ABSTRACT – Background and Objectives: In this study we review the use of the Positive Presentation Management (PPM) and Negative Presentation Management (NPM) scales, two NEO-PI-R derived measures originally devised to control for biased and distorted responses. These scales have been used with normative, job selection and clinical samples, in cross-sectional and experimental studies.

Methods: Web-based and manual searches in personality and psychological assessment journals were conducted, and information on the PPM and NPM scales was systematically recorded. Means, standard deviations and reliability coefficients were summarized and compared between three types of samples: normative, job selection and clinical.

Results: Five studies were performed with normative samples (33%), 3 with employment samples (20%) and 7 with clinical samples (47%). Cross-sectional designs were most common (60%), although there were also experimental studies (40%). Reported reliability coefficients were lower than usually accepted. There were differences in mean PPM and NPM scores in regard to the study sample background.

Conclusions: There were some discrepancies when reporting PPM and NPM results across the reviewed studies. Normative and employment samples scored higher in PPM than clinical samples. Clinical samples scored higher in NPM than normative and employment samples. The PPM and NPM scales could be useful in applied situations, although parallel sources of information should be taken into account to detect distorted responses to the questionnaire. However, the results on these scales should be systematically reported in future studies.

Received 7 July 2008
Revised 5 February 2009
Accepted 10 February 2009
Introduction

The NEO-PI-R is a measure of the five-factor model of personality, a personality inventory that includes the factors of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. This questionnaire has been widely used in basic research, but also in applied settings such as clinical assessment and job selection. The usefulness of the NEO-PI-R for clinical assessment was advocated in the early development stage of the instrument. Nevertheless, this point of view was questioned because of the failure of the instrument to incorporate validity scales designed to detect misleading answers. Clinical patients showing very low self-esteem and lacking of defensive strategies, present personality profiles characterized by an unstable emotionality and a social withdrawal consolidated pattern. Therefore, they would be likely to respond to the NEO-PI-R in accordance with a negative presentation. Moreover, in forensic contexts an individual might exaggerate or create psychopathological symptoms, or underreport symptoms of mental illness. In regard to the particular case of applicants in a job selection situation, meta-analytic studies have shown that those people who complete a personality inventory as part of a job selection process tend to present themselves in a more positive manner, although in a lesser extent than when instructed to fake, and as a function of the particular personality dimension, and the type of job and test. Besides, scores in social desirability were suggested to indicate true individual differences in personality variables such as emotional stability and conscientiousness, that is, job applicants with high scores on social desirability measures also scored high on these big five personality dimensions. This fact might be problematic for personality assessment in employment contexts, considering that conscientiousness and emotional stability are perhaps the best predictors of job performance. However, social desirability was not considered as a consistent predictor, suppressor, or mediator variable between personality and job performance, as controlling for its effects did not increase in any way the predictive power of personality variables in regard to job performance.

A considerable body of research has addressed the answering style in personality self-report instruments. Particularly, seminal personality inventories such as the Minnesota Multidimensional Personality Inventory (MMPI) or the 16PF have included a variety of control scales to detect several forms of distorted responses such as reading or understanding problems, non cooperative, defensive or negative attitudes, a random response, or willingness to appear in a favourable or a non favourable manner. Schinka, Kinder and Kremer designed three 10-item validity scales from the NEO-PI-R items: Positive Presentation Management (PPM), Negative Presentation Management (NPM) and Inconsistence (INC). PPM was intended to identify respondents claiming uncommon virtues and/or denying common faults. In contrast, NPM was intended to identify respondents claiming uncommon faults and/or denying common virtues. The INC scale was basically designed to detect random responding, although in the present study we focus on PPM and NPM scales. While Extraversion and Conscientiousness are significantly positively correlated with NPM, they are significantly negatively correlated with NPM. The opposite is true for Neuroticism, which is positively correlated with NPM but negatively with PPM.

It has been argued that PPM and NPM scales might be useful to detect particular personality profiles addressed to manipulate...
individual presentation in either a positive or negative way, as they have shown from moderate to good discriminant capacity between faking and standard responding instructions\textsuperscript{2,20}. However, the usefulness of these scales has also been questioned, advocating for the comparison of self-reports with independent scores as a plausible alternative to the common use of validity scales in personality assessment\textsuperscript{21}. The PPM and NPM scales have been used with normative, clinical and employment samples. Therefore, the aim of the present study was to collect and review the information that has been generated on the use of the NEO-PI-R, PPM and NPM validity scales. Descriptive basic statistics, reliability and correlational data were summarized and compared in regard to the type of sample background.

**Method**

**Literature search**

In accordance with the design of the PPM and NPM scales\textsuperscript{18}, we looked for studies from 1997 onwards. We conducted a web-based literature search using PsychINFO and Social Sciences Citation Index databases, including several boolean combinations of the keywords “Positive Presentation Management”, “Negative Presentation Management”, “Validity scales” and “NEO-PI-R”. In addition, we performed a manual search in 8 key international journals that address personality assessment issues: *European Journal of Personality, Journal of Personality, Journal of Personality Assessment, Journal of Personality and Social Psychology, Journal of Research in Personality, Personality and Individual Differences, Personality and Social Psychology Bulletin, and Psychological Assessment*.

**The PPM and NPM scales**

The NEO-PI-R Personality Inventory\textsuperscript{1,22} is used to tap the Big Five personality dimensions (Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness), with six facets within each domain allowing for a more fine-grained description of human personality. The instrument contains 240 items to which individuals respond on a 5-point Likert-type scale with response options ranging from “strongly disagree” (0) to “strongly agree”\textsuperscript{4}. Scores on Positive Presentation Management (PPM) and Negative Presentation Management (NPM) are obtained from the NEO-PI-R scales. The items from the NEO-PI-R facets integrated in PPM and NPM scales are shown in Table I. All items are to be recoded in the adequate direction. PPM identifies respondents claiming uncommon virtues and/or denying common faults, whereas NPM is intended to identify respondents claiming uncommon faults and/or denying common virtues.

**Results**

Table II shows the studies using the PPM and NPM scales performed since 1997\textsuperscript{18}, considering its sample size and design, mean scores, standard deviations and reliabilities and whether correlations between PPM and NPM with the NEO-PI-R dimensions were reported or not. In the case of experimental studies, only means belonging to standard instructions were considered. There were 15 studies, 5 with normative samples (33%), 3 with employment samples (20%) and 7 with clinical samples (47%). Most studies employed a cross-sectional design (60%), although 40% conducted experimental studies addressed to deter-
mine differences between groups with standard and faking response instructions. Sample sizes of studies varied greatly from 22 to 21349 individuals. For PPM, raw means and standard deviations were reported in 14 studies (93%), with two studies reporting instead T scores, and one study not reporting any information on basic statistics such as means and standard deviations. For NPM, raw means and standard deviations reporting decreased to 12 studies (80%). PPM reliability coefficients were reported for 8 studies (53%), whereas there were only 6 studies (40%) that reported the reliability coefficients for NPM. Reliabilities in PPM for normative, employment and clinical samples ranged between 0.46 to 0.50, 0.50 to 0.60, and 0.43 to 0.70, respectively, whereas NPM reliabilities ranged between 0.52, 0.52 to 0.57, and 0.60 to 0.75. These coefficients were lower than the usually accepted standards except for the Young and Schinka study (PPM = 0.70, NPM = 0.75). In addition, only 6 studies (40%) reported the intercorrelations between PPM and NPM with the NEO-PI-R dimensions.

Normative, employment and clinical samples PPM mean scores ranged between 18.40 to 23.34, 20.25 to 23.51, and 13.82 to 24.19, whereas for NPM mean scores ranged between 7.57 to 9.80, 7.90 to 8.78, and 10.71 to

Table I
NEO-PI-R facets and corresponding items to PPM and NPM validity scales

<table>
<thead>
<tr>
<th>NEO-PI-R facets</th>
<th>PPM</th>
<th>NPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1: Anxiety</td>
<td>–</td>
<td>31</td>
</tr>
<tr>
<td>N2: Angry Hostility</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N3: Depression</td>
<td>–</td>
<td>161</td>
</tr>
<tr>
<td>N4: Self-Consciousness</td>
<td>196</td>
<td>–</td>
</tr>
<tr>
<td>N5: Impulsiveness</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N6: Vulnerability</td>
<td>146</td>
<td>–</td>
</tr>
<tr>
<td>E1: Warmth</td>
<td>–</td>
<td>62</td>
</tr>
<tr>
<td>E2: Gregariousness</td>
<td>37</td>
<td>–</td>
</tr>
<tr>
<td>E3: Assertiveness</td>
<td>42, 162</td>
<td>–</td>
</tr>
<tr>
<td>E4: Activity</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>E5: Excitement seeking</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>E6: Positive emotions</td>
<td>–</td>
<td>57</td>
</tr>
<tr>
<td>O1: Fantasy</td>
<td>93, 153</td>
<td>–</td>
</tr>
<tr>
<td>O2: Aesthetics</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>O3: Feelings</td>
<td>–</td>
<td>73</td>
</tr>
<tr>
<td>O4: Actions</td>
<td>–</td>
<td>48</td>
</tr>
<tr>
<td>O5: Ideas</td>
<td>113</td>
<td>–</td>
</tr>
<tr>
<td>O6: Values</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>A1: Trust</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>A2: Straightforwardness</td>
<td>–</td>
<td>129</td>
</tr>
<tr>
<td>A3: Altruism</td>
<td>–</td>
<td>104</td>
</tr>
<tr>
<td>A4: Compliance</td>
<td>139</td>
<td>–</td>
</tr>
<tr>
<td>A5: Modesty</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>A6: Tender-Mindedness</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C1: Competence</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C2: Order</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C3: Dutifulness</td>
<td>–</td>
<td>15, 135</td>
</tr>
<tr>
<td>C4: Achievement Striving</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C5: Self-Discipline</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>C6: Deliberation</td>
<td>30</td>
<td>–</td>
</tr>
</tbody>
</table>
Using standardized mean difference ($d^{29}$), moderate differences were noted on PPM between normative and employment ($d = -0.67$) and normative and clinical ($d = 0.55$) samples, while a larger difference was evident between employment and clinical ($d = 0.99$) samples. There were moderate differences on NPM between normative and employment samples ($d = 0.37$), and large differences between normative and clinical ($d = -2.78$), and employment and clinical samples ($d = -2.91$).

Standardized scores were computed for both PPM and NPM ($z = \frac{X - \bar{X}}{sd}$), in order to facilitate the comparison of mean scores in the two scales across the three sample types. Figure 1 shows that PPM $z$ scores were mostly above the mean for both, normative and employment samples and well below the mean for clinical samples. In contrast, the NPM $z$ scores were essentially above the mean for clinical samples and around a standard deviation below the mean for normative and em-

16.40. Using standardized mean difference ($d^{29}$), moderate differences were noted on PPM between normative and employment ($d = -0.67$) and normative and clinical ($d = 0.55$) samples, while a larger difference was evident between employment and clinical ($d = 0.99$) samples. There were moderate differences on NPM between normative and employment samples ($d = 0.37$), and large differences between normative and clinical ($d = -2.78$), and employment and clinical samples ($d = -2.91$).
ployment samples. These results suggest that the studies that have obtained data from the PPM and NPM scales, report that normative and employment samples tend to score higher in PPM than clinical samples, whereas clinical samples tend to score higher in the NPM scale than normative and employment samples.

Discussion

While a body of research has highlighted the usefulness of the PPM and NPM scales to detect distorted responding to the NEO-PI-R\(^3\),\(^1\),\(^9\),\(^2\), other research works have suggested alternative ways of controlling for the validity of responses, such as comparing multiple sources of data in the interpretation of a given questionnaire\(^2\),\(^2\). Independently of this debate, the present study was designed to analyze from a descriptive point of view the research done to date since the development of PPM and NPM validity scales\(^1\). The results of this review suggest some discrepancies in the reporting of PPM and NPM results. In addition, there appears to be a moderating effect of the sample background in regard to the results reported in the literature.

There exist some discrepancies in PPM and NPM results reporting. For instance, some studies reported mean \(T\) scores instead of providing the raw means and standard deviations, do not report the scores in one of the scales, or even both of them. Most notably, experimental studies do not report reliability coefficients, whereas the reported reliability coefficients tend to be lower than accepted standards. This is probably due to the fact that both scales are composed of items from different personality constructs, which precludes an acceptable internal consistency. On the contrary, if these scales were made up from ‘external’ items such as those used by the MMPI\(^1\) or the 16PF\(^17\), better internal consistencies could perhaps be expected\(^3\). Reporting reliabilities is a particularly important issue, because with low reliabilities users of the PPM and NPM scales might not be

![Figure 1. Comparison of mean z scores on PPM and NPM by sample background.](image-url)
sure whether they are measuring in fact a uni-
ified construct, and might not be able to make
valid predictions whatsoever.

Only half of the studies informed about
PPM and NPM intercorrelations with the
rest of NEO-PI-R dimensions. Therefore,
we think that future studies on the PPM and
NPM scales should provide raw mean and
standard deviation scores, inform about the
reliability coefficients, and provide the cor-
relation coefficients with the NEO-PI-R
personality factors, or in any case, make this
information readily available to interested
researchers in performing for instance, a
meta-analytic review. This fact, refrain us
from carry out a meta-analytic review on the
PPM and NPM scales, and constitutes from
our viewpoint an important limitation in
order to perform this sort of analyses. In this
review, we succeed when asking to some re-
searchers for this missing data in published
research reports, but in other cases original
authors could not be reached and therefore,
that information could not be used.

The present work shows that the obtained
outcomes on both, PPM and NPM differ in
regard to the sample background: normative
and employment samples score higher in
PPM than clinical samples, whereas clinical
samples have higher scores in NPM than
normative and employment samples. Never-
theless, there are cases where scores in PPM
were much higher in a clinical sample than
in a NEO-PI-R normative sample24,25. This
finding suggests that individuals included in
normative and particularly in employment
samples tend to present themselves as more
positive than individuals in clinical samples.
Nevertheless, the present results should be
interpreted with caution considering a num-
ber of factors in regard to the available stud-
ies: a) the reduced size of studies; b) the
great diversity of sample size; c) the only re-
liance on published studies; and d) the fact
that some included studies might not be
completely independent.

The application of the NEO-PI-R for clini-
cal purposes has been the origin of a debate
on whether or not contemplating validity
scales in this instrument is sufficiently ade-
quate6,7. As far as we know, the research in-
terested on the PPM and NPM validity
scales18 has been rather limited and on occa-
sions critical. For instance, Piedmont21 ques-
tioned the usefulness of validity scales for re-
search purposes, although suggested that
their results did not consistently support the
invalidity of these validity scales in applied
settings, i.e., clinical or in industrial / organiza-
tional psychology. Further, more robust results
were recently obtained with a relative-scored
personality questionnaire when compared
with a standard likert-type big five question-
naire in regard to faking31. On the other hand,
there is evidence pointing out to the useful-
ness of the PPM scale to screen for potentially
distorted NEO-PI-R responses27. Therefore,
while there is some agreement regarding the
usefulness of the PPM and NPM scales in ap-
plied situations, it has also been recommend-
ed to use parallel sources of information such
as independent ratings, personal interviews,
or relative-scored personality questionnaires.

The results reviewed in the present study
about the PPM and NPM scales indicate that:
a) there have been only a few studies which
are insufficient to reach any consistent con-
clusions about its usefulness; b) results are re-
ported in different forms across studies,
which poses some difficulties in order to per-
form a meta-analytical review of these scales;
and c) there are differences in mean scores
across different sample types. Future studies
might perhaps attempt to report PPM and
NPM results in a uniform way, and to com-
pare their results with the present findings
considering the sample background.
References

Articles preceded by an asterisk were included in the analysis.


27. *Sellbom M, Bagby RM. The validity and utility of the Positive Presentation Management and Negative Pre-
sentation Management scales for the Revised NEO Personality Inventory. Assessment 2008; 15: 165-176.


Address for correspondence:
Angel Blanch
Department of Pedagogy and Psychology
Avda. Estudi General, 4
25001 Lleida (Catalonia)
Spain
Tel: 973706529
Fax: 973706505
Email: ablanche@pip.udl.cat