Social Desirability and the Clinical Self-Report Inventory: Methodological Reconsideration

▼

Chiou-yan Huang, Hsin-ya Liao, and Sue-Hwang Chang
National Taiwan University

This study assesses how social desirability affects responses in clinical self-report inventories. Six hundred items gathered from four normal personality questionnaires were adapted to devise a pre-experimental personality questionnaire (pre-EMHQ). Results obtained from administering Ko's Mental Health Questionnaire (KMHQ) and the pre-EMHQ to separate samples were the proportion of individuals answering "true" to each item (i.e., P(t)) and the social desirability scale value (i.e., SDSV) of each item. The Experimental Mental Health Questionnaire (EMHQ) was established from the pre-EMHQ by closely matching the P(t)s and the SDSVs of the two questionnaires. Administering the KMHQ and the EMHQ concurrently to another sample provided results for factor analysis and other statistical analyses. The SDSVs and the P(t)s for each of the KMHQ items certainly displayed a linearly increasing relation. The two sets of corresponding subscales also correlated significantly. By applying the polynomial regression analysis, the tendency to score might be expressed as a quadratic function of SDSVs. Two iterative principal–factor analyses of the two sets of subscales each resulted in two factors, and Factor 1 is similar in both the KMHQ and the EMHQ. In brief, social desirability plays a critical role in affecting responses in a clinical self-report inventory. The factors involved and suggestions proposed will be of value for further research. © 1998 John Wiley & Sons, Inc. J Clin Psychol 54: 517–528, 1998.

Most items in a given clinical self-report inventory are vulnerable to faking. Some respondents tend to dissemble and give socially desirable or undesirable responses for the sake of appearance. Occasionally, they try to project a good impression and deny having socially undesirable traits, sometimes by intentional deception, sometimes with no clear intention. Previous research

---

This research was supported in part by Grant NSC84-0128-C002-01-187H from the National Science Council, Taiwan, R.O.C.
Correspondence concerning this article should be addressed to Sue-Hwang Chang, PhD, Department of Psychology, National Taiwan University, Taiwan, R.O.C.

JOURNAL OF CLINICAL PSYCHOLOGY, Vol. 54(4), 517–528 (1998)
 CCC 0021-9762/98/040517-12

(Crowne & Marlow, 1964; Frederiksen, 1965) has indicated that such behavior is apt to arise from the need for self-protection, avoiding criticism, social conformity, and social approval, hence "the tendency to give socially desirable responses" in questionnaires. In addition, Nederhof (1985) found that the tendency explains about 10% to 75% of variance (the wide range is related to socially desirable situations or to the importance of such responses). Under other conditions, respondents fake bad for some reason and claim to have socially undesirable traits. Since social desirability is pernicious to the construct validity and can contribute to the differential diagnosis such as Malingering and Factitious Disorders, etc. (DSM-IV, American Psychiatric Association, 1994), researchers have tried to measure or control the faking.

To study the above effects in self-report inventories, Edwards (1957) drew 150 heterogenous items from three validity scales in the Minnesota Multiple Personality Inventory (MMPI) and Taylor Manifest Anxiety Scale. He selected 79 items according to their social desirability, which were finally reduced to 39, to establish a socially desirable scale (SD). The correlation between the SD scale and the correction scale (K[1]) in the MMPI is .81 (Anastasi, 1988). The high correlation is partially because both scales measure a similar construct so that by correlating the scores in the SD scale and in the other personality inventories, researchers can assess to what degree self-report inventories are affected by social desirability.

Relevant findings in previous research (Edwards & Edwards, 1992) are elucidated by the following symbols. Where P(t) is the proportion of individuals answering "true" to an item, SDSV is the average rating of each item on an n-point social desirability scale. HSD and LSD represent two groups of individuals: high scorers and low scorers on the SD scale. The findings are as follows. First, P(t) is a linearly increasing function of the SDSVs of items. Respondents tend to answer "true" to items with high SDSVs and "false" to those with low SDSVs. Second, the proportion of socially desirable responses is the proportion of individuals answering "true" to items for high social desirability and "false" to socially undesirable ones. Therefore, if the proportion of socially desirable responses is plotted against the SDSVs of items, the resulting plot will be V-shaped. Third, the linearly increasing function differs in slope for the HSD and LSD groups, indicating that items with either high or low SDSVs are the most sensitive to individual differences.

The above three results suggest that the SD scale measures a general personality trait (i.e., the tendency to give socially desirable responses to items in standard conditions) and that this trait does not depend on the specific content of the items but on their SDSVs. Items with high SDSVs are more likely to be answered "true" by individuals with high rates of socially desirable responses. Such respondents are more likely than individuals with low rates to answer "false" to items with low SDSVs. It is concluded that the social desirability rating of items and this kind of behavioral tendency in respondents will both affect scale scores.

Edwards and Edwards (1992) assessed the effects of social desirability on scale scores by applying the MMPI according to the following rationale. If two similar structures result from closely matching items with close social desirability in the two scales without a common content, we can then conclude that the variables in social desirability affect the scale scores. Scales used in this study include Wiggins's 13 MMPI scales with a homogeneous content, an Experimental Multiphasic Personality Inventory (EMPI), Edwards's SD scale, Welsh's R (repression) scale and Wiggins's Sd scale. Administering an item pool of 2824 normal personality items to two independent samples of college students provided a batch of the P(t)s and the SDSVs. Next, based on the P(t)s and the SDSVs for each of the MMPI items published by previous researchers, Edwards and Edwards selected items with approximately matching P(t)

[1] The score can indicate the respondent's attitude. A high score may imply self-protection or an attempt to "fake good"; a low score may imply extreme truthfulness, self-criticism or an attempt to "fake bad."

and SDSV and devised an Experimental Multiphasic Personality Inventory (EMPI). Finally, the MMPI and the EMPI were respectively administered to two samples before applying a principal–component analysis to the results. Classifying the scales in both according to factor loadings produced four factors. The two factor structures of these 13 scales and the three marker scales—SD, R and Sd—were extremely similar. The scales, loading each corresponding factor, were the corresponding scales that resulted from the matching procedure. In addition, extreme similarity was indicated by the high correlation and congruence between the two sets of factor loadings. Furthermore, SD, R, and Sd exhibited their highest loadings on Factor 1, Factor 2, and Factor 3, respectively. In spite of those seemingly perfect results, the factor extraction method, principal–component analysis, might be inappropriate. Only the eigenvalue-greater-than-one criterion was employed to determine the number of factors. The factors were then rotated using a varimax rotation. The method is the default in SPSS, and the problems resulting from it are as follows. First, merely using the eigenvalue-greater-than-one criterion to determine the number of factors is inappropriate (Cattell & Vogelman, 1977). Second, principal–component analysis and factor analysis are two different statistical methods, so it is not reasonable to interchange them (Bentler & Kano, 1990; Mulaik, 1986). Third, principal–component analysis tends to overestimate factor loadings (Lee & Comrey, 1979). Fourth, the varimax rotation is not necessarily appropriate for all kinds of tasks (Ford, MacCallum, & Tait, 1986).

These four methodological problems can be further discussed as follows. First, depending only on the eigenvalue-greater-than one criterion to decide the number of factors frequently extracts too many factors. Such a fallacy can be found from the results of the principal–component analysis in Edwards and Edwards' (1992) study: Factor 4 is indicated by only one variable; obviously it is a specific rather than common factor. Therefore, in determining the appropriate number of factors, it is better to simultaneously consider more than one deciding rule (e.g., the scree test and the number suggested by the χ^2 test in the maximum likelihood method) and take account of theory and interpretability in reaching the final decision. An inappropriate number of factors affects the rotated results. Therefore, to avoid possible mistakes, both the eigenvalue-greater-than one criterion and the scree test were applied in this study to determine the number of factors. Second, the principal–component analysis is applied to find components, which are linear composites of the original variables, and to make variances of the individual scores on the components maximum. However, the factor analysis is applied to find latent variables (i.e., elements composing the observed variable) to explain correlations among the observed variables. Restated, the former analysis (with emphasis on variance) results merely in convenient groupings of variables used to describe individual differences rather than theoretical constructs or latent variables which the latter analysis (with emphasis on correlations) may find from correlations between the observed variables (Ford et al., 1986). In addition, while principal–component analysis does not take measurement errors into account, the common factor model does assume that observed scores are affected by measurement errors and only requires the common variance of each variable analyzed, which is an estimation of communalities. This is why when employing the principal–component method to extract factors in factor analysis, factor loadings are overestimated. Although principal–component analysis overestimates the factor structure among the observed variables, factor loadings will still almost equal those obtained by another method (Velicer, Peacock, & Jackson, 1982).

Considering the perspective and methodology, previous research findings still have a certain value, but substituting another method for principal–component analysis is more reliable. In this study, an iterative principal–factor method was employed because it does not require estimating communality to prevent instability from affecting the results (Harman, 1976). Concerning the appropriateness of the rotation method, researchers should consider in advance if correlations exist among the factors. If they do correlate in theory, they may need to be rotated by an oblique rotation method. It is for the researcher to assess the appropriate method, not to leave the task to the default in the statistical software. Besides, if the number of factors is inappropriate, varimax rotation may split one factor into two or distort the factor structure; therefore, determining how many factors should be retained must be a cautious operation. In this study, theoretical considerations (Ko, 1982) suggest that the factors do not correlate, so the varimax rotation method was selected. Finally, the problem here has aspects in common with other factor analyses. Because the factor analysis examines the correlations between the observed variables, factors affecting the correlation coefficient (e.g., the nature of subscales) also affect the result. That is, a result obtained from factor analysis is unstable, so a check by cross-validation must be made to see whether the factor structure is stable among the observed variables.

In Taiwan, little relevant research has been conducted to assess social desirability in self-report inventories. This study aims to investigate the effects with the method described above using KMHQ (Ko, 1985a), which is extensively employed in Taiwan. In this study, the KMHQ and EMHQ are administered to the same sample, in contrast to Edwards and Edwards (1992) where the MMPI and the EMPI were administered to two separate samples, perhaps for fear of their respondents being fatigued by so many items (550×2). It is inappropriate to compare two factor-analysis results from two separate samples, so in this study, the factor-analysis results were from the same sample, and the numbers of items is moderate (228×2). The KMHQ is a clinical self-report inventory whose answers are either “true” or “false” in form, and it has been used for several decades in clinical practice. The scale contains 288 items subdivided into 11 subscales. The first six subscales comprise items, the content of which relates to abnormal individuals’ chief complaints. The seventh subscale termed Obsession–Compulsion contains items measuring not only the obsessive–compulsive personality but the traits related to attitudes toward time management, money, affect, and daily routine. The other four subscales contain items related to abnormal individuals’ personalities (Ko, 1985b). Like the MMPI, the test results can reveal respondents’ mental health state and various kinds of personalities and dispositions. In addition, the scores on the first seven subscales are vulnerable to faking, suggesting kinds of behavior closely related to the concept of mental health. A previous report, Ko (1976), studied the effect of faking on scale scores by comparing the results obtained in standard conditions with those in situations designed to elicit the faking-good motivation. He found that scores for more than half of the subscales are contaminated by test-taking attitudes. Factor analysis has never been applied to this kind of study in Taiwan, but, in fact, with methodological consideration, it can reveal more explicit effects. So, a further study of the effect of the variables in social desirability in clinical self-report inventories is worthwhile.

METHOD

Step 1

Subjects. The subjects were mostly students from the National Taiwan University, and the questionnaires were distributed as follows:

1. Pre-experimental mental health questionnaire (Pre-EMHQ): 95
2. Pre-experimental social desirability 5-point questionnaire: 80
3. KMHQ social desirability 5-point questionnaire: 100
4. KMHQ: 159

The distribution of ages, sex, and nature of majors among the subjects was considered in each sample.

Instruments. A questionnaire comprising 600 items, the pre-EMHQ, was devised, based on the normal personality questionnaire pool of Edwards Personal Preference Scale (Edwards, 1959), Guilford's Inventory of Factors STDCR (Guilford, 1943), Gordon Personal Inventory (Gordon, 1963), and the Sixteen Personality Factor questionnaire (Cattell, Eber, & Tatsuoka, 1970). The pre-experimental social desirability 5-point questionnaire was therefore derived from the pre-EMHQ. The third instrument was the KMHQ social desirability 5-point questionnaire.

Procedure. The first testing employed the pre-experimental social desirability 5-point questionnaire and the KMHQ social desirability 5-point questionnaire, allowing the subjects to evaluate the social desirability of the pre-EMHQ and the KMHQ. The instructions for both tests read "Please evaluate the content of the question in terms of what the general public feels is good, normal, acceptable, and praiseworthy, on a scale from '1', meaning least socially desirable, to '5', most socially desirable."

In the second testing, the subjects filled out "yes" or "no" in the pre-EMHQ and the KMHQ. The instructions for both tests read "If the content of the question fits your current situation, please fill in 'yes', if not, then 'no.' " In this procedure, four samples were tested under the standard conditions.

Statistical Analysis. The test results and the average score for each of the four questionnaires were analyzed. (The "yes" and "no" answers were taken into account; 5-point scales were used to calculate raw scores.) The average of the pre-EMHQ and the KMHQ are represented by the P(t)s. The average of the pre-experimental social desirability 5-point questionnaire and the KMHQ social desirability 5-point questionnaire are represented by the SDSVs.

Then the SDSVs and the P(t)s for each pre-EMHQ item were matched with those for the KMHQ. Based on whether the SDSVs and the P(t)s for each pre-EMHQ and KMHQ item were the same or similar, 288 items were selected to establish the Experimental Mental Health Questionnaire (EMHQ). A check was made to see whether the matching procedure was as ideal as expected by employing a *t* test to detect degrees of difference between corresponding scales in the KMHQ and EMHQ according to the SDSVs and P(t)s.

Step 2

Subjects. The KMHQ and the EMHQ were administered to 160 female and 121 male subjects from the National Taiwan University (including from freshmen and seniors), and 86 of them took the KMHQ social desirability 5-point questionnaire. The distribution of ages and majors in each sample was evenly divided.

Instruments. Based on the pre-EMHQ, 288 items were selected for the EMHQ along with the KMHQ and the KMHQ social desirability 5-point questionnaire for the group testing.

Procedure. The EMHQ follows the format of the KMHQ, so the answer sheets adopt the same method. The subjects then have to respond with a "yes" or "no." Of the 281 KMHQ and EMHQ respondents, 86 also completed the KMHQ social desirability 5-point questionnaire. In this way, the SDSVs and the P(t)s can be derived from the dependent sample. All the tests were conducted under the standard conditions.

Statistical Analysis. The calculated results of the KMHQ employed scoring cards for the KMHQ's, which totaled 11 subscales. Since the EMHQ is based on the results of the KMHQ (the SDSVs and the P(t)s derived in Step 1), the calculated procedure of the 11 subscales and the KMHQ are completely compatible.

The average subscale scores and standard deviation for each respective calculation of the KMHQ and the EMHQ were then derived, and the corresponding subscale scores of both correlated. To confirm the hypothesis of this paper, the P(t) and the SDSV for each item of the KMHQ social desirability 5-point questionnaire were correlated ($n = 86$). In the KMHQ, polynomial regression analysis was applied to the SDSV for each item and to the tendency to answer "yes" toward positive scoring items and "no" toward negative ones.

The iterative principle–factor method was employed to extract factors from the respective subscale scores of the EMHQ and the KMHQ. Next, the two principles of the scree plot of the correlation matrix and the eigenvalue-greater-than-one were used to decide how many factors to retain. Then, the factors were rotated by the varimax rotation method. Finally, the correlation and congruence coefficient in the two sets of factor loadings were found.

RESULTS

Table 1 indicates the means and standard deviations for the 11 subscales scored in the KMHQ and the EMHQ. Large differences in the means of some corresponding subscales may be the result of imperfect matching. This was expected because of the social desirability rating for some items of the KMHQ being lower than the EMHQ. Although the matching procedure was not as ideal as expected, Table 2 reveals that it still achieved the goal set in the present study.

In general, the hypothesis that the P(t) and the SDSV for each KMHQ item correlate significantly was confirmed, $r = .73$, $p = .0001$. Specifically, the relationships between the proportion of socially desirable responses and SDSVs are elucidated as follows. For KMHQ items whose right answers are "true" in the KMHQ, the correlation between the P(t)s and the SDSVs is significant, $r = .73$, $p = .0001$, and for the items whose right answers are "false," the correlation between the (1-P(t))s and the SDSVs is significant, $r = -.55$, $p = .0001$. The results show that "the tendency to score" is a quadratic function of the SDSVs (i.e., the V-shaped quadratic curve). Restated, the regression model of prediction reached the level of significance, $F(2,280) = 70.00$, $p = .0001$.

Table 1. *Means and Standard Deviations of Subscales When Scored in the KMHQ and When Scored in the EMHQ (n = 281)*

	KMHQ		EMHQ	
Subscale	*M*	*SD*	*M*	*SD*
1. Paranoia	5.76	3.52	8.98	4.33
2. Hypochondriasis	7.50	3.84	11.73	3.69
3. Social withdrawal	6.37	3.87	8.78	4.02
4. Confidence	8.07	3.20	9.56	2.91
5. Depression	7.67	4.49	10.03	4.06
6. Anxiety	6.08	4.18	8.45	3.44
7. Obsession–compulsion	8.09	4.09	10.67	3.28
8. Sexual inhibition	10.31	3.54	13.43	2.70
9. Aggression	6.00	3.45	8.65	4.54
10. Ego strength	16.00	3.20	15.17	2.86
11. Independence	13.56	4.73	14.82	3.44

Table 2. *Correlations of the 11 Subscales When Scored in the KMHQ and When Scored in the EMHQ (n* = 281)

Subscale	*r*
1. Paranoia	.58*
2. Hypochondriasis	.36*
3. Social withdrawal	.60*
4. Confidence	.39*
5. Depression	.61*
6. Anxiety	.62*
7. Obsession–compulsion	.30*
8. Sexual inhibition	.17*
9. Aggression	.52*
10. Ego strength	.41*
11. Independence	.52*

*$p < .05$.

Next, a factor analysis was applied to the scores on the subscales in the KMHQ and the EMHQ by an iterative principal–factor method. Applying the scree plot of the eigenvalues of the correlation matrix and the eigenvalue-greater-than-one criterion to decide the number of factors resulted in two factors[2] in the KMHQ with eigenvalues of 4.571 and 1.842, respectively and also in two factors in the EMHQ with eigenvalues of 5.433 and 1.336, respectively. Then, rotating the two factors by a varimax rotation resulted in the factor loadings of the subscales scored in the KMHQ and the EMHQ (see Table 3). The correlation coefficient between the two sets of the first factor loadings is equal to .95, and Tucker's coefficient of congruence (Harman, 1976) is equal to .94, indicating that the first corresponding factors of both are similar (the second factor explains the small variances, and that the correlation coefficient and Tucker's coefficient of congruence are low and insignificant). Finally, the subscales were classified according to the factor on which they have the higher absolute loading. They are arranged in Table 4, which indicates that the subscales whose scores are affected by the first corresponding factors are largely those formed by the matching procedure.

DISCUSSION

Previous research (e.g., Edwards, 1970) has demonstrated that a linearly increasing function of the P(t)s and SDSVs exists in any self-report inventory item. This observation was confirmed in the present study. In positive scoring items, the P(t)s represent the scoring tendency and are positively correlated with the SDSVs; in negative scoring items, the (1-P(t))s represent the scoring tendency and are negatively correlated with the SDSVs. The relationship between the SDSVs and the proportion of social desirability responses can be demonstrated in a V-shaped curve graph. In this research, the polynomial regression analysis was applied to test the interference and found that the scoring tendency of items could be expressed as a quadratic function of the SDSVs. The derived graph was a V-shaped parabola, which confirmed that the inference was correct.

[2] Although similarity between two sets of corresponding subscales in the two questionnaires can be enhanced if three factors are retained, the analytical result indicates that too many have been retained (i.e., only one variable has a very high factor loading on the third factor, and other variables have trivial factor loadings on it).

Table 3. *Varimax Rotated Factor Loadings of the Subscales When Scored in the KMHQ and When Scored in the EMHQ*

	Factor 1		Factor 2	
Subscale	KMHQ	EMHQ	KMHQ	EMHQ
1. Paranoia	.71	.77	−.19	.30
2. Hypochondriasis	.67	.66	−.19	.29
3. Social withdrawal	.81	.80	−.26	.25
4. Confidence	.39	.09	.71	.75
5. Depression	.72	.79	−.39	.20
6. Anxiety	.80	.74	−.28	.30
7. Obsession–compulsion	.59	.43	−.04	.60
8. Sexual inhibition	.14	.37	−.37	.06
9. Aggression	.69	.73	.15	.33
10. Ego strength	−.17	−.38	.58	.23
11. Independence	−.22	−.74	.51	−.05

By testing the matching procedure, the matching result was not as ideal as expected because of the obvious differences in each pair of the corresponding subscales (Appendix 1). Nevertheless, based on the average for each subscale, only a small difference (Appendix 2) appeared. Besides, the corresponding subscales in the KMHQ and the EMHQ all correlated significantly (Table 2). The matching procedure of the P(t)s and the SDSVs to some degree still served the purpose, because the EMHQ consisted of normal personality items, so that the significant correlations were not due to shared content.

The structures of the first corresponding factors in the KMHQ and the EMHQ were similar, except in the 7th, 8th, 10th, and 11th subscales.[3] The rest of the corresponding subscales were all loaded in corresponding factors. Furthermore, both factor loadings are relatively consistent. What caused the similar structures of the two inventories with heterogeneous content?

[3] This may be due to a small matching error or to high scores not necessarily implying psychological pathology.

Table 4. *Classification of the Subscales According to the Factor on Which They Have the Higher Absolute Loading*

	Factor 1		Factor 2	
Subscale	KMHQ	EMHQ	KMHQ	EMHQ
1. Paranoia	.71	.77	—	—
2. Hypochondriasis	.67	.66	—	—
3. Social withdrawal	.81	.80	—	—
4. Confidence	—	—	.71	.75
5. Depression	.72	.79	—	—
6. Anxiety	.80	.74	—	—
7. Obsession–compulsion	.59	—	—	.60
8. Sexual inhibition	—	.37	−.37	—
9. Aggression	.69	.73	—	—
10. Ego strength	—	−.38	.58	—
11. Independence	—	−.74	.51	—

The loadings of both corresponding factors were relatively approximate. The relative high consistency was indicated by the significant correlation and high congruence coefficient between the two sets of factor loadings. Thus, in the KMHQ and the EMHQ, there is one latent variable that has a common effect on subscales loading Factor 1. The latent factor and socially desirable responding tendency are closely related because after a procedure to match the P(t)s and the SDSVs, if the subjects give socially desirable responses in the KMHQ, then the answers for the EMHQ also indicate the same response pattern; because the corresponding subscales of both inventories exhibit similar SDSVs, the social desirability intensity index are practically the same. Theoretically, the similar factor structure for the two inventories with heterogenous content implies that Factor 1 can be explained as a Social Desirability factor.

Based on the factor analysis results, the factor loadings in the KMHQ and the EMHQ did not correspond in some corresponding subscales. They were the Obsession–Compulsion, Sexual Inhibition, Ego Strength, and Independence subscales. The result might arise not only from matching errors but from the common characteristics of the several subscales in the KMHQ. The four subscales, although not psychopathological but used in clinical practice, differ from other psychopathological subscales, so the social desirability rating for each of their items is difficult to maintain at a consistent standard. Besides, if the construct measured by a subscale is multidimensional, it often fails to detect the effect. For example, the content of Sexual Inhibition may register two extremes: healthy (high-SDSV) and unhealthy (low-SDSV). The subjects can not differentiate between the two extremes, and thus the social desirability effect fails to appear. Neither the content of the subscales nor the related social desirability point in any clear direction toward either extreme of high-SDSV or low-SDSV. With this in mind, it would be inappropriate in the present study to examine the whole subscale as a unit. It would be better suited to high-homogeneous subscales (with content consistent with social desirability or social undesirability) and not for low-homogeneous subscales. Apparently, when insufficient information about homogeneity is available, it is far more appropriate to consider each item as a unit.

Edwards and Edwards (1992) attained a perfect result, though some aspects of their procedure are somewhat confusing. If the P(t)s were not considered, would the same results be derived? That the P(t)s affect the scores is generally accepted. Would matching the P(t)s in the procedure induce an extremely close similarity in the factor structure? If so, it would then be doubtful whether the inference drawn from the results was a circular argument. After reconsideration, this was not the case. Matching the P(t)s proved a necessary component, for if not, then it would become a confounding variable and interfere with the results. In the matching procedure, if the subjects do not demonstrate any social desirability response, then it becomes impossible to obtain the same factor structure. In fact, the P(t)s are closely related to the tendency to score, but only indirectly to the scale scores. When the subjects indeed have given socially desirable responses, the P(t)s then directly affect scale scores. Therefore, the procedure exactly matches the purpose of this research. What other methods can match the P(t)s and the SDSVs at the same time? If the P(t)s and the SDSVs can never be matched, how can the research proceed? After factor analysis, different numbers of the subscale produce change. In conducting this research, if the numbers of subscales increase or decrease, can a high similarity in factor structure be attained?

In Step 1, 600 items make up the pre-EMHQ; the subjects have to answer all of them on one occasion. Errors may arise from time-consuming work, haste, exhaustion, or irritation. Any such factors can affect the data to an immeasurable degree. One resolution is to divide the items into halves. However, such an operation entails labor and time.

Some respondents may have found instructions of the pre-experimental social desirability 5-point questionnaire insufficiently clear, so the rating criterion fluctuated. Obviously, the best solution was to improve the instructions by further clarificatory notes, for example, employing the third person singular in place of the first person singular. This research set out to discover how social desirability affects test-taking responses and whether subjects are aware of this effect. This instructional error should have been anticipated earlier and then applied our solution to achieve more accurate research data.

SDSVs for some items in the KMHQ are between 1 and 2 (because most items measure psychopathology); therefore, finding items with approximately equal SDSVs from the pre-EMHQ is unlikely. In addition, compared to the subscales of the KMHQ, the EMHQ scores are slightly higher because of the matching procedure. Certain items should have been added to the pre-EMHQ, or there should have been a separate questionnaire with more items relating to normal personality and social undesirable characteristics such as selfishness and arrogance.

Since social desirability is multidimensional and is affected by the test situation, if the characteristics of the test situation does not induce strong motivation in the subjects, then they will not give the expected responses. Moreover, in the items with high or low social desirability, sometimes subjects will maintain a reserved disposition and select a neutral response, so that they may not exhibit the social desirability effect. Further investigation of this effect requires devising a test to encourage stronger motivation and to represent a more applicable, more valid SD scale as a marker. Thus far, the social desirability effect has had an obvious impact on the subjects. The suggestions above are to promote further research.

In brief, social desirability plays a major factor in affecting test scores in a clinical self-report inventory. If the effects are to be evaluated, it will be necessary to select test materials and appropriate statistical procedures with great care and to control the confounding variables deliberately. From a general overview, such research not only permits an understanding of the social desirability effect on the test-taking process, but provides a pointer for future research.

APPENDIX 1

The Results of the Difference between Corresponding Scales in the KMHQ and EMHQ According to the SDSVs and P(t)s

	SDSV		P(t)	
Subscale	T	Prob > |T|	T	Prob > |T|
1. Paranoia	9.64	.0001*	9.25	.0001*
2. Hypochondriasis	7.56	.0001*	9.96	.0001*
3. Social withdrawal	6.42	.0001*	4.89	.0001*
4. Confidence	2.28	.0339*	4.73	.0001*
5. Depression	8.75	.0001*	6.34	.0001*
6. Anxiety	6.91	.0001*	4.41	.0003*
7. Obsession–compulsion	3.65	.0013*	5.35	.0001*
8. Sexual inhibition	7.67	.0001*	6.01	.0001*
9. Aggression	7.18	.0001*	6.22	.0001*
10. Ego strength	2.09	.0481*	.44	.6649
11. Independence	4.30	.0002*	2.49	.0193*

*$p < .05$.

APPENDIX 2

The Means and Standard Deviations of the SDSVs and P(t)s for Each of the Corresponding Subscales Scored in the KMHQ and EMHQ

	KMHQ		EMHQ	
Subscale	SDSV	P(t)	SDSV	P(t)
1. Paranoia	2.19	.17	2.37	.34
2. Hypochondriasis	2.54	.21	2.66	.40
3. Social withdrawal	2.33	.27	2.47	.42
4. Confidence	3.30	.35	3.32	.48
5. Depression	2.40	.24	2.56	.37
6. Anxiety	2.37	.32	2.49	.44
7. Obsession–compulsion	2.87	.30	2.94	.42
8. Sexual inhibition	2.52	.22	2.65	.36
9. Aggression	2.18	.20	2.34	.37
10. Ego strength	3.57	.61	3.59	.62
11. Independence	2.85	.47	2.90	.51

REFERENCES

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Bentler, P.M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*(1), 67–74.

Cattell, R.B., Eber, H.W., & Tatsuoka, M.M. (1970). *The Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R.B., & Vogelman, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research, 12*(3), 289–325.

Crowne, D.P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.

Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.

Edwards, A.L. (1959). *Edwards Personal Preference Schedule*. New York: Psychological Corporation.

Edwards, A.L. (1970). *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart, & Winston.

Edwards, A.L., & Edwards, L.K. (1992). Social desirability and Wiggins's MMPI content scales. *Journal of Personality and Social Psychology, 62*(1), 147–153.

Ford, J.K., MacCallum R.C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology. A critical review and analysis. *Personnel Psychology, 39*(2), 291–314.

Frederiksen, N. (1965). Response set scores as predictors of performance. *Personnel Psychology*, *18*, 225–244.

Gordon, L.V. (1963). *Gordon Personal Inventory*. New York: Harcourt, Brace, and World.

Guilford, J.P. (1943). *An Inventory of Factors STDCR*. California: Sheridan Supply Co., Beverly Hills.

Harman, H.H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Ko, Y.H. (1976). The effect of the purpose of testing on EPPS and KMHQ responses. *Chinese Psychological Testing*, *23*, 53–56 (in Chinese).

Ko, Y.H. (1982). The factorial research of KMHQ. *Chinese Psychological Testing, 29*, 11–24 (in Chinese).

Ko, Y.H. (1985a). *KMHQ*. Taipei: Chinese Behavior Science Press (in Chinese).

Ko, Y.H. (1985b). *KMHQ Guidance*. Taipei: Chinese Behavior Science Press (in Chinese).

Lee, H.B., & Comrey, A.L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research, 14*(3), 301–321.

Mulaik, S.A. (1986). Factor analysis and psychometrika: Major developments. *Psychometrika, 51*(1), 23–33.

Nederhof, A.J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263–280.

Velicer, W.F., Peacock, A.C., & Jackson, D.N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research, 17*(3), 371–388.