Controlling social desirability may attenuate faking effects:
A study with aggression measures

Cristina Anguiano-Carrasco, Andreu Vigil-Colet and Pere Joan Ferrando
Universidad Rovira i Virgili

Abstract

Background: Several studies have been conducted to better understand what happens with personality scores when faking occurs, but very few use socially undesirable trait measures such as aggression. The aim of the present research is twofold: (a) we aim to apply a General Factorial-Analytic procedure to aggression scales and determine whether it can correct for faking effects; (b) we aim to test the impact that individual differences can have on change scores due to faking. **Method:** Participants were 371 undergraduate students. Of these, 215 answered the questionnaires twice, under neutral conditions and under faking-motivating conditions. 156 were the control group who answered the questionnaires twice, both times under neutral conditions. **Results:** The mean comparison tests as well as the repeated measures ANOVA showed significant results. Individual differences played an important role in all the scales except in physical aggression. **Conclusions:** The results showed that the procedure does correct for faking effects and that individual differences have an important impact on the change scores due to faking, except in the most undesirable Physical aggression measure, which was hardly affected.

Keywords: Physical aggression, verbal aggression, indirect aggression, social desirability, faking.

Resumen

Controlar la deseabilidad social puede atenuar los efectos del falseamiento: un estudio con medidas de agresividad. **Antecedentes:** varios estudios han intentado entender qué sucede con las puntuaciones de las medidas de personalidad cuando se produce falseamiento, pero pocos han utilizado medidas socialmente indeseables como la agresividad. El presente estudio tiene dos objetivos principales: (a) se quiere aplicar el método Analítico Factorial General a las escalas de agresividad para determinar si el método puede corregir los efectos del falseamiento; (b) se quiere comprobar el impacto de las diferencias individuales en las puntuaciones de cambio debido a falseamiento. **Método:** 371 estudiantes universitarios participaron en el estudio. De ellos, 215 respondieron el cuestionario dos veces, bajo condiciones neutras y bajo condición de falseamiento. 156 participantes formaron el grupo control, contestando dos veces el cuestionario bajo condiciones neutras. **Resultados:** tanto las pruebas de comparación de medias como el ANOVA de medidas repetidas obtuvieron resultados significativos. Las diferencias individuales tuvieron un papel importante en todas las escalas excepto en la de agresividad física. **Conclusiones:** los resultados muestran que el método corrige los efectos del falseamiento y que las diferencias individuales tienen un papel importante en las puntuaciones de cambio debido a falseamiento, con la excepción de la medida de agresividad más indeseable, la agresividad física, que casi no se ve afectada.

Palabras clave: agresividad física, agresividad verbal, agresividad indirecta, deseabilidad social, falseamiento.

Although personality questionnaires are widely used, they are far from being a perfect measure of the trait/s they intend to measure. In addition to the problem of estimating the true trait level, another important problem is that some subjects might intentionally distort their answers in order to give a better or worse image, especially when they are under pressure and trying to create a positive or negative impression that they believe will increase their chance of achieving certain goals (e.g., Furnham, 1986; Griffith & Peterson, 2008; McFarland & Ryan, 2000). This behaviour is known as faking.

Social Desirability Scales (SDS) were developed to capture a response style used by respondents that is intended to make them appear more favourable than they really are (Paulhus, 1991). Most SDS measure this tendency by using items that are difficult to endorse in a normative sample (Burns & Christiansen, 2006). Thus, SDS measure the respondent's tendency respond to the socially desirable content of the item instead of its trait content (Kuncel, Borneman, & Kiger, 2012). For this reason, SDS are often used as indicators of faking.

As faking-related measures, some authors (e.g., Furnham, 1986; Eysenck & Eysenck, 1976) have interpreted Social Desirability (SD) scores in two ways. When administered under faking-motivating conditions, these scores are thought to behave as detection measures because they are highly sensitive to faking. When administered under neutral conditions, however, they are thought to measure a substantive personality variable that has

a certain degree of consistency across time and situations (e.g., Furnham, 1986; McFarland & Ryan, 2000).

In many selection and assessment settings, various strategies are used that combine SDS in conjunction with a personality test in order to obtain a more accurate trait estimate under faking-motivating conditions. The earliest technique uses SDS to eliminate candidates with scores over a certain cut-off point. The two main concerns when using this technique are that, although extreme respondents are removed, the participants who are not cannot be said to be free of SD. Furthermore, SD may be related to such traits of interest as conscientiousness or responsibility and, therefore, deleting the candidates with high SD scores usually involves deleting the candidates who have extreme scores on these traits (McCrae & Costa, 1983; Smith & Ellingson, 2002).

Another commonly used method is partialing or correcting questionnaire scores using SDS. In fact, SDS were originally used to remove the effects of faking by regressing the SD scores onto trait scales and computing a residual score (Meehl & Hataway, 1946). Using correction techniques under neutral and selection conditions, Christiansen, Goffin, Johnston, and Rothstein (1994) found that 70% of the sample was affected by corrections based on SDS, and the rank order changed for 85% of candidates. Thus, decisions based on corrected or uncorrected scores would be markedly discrepant. Nevertheless, various studies have shown that correcting or partialing SD decreases validity: i.e., the partialing of variance associated with SDS may remove meaningful variance from the relevant trait and may decrease the validity of the measures (Li & Bagger, 2006; McCrae & Costa, 1983; Ones, Viswesvaran, & Reiss, 1996; Soubelet & Salthouse, 2011). Furthermore, partialing or correcting also assumes that all items are parallel measures of the trait and this is almost never true (Leite & Cooper, 2010).

Recently Ferrando, Lorenzo-Seva and Chico (2009) proposed a general factor-analytic procedure for assessing response bias in questionnaire measures which may be useful in developing a third approach that overrides the limitations of the previous approaches. The procedure has two main steps. The first step identifies a factor related to SD. To this end a set of items related to SD are selected. These items are known as markers. The inter-marker correlation matrix obtained is factor analyzed and the corresponding loading values of each marker on the SD factor are estimated. These loading values are then used to compute the loading values of the content items on the SD factor using an instrumental-variables approach (Hägglund, 1982), and the variance explained by the SD factor is removed from the inter-item correlation matrix. In the second step, the residual inter-item correlation matrix is factor analyzed to identify the content factor or factors of interest which are orthogonal to the SD factor.

The application of this procedure at the item calibration level provides two loading estimates for each item: a loading on the content factor that the test wants to measure, and a loading on an orthogonal factor identified as SD. Thus, SD-free content scores are obtained and there is no need to: (a) assume that items are parallel measurements (which they never are); (b) include SDS in the content scales of interest, which considerably increases the questionnaire's length; or (c) have a non-faked measure for purposes of comparison, which is practically impossible.

Because scores on both the SD marker items and the content items with high loadings on SD are expected to be more prone to change under faking instructions (Furnham, 1986; Eysenck &

Eysenck, 1976), our hypothesis is that under these conditions the SD correction on the content scores will be stronger than under neutral conditions. This is expected to remove (ideally), or at least attenuate, the effects of faking on the content scores.

When faking occurs, it is important to know the extent to which individual differences affect the magnitude of the change in the scores due to similar faking conditions. However, recent reviews (Burns & Christiansen, 2006; Mesmer-Magnus & Viswesvaran, 2006) show that there is very little literature on this issue. What is of most interest is to investigate whether the magnitude and the direction of the change is the same for all subjects or whether the change is specific to every single subject. If all subjects change in exactly the same way, individual differences have no effect on the amount of change, the rank order is not affected by the faking instructions and, therefore, controlling the amount of change would make no difference in selection. On the other hand, if individual differences impact the amount of faking-related change, those subjects who modify their scores in the most appropriate direction will have an unfair advantage over the honest subjects. Ferrando and Anguiano-Carrasco (2011) assessed this issue and found that individual differences have an important impact on the Psychoticism and Neuroticism scales of the Eysenck Personality Questionnaire.

Unlike previous research, the present research uses measures of such highly undesirable behaviours as aggression. Faking is expected to have greater effects on these measures as individuals want to give a good impression. As for the impact of SD, the research generally shows a moderate-to-high relationship between SD and aggression measures. Biaggio (1980) and Selby (1984) reported that most of the correlations between the Buss-Durkee Hostility Inventory scales and the Marlowe-Crowne Social Desirability scale were in the range $r= -.3$ to $-.5$. SD has also been related to measures of violent behaviours and partner abuse (Bell & Naugle, 2007; Devon, Collie, & Walkley, 2004) and those aspects of NEO-PI-R scales most related to aggressive behaviour such as impulsivity and angry hostility (Holden & Passey, 2010). Recently Vigil-Colet, Ruiz-Pàmies, Anguiano-Carrasco and Lorenzo-Seva (2012) used the same method as the one used in the present research in a study based on neutral conditions. Results showed (a) that the items on the aggression questionnaires have moderate-to-high loadings on the SD factor, and (b) that when corrected for this effect, the scores on the aggression scales tended to increase considerably. Conceptually these results suggest that (a) the chosen measures are clearly impacted by SD and (b) the method corrects in the expected direction. Consequently, they are the basis for the present research, which can essentially be considered as an extension of the study by Vigil-Colet et al. (2012) in which the scores are obtained under both neutral and faking-inducing conditions.

The two aggression measures used in Vigil et al. (2012) were: (a) Buss and Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992) which has proved to be useful in assessing various levels and types of direct aggression (e.g., Morales-Vives & Vigil-Colet, 2010), and (b) the Indirect Aggression Scale (IAS; Forrest, Eatough, & Shevlin, 2005). The BPAQ is intended to measure four aggression scales: Physical aggression, Verbal aggression, Anger and Hostility. However, the factorial structure of the BPAQ remains controversial, generally due to the scales intended to measure anger and hostility. As for the IAS, it was included because indirect aggression (see Björkqvist, Osterman, & Kaukiainen, 1992), which has been shown to be the most usual type of aggression in adults, is not considered in the BPAQ.

Overall, the present research used the Physical, Verbal (BPAQ) and Indirect (IAS) Aggression scales. The Anger and Hostility scales of the BPAQ were avoided for two main reasons. On the one hand, the procedure for "cleaning" the content scores of SD uses the residual correlation matrix, so using dimensions which often present unstable solutions would only produce confounding and unsettled results. On the other hand, Anger and Hostility are defined as feelings and cognitions that are strongly related to aggression but they cannot be considered to be aggressive behaviour.

To assess our main hypothesis we proposed a repeated measures design with two factors: condition (neutral vs. faking) and correction (with or without the proposed SD correction). Our hypothesis is that if the proposed correction reduces faking effects, then we will find an interaction between condition and correction in the sense that the content scores are less affected by faking under the correction condition. To assess the second important issue in the present research,—the impact of individual differences on change scores due to faking,—we used the same procedure and statistics as the ones described by Ferrando and Anguiano-Carrasco (2011).

Method

Participants

Participants were 371 undergraduate students from different faculties of the Rovira i Virgili University (Spain). They were randomly assigned in class groups to experimental or control groups. The control group was made up of 156 students and the experimental group of 215 students. The groups were comparable: 85% were women and the mean age was 21 years old in both. The questionnaires were administered in paper and pencil version by the same person in all cases, and completed voluntarily in classroom groups of 25 to 60 students. The administration was anonymous, and the respondents had to provide only three particulars which were used for matching: gender, date of birth and favourite colour.

Procedure

All participants filled in the questionnaires twice. The participants in the control group were asked to respond twice under the standard instructions provided in questionnaires. Among other things, the instructions advise participants to give honest answers. The participants assigned to the experimental group were divided into two subgroups, one of which was first given the faking-motivating instructions and then, on the retest, asked to respond honestly. The other half was first instructed to answer honestly and then, on the retest, given the faking-motivating instructions. The faking-motivating instructions were those listed in Eysenck, Eysenck and Shaw (1974). Respondents are asked to imagine that they are applying for a job that they really want. They should try to give a good impression by answering what they think the employer would like to hear. The re-test interval was six weeks in all cases.

Instruments

The study used the Physical and Verbal Aggression scales (7 and 4 items respectively) of the Spanish short version of the BPAQ (Vigil-Colet, Lorenzo-Seva, Codorniu-Raga, & Morales, 2005) as well as the Indirect Aggression Scale (IAS), in the Spanish short version (10 items) by Anguiano-Carrasco and Vigil-Colet (2011). Overall, given that the procedures in Ferrando et al. (2009) provide content and SD scores for each measure, the analyses that follow are based on five sets of individual scores: Physical Aggression, Verbal Aggression, Indirect Aggression, BPAQ SD and IAS SD.

Data analysis

Only the experimental group was used to assess the first issue. To assess the second issue we used a structural equation model (SEM) in which the amount of individual change is estimated on the basis of a bidimensional invariant model (Ferrando & Anguiano-Carrasco, 2011). Both control and experimental groups were used in the second assessment.
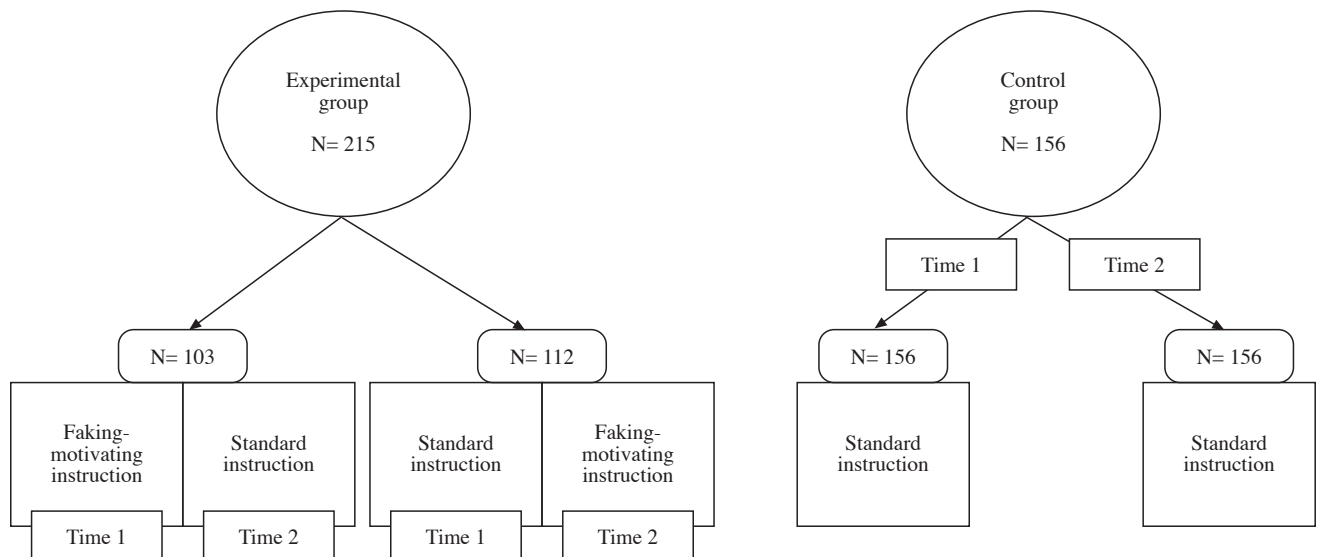

Figure 1. *Experimental design: groups and instructions given*

To assess the first issue, we obtained the effects of the SD corrections under neutral and faking conditions, the content and the SD factor scores using the following procedure. First, we used separate factor analyses and checked that the loading estimates obtained under the neutral and faking conditions were essentially invariant (they were). Second, common estimates were obtained by averaging the loadings obtained in both conditions. Third, Bartlett factor scores (see, e.g., Ferrando, 2007) were obtained based on these common estimates. In the "corrected" conditions, the factor-score estimates were based on the bidimensional (content and SD) solution of Ferrando et al. In the non-corrected conditions they were based on the unidimensional solution. The scores were standardized in the complete dataset, so that results were comparable when the dataset was split into the different conditions.

Results

Table 1 shows the mean scores on the T scale for the four conditions for each of the three scales used in the study and for SD. The table shows that all the scales corrected by the procedure have higher means than the uncorrected ones, as expected given that higher scores imply higher levels of aggression. The neutral scores are also higher than the faked ones for each content scale. The SD scores were expected to be sensitive to faking, and higher in the faking condition. We found that the scores for the two SD measures (SD computed on BPAQ and IAS) were significantly greater ($t= 14.53$, $p= 0.01$; $t= 13.27$, $p= 0.01$; respectively) with effect sizes of $d= 1.17$ and $d= 1.06$. According to Cohen's criteria (1969, p. 23), these may be considered to be large.

Table 2 shows the results of the two by two (corrected vs. uncorrected, neutral vs. faked) factor repeated measures analysis of variance for each scale. Both factors and their interaction showed significant effects on all scales, so the score changes related to faking depended upon the presence or absence of SD correction. The partial Eta squared statistic, is also shown. Figure 2, shows the interaction effects on each scale. As can be seen, the differences between the scores under faking and neutral conditions are always smaller for corrected scores, and are even non-significant in the case of physical aggression ($t= -1.56$, $p= 0.119$). For verbal and indirect aggression, on the other hand, there is a reduction in faking effects but the difference between both conditions is still significant ($t= 7.76$, $p= 0.01$; $t= 6.90$, $p= 0.01$, respectively).

Table 3 shows the correlations between the increments in SD and in the aggression scale scores when subtracting the T scores under faking conditions from the T scores under neutral conditions. All the correlations except those with the Indirect Aggression Scale were significant, showing that in direct aggression measures the change in SD items is related to the change in content measures.

In order to test the hypothesis that individual differences play an important role in scale change scores due to faking, the three scales were compared by fitting them on the bidimensional invariant and non-invariant models. The invariant model indicates that the factor under neutral conditions has exactly the same structure, factor loadings and thresholds as under faking-inducing conditions so the scores obtained under neutral and faking conditions can be compared. Table 4 shows that the fit of the invariant model was

Table 2
Univariate contrast. F statistic, its significance level and partial eta squared for principal factors and their interaction for each scale

		F	*P*	η^2_p
	Answer condition	31.57	0.00	0.06
Physical aggression	**Correction condition**	5452.54	0.00	0.80
	Interaction C × C	125.63	0.00	0.45
	Answer condition	112.02	0.00	0.32
Verbal aggression	**Correction condition**	580.27	0.00	0.88
	Interaction C × C	187.88	0.00	0.47
	Answer condition	191.40	0.00	0.25
Indirect aggression	**Correction condition**	245.96	0.00	0.94
	Interaction C × C	80.65	0.00	0.32

Table 1
Mean T scores for each scale on each condition

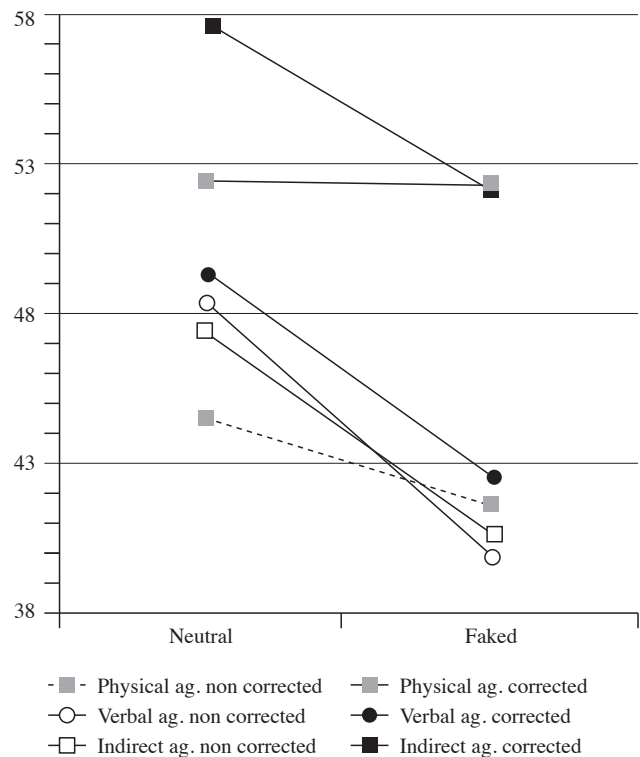|	**Non-Corrected**		**Corrected**	
	Neutral *M (SD)*	**Faked** *M (SD)*	**Neutral** *M (SD)*	**Faked** *M (SD)*
Physical aggression	44.54 (5.52)	41.56 (3.47)	52.43 (5.35)	51.79 (3.40)
Verbal aggression	48.38 (8.72)	39.88 (8.56)	49.32 (8.64)	42.54 (8.43)
Indirect aggression	47.47 (9.96)	40.60 (6.56)	57.62 (9.07)	52.07 (5.91)
SD on BPAQ			66.23 (6.04)	75.43 (5.62)
SD on IAS			67.62 (7.57)	76.10 (5.46)

Figure 2. Interaction effects on the aggression scales

acceptable for all the scales and not substantially worse than the fit of the less restrictive non-invariant model. We therefore consider the invariant model to be acceptable.

We next estimated the amount of relative variance to assess the impact of individual differences (measured by Cohen's d) on the amount of change caused by faking-inducing instructions for each scale. Those results showed that the most impacted measure was Indirect Aggression ($d = 2.95$), followed by Verbal Aggression ($d= 2.24$) and then Physical Aggression ($d= 0.65$).

Discussion

The aim of the study was twofold. First, it aimed to confirm the hypothesis that the procedure proposed by Ferrando et al. (2009) reduces the faking effect, and that the SD factor obtained is highly affected by faking-induced change and may be useful for correcting the scores on the scales that change the most under faking inducing instruction. Second, we assessed the impact that individual differences have on the change scores due to faking on the aggression measures. We were particularly interested in determining whether the scales that are most impacted by individual differences are also the ones in which increments in SD do not correlate with the increments in the scale scores.

The results suggests that, although far from being perfect, the procedure is useful for 'cleaning' the scores and attenuating the effects of faking-inducing instructions because it has a differential effect on the increments caused by faking on the aggression measures. Cohen's d indicates that the SD is very sensitive to faking-inducing instructions and it has a big effect on both questionnaires. The correlations show that in direct aggression measures the amount of change due to faking is related to the increments in the SD scores that the procedure provides.

The results also seem to indicate that SD factor increments are clearly related to increments due to faking-inducing instructions

on the aggression scales with the exception of Indirect Aggression. This result could be explained by the fact that indirect aggression is the most acceptable, socialized type of aggression, so when subjects fake, they do not consistently change their scores on Indirect Aggression in the same direction or magnitude. This conjecture is supported by the result that individual differences have the biggest impact on this scale.

Individual differences explain quite a large amount of the total variance in verbal and indirect aggression, but not so much of the variance in physical aggression. However, although it is clear that individual differences have less impact on physical aggression, according to Cohen's criterion their effect would still be medium. We should point out here that the Physical Aggression scale showed the smallest overall variance. In our opinion, the fact that the overall variance is small is one reason why the relative importance of the individual differences variance appears to have a medium effect size although the direct measure is not very big. Therefore, we consider here that individual differences have a very small, almost negligible, impact.

Physical aggression is considered to be the predominant type of aggression in children but it progressively decreases during the socialization process. Verbal and indirect aggression become more important and peak during adolescence and adulthood (Vaillancourt, 2005; Tremblay & Nagin, 2005). It is, therefore, reasonable to suggest that physical aggression is the most socially undesirable behaviour of all the aggression types assessed in the present research. Taking into account everything explained above, we conjecture that the impact of individual differences on highly undesirable behaviours is negligible in terms of rank order: that is, all the subjects increase or decrease their scores by the about same magnitude and in the same direction. Therefore, rank order in a possible personnel selection, or any situation in which the extreme scoring subjects are to be selected, would not be affected by faking on these types of measure. As can be seen, physical aggression is practically not impacted by individual differences in faking change scores but verbal and indirect aggression, which are more acceptable aggression behaviours, are.

Consequently, it would be of interest to measure how different personality traits are affected by individual differences in faking. If the results obtained here are generalizable to other behaviours considered to be extremely undesirable, the decisions based upon individuals' scores may be correct even though they may be affected by faking.

No study is free of limitations and the present one is no exception. On the one hand, our participants were university

Table 3
Correlations between SD Score Increment and the increments on each aggression scale

	SD
Physical aggression	0.36**
Verbal aggression	0.33**
Indirect aggression	-0.11

Note: ** significance level .01

Table 4
Goodness-of-fit statistics for invariant and non-invariant models of physical aggression, verbal aggression, and indirect aggression

	χ^2		df		CFI		RMSEA	
	Non-invariant	Invariant	Non-invariant	Invariant	Non-invariant	Invariant	Non-invariant	Invariant
Physical aggression control	64.48	99.80	47	57	0.96	0.92	0.04	0.06
Physical aggression faked	71.81	85.89	47	57	0.93	0.92	0.05	0.05
Verbal aggression control	9.71	19.83	5	10	0.97	0.95	0.07	0.07
Verbal aggression faked	6.41	12.93	5	10	0.99	0.98	0.03	0.03
Indirect aggression control	151.74	196.11	125	142	0.96	0.93	0.03	0.05
Indirect aggression faked	174.30	237.27	125	142	0.95	0.92	0.04	0.05

students instructed to fake, not real job applicants or patients. It would be desirable to compare the results obtained here with results from samples of real job applicants or patients. On the other hand, in order to consolidate the procedure and generalize its use it would be of interest to replicate the results of this research on such trait scales as Conscientiousness or Integrity, which have proved to be closely related to SD (McFarland & Ryan, 2000; Muller-Hanson, Heggestad, & Thornton, 2006; Griffith, Malm, English, Yoshita, & Gujar, 2006).

In conclusion, the factor analytic procedure proposed by Ferrando et al. appears to be an important tool for controlling the effect that faking has on personality scale scores. The procedure only needs the four selected markers to be added to the scale of interest and to be administered once. The test, then, is not excessively longer, and there is no need for initial scores to be neutral, which is by no means easy to achieve in such contexts as clinical assessments or personnel selection procedures.

Acknowledgements

References

Anguiano-Carrasco, C., & Vigil-Colet, A. (2011). Assessing indirect aggression in aggressors and targets: Spanish adaptation of Indirect Aggression Scales. *Psicothema, 23*, 146-152.

Bell, K.M., & Naugle, A.E. (2007). Effects of social desirability on students' self-reporting of partner abuse perpetration and victimization. *Violence and Victims, 22*, 243-256.

Biaggio, M.K. (1980). Assessment of anger arousal. *Journal of Personality Assessment, 44*, 289-298.

Björkqvist, K., Osterman, K., & Kaukiainen A. (1992). The development of direct and indirect strategies in males and females. In K. Bjorkqvist & P. Niemela (Eds.), *Of mice and women: Aspects of female aggression* (pp. 51-64). San Diego, CA: Academic Press.

Burns, G.N., & Christiansen, N.D. (2006). Sensitive or senseless: On the use of social desirability measures in selection and assessment. In R.L. Griffith & M.H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 113-148). Greenwich, CT: Information Age Publishing.

Buss, A.H., & Perry, M.P. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.

Christiansen, N.D., Goffin, R.D., Johnston, N.G., & Rothstein, M.G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847-860.

Devon, L.L., Collie, R.M., & Walkley, F.H. (2004). Criminal attitudes to violence: Development and preliminary validation of a scale for male prisoners. *Aggressive Behavior, 30*, 484-503.

Eysenck, H.J., & Eysenck, S.B.G. (1976). *Psychoticism as a dimension of personality*. New York: Crane-Russak.

Eysenck, S.B.G., Eysenck, H.J., & Shaw, L. (1974). The modification of personality and Lie scores by special 'honestly' instructions. *British Journal of Social and Clinical Psychology, 13*, 41-50.

Ferrando, P.J. (2007). Factor-analytic procedures for assessing response pattern scalability. *Multivariate Behavioral Research, 42*, 481-508.

Ferrando, P.J., & Anguiano-Carrasco, C. (2011). A structural equation model at the individual and group level for assessing faking-related change. *Structural Equating Modeling, 18,* 91-109.

Ferrando, P.J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling, 16*, 364-381.

Forrest, S., Eatough, V., & Shevlin, M. (2005). Measuring adult indirect aggression: The development and psychometric assessment of the Indirect Aggression Scales. *Aggressive Behavior, 31*, 84-97.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*, 385-400.

Griffith, R.L., & Peterson, M.H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology, 1*, 308-311.

Griffith, R.L., Malm, T., English, A., Yoshita, Y., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive bias, and individual differences. In R.L. Griffith & M.H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 151-178). Greenwich, CT: Information Age Publishing.

Hägglund, G. (1982). Factor analysis by instrumental variable methods. *Psychometrika, 47*, 209-222.

Holden, R.R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49*, 446-450.

Kuncel, N.R., Borneman, M., & Kiger, T. (2012). Inovative item response process and bayesian faking detection methods. In M. Ziegler, C. MacCann, & R.D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 72-84). New York: Oxford University Press.

Leite, W.L., & Cooper, L.A.(2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioral Research, 45*, 271-293.

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection & Assessment, 14*, 131-141.

McCrae, R.R., & Costa, P.T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.

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

McFarland, L.A., & Ryan, A.M. (2006). Toward and integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*, 979-1016.

Meehl, P.E., & Hathaway, S.R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30*, 526-564.

Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review of research designs and analytic strategies. In R.L. Griffith & M.H. Peterson (Eds.), *A closer examination of applicant faking behaviour* (pp. 85-113). Greenwich, CT: Information Age Publishing.

Morales-Vives, F., & Vigil-Colet, A. (2010). Are there sex differences in physical aggression in the elderly? *Personality and Individual Differences, 49*, 659-662.

Mueller-Hanson, R.A., Heggestad, E.D., & Thorton, G.C. (2006). Individual differences in impression management: An exploration of the psychological process underlying faking. *Psychology Science, 48*, 288-312.

Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679.

Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Selby, M.J. (1984). Assessment of violence potential using measures of anger, hostility, and social desirability. *Journal of Personality Assessment, 48*, 531-544.

Smith, D.B., & Ellingson, J.E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*, 211-219.

Soubelet, A., & Salthouse, T.A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality, 79*, 741-762.

Tremblay, R.E., & Nagin, D.S. (2005). The developmental origins of physical aggression in humans. In R.E. Tremblay, W.W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 83-106). New York: Guilford Press.

Vaillancourt, T. (2005). Indirect aggression among humans: Social construct or evolutionary adaptation? In R.E. Tremblay, W.W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 158-177). New York: Guilford Press.

Vigil-Colet, A., Lorenzo-Seva, U., Codorniu-Raga, M.J., & Morales, F. (2005). Factor structure of the Aggression Questionnaire among different samples and languages. *Aggressive Behavior, 31*, 601-608.

Vigil-Colet, A., Ruiz-Pàmies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema, 24*, 310-315.