

# Vividness of visual imagery and social desirable responding: correlations of the vividness of visual imagery questionnaire with the balanced inventory of desirable responding and the Marlowe–Crowne scale

John Allbutt · Jonathan Ling · Martin Rowley ·  
Mohammed Shafiullah

Published online: 15 April 2011  
© Psychonomic Society, Inc. 2011

**Abstract** Correlational research investigating the relationship between scores on self-report imagery questionnaires and measures of social desirable responding has shown only a weak association. However, researchers have argued that this research may have underestimated the size of the relationship because it relied primarily on the Marlowe–Crowne scale (MC; Crowne & Marlowe, *Journal of Consulting Psychology*, 24, 349–354, 1960), which loads primarily on the least relevant form of social desirable responding for this particular context, the moralistic bias. Here we report the analysis of data correlating the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, *Journal of Mental Imagery*, 19, 153–166, 1973) with the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002) and the MC scale under anonymous testing conditions. The VVIQ correlated significantly with the Self-Deceptive Enhancement (SDE) and Agency Management (AM) BIDR subscales and with the

MC. The largest correlation was with SDE. The ability of SDE to predict VVIQ scores was not significantly enhanced by adding either AM or MC. Correlations between the VVIQ and BIDR egoistic scales were larger when the BIDR was continuously rather than dichotomously scored. This analysis indicates that the relationship between self-reported imagery and social desirable responding is likely to be stronger than previously thought.

**Keywords** Imagery · Questionnaires · Social desirable responding

Mental images are quasi-perceptual mental representations that can occur in all sensory modalities. Visual imagery has received the most attention from psychologists and refers to the experience of mentally visualising the appearance of something, usually without it being present. Individual differences in imagery experience, such as its vividness, are often reported by individuals. Psychologists have commonly assessed these differences using subjective self-report imagery scales that request participants to form mental images and then rate some aspect of their imagery experience. Self-report imagery questionnaires have appeal because they are quick and easy to use and, for some aspects of imagery experience, such as image vividness, appear to be the only measurement option available.

An important concern over the results obtained from self-report imagery questionnaires has been the extent to which they might be influenced by social desirable responding. Paulhus (2002, p.50) defines social desirable responding as “...the tendency to give overly positive self-descriptions.” There has been much debate over how best to conceptualise social desirable responding. Domino and Domino (2006) stated that there are differing opinions over

---

J. Allbutt (✉)  
Directorate of Psychology,  
Counselling and Public Health,  
University of Salford,  
Salford, Greater Manchester, UK  
e-mail: j.d.allbutt@salford.ac.uk

J. Ling  
Department of Pharmacy, Health and Wellbeing,  
University of Sunderland,  
Sunderland, England, UK

M. Rowley  
School of Psychology, Keele University,  
Keele, England, UK

M. Shafiullah  
Psychology Division, De Montfort University,  
Leicester, England, UK

whether such responding should be thought of as distortion or error that should be eliminated or minimised, or whether it is best thought of as a personality trait related to other positive traits, such as psychological adjustment or conscientiousness, and should be treated as substantive or valid responding. Leite and Cooper (2010) argued that there is evidence (Li & Bagger, 2006; Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkaslo, 2007; Ones, Viswesvaran, & Reiss, 1996) that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the nature of the items making up the focal scale, the respondent's characteristics, and the situation that the testing takes place in.

Within the mental imagery literature, social desirable responding has been conceptualised as distortion/error. McKelvie (1995b) drew on considerations such as factors that affect reliability and validity to suggest that correlations in excess of .25 between imagery questionnaires and measures of social desirable responding would indicate unacceptable "contamination." In a meta-analytic review article, McKelvie (1995a) estimated the size of correlation between the most widely used imagery scale, the VVIQ (Vividness of Visual Imagery Questionnaire, Version 1; Marks, 1973), and social desirable responding. McKelvie made two estimates of the size of this relationship. The analysis that produced the largest estimate drew only on data from the most widely used test of social desirable responding, the Marlowe–Crowne scale (MC; Crowne & Marlowe, 1960). The analysis estimated the size of this relationship to have a mean  $r$  of .189, with 95% confidence intervals of .078 and .296, respectively, based on data from 10 studies. McKelvie concluded that the VVIQ was not

seriously, if at all, affected by social desirable responding. However, in previous work (e.g., Allbutt, Ling, Hefferman, & Shafiullah, 2008; Allbutt, Ling, & Shafiullah, 2006; Allbutt, Shafiullah, & Ling, 2006), we have argued that McKelvie's (1995a) estimate of the size of the relationship between the VVIQ and social desirable responding may have been limited because previous research (1) had theorised social desirable responding to be a unitary construct and (2) had predominantly used the MC scale as its measure of social desirable responding, which may not be the most appropriate measure to use with self-report imagery tests.

Early research into social desirable responding conceptualised the phenomenon as a unitary construct. However, observations of low intercorrelations between different measures and the results of factor analysis led several researchers to argue that social desirable responding could occur in several forms (e.g., Edwards, Diers, & Walker, 1962). Work by Paulhus (e.g., 1984) has been particularly influential and has led to the development of a "two-tier model" of the different forms of social desirable responding, as well as a questionnaire to measure them called the Balanced Inventory of Desirable Responding, Version 8 (BIDR; Paulhus, 2002; see Fig. 1).

In the first tier of the model, social desirable responding is divided into egoistic and moralistic biases. *Egoistic bias* refers to the tendency to claim positive social and intellectual qualities, while *moralistic bias* is the claiming of positive moral qualities and the refutation of negative, socially deviant qualities. In the second tier, both biases are split into unconscious and conscious forms. The egoistic bias becomes self-deceptive enhancement (SDE) and agency management (AM), while the moralistic bias

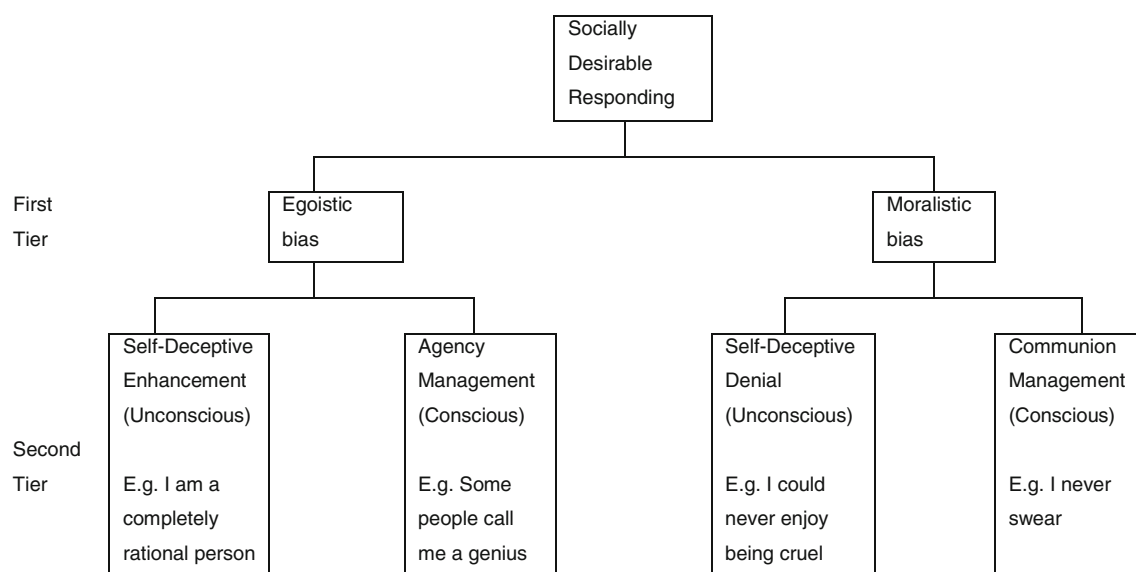


Fig. 1 Paulhus's two-tier model (based on Paulhus, 2002)

becomes self-deceptive denial (SDD) and communion management. The four scales of the BIDR map straightforwardly onto the model, apart from the fact that in the BIDR communion management is referred to by a different term, *impression management* (IM, the term we will use throughout). The nature of the differentiation of unconscious from conscious levels differs between the two kinds of biases. For the egoistic bias, SDE scale items differ from AM items in the degree of exaggeration made about ability claims, such that AM claims are very extreme. In contrast, for the moralistic bias, SDD scale items cover more sensitive and emotionally charged themes than IM items. These include the admission of negative emotions such as anger, jealousy, or sexual feelings and have a defensive tone similar to that of psychoanalytic denial, while IM items cover compliance with norms relating to socially acceptable behaviour, such as not telling lies or not covering up one's mistakes.

Paulhus (2002) views SDE and SDD as best suited to measuring response styles, long lasting and consistent biases across time and questionnaires, while AM and IM are best suited to measuring responses to instructional sets, which are more short-lived biases caused by temporary factors such as the desire to impress a particular audience. Paulhus's conceptualisation of social desirable responding, and consequently the BIDR, has evolved over time. Early versions of the BIDR only included the SDE and IM scales, and the addition of the AM scale has been the most recent development to it. Research has supported the validity of Paulhus's two-tier model (e.g., John & Paulhus, 2000).

It is possible to conceptualise Paulhus's model in light of the most recent thinking on social desirable responding—that is, that social desirable responding is best thought of as a three-way interaction (between the nature of the items making up the focal scale, the respondent's characteristics, and the situation in which the testing takes place), and that it might be distortion/error, substantive, or a combination of both. In this view, correlations between a focal scale and either of the measures of unconscious bias (SDE or SDD) would equate to three-way interactions, but with respondent characteristics having greater weight in the interaction than situational factors. In contrast, correlations between either of the measures of conscious bias (AM or IM) would also equate to three-way interactions, but with situational factors having equal or greater weight in the interaction than respondent characteristics. "Respondent characteristics" correspond to the expression of traits that lead to distortion/error, substantive responding, or a combination of both. Situational influences lead to distortion/error. Research by Lönnqvist et al. (2007) on Finnish versions of the SDE and IM scales offers some support for this argument. They found that neither scale was a "pure" measure and that scores on both were affected by "style," "set," and substantive individual differences, but to differing degrees.

When published in 1960, the MC scale was intended by its authors to be a measure of social desirable responding in self-reports. However, Crowne and Marlowe (1964) later suggested that the scale taps a more general motive of a person's need for approval from others, which Crowne (1979) refined further as a drive to avoid disapproval from others. Paulhus (1991, p.29) notes that the scale sustains "a dual existence as a social desirable responding scale and a measure of the approval-dependent personality." McKelvie's (1995a) conclusion that the VVIQ was not seriously affected by social desirable responding was derived mainly from data involving the MC scale.

A number of specific weaknesses have been identified in the MC scale. Critics have suggested that some of the items on the scale are out of date (Stöber, 2001); that the scale may confound the type of item, attribution of positive attributes and denial of negative attributes, with direction of item scoring (Helmes, 2000); and that confirmatory factor analysis has not supported either single- or two-factor models of the scale (Leite & Beretvas, 2005). Further, the MC scale was originally developed on the assumption that social desirable responding is a unitary construct—a problematic assumption, given evidence suggesting that social desirable responding occurs in more than one form.

There is also evidence that McKelvie (1995b) underestimated the size of the relationship between the VVIQ and social desirable responding, because the MC scale is predominantly a measure of moralistic bias. We would argue that respondents answering self-report imagery questionnaires will most likely view imagery as an ability being tested rather than as a "value neutral" aspect of their cognition, and therefore will value responses they perceive to indicate "better" imagery ability. This would be particularly true when data are collected in settings that may emphasise ability and competition, such as data collected from groups of students in academic settings in the presence of a tutor. Given this, the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, would seem to be more likely to correlate with self-report imagery questionnaires than the moralistic bias. Paulhus (1984) factor-analysed scores from several measures of social desirable responding and found that the MC scale loads .68 on a factor labelled "impression management" and .40 on a factor labelled "self-deception." Interpreted in terms of Paulhus's (2002) two-tier model, it is not possible to know for sure whether the MC loaded more highly on the Impression Management factor because it is predominantly a measure of conscious forms of social desirable responding, because of the moralistic bias, or because of a mixture of both, since the factors differed in terms of content (moralistic bias vs. egoistic bias) and level of consciousness (conscious vs. unconscious). However, from studying the items of the MC and noting that they do

not appear to tap into the emotionally charged themes characteristic of SDD, we argue that it is most likely that the key distinguishing feature between the factors is content and that the MC scale is predominantly a measure of moralistic bias.

Previous work investigating the relationship between a variety of self-report imagery scales and social desirable responding (Allbutt et al., 2008; Allbutt et al., 2006; Allbutt, Shafiullah, & Ling, 2006) supports our critique of the reliance on data from the MC scale. In this research, the SDE and IM scales of the BIDR have been used as measures of social desirable responding. The SDE and IM scales were selected because they are the best established of the BIDR scales and because they include one measure of egoistic bias and one measure of moralistic bias. Together, these studies found the following: (1) “greater” imagery ability was always associated with higher social desirable responding; (2) correlations with self-report imagery scales and SDE were always higher than those with IM; (3) the magnitude of correlations between imagery scales and SDE sometimes exceeded McKelvie’s (1995b) .25 criterion; (4) significant correlations between imagery properties and SDE were seen for scales measuring vividness of visual imagery and for some other visual properties of images, such as ease of image generation and ease of image maintenance, but not for measures of vividness of visual thinking style and auditory imagery; (5) the size of imagery scale–SDE correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which the imagery property was valued; and (6) the form of BIDR subscale scoring did not appear to greatly affect the pattern of results obtained with the SDE and IM scales.

In the present study, participants completed the VVIQ, all four scales of the BIDR, and the MC. The study had three main aims, which extend previous work:

1. A direct comparison of how the VVIQ correlates with the scales of the BIDR and the MC scale in a single sample of participants. Earlier work argued that McKelvie (1995a) underestimated the degree of relationship between the VVIQ and social desirable responding because of a reliance on data from the MC scale, which may primarily be a measure of moralistic bias. However, we have not directly tested the validity of our argument using the MC scale in our previous work. If our critique is correct, we would expect the VVIQ to correlate more highly with BIDR egoistic scales than with the MC scale, and for the MC scale to correlate more highly with BIDR moralistic scales than egoistic scales.
2. To investigate whether the VVIQ correlates significantly and independently with both BIDR egoistic scales. Our

previous work has shown a consistent relationship between the VVIQ and SDE, but we have not yet tested for a relationship between the VVIQ and AM. We would normally expect anonymous testing conditions to generate minimal social desirable situational pressures. However, data collected from groups of students as part of an experiment carried out in an academic setting and with a tutor present acting as the experimenter potentially could create situational demands to respond consciously in a socially desirable way. If so, AM might correlate with self-report imagery questionnaires independently from SDE, and the relationship with socially desirable responding seen under our testing conditions could be greater than the estimate made from our previous studies, which were based on SDE alone.

3. Finally, investigate the effect of the BIDR scoring system on the pattern of results obtained. There are two ways to score the BIDR subscales, continuously or dichotomously. After recoding of negative items, in the continuous approach the scores are summed to create a total for each subscale, while in the dichotomous approach, recoded extreme scores (either six or seven on an item), are counted as one point and are added to create a total for each subscale. Paulhus (1994) favours the dichotomous approach to scoring, arguing that it has the advantage of only scoring clearly exaggerated or managed high responses. Other authors, however, have challenged this assertion on the basis of findings with the SDE and IM subscales, in which better performance has been observed using continuous scoring (higher Cronbach’s alphas and higher convergence correlations; Stöber, Dette, & Musch, 2002), and better fit for the data and their student sample (Cervellione, Lee, & Bonanno, 2009). Both Stöber et al. and Cervellione et al. suggested that one reason why performance may be superior with continuous scoring is that, unlike dichotomous scoring, it can tap into exaggerated low responses as well as exaggerated high responses. Allbutt et al. (2008) tested the effect of scoring method for the SDE and IM subscales for three self-report imagery subscales and found little difference in the patterns of results obtained; however, the AM and SDD subscales remain to be tested in this way.

## Method

### Participants

Our participants were 123 undergraduate psychology students (23 males, 100 females). Their ages ranged from 18 to 54 years, with a mean age of 19.6 years. Students

received course credit for participation, although they had the option to choose which studies they took part in.

## Materials

*Vividness of visual imagery scale (VVIQ; Marks, 1973)* The VVIQ assesses the vividness of visual imagery using 16 items. Participants form a series of images, of a friend or relative's face, the rising sun, a shop, and a country scene. Ratings of image vividness are made on a five-point scale, ranging from 1, *Perfectly clear and vivid as normal vision*, to 5, *No image at all, you only "know" you are thinking of the object*. Traditionally, participants make 16 ratings with their eyes open, then 16 ratings with their eyes closed, and the two sets of scores are added together. However, McKelvie's (1995a) meta-analytical review found no difference in the ratings made with eyes open and eyes closed. Thus, mention of imaging with eyes open was edited out of the VVIQ instructions, and the participants were asked only to complete the questionnaire once with their eyes closed. Lower total scores indicate more vivid imagery. McKelvie's (1995a) meta-analytical review estimated internal consistency, as measured by Cronbach's alpha, to be .88. Cronbach's alpha from the current study was .82.

*The balanced inventory of desirable responding, Version 8 (BIDR; Paulhus, 2002)* The BIDR has four 20-item subscales (SDE, AM, SDD, and IM) that assess four kinds of social desirable responding. All scales are made up of alternating positively and negatively phrased statements, 10 of each type. The responses are made on a seven-point scale ranging from 1 (*Not true*) to 7 (*Very true*). Both egoistic scales have statements about the respondent's abilities across a range of situations. The four subscales were scored separately, and all were scored both continuously and dichotomously. Paulhus (1994) recommends omitting items from the BIDR scales if they are not appropriate or cause confusion for a particular sample—for example, where an item refers to being a car driver, yet most participants in a sample are likely not to own a car. Because of this, SDE Items 8 and 14 and IM Item 13 were omitted. Paulhus (1999) also noted that some of the items of the SDD scale make its use ethically problematic and may be offensive to participants—for example, "I have never felt like I wanted to kill someone." Because of this concern, SDD Items 6, 12, and 13 were also omitted. Paulhus (1994) reported Cronbach's alphas for SDE scored dichotomously to range from .65 to .75, and alphas for IM scored dichotomously to range from .75 to .80. We are not aware of any currently published reliability data for the AM or SDD scales. In the present study, Cronbach's alphas for

the SDE, AM, SDD, and IM subscales, scored continuously, were .75, .82, .75, and .72, and scored dichotomously, were .70, .69, .64, and .60.

*The Marlowe–Crowne scale (MC; Crowne & Marlowe, 1960)* This scale assesses social desirable responding, specifically the "need" to avoid the disapproval of others. It consists of 33 statements to which respondents reply either "true" or "false." Of the statements, 18 are positively phrased and 15 negatively phrased, and they are randomly ordered across the questionnaire. Positively-phrased-item responses of "true" score one point, and negatively-phrased-item responses of "false" score one point. Higher total scores indicate greater social desirable responding/need for approval. Item 27, "I never make a long trip without checking the safety of my car," was omitted for reasons similar to those described above for BIDR item omission. Paulhus (1991) reported Cronbach's alphas to range from .73 to .88. Cronbach's alpha from the present study was .77.

*Ethical rights form* A form was used that notified participants of the rights they possessed under the British Psychological Society's code of ethics.

The ethics form and three questionnaires were combined together into a booklet, starting with the ethics form, followed by the VVIQ, then the BIDR and ending with the MC.

## Procedure

Participants answered the questionnaires in groups of about 20, with a tutor present who acted as the experimenter. They were asked to seat themselves as far apart as possible in the room. Responses were made in silence, and participants were told not to put their names anywhere in the booklet. Participants answered the questionnaires in the order in which they appeared in the booklet.

## Results

Descriptive statistics were calculated for all variables and can be seen in Table 1. The VVIQ, all BIDR subscales scored dichotomously, and the MC scale were all found to be mildly positively skewed, so the median and interquartile range (IQR) are presented in Table 1 as well as the mean and standard deviation (*SD*). Note that throughout all of the results tables, continuous scoring of BIDR scales is indicated by the suffix "c" and dichotomous scoring by the suffix "d."

One feature of note shown by the descriptive statistics is the effect that the scoring system had on the spread of AM scores. When scored continuously, these scores showed the

**Table 1** Descriptive statistics for scales

Variable	Number of Scale Items	Mean	SD	Median	IQR	Max	Min
VVIQ	16	38.52	10.13	38	15	18	66
SDEc	18	66.68	13.13	65	16	107	40
SDEd	18	3.37	2.79	3	4	13	0
AMc	20	59.87	15.16	59	19	93	30
AMd	20	2.19	2.30	1	3	9	0
SDDc	17	59.15	13.92	59	18	107	27
SDDd	17	3.46	2.53	3	3	13	0
IMc	19	68.29	14.26	67	17	112	22
IMd	19	4.81	2.65	4	3	14	0
MC	32	12.72	5.01	12	6	26	2

$N = 123$ . VVIQ, Vividness of Visual Imagery Questionnaire; SDE, Balanced Inventory of Desirable Responding (BIDR) Self-Deceptive Enhancement subscale; AM, BIDR Agency Management subscale; SDD, BIDR Self-Deceptive Denial subscale; IM, BIDR Impression Management subscale; MC, Marlowe–Crowne scale. The letters “c” and “d” after the BIDR subscales indicate continuous and dichotomous scoring, respectively

largest spread in all of the BIDR subscales; however, when scored dichotomously, they showed the smallest spread.

A series of Pearson product–moment correlations between the VVIQ, the BIDR subscales (scored both continuously and dichotomously), and the MC were carried out. Prior to carrying out the correlations, the positive skews seen in the VVIQ, the BIDR subscales scored dichotomously, and the MC were corrected with square root transformations to make their distributions normal. Four multivariate outliers were removed, reducing the sample size to 119. The results of the correlations are shown in Table 2. The reader should note that a large number of statistical tests have been carried out, and as such, some correlations may be due only to chance. However, the facts that most significant correlations observed indicated large effect sizes, the pattern of correlations is meaningful when viewed in terms of Paulhus’s (2002) model, and the correlations are generally consistent with the results seen in our previous three studies argue against the effects having been caused by chance alone.

The VVIQ showed significant low negative correlations (reflecting the opposite keying of the VVIQ to the other

scales) with some of the social desirability scales, indicating that more vivid imagery was associated with greater social desirable responding. VVIQ correlations were typically largest with SDE and when BIDR scales were continuously scored. The VVIQ correlated significantly with both SDE and AM when they were scored continuously, and with SDE only when scored dichotomously. The VVIQ also correlated with the MC, but to a lesser extent than with the BIDR egoistic scales (apart from the AM dichotomously scored), and showed no significant correlations with the BIDR moralistic subscales.

The pattern of correlations between the BIDR social desirability scales in general supported the two-tier model, though with continuous scoring, the SDE subscale did not clearly correlate to a higher level with AM than with the BIDR moralistic subscales. The SDD and IM subscales were more highly correlated than were SDE and AM. Correlations between the same BIDR subscales when scored continuously and dichotomously were high, ranging from .74 for AM to .86 for IM. The MC correlated significantly with SDE, SDD, and IM, but notably more highly with the BIDR moralistic scales

**Table 2** Matrix for Pearson’s correlations between measures

	VVIQ	SDEc	SDEd	AMc	AMd	SDDc	SDDd	IMc	IMd
VVIQ									
SDEc	-.35**								
SDEd	-.25**	.76**							
AMc	-.27**	.34**	.29**						
AMd	-.14	.25**	.34**	.74**					
SDDc	-.05	.31**	.13	-.09	-.15				
SDDd	-.03	.20*	.13	-.17	-.13	.84**			
IMc	-.11	.38**	.21*	.15	.04	.66**	.51**		
IMd	-.05	.29**	.17	.07	.03	.60**	.53**	.86**	
MC	-.18*	.38**	.30**	.15	.12	.66**	.58**	.68**	.62**

\*  $p < .05$ . \*\*  $p < .01$  (two-tailed)

than with SDE, and only marginally more highly with IM than with SDD. Intercorrelations between BIDR subscales tended to be larger with continuous scoring, and BIDR subscale correlations with the MC also tended to be slightly larger with continuous scoring.

While SDE, AM, and the MC all correlated significantly with the VVIQ, considerations of Paulhus's (2002) model together with the pattern of correlations observed (SDE, AM, and MC showed only small-sized correlations between them, and the MC correlated most highly with the BIDR moralistic scales) raise the possibility that these scales might share unique variance with the VVIQ. Thus, two stepwise multiple regressions were conducted to investigate the degree of independence of SDE, AM, and the MC correlations with the VVIQ. In both regressions, the VVIQ acted as the criterion variable. In the first regression, the predictor variables were SDE scored continuously, AM scored continuously, and the MC. In the second regression, the predictor variables were SDE scored dichotomously and the MC. The results of the regressions are shown in Table 3. Neither regression proceeded beyond Step 1, and both had only SDE in the equation, indicating that the addition of variables beyond SDE did not significantly improve prediction of VVIQ scores, although in the first regression AM came close to satisfying entry requirements at Step 2.

Finally, further Pearson's correlations tested the possibility that dichotomously scored BIDR subscales might underestimate the size of the correlations with the VVIQ by not tapping into exaggeratedly low responses. To do this new analysis, SDE and AM scores were calculated in an analogue of the dichotomous scoring method, but transposed to the bottom end of the scoring range, with item responses of 1 or 2 scored as one point and all other responses scored zero. Distributions of these new variables were normally distributed and so did not require transformation. The resultant new SDE and AM scores were correlated with the VVIQ. Both VVIQ correlations were not significant [for VVIQ–SDE,  $r(117) = -.05$ ,  $p = .62$ , two-tailed; for VVIQ–AM,  $r(117) = -.10$ ,  $p = .30$ , two-tailed].

## Discussion

This study marks the first time a direct comparison has been made between the BIDR subscales and the MC. Previously, we argued that the MC scale is most likely primarily a measure of the moralistic bias, and so is not ideally suited to assess the relationship between scores on the VVIQ and social desirable responding, and that the BIDR is superior to the MC scale as a measure of social desirable responding because of its ability to treat social desirable responding as a multifactorial phenomenon. The VVIQ correlation with the MC scale was lower than its correlation with the BIDR egoistic scales (apart from AM scored dichotomously). The coefficient value of .18 for the VVIQ–MC correlation matches McKelvie's (1995a) meta-analysis estimate from 10 studies. The MC correlated significantly with SDE, SDD, and IM, but notably more highly with the BIDR moralistic scales than with SDE, and only marginally more highly with IM than with SDD. The results showed several significant correlations between the VVIQ and the measures of social desirable responding, in which greater imagery ability was always associated with higher social desirable responding. The largest of these correlations were with the BIDR measures of egoistic bias, particularly SDE. This pattern of results supports previous research (e.g., Allbutt et al., 2008).

Regression analyses investigated the relationship between the VVIQ, the BIDR egoistic subscales, and the MC. We were concerned that, despite the use of anonymous responding, factors such as participation in an experiment, group testing in an academic setting, and the presence of a tutor in the role of the experimenter might create the conditions for the VVIQ to correlate with the AM independently of SDE. However, the stepwise regression analyses showed no conclusive evidence that correlations between the VVIQ and SDE were enhanced by adding other social desirable responding scales.

The continuous and dichotomous BIDR scoring methods were also compared. Allbutt et al. (2008) found no major differences between BIDR scoring methods in studies investigating the relationship between a new, longer version

**Table 3** Results of stepwise regressions

Regression	Variable	<i>B</i>	<i>SE B</i>	$\beta$	<i>R</i> <sup>2</sup>	<i>T</i>	Sig
Regression 1, Step 1							
Variables in the equation	Constant	7.61	.67				
	SDEc	-.02	.00	-.35	.12	-4.02	.00
Variables not in the equation	AMc			-.17		-1.86	.06
	MC			-.06		-.64	.52
Regression 2, Step 1							
Variables in the equation	Constant	6.52	.15				
	SDEd	-.23	.08	-.25	.06	-2.77	.01
Variables not in the equation	MC			-.12		-1.29	.20

*F* to enter (d) = .05, *F* to remove (e) = .10

of the VVIQ called the VVIQ Version 2 (Marks, 1995), the Individual Differences Questionnaire–Imagery Habit Scale (Cohen & Saslona, 1990), the Auditory Imagery Scale (Gissurason, 1992), and the SDE and IM subscales of the BIDR. However, several notable differences in the patterns of results were seen in this study between the two scoring methods: VVIQ correlations were higher with SDE and AM when continuously scored. The MC correlated more highly with BIDR subscales when they were continuously scored. Also when they were continuously scored, BIDR subscales showed higher Cronbach's alphas, were normally distributed as opposed to positively skewed, and showed higher correlations with each other. This pattern of results is in agreement with those of Stöber et al. (2002) and Cervellione et al. (2009).

In contrast, the pattern of correlations seen with dichotomous BIDR scoring showed Paulhus's (2002) two-tier structure more clearly. The fact that the correlations between each BIDR subscale scored using the two scoring methods were around .7–.8, and the differences in distribution normality between the scoring approaches, suggests that the different scoring methods measure related but not identical constructs. The possibility that dichotomous scoring might underestimate the size of correlations with the VVIQ by not tapping into exaggerated low responses was explored by scoring SDE and AM scales using a “mirror image” analogue of the dichotomous approach. However, no significant correlations were found between the VVIQ and the BIDR egoistic subscales scored in this way, which suggests that no relationship between VVIQ and exaggerated low responses exists. The scoring method may also have had an effect on VVIQ correlations with the AM subscale. The VVIQ–AM correlation reached significance when AM was scored continuously but not when it was scored dichotomously. AM scale items differ from SDE items in being particularly strong egoistic claims. Under our anonymous testing conditions, descriptive statistics showed that AM dichotomous average scores were low and not widely spread. This restricted spread of AM scores when scored dichotomously may have contributed to the difference in VVIQ–AM correlations when scored using the continuous and dichotomous approaches.

In our introduction, we noted that there is debate over whether correlations with social desirable responding should best be viewed as distortion/error, substantive responding, or a mixture of both. We have argued that respondents answering self-report imagery questionnaires will most likely view imagery as an ability being tested and will value responses that they perceive as indicating “better” imagery ability. The results of previous studies have provided some support for this assertion, in that the sizes of imagery scale–SDE correlations were correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which the imagery property was valued. If social

desirable responding is best viewed as distortion/error or as different on a conceptual level from mental imagery processes, then McKelvie's (1995b) suggestion that correlations with measures of social desirable responding in excess of .25 (6.25% overlap of variance) indicate an unacceptable degree of contamination becomes relevant. The correlations between the VVIQ and SDE either exceeded this criterion when SDE was scored continuously ( $r = -.35$ ), or matched the criterion when SDE was scored dichotomously ( $r = -.25$ ). However, like all cutoff points, the exact value chosen is somewhat arbitrary, and even assuming that social desirable responding reflects distortion, judgements of the validity of imagery questionnaires would also need to be made in light of data on other aspects of validity.

If social desirable responding is best viewed as valid substantive responding in its own right, the crucial questions shift from what is an unacceptable level of contamination and how can we control the influence of SDE to why the VVIQ and other scales from our previous work should correlate with the egoistic bias. One possible answer is that the relationship between the VVIQ and the egoistic bias is mediated by anxiety levels. Although no study has directly correlated VVIQ total scores with anxiety levels, Bent and Wick (2006) measured participants' vividness of visual imagery using Ahsen's adapted VVIQ (Ahsen, 1985) when imaging their mothers, fathers and other VVIQ scenes while keeping either their mother or father in mind. Anxiety levels were also measured using Spielberger's State–Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983) while keeping either their mother or their father in mind. Bent and Wick found that relationships between vividness and anxiety scores. For example vividness difference scores for imaging mothers and fathers correlated significantly with trait anxiety difference scores ( $r = .27$ ), and the vividness difference scores for VVIQ scenes while keeping mothers and fathers in mind correlated significantly with both state ( $r = .43$ ) and trait ( $r = .60$ ) anxiety. Similarly, Paulhus (1994) reported the highest correlates of the SDE scale to be trait anxiety ( $r = -.52$ ) and self-esteem ( $r = +.50$ ), and Davies, French, and Keogh (1998) reported the correlation between the SDE and Eysenck's neuroticism scale, from the Eysenck Personality Questionnaire–Revised (Eysenck & Eysenck, 1991), to be  $-.53$ .

A priority for future work should be to test the assertion that social desirable responding is a construct independent from mental imagery processes, and thus that its control would aid the assessment of self-reported imagery. Testing the relationship between VVIQ scores, egoistic bias scores, and anxiety levels would help determine the extent to which VVIQ–egoistic bias correlations might be mediated by anxiety. Also, McKelvie's (1995a) meta-analytical review reported relationships between the VVIQ and other self-report imagery questionnaires, perceptual measures,



and memory measures that he argued provide evidence to support the construct validity of the VVIQ. Repeating some of these studies and observing the effect of controlling social desirable responding would provide valuable data to help resolve this issue. Meta-analytic reviews (Li & Bagger, 2006; Ones et al., 1996) have found partial correlation approaches to be unable to effectively control for social desirability effects. However, more recently, approaches based on confirmatory factor analysis, which have the capacity to control for item measurement error but require large sample sizes and specialist statistical software, have shown greater promise and suggest that social desirable responding can act as both a suppressor and a moderator variable (Konstabel, Aavik, & Allik, 2006; Leite & Cooper, 2010; Ziegler & Buehner, 2009).

## References

- Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery*, 9, 1–18.
- Allbutt, J., Ling, J., Heffernan, T. M., & Shafiqullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and visual thinking style. *Journal of Individual Differences*, 29, 181–188.
- Allbutt, J., Ling, J., & Shafiqullah, M. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery. *Imagination, Cognition and Personality*, 25, 337–353.
- Allbutt, J., Shafiqullah, M., & Ling, J. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, 30, 1–20.
- Bent, N. A., & Wick, E. (2006). Beyond vividness: Parental filters as moderators in mental imagery and measured anxiety level. *Journal of Mental Imagery*, 30, 21–38.
- Cervellione, K. L., Lee, Y.-S., & Bonanno, G. A. (2009). Rasch modeling of the self-deception scale of the balanced inventory of desirable responding. *Educational and Psychological Measurement*, 69, 438–458. doi:10.1177/0013164408322020
- Cohen, B. H., & Saslona, M. (1990). The advantage of being an habitual visualiser. *Journal of Mental Imagery*, 14, 101–112.
- Crowne, D. P. (1979). *The experimental study of personality*. Hillsdale: Erlbaum.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- Davies, M. F., French, C. C., & Keogh, E. (1998). Self-deceptive enhancement and impression management correlates of EPQ-R dimensions. *The Journal of Psychology*, 132, 401–406.
- Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge: Cambridge University Press.
- Edwards, A. L., Diers, C. J., & Walker, J. N. (1962). Response sets and factors loadings on sixty-one personality scales. *The Journal of Applied Psychology*, 46, 220–225.
- Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck personality scales (EPS Adult)*. London: Hodder & Stoughton.
- Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp.21–40). Boston: Kluwer
- John, O. P., & Paulhus, D. L. (2000). *The structure of self enhancement*. Unpublished manuscript. Berkeley: University of California.
- Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality*, 20, 549–566.
- Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe–Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and Psychological Measurement*, 65, 140–154.
- Leite, W. L., & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioural 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 and Assessment*, 14, 131–141.
- Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkaslo, M. (2007). Substance and style in social desirable responding. *Journal of Personality*, 75, 291–322.
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, 64, 17–24.
- Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, 19, 153–166.
- McKelvie, S. J. (1995a). Response to commentaries: The VVIQ and beyond. Vividness and its measurement. *Journal of Mental Imagery*, 19, 197–251.
- McKelvie, S. J. (1995b). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery*, 19, 1–106.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *The Journal of Applied Psychology*, 81, 660–679.
- Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- 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.
- Paulhus, D. L. (1994). *Balanced Inventory of desirable responding: Reference manual for BIDR Version 6*. Unpublished manual. Vancouver: University of British Columbia.
- Paulhus, D.L. (1999). *BIDR Self-Deceptive Denial (SDD)*. Retrieved 1 August, 2003, from <http://neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm>
- Paulhus, D. L. (2002). Social desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah: Erlbaum.
- Spielberger, C. D., Gorsuch, R., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-Trait anxiety inventory (Form Y)*. Palo Alto: Consulting Psychologists Press.
- Stöber, J. (2001). The social desirability scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222–232. doi:10.1027/1015-5759.17.3.222
- Stöber, J., Dette, D. E., & Musch, T. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment*, 78, 370–389.
- Ziegler, M., & Buehner, M. (2009). Modelling social desirable responding and its effects. *Educational and Psychological Measurement*, 69, 548–565.