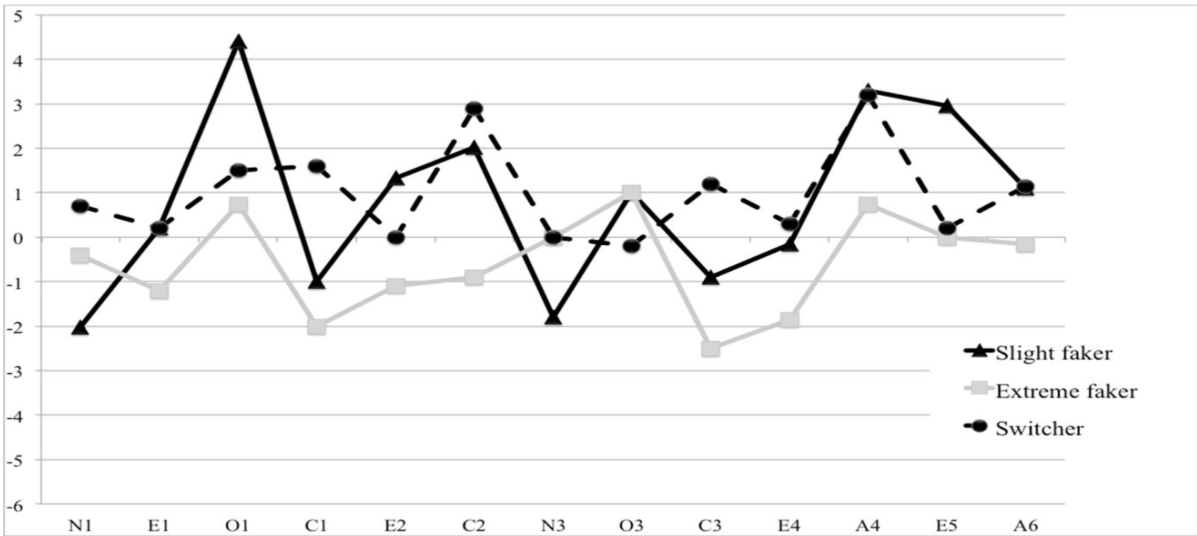

Fig. 13.2: Item locations for the switchers, slight fakers and extreme fakers of the 3-class solution of 13 facets in the LCA; N1 = Anxiety; E1 = Warmth; O1 = Fantasy; A1 = Trust; C1 = Competence; E2 = Gregariousness; C2 = Order; N3 = Depression; O3 = Feelings; C3 = Dutifulness; E4 = Activity; A4 = Compliance; C4 = Achievement striving; N5 = Impulsiveness; E5 = Excitement seeking; O5 = Ideas; A6 = Tender-mindedness.

Table 13.2: Latent class analysis with 13 facets of the NEO-PI-R for Occasion 2

Number of classes	class nr.	class size	sum score	Information Criteria		
				AIC	BIC	CAIC
1	1	1.00	5.25	5028.01	5076.67	5089.67
2	1	0.67	4.30	4839.51	4940.57	4967.57
	2	0.33	7.18			
3	**1**	**0.17**	**4.17**	4754.87	**4908.34**	**4949.34**
	2	**0.53**	**5.13**			
	3	**0.30**	**7.88**			
4	1	0.37	4.36	4722.98	4928.84	4983.84
	2	0.23	5.45			
	3	0.23	4.31			
	4	0.16	8.31			
5	1	0.25	5.49	**4707.36**	4965.63	5034.63
	2	0.24	4.47			
	3	0.21	5.83			
	4	0.20	3.27			
	5	0.12	8.69			

Notes: sum score = sum of the means of the facets included in the LCA; bold class solutions indicate the best solution according to the particular criteria

The findings support a 3-class solution. As can be seen in Figure 13.2, two classes again have almost parallel item locations indicative of a response style. One class however has item locations that are not parallel. The most likely explanation is that this class used a response style for some of the facets (here parallel item locations occur) but did not use a response style for other facets. In other words, this class did not fake on all of the facets. Based on MRM classifications, 53% of participants were consistently classified as SF and 30% as EF by the Latent Class Analysis. 17% of the participants switched from responding honestly for some facet to responding dishonestly for others. This is most likely due to differences in interpreting the meaning of the facet with regard to the utility of the trait for the imagined student selection procedure (Ziegler, 2011).

Summing up, the first part of the analyses, participants could be classified consistently as using ERS or MPR for most facets based on data from Occasion 1. Classification as SF or EF with data from Occasion 2 (faking-good) was consistent for most participants but not for all. These switchers were excluded from the following comparison between ERS/MPR and SF/EF as our main interest was the statistical relation between these response biases. Thus, n = 129 participants remained for the McNemar-Test. Table 13.3 contains a cross-classified table showing the change of the classification from Occasion 1 to Occasion 2.

Table 13.3: Cross-classified table for the change of classification from Occasion 1 to Occasion 2

	Faking (Occasion 2)		Σ
Response Style (Occasion 1)	SF	EF	
MPR	64	24	88
ERS	18	23	41
Σ	82	47	129

The McNemar-Test yielded a small and nonsignificant result ($\chi 2= .60$, $p = .44$, Cohen's $w = .07$). There was no substantial change in classification between the measurement occasions. Thus, there is no difference between the two response biases ERS/MPR and SF/EF that could not be explained by chance.

Discussion

The present analyses were conducted to shed light on the relationship between two well investigated response biases. ERS and MPR have repeatedly been found in research data collected in low stakes settings using Mixed Rasch Models. With the same technique, slight and extreme faking (SF/EF) have been found in data derived from high stake settings. Despite the overlap in the interpretation of the underlying statistical analyses, the relationship between the two response biases is unknown.

The findings presented here suggest that the supposedly distinct response biases occurring under low stake (ERS/MPR) and high stake conditions (SF/EF) are not that distinct after all. In fact, the results strongly suggest that people use the same response style regardless of the demand they face in specific situations, e.g. when responding to a survey questionnaire in a low stake situation (computer assisted web interview) or a high stake situation (face-to-face interview with family members or peers present). An exception in our study was the small class of participants who switched from responding honestly to dishonestly only for some facets.

These results bear little hope to use the MRM to correct for faking in survey data or even to identify possible faked data sets. However, the method might be a valid and reliable approach to capture ERS and MPR and to correct for these biases. A limiting factor of the present study is that only laboratory data from a student sample and one specific questionnaire could be used. Thus, replications with real-life data from applicant settings and representative data from large-scale surveys including other measures than the NEO-PI-R are necessary. Moreover, future research should investigate how scores of multi-item scales corrected for response bias in MRM analyses can improve data quality, e.g. by purging irrelevant variance from measures and improving their statistical relationship with important outcome variables.

Nevertheless, the finding that ERS and MPR also exist on the level of the interviewee within a survey bears implications for survey research in general and identification of protocols faked by the interviewer in particular. Methods counting on ERS as being indicative of wrong doing on the side of the person conducting the interview have to take into account that such a response style can and most likely will also occur due to the person being interviewed. Thus, MRM as used here might still be a helpful tool in order to ascertain that a response style is due to the interviewee and not the interviewer. Future research could therefore use data sets including faked protocols and apply MRM to potentially differentiate between response styles on the side of the interviewee and wrong doing by the interviewer.

Summing up, the present study indicates that MRM are not suitable to detect faking. However, identifying ERS and MPR reliably seems feasible and offers the opportunity to correct data for the detrimental effects such response styles might have and to improve the quality of survey data based on multi-point rating scale items.

Bibliography

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A.: I-S-T 2000 R (Intelligenz-Struktur-Test 2000 R) [Intelligence-Structure-Test 2000 R], Göttingen, Hogrefe, 2001.

Austin, E. J., Deary, I. J., & Egan, V.: Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items, Personality And Individual Differences, 40, 1235-1245, 2006.

Berg, I. A., & Collier, J. S.: Personality and group differences in extreme response sets, Educational and Psychological Measurement, 13, 164-169, 1953.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A.: A meta-analytic investigation of job applicant faking on personality measures, International Journal of Selection and Assessment, 14, 317-335, 2006.

Bolt, D. M., & Johnson, T. R.: Addressing score bias and differential item functioning due to individual differences in response style, Applied Psychological Measurement, 33, 335-352, 2009.

Bozdogan, H.: Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, Psychometrika, 52, 345-370, 1987.

Bredl, S., Winker, P., Kötschau, K.: A statistical approach to detect interviewer falsification of survey data, Survey Methodology, 38, 1-10, 2012.

Costa, P. T., & McCrae, R. R.: Revised NEO personality inventory (NEO-PI-R) and NEO Five-Factor inventory (NEO-FFI): Professional Manual.: Odessa, Psychological Assessment Resources, 1992.

Eid, M., & Rauber, M.: Detecting measurement invariance in organizational surveys, European Journal of Psychological Assessment, 16, 20-30, 2000.

Eid, M., & Zickar, M.: Detecting response styles and faking in personality and organizational assessments by mixed Rasch models, In von Davier, M. & Carstensen, C. H. (eds.), Multivariate and mixture distribution Rasch models (255-270), New York, NY, Springer, 2007.

Ellingson, J. E.: People fake only when they need to fake. In Ziegler, M., MacCann, C. & Roberts, R. (eds.), New perspectives on faking in personality assessment (19-33), New York, NY, Oxford University Press, 2011.

Gerber-Braun, B.: The Double Cross. Dissertation, LMU München, Fakultät für Psychologie und Pädagogik, 2010.

Gollwitzer, M., Eid, M., & Jürgensen, R.: Response styles in the assessment of anger expression, Psychological Assessment, 17, 56-69, 2005.

Greenleaf, E. A.: Measuring extreme response style, Public Opinion Quarterly, 56, 328-351, 1992.

Grubb, W. L., & McDaniel, M. A.: The fakability of Bar-On's Emotional Quotient Inventory Short Form: Catch me if you can, Human Performance, 20, 43-59, 2007.

Heggestad, E. D., George, E., & Reeve, C. L.: Transient error in personality scores: Considering honest and faked responses, Personality and Individual Differences, 40, 1201-1211, 2006.

Hinz, A., Brahler, E., Geyer, M., & Korner, A.: Urteilseffekte beim NEO-FFI / Response sets measures with NEO-FFI, Diagnostica, 49, 157-163, 2003.

Kemper, C.: Das Persönlichkeitsmerkmal Angstsensitivität: Taxon oder Dimension? – Eine Analyse mit dem Mischverteilungs-Raschmodell, Kovacs, 2010.

Kruglanski, A. W.: Lay epistemics and human knowledge: Cognitive and motivational bases, Plenum Press, 1989.

Kulas, J. T., & Stachowski, A. A.: Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings, Journal of Research in Personality, 43, 489-493, 2009.

MacCann, C., Ziegler, M., & Roberts, R.: Faking in personality assessment: Reflections and recommendations, In Ziegler, M., MacCann, C. & Roberts, R. (eds.), New perspectives on faking in personality assessment (309-329), New York, Oxford University Press, 2011.

McFarland, L. A., & Ryan, A. M.: Variance in faking across noncognitive measures, Journal of Applied Psychology, 85, 812-821, 2000.

Meisenberg, G., & Williams, A.: Are acquiescent and extreme response styles related to low intelligence and education? Personality and Individual Differences, 44, 1539-1550, 2008.

Meiser, T., & Machunsky, M.: The personal structure of personal need for structure - A mixture-distribution Rasch analysis, European Journal of Psychological Assessment, 24, 27-34, 2008.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C.: Individual differences in impression management: an exploration of the psychological processes underlying faking, Psychology Science, 3, 288-312, 2006.

Naemi, B. D., Beal, D. J., & Payne, S. C.: Personality predictors of extreme response style, Journal of Personality, 77, 261-286, 2009.

Ostendorf, F., & Angleitner, A.: NEO-PI-R. NEO Persönlichkeitsinventar nach Costa und McCrae. Revidierte Fassung. [NEO-PI-R. NEO Personality Inventory], Göttingen: Hogrefe, 2004.

Paulhus, D. L.: Measurement and control of response bias. In Robinson, J.P. & Wrightman, L.S. (eds.), Measures of Personality and Social Psychological Attitudes (17-59), San Diego, CA , Academic Press, 1991.

Paulhus, D. L.: Socially desirable responding: The evolution of a construct. In Braun, H. I., Jackson, D. N. & Wiley, D. E. (eds.), The role of constructs in psychological and educational measurement (49-69), Mahwah, NJ, Lawrence Erlbaum Associates, 2001.

Pauls, C. A., & Crost, N. W.: Cognitive Ability and Self-Reported Efficacy of Self-Presentation Predict Faking on Personality Measures, Journal of Individual Differences, 26, 194-206, 2005.

Read, T. R. C., & Cressie, N.: Goodness-of-Fit Statistics for Discrete Multivariate Data, New York, NY, Springer-Verlag, 1988.

Robie, C., Brown, D. J., & Beaty, J. C.: Do people fake on personality inventories? A verbal protocol analysis, Journal of Business and Psychology, 21, 489-509, 2007.

Rogers, R.: Researching Dissimulation. In Rogers, R. (ed.), Clinical assessment of malingering and deception (2 ed., 309-327), New York, NY, Guilford, 1997.

Rost, J., Carstensen, C. H., & Von Davier, M.: Applying the mixed Rasch model to personality questionnaires. In Rost, J. & Langeheine, R. E. (eds.), Applications of latent trait and latent class models in the social sciences, New York, NY, Waxmann, 1997.

Rost, J., Carstensen, C. H., & Von Davier, M.: Are the Big Five Rasch scalable? A reanalysis of the NEO-FFI norm data, Diagnostica, 45, 119-127, 1999.

Von Davier, M.: WINMIRA 2001, Groningen, The Netherlands, ASC-Assessment Systems Corporation USA and Science Plus Group, 2001.

Zickar, M. J., & Burnfield, J. L.: A Monte Carlo investigation of the polytomous mixed Rasch model, unpublished manuscript, 2003.

Zickar, M. J., Gibby, R. E., & Robie, C.: Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory, Organizational Research Methods, 7, 168-190, 2004.

Ziegler, M.: Situational demand and its impact on construct and criterion validity of a personality questionnaire: State and trait, a couple you just can't study separately! Dissertation, LMU München, Fakultät für Psychologie und Pädagogik, 2007.

Ziegler, M.: Applicant Faking: A Look Into the Black Box, The Industrial and Organizational Psychologist, 49, 29-36, 2011.

Ziegler, M., & Bühner, M.: Modeling Socially Desirable Responding and Its Effects, Educational and Psychological Measurement, 69, 548-565, 2009.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M.: Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked, European Journal of Personality, 24, 341-355, 2010.

Ziegler, M., MacCann, C., & Roberts, R. R.: Faking: Knowns, unknowns, and points of contention. In Ziegler, M., MacCann, C. & Roberts, R. (eds.), New perspectives on faking in personality assessment (309-329), New York, Oxford University Press, 2011.