

11-1-2005

The Influence of Reliability on Four Rules for Determining the Number of Components to Retain

Gibbs Y. Kanyongo

Duquesne University, kanyongog@duq.edu

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Kanyongo, Gibbs Y. (2005) "The Influence of Reliability on Four Rules for Determining the Number of Components to Retain," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 2 , Article 7.

DOI: 10.22237/jmasm/1162353960

Available at: <http://digitalcommons.wayne.edu/jmasm/vol5/iss2/7>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

The Influence of Reliability on Four Rules for Determining the Number of Components to Retain

Gibbs Y. Kanyongo
Duquesne University

Imperfectly reliable scores impact the performance of factor analytic procedures. A series of Monte Carlo studies was conducted to generate scores with known component structure from population matrices with varying levels of reliability. The scores were submitted to four procedures: Kaiser rule, scree plot, parallel analysis, and modified Horn's parallel analysis to find if each procedure accurately determines the number of components at the different reliability levels. The performance of each procedure was judged by the percentage of the number of times that the procedure was correct and the mean components that each procedure extracted in each cell. Generally, the results show that when component loading was high, an increase in reliability resulted in an improvement of the accuracy of parallel analysis and modified horn's parallel analysis.

Key words: Monte Carlo, principal components analysis, factor analysis, parallel analysis, modified Horn's parallel analysis, scree plot, Kaiser rule.

Introduction

When dealing with a large number of variables, it is possible that some of those variables are highly correlated with one another such that they account for the same variance in the dependent variable. In such cases it may be possible to combine several variables into one factor without any substantial loss of information. This reduces the number of variables thereby facilitating the interpretation of the data and is achieved through several factor analytic procedures.

Factor analysis is a term used to refer to statistical procedures used in summarizing relationships among variables in a parsimonious but accurate manner. It is a generic term that includes several types of analyses, including

(a) common factor analysis, (b) principal component analysis (PCA), and (c) confirmatory factor analysis (CFA). Common factor analysis may be used when a primary goal of the research is to investigate how well a new set of data fits a particular well-established model (Merenda, 1997). Principal component analysis is usually used to identify the factor structure or model for a set of variables (Stevens, 2002). In contrast, confirmatory factor analysis is based on a strong theoretical foundation that allows the researcher to specify an exact model in advance. In this study, PCA will be of primary interest. Under PCA, the focus is going to be on procedures by which the number of components is determined. Specifically, two forms of parallel analysis, Horn's (1965) parallel analysis (HPA) and modified Horn's parallel analysis (MHPA), are going to be compared to the Kaiser (1960) rule and scree plot procedures under systematically varied conditions of reliability, component loading and variable-to-component loading.

Users of PCA are required to make decisions on a number of technical issues, including the number of components to retain, method of extraction and rotation techniques. Perhaps, the most important decision is that of determining how many components to retain

Gibbs Y. Kanyongo is Assistant Professor in the School of Education, Department of Foundations and Leadership at Duquesne University. He teaches courses in Educational Statistics and Research Methods.

(Merenda, 1997). Problems emerge when non-optimal numbers of components are extracted. Under-extraction compresses variables into a smaller number of components than what actually exists in the data, resulting in loss of important information, a neglect of potentially important components, a distorted fusing of two or more components, and an increase in error loadings (O'Connor, 2000). Over-extraction diffuses variables into a larger number of components than what actually exists in the data, potentially resulting in components with few components loadings and researchers' attributing excessive substantive importance to trivial components (O'Connor). Fava and Velicer (1992) emphasize that researchers should employ the most accurate procedures for determining the correct number of components in order to minimize such problems.

Methods for Determining the Number of Components

The four methods covered in this study are Kaiser rule, scree plot, Horn's parallel analysis procedure and modified Horn's parallel analysis procedure. Each of these methods is covered in detail below.

Kaiser rule.

The easiest and most commonly used method is to retain all components with eigenvalues greater than 1.0 procedure, which is known as the Kaiser rule. This method only provides a rough estimate of the optimal number of components that can be used to describe the data (Hutcheson & Sofroniou, 1999). The suggestion by Kaiser (1960) was based on a commonly used formula for the reliability of a total score (Cliff, 1988). This formula was first suggested by Kuder and Richardson (1937), and is called Kuder-Richardson formula 20 (K-R 20). The formula estimates the parallel form reliability of a total score from the internal consistency among subscores or items. The Kaiser rule uses the rationale that there are as many reliable components as there are eigenvalues greater than one. The reasoning is that an eigenvalue less than one implies that the scores on the component would have negative reliability.

Cliff (1988) argued that Kaiser's rationale for relating the reliability of components to the number of eigenvalues greater than one was based on a misapplication of a common formula for the reliability of a composite. The reliability of a principal component depends on the reliability of the measures. He pointed out that reliability of components cannot be deduced from the size of the eigenvalues, and that the Kaiser rationale for retaining as many components as there are eigenvalues greater than one does not have any logical basis.

The Kaiser rule continues to be one of the most widely employed, largely because the most widely used statistical packages (i.e., SPSS, SAS) continue to include this method as the default option. While this method continues to be widely used, several studies have shown it to be highly inaccurate and to tend to overestimate the number of components to retain (Velicer et al., 2000; Zwick & Velicer 1986). Linn (1968) performed a Monte Carlo study of the Kaiser rule and the degree of overestimation observed was extreme. On average the Kaiser rule overestimated the correct number of components by 66 percent.

In a simulation study conducted by Zwick and Velicer (1986) to compare different component extraction procedures, the Kaiser rule showed the poorest performance, indicating the correct number of components only 22 percent of the time. In some cases the Kaiser rule may sometimes lead to the extraction of fewer components than should have been extracted (Guttman 1954).

Cattell and Jaspers (1967) conducted a study to evaluate the accuracy of the Kaiser rule. In that study, the Kaiser rule was accurate when the number of variables was small (10 to 15) or moderate (20 to 30) and the communalities are high (.80). The communality of a variable is the amount of variance on a variable accounted for by the set of factors. However, the rule showed overestimation with a large number of variables (40) and low communalities (about .40). Stevens (2002) recommended that the Kaiser rule be used only when the number of variables is less than 30, sample size is more than 250, and when the mean communality is about .60. In summary, the Kaiser rule, although commonly used, is

believed by many researchers to overestimate (e.g., Velicer et al., 2000; Zwick & Velicer 1986) and by others to sometimes underestimate (e.g., Guttman 1954).

Cattell's scree plot

This procedure was proposed by Cattell (1966). With this procedure eigenvalues are plotted against their ordinal numbers and one examines to find where the break or a leveling of the slope of the plotted line occurs. Tabachnick and Fidell (2001) referred to the break point as the point where a line drawn through the points changes direction. The number of components is indicated by the number of eigenvalues above the point of the break. The eigenvalues below the break indicate error variance (Velicer et al., 2000). The graph for scree plot is available as an option in SPSS, and most other statistical programs.

Many studies have found this method to be reasonably effective in suggesting the correct number of components to retain (Cattell & Jaspers, 1967; Cattell & Vogelmann, 1977; Cliff, 1970; Linn, 1968; Tucker, Koopman & Linn, 1969; Velicer et al., 2000; Zwick & Velicer, 1982). Hakstian, Rogers and Cattell (1982) noted that when sample size is more than 250 and mean communality is .60, the scree plot extracts the correct number of components. However, Hakstian et al. found that the scree plot is less accurate with low communality data that resulted in the overestimation of the number of components to retain.

Zwick and Velicer (1982) found that the scree plot was especially effective with large sample sizes and with strong components. Stevens (2002) suggested that with this procedure, sample sizes greater than 200 are reasonable provided most of the communalities are large. Cliff (1970) and Linn (1968) found the method to be less accurate with small sample sizes.

Zwick and Velicer (1986) conducted a later study in which they found that the scree test was less accurate than several other methods they investigated. The visual rationale has potential to avoid some of the over-extraction problems because the trivial factors will not be visually compelling (Velicer et al., 2000). The scree procedure is recommended for use with

other procedures, not as a stand-alone procedure. In their studies Velicer et al. evaluated this procedure to be easy to implement because computer programs typically produce the eigenvalues. On accuracy of the procedure, they found mixed results; that is, some results indicated that the procedure extracted the correct number of components and others showed the procedure to over-extract.

Horn's parallel analysis.

Horn (1965) introduced HPA method for determining the number of components as an alternative to the Kaiser procedure. Horn's parallel analysis procedure involves the generation of a set of random data correlation matrices with the same number of variables and participants as the observed data (Velicer et al., 2000). Horn (1965) proposed that a number, say 50, of correlation matrices of p uncorrelated random normal variables and a sample size equal to n , where p and n are the same as the corresponding entries in the data set under study, be constructed and their mean eigenvalues across all the replications be determined. The eigenvalues of the observed data are then compared to the mean eigenvalues of the random data across the replications. The components with eigenvalues of the observed data that exceed the mean eigenvalues of the corresponding components of random data are retained because the observed eigenvalue is considered a real effect that is not likely to be due to chance sampling variability. Actual eigenvalues less than or equal to the mean random eigenvalues would be considered as due to random sampling variability (Glorfeld, 1995). Horn's parallel analysis can be taken to be a way of simulating the scree plot, while on the other hand the Kaiser rule is the theoretical scree plot.

Previous simulation studies have shown that Horn's parallel analysis procedure is accurate in determining the number of components to retain in PCA (Glorfeld, 1995; O'Connor, 2000; Velicer et al. 2000; Zwick & Velicer, 1986). Since this procedure makes use of mean eigenvalues, it requires a comparison of observed eigenvalues and mean eigenvalues of random data (Horn, 1965).

Modified Horn's parallel analysis

Glorfeld (1995) suggested a modification of Horn's parallel analysis procedure to come up with another procedure; modified Horn's parallel analysis which allows identification of any desired upper $(1-\alpha)$ percentile, such as the 95th percentile of the set of distribution (Glorfeld, 1995). This percentile is then used to determine whether the eigenvalue obtained is larger than what could be expected by chance.

Glorfeld (1995) argued that the use of mean eigenvalues is equivalent to conducting the analysis at the 50 percent significance level in conventional hypothesis testing. At this significance level, Horn's parallel analysis tends to over-extract by including minor components in the extraction. He noted that although Horn's parallel analysis is relatively accurate; it still tends to extract one or two more factors than is actually warranted and therefore sometimes retains poorly defined components. The modified Horn's parallel analysis compares real data eigenvalues to the eigenvalues at the desired percentile (typically the 95th) of the distribution of random data eigenvalues (Cota, Longman, Holden, Fekken & Xinaris, 1993; Glorfeld, 1995).

Buja and Eyuboglu (1992) advanced the same argument as Glorfeld (1995) when they pointed out, "the use of null averages as thresholds implies that the classical version of parallel analysis performs approximately at the .5 significance level" (p. 511). They suggested that a higher percentile be used rather than the mean. Buja and Eyuboglu's (1992) suggestion was also in line with the findings by Harshman and Reddon (1983), who pointed out that using the mean eigenvalue represented a potential flaw in the HPA procedure.

The Effects of Extracting the Wrong Number of Components

Over-extraction

Over-extraction is a situation where more components are extracted than the actual number of components in the data. Fava and Velicer (1992) noted, "A primary example of misspecification occurs as a result of employing Kaiser rule, it is likely to over-extract components ..." (p. 388). Moiser (1939)

conducted a simulation study to investigate the impact of over-extraction on the overall component structure. His findings provided some empirical evidence that supports the notion that the error of over-extraction is less harmful than the error of under-extraction (Gorsuch, 1983, Fava & Velicer, 1992).

There are two theoretical justifications that support the idea that over-extraction may not be as serious problem as under-extraction (Fava & Velicer, 1992). The first concerns the fact that the amount of variance explained by a component decreases for each succeeding component extracted (Cattell, 1958, Comrey, 1973), for example, the fifth extracted component will account for less variance than the fourth. This means if a true solution were a four component solution, it would be a more serious error to only extract three components than to over-extract five, because the information lost by the fourth component will be greater than the error added by the fifth component (Fava & Velicer, 1992). The second theoretical justification concerns the idea that, upon rotation, it is relatively easy to discard extra components that have been retained as trivial components without changing the substantive components (Cattell, 1958, Comrey, 1973).

The work by Fava and Velicer (1992) indicated that over-extraction does have negative effect, especially if component loading is low or sample size is low. They found that the worst effects occurred for cases of combined low component loading and low sample size during maximal over-extraction. In their results, they also noted, "there was also strong support for the hypothesis that over-extraction of a few (one or two) components will not cause major negative effects" (p. 413).

Under-extraction

Retaining too few components is another form of misspecification that occurs in PCA. Researchers point that logical arguments generally support the idea that under-extraction is a more serious problem than over-extraction. One argument is that the amount of explained variance decreases for each succeeding component extracted (Cattell, 1958, Comrey, 1973). This suggests that increasing the

dimensionality of a correct solution by one would result in less error variance relative to the amount of true variance that would be removed by decreasing the dimensionality by one (Fava & Velicer, 1996). It appears the general notion is that under-extraction is a more serious problem than over-extraction (Cattell, 1958, Comrey, 1973, Fava & Velicer, 1996). Although under-extraction is generally considered a worse problem than over-extraction, it has been studied less than over-extraction (Fava & Velicer, 1996). This means there is not much evidence to support the opinions that under-extraction is more serious.

Reliability

The decision to include reliability in this study is mainly because previous Monte Carlo studies to compare PCA procedures did not take reliability into account (e.g., Glorfeld, 1995, Velicer et al., 2000, Zwick & Velicer, 1986). These Monte Carlo studies generated the scores with the implicit assumption that the scores had perfect reliability. However, in this study, the scores were generated at varying levels of reliability to see whether the PCA procedures performed differently at each level of reliability.

Reliability is one of the most important considerations when selecting variables for analyses (Tabachnick & Fidell, 2001). Tabachnick and Fidell pointed out that unreliable variables degrade an analysis while reliable variables enhance the analysis. If variables are unreliable, the entire solution may not be trusted as it may contain a lot of measurement error. Gorsuch (1983) noted that if the variables have low reliabilities, then many more individuals would be needed for PCA. It is important that reliability be considered in studies dealing with PCA because in real world phenomenon perfectly reliable data rarely exist.

Reliability is important in principal components analysis because PCA studies correlations among variables and these correlations are impacted by reliability. When reliability is low, these correlations are reduced and consequently the magnitude of the components is also diminished. Also, PCA assumes perfect reliability, since it has 1.0 on the diagonal of the population correlation matrix. On the other hand, principal axis

factoring accommodates reliability and does not get impacted that much by a change in reliability.

McMillan and Schumacher (2001) advised that in any study, the reliability of scores should be established before the research is undertaken, and the type of reliability should be consistent with the use of the results. Reliability is a function of the trait being measured. This means some variables, such as most measures of achievement, provide highly reliable scores; whereas scores from measures of personality traits have lower reliability. It is common to have reliability values of .80 or above for achievement tests, whereas values of .70 may be acceptable for measuring personality traits (McMillan & Schumacher, 2001). What this means is that an achievement test with a reliability of .70 may be seen to be weak whereas a personality instrument with a reliability coefficient of .90 is considered excellent.

Tuckman (1999) provided the following factors that affect reliability of a measurement instrument, (a) familiarity with the particular measurement instrument, (b) participant fatigue, (c) emotional strain, (d) physical conditions of the room in which the test is administered, (e) participant's health, (f) participant's practice or experience in the specific skill being measured, and (g) specific knowledge gained outside the experience being evaluated by the measurement instrument. McMillan and Schumacher (2001) noted that to enhance reliability, it is best to establish standard conditions of data collection. For example, all participants should be given the same directions and have the same time frame in which to answer questions at the same time during the day. Error is often increased if different people administer the instrument.

Methodology

Monte Carlo Procedure

Monte Carlo simulations perform functions empirically through the analysis of random samples from populations whose characteristics are known to the researcher (Brooks, Barcikowski & Robey, 1999). That is, Monte Carlo methods use computer assisted simulations to provide evidence for problems

that cannot be solved mathematically, such as when the sampling distribution is unknown or hypothesis is not true.

The principle behind Monte Carlo simulation is that the behavior of a statistic in a random sample can be assessed by the empirical process of actually drawing many random samples and observing this behavior (Mooney, 1997). The idea is to create a pseudo-population through mathematical procedures for generating sets of numbers that resemble samples of data drawn from the population.

Mooney (1997) noted that other difficult aspects of the Monte Carlo design are writing the computer code to simulate the desired data conditions and interpreting the estimated sampling plan, data collection, and data analysis. An important point to note is that a Monte Carlo design takes the same format as a standard research design. This was noted by Brooks, Barcikowski, and Robey, (1999) when they wrote "It should be noted that Monte Carlo design is not very different from more standard research design, which typically includes identification of the population, description of the sampling plan, data collection and data analysis" (p. 3).

Like any methodology, Monte Carlo studies are not without disadvantages. Their usefulness depends in large part on the realism of the conditions that are modeled (Hutchinson & Bandalos, 1997). This means in setting up a Monte Carlo design, conditions set should resemble those found in practice otherwise results obtained will be of less utility. In this study, the conditions that are modeled are those that are encountered in real life situations (Zwick & Velicer, 1986). Another factor influencing results in a Monte Carlo study is the number of replications. This means in order to have a good model of a sampling distribution several replications should be performed.

Pseudo-Population

The conditions in the pseudo-population that were manipulated in this study are sample size, reliability, number of components, and variable-to-component ratio. These independent variables were selected on the basis of their importance in applied research and their ability to distinguish the accuracy of the PCA

procedures (Velicer et al., 2000, Zwick & Velicer, 1986). The conditions used in this study and the levels within them are given below:

Conditions Investigated

Component loading (a_{ij}).

The component loading is the Pearson correlation between a component and a variable. It is a measure of the degree of generalizability found between each variable and each component. This condition had two levels; a moderate coefficient of .50 and a very strong coefficient of .80 to represent small and large component loading respectively. These values were used by Zwick and Velicer (1986) because they generally bridge the range found in applied research situations and were shown to differentially affect the accuracy of the component extraction procedures. Component loading has been found to be one of the factors having the greatest effect on the accuracy of PCA procedures (Guadagnoli & Velicer, 1988, Velicer et al., 2000, Zwick & Velicer, 1986).

Reliability (ρ_{xx}).

Three levels of this variable were included. These levels are .60, .80 and 1.0. The value of .60 was included because it represents the near minimum acceptable reliability in applied research. The value of .80 was included because a reliability of .80 is generally considered good for most measures (McMillan & Schumacher, 2001). Previous Monte Carlo studies on PCA have assumed a reliability of 1.0. So this value was included in this study for comparison purposes with earlier studies. In this study each of the 24 variables had uniform reliability. That is they all had either reliability of .6, .8 or 1.0.

Variable-to-Component Ratio ($p:m$).

The number of components chosen was three and six, but because the number of variables was chosen to be constant at 24, this leads directly to the variable-to-component ratios of 8:1 and 4:1, respectively. These values were chosen to reflect those reported in literature (Velicer, Eaton, & Fava 2000; Zwick & Velicer, 1986). The number of variables was taken as a constant at a value of $p = 24$ across all the conditions. This value represents a moderately

sized data set (Stevens, 2002). It should be noted that other studies on principal component analysis (PCA) regard a value of $p = 36$ to be a small- to moderate-sized data set (e.g., Fava & Velicer, 1996, Zwick & Velicer, 1982, 1986).

Generation of Population Correlation Matrices

The underlying population correlation matrices were generated for each possible $p, p:m$ and a_{ij} combination in the following manner using RANCORR program by Hong (1999):

1. The component pattern matrix was specified based on the combination of values for $p:m$ and a_{ij} .
2. The population correlation matrix was produced from the pattern matrix.
3. The program was executed four times to yield four different population correlation matrices, one correlation matrix for each combination of conditions.

Generation of samples

After the population correlation matrices were generated as described in the above section, the MNDG program (Brooks, 2002) was then used to generate samples from the population correlation matrices. This program generated multivariate normally distributed data, and reliability was systematically varied in the program to create unreliable scores based on the classical test theory. That is, reliability is defined as the proportion of raw score variance explained by true score variance, σ_T^2/σ_X^2 , or equivalently $1 - \sigma_E^2/\sigma_X^2$, with each raw score generated taken to be a total score.

In the program, the reliability estimate was entered and was set to be the same for all the variables. For example, the value for reliability for .80 was provided as input to the program and was set to be the same for all the 24 variables. The population mean and standard deviation were set to be equal to 0.0 and 1.0 respectively. The data sets were generated for each of the combinations of the $3 \times 2 \times 2$ design representing three levels of reliability, two levels of component loading and two levels of variable-to-component ratio yielding 12 different cells.

Data Analysis

One of the decisions in Monte Carlo studies is determining the number of replications to be run to obtain individual samples from the pseudo-population. There seem to be no clear guidelines on how to select the optimum number of replications. The specific number depends on the type of phenomenon being studied, the extent to which the steps of the simulation can be automated, as well as available computer resources (Hutchinson & Bandalos, 1997). Guadagnoli and Velicer (1991) recommended that five replications per cell are adequate for PCA studies. This was also supported by Velicer et al. (2000) who used five replications in their study to compare various PCA procedures. However, this study used 10 replications (samples) per cell since this was feasible given the computer capabilities available. This yielded a total of 120 samples for the 12 cells.

Each of the samples was submitted to the four procedures to determine the number of components extracted. The number of times that each procedure was correct in determining the number of components was recorded. The accuracy was then measured by the percentage of the correct components extracted by each procedure. This was done by counting the number of times each procedure extracted the exact number of components in each cell, and then expressing it as a percentage of the total number of samples in that cell. The mean component extracted in each cell was calculated by summing the number of components across the 10 samples and then divide by 10.

Results

The issue in this study was whether reliability influences the accuracy of the Kaiser rule, scree plot, Horn's parallel analysis and modified Horn's parallel analysis procedures. This issue was investigated by calculating the percentage of times that each procedure was correct in determining the correct number of components. It became apparent in this study that reliability has an influence in determining the number of components to retain under certain conditions.

However, under other conditions, the influence of reliability on the performance of the procedures is not very apparent.

Figure 1 shows the performance of each procedure in terms of percentages at each level of reliability for variable-to-component ratio of 4:1 when the cells were collapsed across sample size and component loading. At the reliability of .60, all the procedures were less than 60 percent accurate. As reliability increased, so was the accuracy of the procedures. At lower reliability levels, Horn's parallel analysis procedure performed slightly better than the all the other procedures. However, at a reliability of 1.0, modified Horn's parallel analysis was the best procedure.

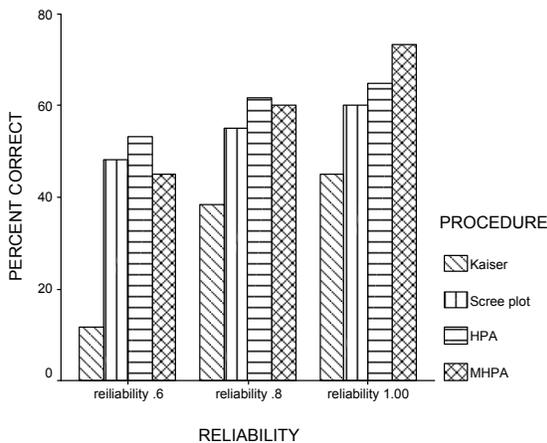


Figure 1. The overall percentage of the correct number of components extracted by each procedure collapsed across component loading for 4:1 variable-to-component ratio.

When the information is presented collapsed across sample size only (as in Figure 2) and when component loading was high, (.80), and variable-to-component ratio was 4:1, the scree plot, HPA, and MHPA were at least 70 percent correct even at the reliability of .60. However, the picture is the same as in Figure 1 that as reliability increased, so does the performance of the procedures as measured by the percentage correct. The performance of the

Kaiser rule is something that stood out in these results especially considering that in most previous studies; the Kaiser rule demonstrated poor performance. In this study, generally the procedure was poor as well, but a look at Figure 2 shows that when variable-to-component ratio was 4:1, and component loading was .80, the Kaiser rule did very well especially at reliability of 1.00. At this reliability level, the Kaiser rule was about 90 percent correct in determining the number of components.

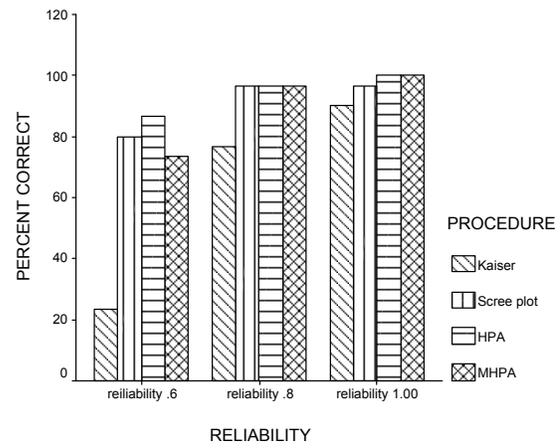


Figure 2. Percentage of the correct number of components extracted by each procedure when the variable-to-component ratio is 4:1, component loading is .80.

Figure 3 shows that when variable-to-component ratio was 8:1; and component loading was .80, reliability did not have any influence on the accuracy of the scree plot, HPA, and MHPA. These three procedures were 100 percent accurate at all the reliability levels. However, reliability had a slight influence on the performance of the Kaiser rule. Under these conditions (with few variables per component), the Kaiser was very poor even though the other procedures were very accurate at all the levels of the reliability.

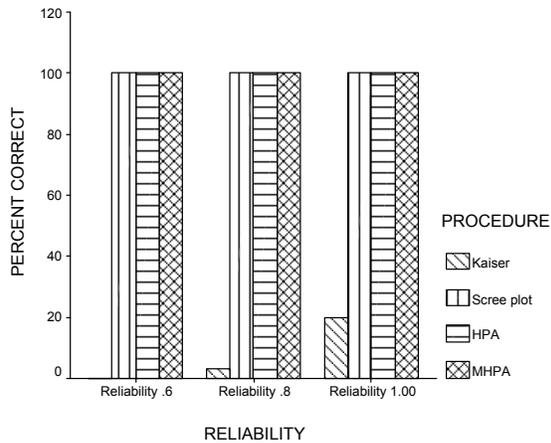


Figure 3. Percentage of the correct number of components extracted by each procedure when the variable-to-component ratio is 8:1, component loading is .80.

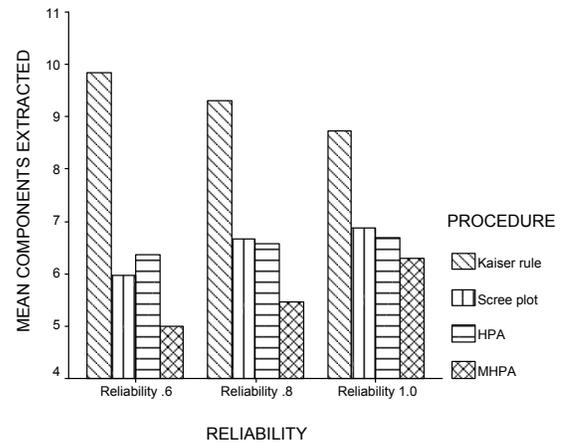


Figure 4. The mean number of components extracted by each procedure for variable-to-component ratio of 4:1 and component loading of .50.

Figure 4 shows the relationship between the mean number of components extracted by each procedure and reliability. The exact number of components to be extracted here is six, and is indicated on the graph by a horizontal line. Under conditions of component loading of .50, only the Kaiser rule appeared to be responsive to a change in reliability. It is clear from Figure 4 that most of the procedures did not extract the correct number of components under these conditions. Only the scree plot was accurate at reliability of .60, but as reliability increased, the scree plot displayed a tendency to over-extract. On the other hand, MHPA under-extracted at lower reliability levels, but as reliability increased, it displayed a tendency to over-extract. The bottom line is, when component loading is low, the pattern of the performance of the procedures is not clearly defined.

In Figure 5, the actual number of components was three and that is shown by a horizontal line at 3. The graph shows an almost similar picture as that in Figure 4 where the performance of the Kaiser rule improved with an increase in reliability. However, the pattern of the performance for the scree plot, HPA, and MHPA as reliability increased is not clearly defined. These results are consistent with previous research that found that when component loading is low, the component structure is more diffuse, and the performance of the different procedures cannot be clearly distinguished under these conditions.

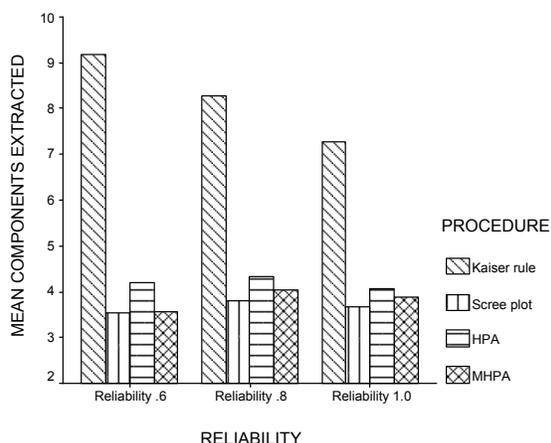


Figure 5. The mean number of components extracted by each procedure for variable-to-component ratio of 8:1 and component loading of .50.

Practical implications

It is important for practitioners to know that imperfect scores impact the performance of the procedures they use to determine the number of components to retain. A stronger reliability is mostly desirable if the results will be used to make decisions about individuals. Although most practitioners and researchers use instruments with high reliability (at least reliability of .80); sometimes instruments with low reliability are unavoidable. For example, studies of groups can tolerate a lower reliability, sometimes as low as .50 in exploratory research (McMillan & Schumacher, 2001). The same authors also noted that measures of young children are usually less reliable than those of older participants. This could be because of problems in the reading and language level, as well as lack of clear instructions. Under these conditions of low reliability, it becomes important for practitioners to understand how that might impact their decisions on retaining the correct number of components.

What also became clear in this study is that practitioners should be cautious in determining the number of components to retain in PCA. This is especially critical when the data have low component loading. Under these conditions, almost all procedures are inaccurate

in determining the correct number of components. It would be best for practitioners to use more than one procedure and then compare the results. The scree plot can be a very useful procedure to use as an adjunct, but probably not as a stand-alone procedure.

It is important to mention that the best way to make a decision about how many components to extract should be based on the knowledge that the practitioner or researcher has about the data. These various PCA procedures we use are just tools that help us extract components but should never substitute the knowledge that the researcher has about the data. In other words, components that are extracted should have some practical significance, they should have meaning to the researcher and the tools should not be used blindly to determine the number of components to extract.

Conclusion

Although this study provided some important insights into how reliability influences the number of components to extract especially for high component loading, what seemed unclear is whether the differences between the different procedures are large enough to recommend one procedure over the other at a given reliability level. For example, can one say at low reliability level, the scree plot is a better procedure to use than HPA or MHPA? Such a recommendation is not possible in this study. Instead, researchers are recommended to take into account the reliability of the scores when they interpret the number of components extracted. This is supported by Henson (2001) who pointed, "... for tests that consist of scales measuring different constructs, internal consistency should be assessed separately for each scale" (p. 181).

References

- Brooks, G. P. (2002). MNDG. <http://oak.cats.ohiou.edu.edu/~brooksg/mndg.htm>.

- Brooks, G. P., Barcikowski, R. S., & Robey, R. R. (1999). *Monte Carlo simulation for perusal and practice*. A paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED449178).
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509-540.
- Cattell, R. B. (1958). Extracting the correct number of factors in factor analysis. *Educational and Psychological Measurement, 18*, 791-838.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.
- Cattell, R. B., & Jaspers, J. (1967). A general procedure for factor analytic exercises and research. *Multivariate Behavioral Research Monographs, 3*, 1-212.
- Cattell, R. B., & Volgemann, S. A. (1977). A comprehensive trial of the scree and Kaiser-Guttman criteria for determining the number of factors. *Multivariate Behavioral Research, 12*, 289-325.
- Cliff, N. (1970). The relation between sample and population characteristic vectors. *Psychometrika, 35*, 163-178.
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin, 103*, 276-279.
- Comrey, A.L. (1973). *A first course in factor analysis*. New York: Plenum Press.
- Cota, A. A., Longman, R. S., Holden, R. R., & Fekken, G. C. (1993). Comparing different methods for implementing parallel analysis: A practical index of accuracy. *Educational and Psychological Measurement, 53*, 865-877.
- Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 387-415.
- Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement, 56*, 907-29.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*, 265-275.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika, 19*, 149-162.
- Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research, 17*, 193-219.
- Harshman, R. A., & Reddon, J. R. (1983). Determining the number of factors by comparing real with random data: A serious flaw and some possible corrections. *Applied Psychological Measurement, 9*, 14-15.
- Henson, R.K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.
- Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments & Computers, 31*, 727-730.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. Thousand Oaks, CA: Sage Publications.
- Hutchinson, S. R., & Bandalos, D. L. (1997). A guide to Monte Carlo simulation research for applied researchers. *Journal of Vocational Education Research, 22*, 233-245.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151-160.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, 33, 37-71.

McMillan, J. H., & Schumacher, S. (2001). *Research in Education: A conceptual introduction*. (5th ed.). San Francisco, CA: Addison Wesley Longman.

Merenda, F. P. (1997). A Guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counselling and Development*, 30, 156-164.

Moiser, C. I. (1939). Influence of chance error on simple structure: An empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika*, 4, 33-44.

Mooney, C. Z. (1997). *Monte Carlo simulation* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-116). Thousand Oaks, CA: Sage.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32, 396-402.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights., MA: Pearson.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34, 421-459.

Tuckman, B. W. (1999). *Conducting educational research* (5th ed.). Orlando, FL: Harcourt Brace College.

Velicer, F. W., Eaton, C. A., & Fava, J. L. (2000). Construct Explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, & E. Helmes (Eds.), *Problems and Solutions in Human Assessment* (pp. 42 -71). Boston: Kluwer Academic Publishers.

Zwick, R. W., & Velicer, W. F. (1982). Factors influencing four rules of determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269.

Zwick, R. W., & Velicer, F. V. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.