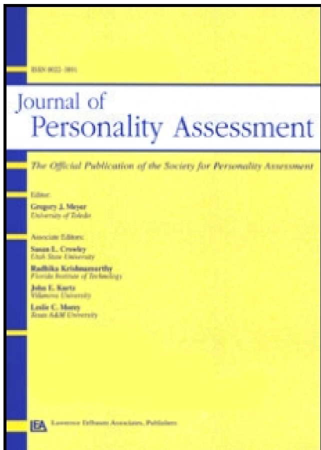

Journal of Personality Assessment
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653663

A Comparative Study of the Dimensionality of the Self-Concealment Scale Using Principal Components Analysis and Mokken Scale Analysis

Andreas A. J. Wismeijer [a]; Klaas Sijtsma [a]; Marcel A. L. M. van Assen [a]; Ad J. J. M. Vingerhoets [a]
[a] Tilburg University, Tilburg, The Netherlands

Online Publication Date: 01 July 2008

To cite this Article: Wismeijer, Andreas A. J., Sijtsma, Klaas, van Assen, Marcel A. L. M. and Vingerhoets, Ad J. J. M. (2008) 'A Comparative Study of the Dimensionality of the Self-Concealment Scale Using Principal Components Analysis and Mokken Scale Analysis', Journal of Personality Assessment, 90:4, 323 — 334

To link to this article: DOI: 10.1080/00223890802107875
URL: http://dx.doi.org/10.1080/00223890802107875

Journal of Personality Assessment, *90*(4), 323–334, 2008

ISSN: 0022-3891 print / 1532-7752 online
DOI: 10.1080/00223890802107875

STATISTICAL DEVELOPMENTS AND APPLICATIONS

A Comparative Study of the Dimensionality of the Self-Concealment Scale Using Principal Components Analysis and Mokken Scale Analysis

Andreas A. J. Wismeijer, Klaas Sijtsma, Marcel A. L. M. van Assen, and Ad J. J. M. Vingerhoets

Tilburg University, Tilburg, The Netherlands

We discuss and contrast 2 methods for investigating the dimensionality of data from tests and questionnaires: the popular principal components analysis (PCA) and the more recent Mokken scale analysis (MSA; Mokken, 1971). First, we discuss the theoretical similarities and differences between both methods. Then, we use both methods to analyze data collected by means of Larson and Chastain's (1990) Self-Concealment Scale (SCS). We present the different results and highlight the instances in which the methods complement one another so as to obtain a stronger result than would be obtained using only 1 method. Finally, we discuss the implications of the results for the dimensionality of the SCS and provide recommendations for both the further development of the SCS and the future use of PCA and MSA in personality research.

The decision to keep certain information secret often depends on situational demands, but personality and cultural and socioeconomic factors may also be relevant: Some people are more prone to secrecy than others. That is, some people feel relief when they disclose highly personal information, whereas others avoid to reveal the same information at all cost. As a consequence, part of secret keeping can be understood in terms of stable personality traits (Larson & Chastain, 1990; Pennebaker, 1989). A well-known trait in this context is *self-concealment,* which is defined as the "predisposition to actively conceal from others personal information that one perceives as distressing or negative" (Larson & Chastain, 1990, p. 440). Self-concealed personal information is a subset of private personal information that is conciously accessible to the individual and actively kept from the awareness of others. It is negative in valence and, if disclosed at all, usually confided to only a small number of persons because of its highly intimate content (Larson & Chastain, 1990).

To measure self-concealment, Larson and Chastain (1990) developed the Self-Concealment Scale (SCS). According to Larson and Chastain, the SCS items refer to three different facets of withholding personal information: (a) tendency to self-conceal, (b) possession of a personally distressing secret, and (c) apprehension about disclosure. The reported psychometric properties of the SCS such as internal consistency and test–retest reliability are satisfactory (Cramer & Barry, 1999; Larson & Chastain, 1990), and the total scale score has repeatedly been shown to have a positive significant correlation with depression, anxiety, and physical discomfort (Ichiyama et al., 1993; King, Emmons, & Woodley, 1992).

Received May 10, 2007; Revised July 11, 2007.

Address correspondence to Andreas A. J. Wismeijer, Tilburg University, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands; Email: a.a.j.wismeijer@uvt.nl

Although Larson and Chastain (1990) claimed that the SCS taps three distinct aspects of self-concealment, they did not explicitly design their questionnaire to measure these aspects. This leaves doubt with respect to the interpretation of the dimensionality of the SCS. In fact, Larson and Chastain were the first to report that the SCS could be viewed as bidimensional if Kaiser's (1960) eigenvalue-greater-than-1 criterion was used and as unidimensional because the first factor explained the lion's share of the total variance (i.e., 65%). Additionally, Larson and Chastain argued that the second factor was uninterpretable even after application of several rotation algorithms. More recent studies using the SCS have all reported similar results (Cramer & Lake, 1998; Ichiyama et al., 1993; King et al., 1992; Vögele & Steptoe, 1992).

In an attempt to confirm Larson and Chastain's (1990) proposed unidimensionality of the SCS, Cramer and Barry (1999) designed two consecutive studies using independent student samples. In the first study, Cramer and Barry assessed the total-score reliability (estimated by means of Cronbach's alpha) and explored the SCS's factor analytic structure; and in the second study, they analyzed the factor structure by means of a confirmatory factor analysis and evaluated both the total-score reliability and the retest reliability. Cramer and Barry's conclusions were essentially the same as those of Larson and Chastain: Exploratory factor analysis suggested two subscales, whereas confirmatory factor analysis and reliability analysis suggested a single-factor solution. Cramer and Barry (1999) concluded that "whereas the present study supported a unidimensional conceptualization, it does not preclude researchers from adopting the multifactorial model" (p. 636). In other words, the dimensionality of the SCS still remained unclear.

To determine the dimensionality of a scale, many statistical methods are available of which principal components analysis (PCA) probably is the most popular. PCA is an exploratory method that is used in situations in which little prior knowledge is available about the dimensionality of the data. This situation is typical of much test and questionnaire construction in personality research, and questionnaire construction with respect to self-concealment is no exception. Given the uncertainty about the true dimensionality of the data, a second opinion provided by a conceptually different method may be very useful. Mokken scale analysis (MSA; Mokken, 1997; see also Hemker, Sijtsma, & Molenaar, 1995; Sijtsma, 1998; Sijtsma & Molenaar, 2002) can be used for this purpose. MSA is a method from item response theory (Embretson & Reise, 2000; Van der Linden & Hambleton, 1997). Item response theory defines models for the measurement of, for example, personality traits. These models put forward assumptions about the dimensionality of measurement and offer techniques to explore and test hypotheses about dimensionality.

Our aim in this article is twofold: (a) to provide a theoretical introduction to MSA and compare PCA with MSA and (b) to apply PCA and MSA to the SCS to scrutinize its dimensionality. Because our goal is to compare techniques that can be used to explore a scale's dimensionality, we do not compare the results of PCA and MSA to confirmatory factor analysis. In confirmatory factor analysis, based on a hypothesis about the composition of the scale, one specifies both the dimensionality of the scale and which items group on the same dimension. In this study, however, we did not start from such explicit hypotheses but instead explored the dimensionality and the item assignment to the dimensions.

We start following with a description of the typical situation in which personality questionnaires are constructed. Then we provide outlines of PCA and MSA, enumerate the properties of both methods, and compare the methods with respect to these properties. Subsequently, we apply both methods to the SCS data and discuss the results, highlighting the instances in which the methods complement one another so as to obtain a stronger result than would be obtained using only one method. Finally, we discuss the implications of the results for the dimensionality of the SCS and provide recommendations for both the further development of the SCS and the future use of PCA and MSA in personality research.

Preliminary Considerations on Personality Measurement

Ideally, the construction of a personality inventory rests on a well-established theory about the trait to be measured. However, for many traits, such a theory has not been well developed and sometimes a "theory" consists only of a set of intuitions and experiences that have yet not been articulated well enough to have been exposed to rigorous hypothesis testing. Also, in the absence of a theory, test construction may lean heavily on the experience of other researchers and existing questionnaires that have the status of "try-out" instruments. Additionally, it may be an open issue whether an existing instrument, even if it has been based on sound theory, can be used in a population in which it has not been used before. All these instances have in common that it is unknown or at least uncertain what the test measures exactly and whether one or more traits are involved in driving responses to the items, that is, whether the inventory induces responses that are driven by one trait (hopefully the intended trait) or a multitude of traits.

Applied to self-concealment, it may thus be hypothesized that self-concealment is a unitary trait that drives responses to the items from a questionnaire such as the SCS; but because little is known about this trait, it may be difficult to select a collection of items that cover only this trait. Actually, because of this uncertainty, the researcher may have selected items that appear to be good choices at first sight but that turn out to be ambiguous stimuli when the data collected by means of these items are analyzed by means of PCA, MSA, or another method. Because this particular outcome may not have been anticipated, as is typical of exploratory research, the use of several statistical methods for assessing the dimensionality of the data may shed a new and different light on the trait structure. This may help to formulate hypotheses on the trait structure that are tested in future research and to construct better measures for the trait(s).

Principal Components Analysis

Although quite well known, we first explain PCA in some detail so that it can be compared with MSA later on. Suppose a trait is measured by means of rating scale items. These items typically consist of a statement, for example, about self-concealment, and following this statement a small number of boxes that represent different ordered levels of endorsement with the statement. The direction of scoring the items depends on whether the statement is positive or negative with respect to the trait.

Let the questionnaire contain J items, and let items be indexed $j = 1, ..., J$. If the number of ordered boxes in each item is $m+1$, we assign scores $0, \ldots, m$ such that a higher score reflects a higher trait level. Let X_j be the random variable for the score on item j, and let x_j be a score on this item; here, the scores are $x_j = 0, ..., m$. The total score is defined as the sum of the J item scores:

$$X_+ = \sum_{j=1}^{J} X_j;$$

notice that the total score can vary from 0 to $J \times m$. High values for the total score are assumed to reflect higher levels on the trait, but this assumption is only correct if the items are all driven by the same trait that represents, in our case, self-concealment. PCA may be used to investigate whether this assumption is correct.

PCA basically does the following (see, e.g., Gorsuch, 1983; Nunnally, 1978, for more details). A weighed sum of the J item scores, technically also known as a linear combination, is constructed. This weighted sum is denoted by C_1 (subscript 1 indicates that this is the first weighed sum score and that others will follow), and item weights are denoted by w_{j1} (subscript j says that this is the weight for item j, and subscript 1 that it is the weight that is used to construct C_1). Then, we have that

$$C_1 = \sum_{j=1}^{J} w_{j1} X_j.$$

The weights w are sought such that the variance of sum score C_1 is maximal (given a constraint on the weights that serves to find a unique solution to the problem; this need not bother us here). Weighed sum C_1 is the first principal component. After C_1 has been determined, a second weighed sum (the second principal component) is sought, this time taking the weighed sum of the residuals of the item scores after their regression on C_1 has been subtracted. This sum is denoted C_2, and the use of the item score residuals ensures that the correlation between the two principal components equals 0. Like the first time, the weights are chosen such that the variance of C_2 is maximal. Similarly, a third weighed sum, a fourth, and so on, are constructed. Typically, each next principal component explains less variance than its predecessor. The maximum number of principal components that can be determined equals the maximum number of items, J.

The J principal components together explain all the variance in the J items; thus, the J principal components and the J items contain exactly the same amount of information. However, PCA "piles up" as much of the variance from the items as possible in the first principal component, and then it piles up as much as possible from the remaining variance in the second principal component and so on. Looked at PCA this way, it is a method to summarize as much information from a set of items in as few principal components as possible.

Piling up as much information in as few principal components as possible does not automatically result in well-interpretable components. To obtain a solution that is well interpretable, usually the first few, say M, principal components that explain much variance relative to the amount explained by one item are subjected to further analysis. Several techniques may be used for selecting these M principal components. One well-known method is a graphical method proposed by Cattell (1966), called the scree test. The scree plot shows the amount of explained variance as a function of the rank number of the principal component. What one usually sees in the plot is that these successive amounts decrease rapidly (steep descent) and then tend to level off. M is determined by retaining those principal components in the steep descent before the first one on the line where the magnitude tends to level off. The scree plot requires subjective judgement especially if the elbow is not sharp.

Kaiser's (1960) eigenvalue-greater-than-1 criterion is another well-known method. Here, the number M is determined by the number of principal components that explain more variance than the variance of one individual item. Unlike the scree plot, this is an objective criterion but one that can easily lead to the selection of too many principal components, and the result may be the overestimation of the dimensionality, as factors are sometimes split into bloated specifics (e.g., Kline, 1987; Rummel, 1970).

The analysis to which the M selected components are subjected next is called *rotation*. Algebraically, rotation means that M principal components are factor analyzed—rotated, in technical parlance—to obtain M new factors that are more interpretable. Geometrically, rotation means the following. The J items can be displayed as points in an M-dimensional orthogonal space. The M axes represent the M principal components, and the coordinates of the items are the loadings of the items on the principal components. These loadings are denoted by a_{jc} (c indexes principal components; $c = 1, \ldots, M$). Indeed, rotation means that the axes are rotated around the origin into a new position, whereas the items maintain their position. So one obtains new axes, now called *factors,* and for each item a new set of coordinates, which are their loadings on the factors. Rotation is done in such a way that the resulting sets of loadings facilitate the interpretation of the factors. Many rotation methods exist (e.g., Gorsuch, 1983, pp. 203–204). An important distinction is between rotation resulting in orthogonal/uncorrelated factors (these are the geometrical and algebraic interpretations, respectively), as with varimax rotation, and oblique/correlated factors as with oblimin rotation.

PCA in combination with a rotation method is typically used for exploring the dimensionality of a data set collected by means of a personality questionnaire. In the absence of pronounced expectations, the method provides the most efficient summary of the data by means of principal components. The number of principal components retained for rotation is often determined by criteria that are not perfect and invite some trial and error. Whenever researchers believe that correlated factors are more realistic than uncorrelated factors, oblique rotation is preferred over orthogonal rotation. Although correlated factors may more realistically reflect a trait structure, the interpretation of the pattern of loadings may be more complicated due to their conditioning on the correlations between the factors. This means that the proportions of item variance explained by each of the factors can no longer be added across factors and that squared (semi) partial correlations are needed instead (cf. regression analysis).

PCA has gained enormous popularity among personality researchers (see, e.g., the volumunous literature with respect to the Big Five; e.g., De Raad & Perugini, 2002). This is probably due to its quick and easy applicability using one of the standard statistical software packages. At the basis of much personality research often stands a theory about the structure of the personality aspects of interest, but the final structure of the questionnaire regularly is the result of PCA plus rotation or other factor analysis methods (e.g., Cattell's, 1956–1957, 16 PF questionnaire, which allows measurement of 16 bipolar dimensions of personality summarized by five factors). The respondent is assigned sum scores on each item sets that load on a particular factor and thus identify a particular trait or an aspect of a more comprehensive trait. These scores can be weighed sums of item scores with weights resulting from PCA plus the rotation method used. However, because such weighed sum scores tend to correlate high in the 90s with unweighed sum scores like X_+, in practical test use, often the unweighed sum scores are used. The resulting profile of scores is then used for personality diagnosis.

MOKKEN SCALE ANALYSIS

The Monotone Homogeneity Model

Assumptions of the Monotone Homogeneity Model. Like PCA, MSA can be used to identify one or more dimensions in the data, but in addition to PCA, the method does this in such a way that the items selected in one cluster satisfy a measurement model known as the monotone homogeneity model (MHM; Mokken & Lewis, 1982; Sijtsma & Molenaar, 2002). This model implies that the persons can be ordered on a scale using the items in a selected cluster. Thus, MSA provides a method for dimensionality investigation and a measurement model in one technique. MSA has been used frequently in psychology (e.g., Michielsen, De Vries, Van Heck, Van de Vijver, & Sijtsma, 2004) but also in political science research (e.g., Van Schuur, 2003), marketing research (e.g., Paas & Molenaar, 2005), and social-medical research (e.g., Roorda et al., 2005) for assessing the dimensionality of the data and constructing scales.

The MHM is an item response theory model (Junker, 2001; Sijtsma & Meijer, 2007; Sijtsma & Molenaar, 2002; Stout, 2002). Item response theory models are based on assumptions that restrict relationships between items and underlying traits. In particular, the MHM is based on two assumptions about the dimensionality of the data and one about the relationship of items with underlying traits:

1. *Unidimensionality;* that is, all items measure the same latent trait (denoted θ). An example is that all items measure self-concealment, assuming this is a unitary construct. Unidimensionality is generally considered a desirable property of measurement primarily because it simplifies the interpretation of answers to the items in a questionnaire.
2. *Monotonicity;* that is, the higher a respondent's disposition on the latent trait the more likely it is that (s)he obtains higher scores on the items measuring that latent trait. For example, people who have a stronger tendency to self-conceal (θ) are more likely to obtain higher scores on positively phrased rating-scale statements with respect to secrecy. More precisely, for item score X_j (e.g., $X_j = 0, ..., 4$, for many rating scales) the mathematical function that expresses this monotone relationship is called the item step response function (ISRF), and it is given by conditional probability $P(X_j \geq x|\theta)$. For $x = 0$, the ISRF $P(X_j \geq 0|\theta)$ equals 1 by definition, only expressing that everybody has one of the five possible item scores; but for $x \geq 1$, the ISRFs are assumed to be monotone, nondecreasing in the latent trait θ.
3. *Local independence;* that is, an individual's response to item j is not influenced by his or her responses to the other items in the same questionnaire. This assumption rules out dependence among items due to an answer to one item influencing the answers to the next items. This could happen, for example, because respondents develop new ideas or personal hypotheses about self-concealment from having answered previous items in the same questionnaire. Although developing such ideas as one goes through the questionnaire is not entirely unlikely, psychometricians consider this a case of "obtrusive" measurement, which is considered undesirable.

These assumptions together define three desirable properties of a measurement instrument. Many researchers will want their instrument to measure one latent trait—unidimensionality—and not an obscure or complicated mixture of influences on item responses leading to uninterpretable measurement results. Also, they will likely find it reasonable that a higher trait level increases the probability that people obtain higher item scores—monotonicity—and they may also agree that respondents should preferably approach each new item independent of the previous items so as not to deliberately construct a particular, possibly invalid image of themselves: local independence.

The three assumptions of the MHM together imply person measurement. That is, if the MHM can be shown to fit the data well, then the total score X_+ can be used to order people on the latent trait θ. Due to the limited number of items usually included in a test, this ordering is liable to random error. The MHM also provides a method to assess the accuracy of this ordering. We discuss this method later on. Ordinal measurement is important because the practical use of many tests and questionnaires requires that people can be ordered on a scale as in "Mary has a higher level of self-concealment than Judy," and "On average women keep more secrets to themselves than men."

The MHM and PCA

An important distinction between MSA and PCA is that the link between the underlying traits and the item scores is less explicit in PCA than in MSA. Mathematically, PCA is based on solving a problem in matrix algebra that finds the J eigenvalues (corresponding to the percentages of explained variance of each of the principal components) of the correlation matrix of the item scores. Because it uses the interitem correlations, PCA assumes that relationships between item scores are linear. It is well known (e.g., Nunnally, 1978, pp. 141–146) that correlations are heavily distorted as soon as the items have different distributions of item scores, in particular when items have only a few (less than, e.g., five) answer categories. Then PCA results in so-called difficulty factors; these are factors that are due to the differences between item-score distributions in addition to what the items measure in common. This problem is absent when item-score distributions are equal, but this is a rare situation and in fact one that is actively avoided by test constructors when they seek the items to represent different intensity levels with respect to the trait of interest. It is sometimes advocated to use tetrachoric or polychoric correlations, but it has also been noted that this introduces other methodological problems (Van Abswoude, Van der Ark, & Sijtsma, 2004).

The near absence of assumptions on the relation between items and underlying traits—only linearity is assumed—means that PCA is mainly an instrument to assess the dimensionality of the data. Thus, the outcome of a PCA followed by some rotation logically results in dimensions and not in scales. Of course, we know that the practice of test construction assumes that a PCA also results in scales, but the absence of assumptions that restrict relationships between items and the underlying traits does, strictly speaking, not allow such conclusions. Another interesting observation in this context is that PCA uses the eigenvalues of the interitem correlation matrix, thus assuming that the 1's on the diagonal represent true-score variance only. Thus, an implicit assumption of PCA is that error variance is 0, an assumption that is usually seen as unrealistic and that is not shared by factor analysis models.

Fit and Misfit of the MHM to the Data

The MHM and its three assumptions seem to be intuitively appealing, but this does not mean that they are congruent with the true mental processes. Thus, the data collected by means of a questionnaire may sometimes be multidimensional; the relation between item score and latent trait may not always be monotone; and while filling out the inventory, some respondents may feel inclined to construct a particular coherent, perhaps also socially desirable image of themselves that only partly reflects their true level of self-concealment. In other words, other sources that drive item responses, such as social desirability, might be active, thus introducing dependence among items in addition to that caused by the trait of interest. This violation of local independence tends to manifest itself in a multidimensional data structure (e.g., Stout, 2002).

Because there is always the possibility of a discrepancy between the model on one hand and the data on the other hand, it must be checked whether a particular model indeed fits the data. If it fits, by implication, the properties of the model, such as an ordinal person scale, hold for the data. If it does not fit, a useful alternative course of action is to investigate the possibility of multidimensionality; that is, the alternative that several latent traits drive item responses and that different sets of items can be identified, each of which measures a different latent trait.

The MHM and the Rasch Model

Finally, it may be interesting to briefly compare the MHM to the Rasch model (e.g., Embretson & Reise, 2000; Van der Linden & Hambleton, 1997), which is a related and popular item response model. The simple Rasch model is defined only for dichotomous item scores (yes–no, correct–incorrect; implying only two different item scores), but generalizations to rating scale item scores also have been proposed (e.g., Masters & Wright, 1997). For simplicity, we restrict the comparison of the MHM and the Rasch model to dichotomous items.

Like the MHM, the Rasch model assumes unidimensional and locally independent measurement; but unlike the MHM, the Rasch model restricts the monotonicity assumption to logistic curves with equal positive slopes (equal discrimination, in technical parlance). Thus, the Rasch model is a special case of the MHM (for dichotomous items): All logistic response curves with equal positive slopes obviously are monotone (thus the Rasch model is also an MHM model, albeit a specialized one), but not all monotone response curves are logistic, let alone have equal positive slopes (thus the MHM model does not imply the Rasch model; the MHM is a more general item response model; see Sijtsma & Molenaar, 2002, chap. 2).

An advantage of the more restrictive Rasch model is that persons can be measured on an interval scale. This may be interesting in an educational measurement context in which different item subsets from item banks (i.e., large numbers of calibrated items all measuring, e.g., the same ability) are used to measure achievement in different groups, thus afterwards necessitating the equating of the different scales for comparing the performance of the different groups. The disadvantage of the Rasch model is that requiring the items to all have the same discrimination seriously restricts the likelihood that the Rasch model will fit the complete data. This may even be too great a limitation in achievement testing (in which different items may be relatively similar), and for that reason other, less restrictive item response models are often used such as the 2- and 3-parameter logistic models (e.g., Embretson & Reise, 2000). In personality testing, items are often intentionally very different so as to adequately cover the personality trait of interest. Here, there is no reason to require response curves to have equal positive slopes. Such a requirement would lead to the rejection of many items that cover an interesting aspect of the trait and have monotone response curves, thus contributing to ordinal person measurement.

Evaluating the Fit of the MHM to Data

The investigation of whether the MHM fits the data well is often captured under the name of MSA (Sijtsma & Molenaar, 2002). Researchers tend to narrow an MSA down to the application of an automated item selection algorithm that serves to identify subsets of items that each are driven by different traits and that each allow the ordering of people with a user-specified degree of accuracy. We also use this algorithm to investigate the dimensionality of the data. In a separate subsection, we also explain the investigation of monotonicity so as to be able to evaluate the fit of the MHM to our SCS data.

Investigating Dimensionality of Data

Scalability coefficients. A typical MSA starts with the investigation of the dimensionality of the data and the identification of subsets of items that constitute a scale of which the accuracy is controlled by the researcher. The tool for this is the scalability coefficient for pairs of items, j and k, which is denoted as H_{jk}. Coefficient H_{jk} equals the ratio of the items' covariance and their maximum covariance given the items' univariate score-frequency distributions (Molenaar, 1997). This definition avoids the problems with respect to the distorting effect of different item-score distributions on the interitem correlations alluded to previously, and as a result, an MSA does not yield artifactual "difficulty factors."

Based on the item-pair coefficients, H_{jk}, item coefficients, denoted as H_j, are defined that express the degree to which an item is related to the other items in the scale. Under the MHM, a stronger relationship expressed by a higher H_j value means that the ISRFs are steeper. Thus, a higher H_j value means that people with relatively low latent trait values and people with relatively high latent trait values are better separated: Almost all people with low trait values score low on the item, whereas almost all people with high trait values score high on that item. This is seen as a desirable property of a measurement instrument: Items that do not or only weakly separate people on the scale of interest contribute only little to an accurate ordering.

Finally, the total-scale coefficient, denoted as H, expresses the degree to which the total score X_+ accurately orders persons on the latent trait scale θ. Given a certain number of items, higher H values express a more accurate person ordering that is, as was seen earlier, a desirable property of measurement instruments.

Each of the item-pair, individual-item, and total-scale coefficients are related as follows to the MHM. The MHM implies that each coefficient has a value between 0 (minimum) and 1 (maximum; i.e., $0 \leq H_{jk}, H_j, H \leq 1$), so that negative values (i.e., $H_{jk}, H_j, H < 0$) are in conflict with the model. In particular, negative H_j values may lead to the identification and the rejection of one or more items from the scale.

Definition of a unidimensional scale. The H_{jk}, H_j, and H coefficients are the basis of the definition of a scale (due to Mokken, 1971, p. 184; also see Sijtsma & Molenaar, 2002, pp. 67–68). Unlike Mokken, in this definition, we use item-pair coefficient H_{jk} instead of product moment-correlation ρ_{jk} (although formally this choice is unimportant, but using H_{jk} fits better in with the discussion thus far). Positive item-pair coefficients H_{jk} are implied by the MHM, and negative values are often interpreted as a sign of multidimensionality. For item coefficient H_j, we introduce a positive lower bound denoted by c, which ascertains that items that belong to a scale provide a contribution to the accuracy of person ordering that exceeds this lower bound. Thus, lower bound c should be seen as a safeguard against weak items that contribute little to accurate person ordering. Then, a scale is defined as a set of items for which, given a value of c, the next two conditions are satisfied:

1. $H_{jk} > 0$, all item pairs $j, k; j \neq k$; and
2. $H_j \geq c > 0$, all items $j = 1, ..., J$.

The researcher can control the lower bound c; and the higher c, the better an item separates persons with low trait values from those with high values and the more accurate the person ordering given that number of items.

An important relationship between item coefficients H_j and total-scale coefficient H is that $\min(H_j) \leq H \leq \max(H_j)$; thus, by requiring that items can only be admitted to the test if they have item scalabilities of at least c, the researcher also controls the minimum of the H coefficient.

Choosing a minimal lower bound of $c = 0$ poses almost no demands on the scalability of the items except that all selected items have a positive H_j. This may result in a single scale with an unacceptably low total-scale H value because several low-quality items have been selected that have H_js close to 0, meaning that they hardly contribute to an accurate person ordering. Increasing the lower bound c has the effect of admitting fewer items to the scale, but the selected items are better indicators of the latent trait.

Positive values of H are interpreted using the following rules of thumb. A set of items is considered unscalable for practical purposes if $H < .3$ (e.g., Sijtsma & Molenaar, 2002, p. 60); that is, although $0 \leq H < .3$ agrees with the MHM, from a practical point of view, such low values suggest that person ordering is inaccurate. Other rules of thumb are that $.3 \leq H < .4$ indicates a weak scale, $.4 \leq H < .5$ a medium scale, and $H \geq .5$ a strong scale (Mokken, 1971, p. 185; Sijtsma & Molenaar, 2002, p. 60).

Automated item selection. The dimensionality of an item set and the scalability (expressed by H) for each item subset that represents a dimension can be investigated by means of the statistical package MSA for Polytomous items (MSP; Molenaar & Sijtsma, 2000). A lower bound value c for H_j is chosen optionally (its default value equals 0.3). The algorithm used sequentially clusters items. The items are selected one by one into disjoint clusters, each of which is relevant for the assessment of a different latent trait. The algorithm consists of three steps:

Step 1: Select as starting pair the two items that have the greatest, positive significant H_{jk} value of all item pairs among the J items in the item pool, that is also greater than the user-specified constant c ($0 \leq c \leq 1$). Hemker et al. (1995) recommended trying several values for c, starting at $c = 0$, and then using increments of .05 in each step until $c = .6$ (or higher). The development of clusters as c increases provides more insight into the scalability of the items than the evaluation of item clustering for only one c value. Denote the starting pair (j, k_1).

Step 2: From the remaining J-2 items, select the item, say, k_2, that (a) correlates positively with items j and k_1; (b) has an H_{k_2} value with respect to items j and k_1 that is significantly greater than 0 and also greater than lower bound c; and (c) maximizes the H value of items j, k_1, and k_2 together given all possible choices of a third item from the remaining J-2 items.

Step 3: Repeat Step 2 for the selection of a fourth item, say, k_3, from the remaining J-3 items; and so on.

The formation of the first cluster of items stops if no more items remain that can be selected so as to satisfy the requirements in Step 2. The items selected up to that point constitute the first scale. If items remain that do not satisfy the requirements in Step 2, the selection procedure goes on to find a second scale

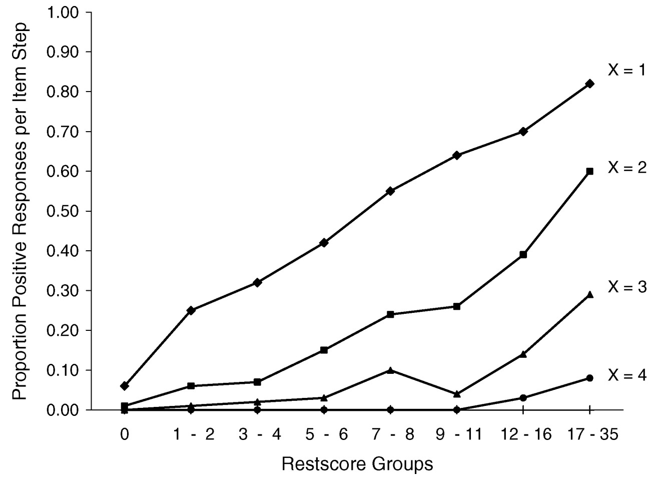

FIGURE 1.—The four estimated item step response functions of Item 10 for minimum group size of 150.

from these unselected items using the same algorithm and, if possible, a third scale, and so on. Items that cannot be selected in one of the clusters remain nonscalable. Instead of letting the algorithm select the starting pair in a cluster, the researcher may define his or her starting set of at least two items, for example, if theory would strongly suggest such a choice. See Van Abswoude et al. (2004) for more information on scalability and dimensionality analysis in the context of item response theory.

Investigating Monotonicity of ISRFs

Selecting items with item scalabilty values of at least c (i.e., $H_j \geq c$) will usually select items in scales of which the ISRFs tend to be monotone; but for smaller values of c, the selection of items with small H_j values may result in the admission of ISRFs with serious nonmonotonicities. An example of nonmonotonicity is depicted in Figure 1. Figure 1 shows the ISRFs of the SCS's Item 10. The horizontal axis represents the scale's rest score, that is, the total score (X_+) minus the score on Item 10, in eight different intervals. The vertical axis shows the proportion $P(X_{10} \geq x|\theta)$ for $x = 1$ (highest curve) through 4 (lowest curve). There is one small nonmonotonicity for $x = 3$ from rest score interval 7 through 8 to interval 9 through 11. This nonmonotonicity does not much impair this item's contribution to the person ordering based on the total score from all SCS items (the item's scalability coefficient still is $H_{10} = 0.39$).

ISRFs of items may show more serious violations of the monotonicity assumption, which result in low H_j values and little if any contribution of the item to accurate person ordering. The combination of a positive item-scalability coefficient and nonmonotone ISRFs can be compared with a positive regression coefficient in a linear regression model that is fitted to data from a nonlinear model: The positive sign of the regression coefficient does not imply that the underlying model is linear, only that the relationship between the variables shows a positive tendency.

Despite the positive H_j value, nonmonotone ISRFs may disturb the ordering of persons using the test total score (Junker & Sijtsma, 2000; Sijtsma & Meijer, 2007). These nonmonotonicities can occur with items that satisfy the formal definition of a scale because item selection is based on scalability coefficients exceeding lower bound c, and although this restriction alone already excludes many items of which the ISRFs show serious nonmonotonicities, it does not rule them out. Thus, an MSA also has to consider the monotonicity assumption of the ISRFs. For this purpose, the MSP statistical program offers the possibility to estimate the complete ISRFs of an item from the data and to assess their monotonicity. MSP shows the graphical display of the ISRFs, such as those in Figure 1, and tests observed nonmonotonicities for significance.

Comparison of PCA and MSA using the SCS

We already mentioned three differences between PCA and MSA in passing (also see Scheirs & Sijtsma, 2001):

1. PCA is suited for dimensionality analysis, but it is not a measurement model that implies particular scale properties. In practice, for each separate factor, the factor scores or the unweighed sum scores X_+ are used to order people, but this ordering does not follow from assumptions on which PCA is based. On the other hand, MSA is based on a measurement model (the MHM) and is aimed at fitting this model to the data. In doing so, MSA therefore not only allows for dimensionality analysis but also investigates the fit of a measurement model to the data. This also entails the investigation of the monotonicity assumption, and an MSA sometimes also involves other analyses that were beyond the scope of this study.
2. Because it is purely a computational technique, PCA always results in J principal components for any set of items irrespective of whether these components are deemed useful or not. On the other hand, MSA is based on assumptions about trait dimensionality and relationships between items and traits. These assumptions may either be supported or refuted by the data. In the latter case, we have failed to construct a scale. The possibility to reject a set of items as a measurement scale is an advantage of MSA that helps moving measurement and theory building forward.
3. PCA can be based on tetrachoric or polychoric correlations, thus avoiding artifacts in factor solutions (known as *difficulty factors*) by eliminating undesirable effects of the difference in discrete item-score distributions on interitem correlations. MSA solves this problem by using the H_{ij} coefficients that normalize interitem covariances against the maximum covariances given the discrete item-score distributions (Michielsen et al., 2004; Van Abswoude et al., 2004).

Two additional important differences are the following:

4. PCA constructs a weighted linear combination of the J item scores by considering the association of all J items simultaneously irrespective of whether they are driven by the same trait. This almost always results in components that do not reflect the dimensionality of the data, thus necessitating rotation of a smaller number of components to obtain better interpretation. MSA selects items one by one, but this approach may result in suboptimal outcomes. This happens when an item has been selected into a cluster because it fitted well together with the items already selected (in terms of H_j), but items selected in later steps had the effect of reducing H_j more and more until it was unacceptably low when the final cluster was assembled (although this effect is known to be small in practice). Of course, the researcher may then decide to remove this item after all.
5. The choice of a lower bound for c in MSA is done before the item selection algorithm starts and strongly affects the outcome of item selection. For example, a low c value admits many items to the same cluster even though they are driven by different traits. Thus, several choices of c should be tried, and conclusions about dimensionality should be based on the pattern of cluster outcomes that emerges as c increases. With PCA (and rotation), the choice of the minimally acceptable factor loading a does not affect the computations, but it does affect the interpretation of the factor solution. Here, the decision about the number of components to be retained and the rotation method to be chosen and executed has to be taken prior to the choice of minimum a. A similarity with MSA could be the following: Just as MSA will come up with several solutions for different choices of c, different item clusters based on varying choices of a can be tried in PCA. However, unlike MSA, trying varying choices of a does not involve new computations. Many researchers actually try different values of a in an effort to find satisfying solutions for item clustering.

In the next section, we present a practical case using both PCA and MSA to analyze test data. These data are collected by means of the SCS (Larson & Chastain, 1990). In this analysis, we highlight differences and similarities of both methods and how they can be used to complement one another.

Method

Analysis Methods

First, we analyzed SCS data using PCA. We used both the scree plot and the eigenvalue-greater-than-1 criterion to determine the number of components to be retained for rotation. Then, because little was known about the self-concealment trait, we used both varimax rotation and oblimin rotation to obtain factors that were well interpretable. We tried several choices of minimum loading for this purpose, and we compared the results. Second, we used MSA for two purposes. Initially, we used the automatic item selection procedure from the MSP program to find the item clustering for several lower bound values c for scalability, starting with $c = 0$, and increasing c with steps of 0.05 until $c = 0.7$. Finally, after a decision had been made with respect to the dimensionality of the SCS data, we investigated the monotonicity assumption at several levels of precision and drew a conclusion about the accuracy with which the scale(s) order(s) persons accurately.

SCS

The SCS (Larson & Chastain, 1990) consists of 10 positively phrased items. Each item is evaluated on a 5-point rating scale (0 = *does not apply to me;* 1 = *somewhat applies to me;* 2 = *moderately applies to me;* 3 = *strongly applies to me;* and 4 = *completely applies to me*), with a higher total scale score (the sum of scores on all items of the scale, X_+) suggesting a higher level of self-concealment. The 10 items of the SCS are shown in the Appendix. The items refer to the kind of information that the self-concealer keeps secret from others without explicitely referring to the content of that information. Some items measure either the possession of secrets or the tendency to have

TABLE 1.—Item descriptives for total sample and men and women separately and results of two-tailed *t* tests and Cohen's *d* values comparing scores of men and women.

	Total (n = 1,503)			Males (n = 829)			Females (n = 674)			
Item	*M*	*SD*	Skew	*M*	*SD*	Skew	*M*	*SD*	Skew	*d* for Gender
1	.82	1.25	1.34	.76	1.21	1.22	.88	1.29	1.23	−.10
2	.58	.95	1.73	.58	.93	1.44	.58	.97	1.82	.00
3	1.11	1.12	.75	1.11	1.11	.73	1.11	1.13	.78	.00
4**	.65	1.10	1.69	.56	1.05	1.89	.75	1.16	1.48	−.17
5***	1.43	1.19	.35	1.53	1.20	.23	1.31	1.18	.50	.18
6*	.84	1.04	1.11	.79	1.01	1.20	.91	1.07	1.01	−.12
7	.44	.88	2.15	.40	.80	2.14	.49	.96	2.09	−.10
8	.73	1.20	1.51	.70	1.13	1.51	.77	1.28	1.48	−.06
9	.47	.92	2.06	.44	.87	2.15	.52	.98	1.94	−.09
10*	.72	.98	1.27	.67	.96	1.39	.78	.99	1.14	−.11
Total score	7.80	7.19	1.16	7.56	6.91	1.22	8.10	7.51	1.09	−.08

*$p < .05$. **$p < .01$. ***$p < .001$.

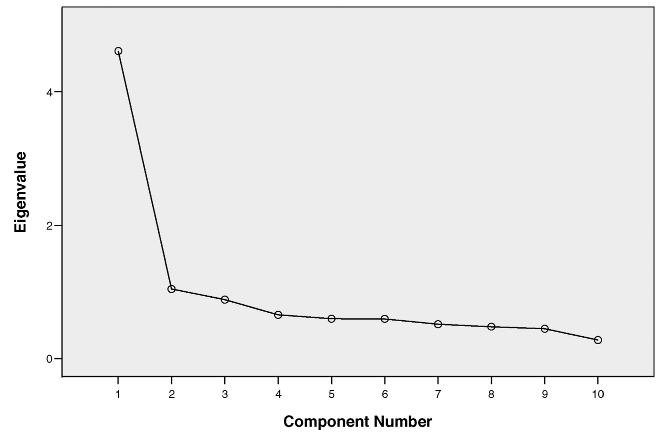

FIGURE 2.—Scree plot of the Self-Concealment Scale.

secrets and the emotional consequences of this tendency, some items refer to the way respondents deal with social interactions while having a secret, and other items focus on the perception self-concealers have of the consequences should others know their secret.

Study Participants

The sample consisted of 1,503 participants representative of the Dutch general population, of which 829 were men (55%) and 674 women (45%). The mean age was 44.19 years (*SD* = 15.10), and age ranged from 16 to 85 years. All participants were members of an internet-based telepanel and completed a computer-administered test battery. There were no missing values.

RESULTS

SCS Item Descriptives

The mean total SCS score was 7.56 (*SD* = 6.91) for men and 8.10 (*SD* = 7.51) for women. This difference was not statistically significant, $t(1501) = -1.45$, $p = .15$, two-tailed, Cohen's $d = -.08$. Table 1 shows the mean, standard deviation, and skewness for all items for the total sample and for men and women separately. Note that although total scores did not differ significantly, men scored higher on average than women on one item on tending to keep bad things for oneself (Item 5) and lower on three items (4, 6, and 10).

Application of PCA to SCS Data

First, Cronbach's alpha for the 10 items together was .86. Such a high value may indicate scale unidimensionality (Nunnally, 1978, chap. 8). PCA yielded a first principal component that explained 46.1% of the total variance (eigenvalue $\lambda_1 = 4.61$) and a second principal component that explained 10.3% ($\lambda_2 = 1.03$). The other principal components had eigenvalues smaller than 1. The scree plot (Figure 2) shows a sharp bend at λ_2, suggesting that only the first component should be retained. The eigenvalue-greater-than-1 criterion identifies two components to be retained, although it must be noted that the second eigenvalue is only just larger than 1.0. These results agree with those that have been reported in the literature; thus, we consider both a single-factor solution and a two-factor solution.

To interpret the one-factor solution, we evaluated the corrected item-total correlation for each item (Nunnally, 1978, chap. 8), that is, the correlation of each item with the total score on the remaining nine items. The corrected item-total correlations are shown in the second column of Table 2. The substantial corrected item-total correlations (ranging from .45–.70) support a single dimension. Because we did not have a sound theoretical basis to expect either orthogonal or oblique dimensions (Nunnally, 1978, chap. 12), the two-factor solution was studied by means of both varimax (orthogonal) and oblimin (oblique) rotation. These rotation methods yielded comparable pattern coefficients. Oblimin rotation yielded factors that correlated .54. Table 2 (third and fourth columns) shows the pattern coefficients after oblimin rotation. The first factor was composed of six items with coefficients between .46 and .93 (and Cronbach's

TABLE 2.—Corrected item-total correlations (second column) and pattern coefficients after oblimin rotation of two principal components (third and fourth columns).

Item No.	Corrected Item-Total Correlations	F1	F2
8 (secret so private I'd lie when asked)	.64	**.93**	–.15
9 (secrets too embarrassing to share)	**.70**	**.88**	–.04
1 (important secret not shared with anyone)	.62	**.76**	.01
4 (secrets tormented me)	.62	**.65**	.13
2 (friends like me less)	.59	**.52**	.25
7 (telling secret backfires, regret)	.52	**.46**	.23
5 (tend to keep bad things for myself)	**.45**	–.15	**.84**
6 (afraid to reveal without wanting)	.55	.09	**.69**
10 (negative thoughts about myself not shared)	.52	.10	**.65**
3 (many things about me I keep to myself)	.63	.26	**.58**

Note. Lowest and highest corrected item-total correlations (second column) are printed in boldface as well as the highest pattern coefficients for each item (third and fourth columns).

TABLE 3.—Mean item scores and item H values for the total group and for men and women separately.

	Item H		
Item No.	Total	Men	Women
5	.37	.38	.37
7	.39	.42	.37
10	.39	.40	.38
6	.41	.41	.41
2	.44	.42	.46
1	.45	.42	.48
4	.45	.42	.48
8	.46	.45	.47
3	.47	.48	.47
9	.52	.50	.54

$\alpha = .84$), and the second factor was composed of four items with coefficients between .65 and .84 (and Cronbach's $\alpha = .72$).

Application of MSA to SCS Data

Results for all items considered as one scale. First, the 10 items from the SCS were considered as one scale. As can be seen in Table 3, the ordering of the items according to scalability as measured by H_j were comparable although somewhat different for men and for women relative to the total group and between men and women. For both groups, the H_j coefficients were all greater than .3. The total H coefficients were nearly equal for men and women: $H = .43$ for men and $H = .44$ for women (and $H = .44$ in the total group), indicating medium scalability.

Dimensionality analysis. Table 4 shows the results of automated item selection for the entire group of participants, starting with lower bound $c = 0$ and progressively increasing c with steps of .05 in each next analysis until $c = .70$, following advice from Hemker et al. (1995). For $0 \leq c \leq .35$, all items were selected in one scale. For $c = .40$, one scale with six items and another with three items were formed, and one item remained unscalable. In the next step ($c = .45$), the second scale lost one item; and for $c = .50$, the first scale was reduced to four items, whereas the two items from the second scale were replaced by two different items than in the previous steps. For higher c values, the second scale had disappeared, whereas the first scale further crumbled down to three, two, and finally no items.

The pattern of results made visible here resembles that which Hemker et al. (1995) considered typical of a unidimensional scale: First, all items are in one scale for small values of c ($\leq$ 0.3), and then as c grows more and more, this scale slowly starts to lose items until finally all items are unscalable. One could argue that the appearance of the three-item scale when $c = .40$ is somewhat at odds with the results of Hemker et al. and may warrant the existence of a second dimension, albeit a weak one.

TABLE 4.—Item numbers for selected scales (and their H values) when c values increase in steps of .05 and unscalable items.

c	Scale 1	Scale 2	Unscalable
.00–.35	1–10 (.44)		
.40	1–4, 8, 9 (.52)	6, 7, 10 (.44)	5
.45	1–4, 8, 9 (.52)	6, 10 (.45)	5, 7
.50	1, 4, 8, 9 (.57)	3, 5 (.52)	2, 6, 7, 10
.55	1, 8, 9 (.62)		2–7, 10
.60-.70	8, 9 (.74)		1–7, 10

Note. For men and women separately, approximately the same results were found.

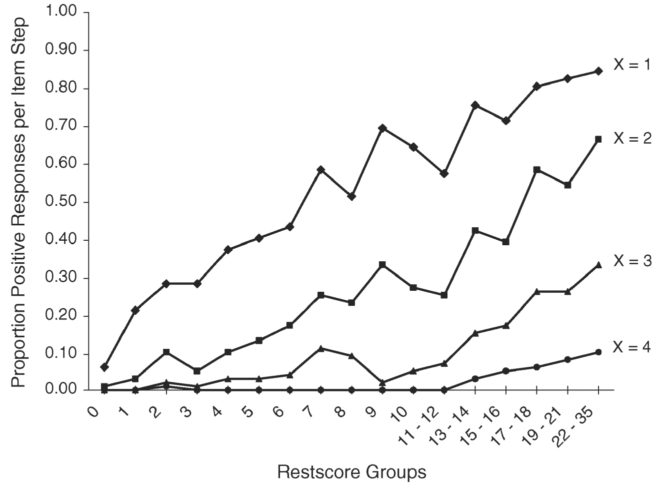

FIGURE 3.—The four estimated item step response functions of Item 10 for minimum group size of 50.

Investigating monotonicity of ISRFs. Monotonicity of the ISRFs was investigated by estimating for each item in the 10-item scale the four ISRFs (for item scores at least equal to 1, 2, 3, and 4, but not 0 because then the ISRF equals 1 by definition), assuming minimum group sizes (abbreviated *minsize*) of 150 and 50 respondents in two separate analyses. These group sizes represent the minimum numbers of respondents used for estimating a discrete point of the ISRFs. Each group consists of respondents with the same θ or neighboring θ values and can be considered homogeneous in this sense. To estimate the ISRF for item score x of item j, that is, $P(X_j \geq x|\theta)$, the fraction is determined that has at least a score of x on the item in each homogeneous group. Across adjacent groups, connecting these fractions gives an estimate of the ISRF. Relatively large groups (e.g., *minsize* $= 150$) produce accurate estimates of the ISRFs but for wide intervals of θ. For example, if *minsize* $= 150$, then the ISRF is estimated for at most 10 intervals of θ given the sample size of 1,503. Figure 1 depicts the ISRFs for *minsize* $=$ 150, resulting in 8 intervals of the rest score group of Item 10 for which the rest score is considered an approximation of θ. Relatively small values (e.g., *minsize* $= 50$; at most 30 points are estimated) produce less accurate estimates but show much more detail, thus revealing possible violations more easily. Figure 3 depicts the ISRFs for *minsize* $= 50$, resulting in 17 intervals. Thus, there is a trade-off between accuracy and bias: Either one sees little of the ISRFs (meaning more bias), but what is visible is also relatively precise, or one sees more detail, but what one sees is relatively less trustworthy (smaller precision). In the latter situation, local statistical tests for decreases in the ISRFs should protect the data analyst from drawing false conclusions.

The MSP program scans each estimated ISRF for decreases. For each ISRF, each pair of fractions in which the second fraction is smaller than the first constitutes a potential violation of the monotonicity assumption (notice that the total number of pairs per ISRF depends on the choice of *minsize* and considerations beyond the scope of this study; see Molenaar & Sijtsma, 2000, for more information). MSP controls for chance capitalization in three ways: (a) by forming homogeneous groups of at least *minsize*, unstable proportion estimates are avoided; (b) by ignoring sample violations of monotonicity smaller than a user-specified value (.03 by default) so that small and unimportant oscillations are not counted; and (c) by testing each decrease larger than the minimum value for significance at a user-specified significance level (.05 by default; using a standard normal approximation to a hypergeometric distribution).

For *minsize* = 150, in 40 ISRFs (i.e., 4 ISRFs for each of the 10 items), only one sample violation of monotonicity greater than .03 was found. It had size .06, and it was statistically significant ($p = .03$, one-tailed test). This violation occurred in the ISRF for $X_{10} \geq 3$, that is, $P(X_{10} \geq 3|\hat{\theta})$, with $\hat{\theta}$ estimated from the data using *minsize* = 150. Note that at this point the ISRF does not increase but decreases (see Figure 1). For *minsize* = 50, a total of 98 sample violations greater than .03 were found. The smallest number (3 sample violations) occurred with Item 3, and the greatest number (22) occurred with Item 5. Out of these 98 sample violations, 6 were significant. Two significant violations occurred with Item 1 (p values were .03 and .04; in both cases, the size of the violation was .16), 1 with Item 5 ($p = .04$; size was .17), and 3 with Item 6 (p values were .03, .01, and .04; sizes were .18, .22, and .07). None occurred with Item 10 depicted in Figure 1. For reasons of comparison, Figure 3 shows the estimated ISRFs of Item 10 for *minsize* = 50. Comparing both figures, one sees the precision-bias trade-off nicely illustrated.

An important question is what one should conclude on the basis of the monotonicity analysis. It is advisable not only to look at detailed results but keep in mind what the analysis is about in the first place. To construct a scale on which all individuals are ordered, it is important to collect enough evidence on the fit of the model to the data, and investigating the monotonicity assumption is an important part of this fit investigation. In the data, only 6 out of 98 violations (i.e., 6.12%) greater than .03 were found to be significant at a 5% level, but the number of 6 significant violations is negligible if one considers all possible pairs of fractions in which a violation of monotonicity could occur. Also, inspecting Figure 3 and the graphs for the other nine items (not provided here), shows that despite the many decreases, the curves show a strong tendency to increase with increasing latent trait value even though there is no exactitude here. These observations are taken as convincing evidence of monotonicity, and this supports the contention that participants can be ordered using the items of the SCS.

Discussion

The aim of this article was twofold. First, we wanted to provide a theoretical introduction to MSA and compare PCA with MSA, and the second aim was to give an empirical example on how MSA can be applied to determine scale dimensionality by using data on the SCS, a questionnaire of which the exact dimensionality is still not unequivocally determined.

Both PCA and MSA aim at determining the dimensionality of an item set. In our study, both methods resulted in different and complementary considerations regarding the dimensionality of the SCS. Executing the PCA, we investigated both a single-factor solution (all 10 items had high corrected item-total correlations) and a two-factor solution (PCA followed by oblimin rotation). We found a strong factor of 6 items and a second, albeit weak, factor of 4 items in the two-factor solution. We conclude that PCA/oblimin favor the single-factor solution: In particular, the high correlations between the oblique factors ($r = .54$) and the substantial corrected item-total correlations support a unidimensional structure.

MSA provided a detailed analysis of the items' scalability and dimensionality structure. By progressively increasing the lower bound c for scalability and thus placing stronger demands on the data structure, MSA provided alternating ways of forming scales. Studying the pattern of cluster outcomes with increasing lower bounds provides detailed information on the most appropriate conclusion with respect to scalability and dimensionality. Overall, the MSA results suggested a medium to strong unidimensional solution of six (for $.40 \leq c \leq .50$) to four items (at $c \geq .50$), respectively. This solution is centered around Items 1, 4, 8, and 9 that explicitly refer to the possession of at least one personally distressing secret and the reluctance to share this secret with other people.

The congruence between PCA and MSA with regard to a unidimensional structure strongly suggests that the SCS is indeed unidimensional. Yet, there are some important differences between the PCA and MSA results that show the merit of applying MSA to determine dimensionality. First, PCA incorporated Items 2 and 7 in the first factor even though the content of these items (apprehension about social consequences following sharing secrets) is clearly different from the other items of the first factor. MSA eliminated Item 7 at $c = .45$ and Item 2 at $c = .50$, revealing that a strong scale could be formed only with the elimination of these two items. Thus, although MSA reached a similar conclusion regarding the SCS dimensionality as PCA did, it made clear that this single dimension consisted of fewer items than suggested by PCA plus rotation. That is, MSA uncovered a core of four items (Items 8, 9, 1, and 4), whereas PCA plus rotation suggested six items (Items 8, 9, 1, 4, 2, and 7).

A second difference is that PCA does not produce a scale. On such a scale, one would expect respondents who score low on items that measure a mild behavioral expression of the trait to obtain a low position. That is, these respondents are expected to have a small probability that higher scores are obtained on items that measure more explicit expressions of the trait. In addition to assessing dimensionality, MSA allows for scrutinizing the monotonicity of the ISRFs, which not only enables constructing scales but also determining the scale's reliable ordinal person measurement capacity. In the SCS, the ISRFs show a strong tendency to increase with increasing latent trait value, suggesting the SCS can adequately be used to order respondents. This is an important conclusion given that several authors have sought specific characteristics of secrets that would be detrimental for physical or mental well-being, yet to no avail (see, e.g., Finkenauer, Engels, & Meeus, 2002). Given that the SCS can be used to provide a scale for respondents, the SCS can adequately distinguish between high versus low self-concealers. This distinction can then be used to analyze the similarities or differences of the secrets that both groups of respondents have.

The results show that Larson and Chastain's (1990) original purpose to design a multidimensional instrument that measures self-concealment only partly succeeded. From the three separate dimensions the SCS was designed to measure (tendency to self-conceal, possession of a personally distressing secret, and apprehension about disclosure), only a single dimension representing the possession of at least one personally distressing secret and the reluctance to share these secrets was found. At best, this dimension can be seen as a hybrid between two from the three separate dimensions Larson and Chastain originally envisioned.

In addition, maintaining a serious secret has emotional and cognitive consequences such as increased levels of anxiety and depression (Larson & Chastain, 1990; Pennebaker, Colder, & Sharp, 1990) and rumination or perseverative thinking (Lane & Wegner, 1995) that must be addressed in any instrument that measures secrecy. The unidimensional character of the SCS therefore seems not to do right to the multidimensionality of secrecy. We therefore call for a revision of the SCS for which our MSA results form a good starting point. For example, the items with lowest scalability (in this case, Items 5 and 7) could be discarded.

CONCLUSION

We showed the merit of MSA as a complementary tool to PCA to determine the dimensionality of an item set. MSA circumvents the problem of difficulty factors often encountered by PCA by eliminating effects of the difference in individual item-score frequency distributions (Michielsen et al., 2004). Further, the detailed output provides a clear view on the items' scalability that can not be obtained with PCA. Monitoring the emergence and breakdown of the subscales at various lowerbound c levels offers a unique possibility to determine which items to retain and which to discard. We therefore recommend that MSA should be used more often in addition to traditional factor-analytic methods such as PCA. Finally, using both PCA and MSA, we provide some recommendations on how to improve the SCS as a scale to assess the general tendency to conceal personally relevant information.

ACKNOWLEDGMENTS

We thank Dr. Marcel Das and Corry Vis from CentERdata at Tilburg University, Tilburg, The Netherlands, for generously sharing their data. We also thank Carmina Puig Sobrevals for her logistical help in preparing this article, and Wilco Emons for preparing Figures 1 and 3.

REFERENCES

Cattell, R. B. (1956–1957). *Sixteen Personality Factors Questionnaire.* Yonkers-on-Hudson, NY: World Book.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Cramer, K. M., & Barry, J. E. (1999). Psychometric properties and confirmatory factor analysis of the Self-Concealment Scale. *Personality and Individual Differences, 27,* 629–637.

Cramer, K. M., & Lake, R. P. (1998). The Preference for Solitude Scale: Psychometric properties and factor structure. *Personality and Individual Differences, 24,* 193–199.

De Raad, B., & Perugini, M. (Eds.). (2002). *Big Five assessment.* Seattle, Washington: Hogrefe & Huber.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Finkenauer, C., Engels, R. C. M. E., & Meeus, W. (2002). Keeping secrets from parents: Advantages and disadvantages of secrecy in adolescence. *Journal of Youth and Adolescence, 31,* 123–136

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional itembank in the polytomous Mokken IRT model. *Applied Psychological Measurement, 19,* 337–352.

Ichiyama, M. A., Colbert, D., Laramore, H., Heim, M., Carone, K., & Schmidt, J. (1993). Self-concealment and correlates of adjustment in college students. *Journal of College Student Psychotherapy, 7,* 55–68.

Junker, B. W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 247–276). New York: Springer-Verlag.

Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement, 24,* 65–81.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151

King, L. A., Emmons, R. A., & Woodley, S. (1992). The structure of inhibition. *Journal of Research in Personality, 26,* 85–102.

Kline, P. (1987). Factor analysis and personality theory. *European Journal of Personality, 1,* 21–36.

Lane, D. J., & Wegner, D. M. (1995). The cognitive consequences of secrecy. *Journal of Personality and Social Psychology, 69,* 237–253.

Larson, D. G., & Chastain, R. L. (1990). Self-concealment: Conceptualization, measurement, and health implications. *Journal of Social and Clinical Psychology, 9,* 439–455.

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–121). New York: Springer-Verlag.

Michielsen, H. J., De Vries, J., Van Heck, G. L., Van de Vijver, F. J. R., & Sijtsma, K. (2004). Examination of the dimensionality of fatigue. *European Journal of Psychological Assessment, 20,* 39–48.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* The Hague: Mouton.

Mokken, R. J. (1997). Nonparametric models for dichotomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351–367). New York: Springer.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6,* 417–430.

Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). New York: Springer.

Molenaar, I. W., & Sijtsma, K. (2000). *User's manual MSP5 for Windows.* Groningen, The Netherlands: iecProGAMMA.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Paas, L. J., & Molenaar, I. W. (2005). Analysis of acquisition patterns: A theoretical and empirical evaluation of alternative methods. *International Journal of Research in Marketing, 22,* 87–100.

Pennebaker, J. W. (1989). Confession, inhibition, and disease. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 211–244). New York: Academic.

Pennebaker, J. W., Colder, M., & Sharp, L. K. (1990). Accelerating the coping process. *Journal of Personality and Social Psychology, 58,* 528–537.

Roorda, L. D., Roebroeck, M. E., Van Tilburg, T., Molenaar, I. W., Lankhorst, G. J., Bouter, L. M., et al. (2005). Measuring activity limitations in walking: Development of a hierarchical scale for patients with lower-extremity disorders who live at home. *Archives of Physical Medicine and Rehabilitation, 86,* 2277–2283.

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Scheirs, J. G. M., & Sijtsma, K. (2001). The study of crying: some methodological considerations and a comparison of methods for analyzing questionnaires. In A. J. J. M. Vingerhoets & R. R. Cornelius (Eds.), *Adult crying: A biopsychosocial approach* (pp. 277–298). Hove, England: Taylor & Francis.

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement, 22,* 3–31.

Sijtsma, K., & Meijer, R. R. (2007). Nonparametric item response theory and related topics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 719–746). Amsterdam: Elsevier.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Stout, W. F. (2002). Psychometrics: From practice to theory and back. *Psychometrika, 67,* 485–518.

Van Abswoude, A. A. H., Van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28,* 3–24.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis, 11,* 139–163.

Vögele, C., & Steptoe, A. (1992). Emotional coping and tonic blood pressure as determinants of cardiovascular responses to mental stress. *Journal of Hypertension, 10,* 1079–1087.

APPENDIX.—Items of the Self-Concealment Scale (Larson & Chastain, 1990).

1.	I have an important secret that I haven't shared with anyone.
2.	If I shared all my secrets with my friends, they'd like me less.
3.	There are lots of things about me that I keep to myself.
4.	Some of my secrets have really tormented me.
5.	When something bad happens to me, I tend to keep it to myself.
6.	I'm often afraid I'll reveal something I don't want to.
7.	Telling a secret often backfires and I wish I hadn't told it.
8.	I have a secret that is so private I would lie if anybody asked me about it.
9.	My secrets are too embarrassing to share with others.
10.	I have negative thoughts about myself that I never share with anyone.

Note. From "Self-Concealment: Conceptualization, Measurement, and Health Implications," by D. G. Larson and R. L. Chastain, 1990, *Journal of Social and Clinical Psychology, 9,* pp. 439–455.