Measuring paternal involvement in childcare: a critical analysis of three data reduction methods

Helen Norman and Mark Elliot

Abstract

Three data reduction methods - *Principal Components Analysis (PCA), Principal Axis Factoring (PAF), Confirmatory Factor Analysis (CFA)* – are used on a sample of households from sweep one of the Millennium Cohort Study to derive two quantitative measures of paternal involvement in childcare and housework, respectively defined as *engagement* and *responsibility*. The purpose of using all three methods, aside from developing the most accurate and robust measure, was to explore the differences within each technique and the effect that these differences would have on results and further analyses that use the latent variables. All three methods produce two, moderately correlated latent variables, which could be used in further analyses to measure the conditions to model paternal involvement, and there are no major differences to the component/factor structure. This challenges criticisms directed mainly to the technique of PCA, which has often been labelled an inferior and more biased method compared to PAF.

www.ccsr.ac.uk

1. Introduction

How to measure paternal involvement is a matter of wide debate (e.g. Dermott 2008, 2003; Williams 2008; Mikelson 2008; Sanderson and Sanders-Thompson 2002; Cabrera et al 2000; McBride and Mills 1993; Lamb 1986). The concept is challenging to define and measure, with its definition and utility contested by scholars who emphasise its varied and subjective nature (e.g. Sanderson and Sanders-Thompson 2002; Coltrane and Parke 1998). For this reason, previous explorations of paternal involvement in childcare have been mainly pursued through more qualitative methods of research (e.g. Doucet 2006; Dermott 2003). What has been lacking is a more precise operationalisation of the concept. Morman and Floyd (2006: 116) assert that there is currently no 'quantitative tool' for measuring 'involved' fatherhood and claim that such an indicator would be useful in creating a benchmark for further research and conceptual elaboration as well as providing a method for operationalising the meaning of being a 'good father'. McBride and Mills (1993) also point out that the lack of a clear and consistent definition of 'involved fathering' has hindered research on the paternal role to date. This paper intends to address this shortfall by deriving a quantitative measures based on Michael Lamb's (1986) classification of paternal involvement (accessibility, engagement and responsibility) through three data reduction techniques.

There are several ways to derive a latent measure, or reduce data into a smaller number of dimensions, but each method comes with its benefits and limitations. Most data reduction methods stem from the technique known as factor analysis. However, for many studies the most appropriate type of factor analysis to use has been subject to debate so, as well as the need to apply such techniques to produce a robust operationalisation of the latent construct of interest, there is also a need for methodological work to identify which method is the most appropriate. Connecting these two points together a construct which is reliably extracted across different methods is likely to be more robust than one which does not. For these reasons, we employ three different data reduction methods on a sample of 11,767 households from sweep one of the Millennium Cohort Study

(MCS)[1] to derive two quantitative measures of paternal involvement in childcare and housework, respectively defined as *engagement* and *responsibility*[2]. The first two methods - *Principal Components Analysis (PCA)* and *Principal Axis Factoring (PAF)* - are exploratory so are utilised to initially explore the structure of the data. The third method - *Confirmatory Factor Analysis (CFA)* – is used to test both the existing theory that involvement is definable in terms of Lamb's (1986) two dimensions, and the reliability of the results produced from the PCA and PAF, which is particularly pertinent given that the former method has been subject to much critique (e.g. Tinsley and Tinsley 1987; Floyd and Widaman 1995; Fabrigar et al 1999; Preacher and MacCallum 2003). This stage-by-stage process will show the different ways in which a latent variable can be explored and subsequently derived. By taking this approach we can assess whether the different techniques produce results that are similar or diverge in any salient way, which will highlight the implications of using any one method interchangeably.

We conclude by recommending the use of the factors produced from the CFA as these align with existing theory. However, the factors produced by PCA and PAF are not too dissimilar suggesting

---

[1] The MCS is a nationally representative survey following a cohort of children born around the year 2000 in Great Britain. The first sweep of data was carried out in 2001/2 and covers a cohort of 18,819 babies aged nine months (brought up in 18,552 families) over a twelve month period starting in 2000. The sample for this analysis has been filtered to include heterosexual couples only and all fathers who were not employed at sweep one were filtered out (10.5 per cent of households) in order to focus the study on men who all start with similar commitments to paid work (although this will vary by hours worked, type of job etc) and family. Fathers who did not take part in the survey were also filtered out (8.7 per cent of households) since no information about their parenting practices could be obtained. The final sample derived for this study therefore amounted to 11,767, representing 63 per cent of the original MCS sample.

[2] Lamb's third dimension of *accessibility* was not explored in this study due to data restrictions. See section 3. Only one indicator - fathers' weekly work hours – was identified as measuring this in the MCS on the basis that employment patterns impact on how often fathers can contribute to their child's care and long hours of work lead to less time at home and therefore less time with children. Although this is an appropriate proxy measure of accessibility, it cannot be used alone since further analyses on a single-item component in factor analysis would be equivalent to analysing the conditions that influence fathers' work hours, which is not the focus here. There were other potential variables to measure paternal accessibility alongside work hours such as frequency of evening, weekend and night work, however, these variables were unsuitable because they measured paternal accessibility at a certain point in time only. Since no other suitable variables were available, the decision was taken to remove the third component of accessibility. Although this disrupts Lamb's three dimensional classification of paternal involvement, accessibility is considered the least important out of the three dimensions for this particular age of cohort child. At nine months old, the child will be unaware as to whether the father is 'there' for her or not. Accessibility is perhaps more pertinent when children are older and do not require the same level of engagement from their fathers, which at nine months, is more important in terms of the child's development.

that the selection of one method over the other would have virtually no impact on the results produced (e.g. Floyd and Widaman 1995; Preacher and MacCallum 2003).

2. Defining paternal involvement: three dimensions

This paper builds on the dimensions of involvement originally defined by Michael Lamb in 1986. Lamb argued that there are three components of parental involvement: accessibility, engagement and responsibility; these distinctions have since been widely used and cited (e.g. Dermott 2008; Morman and Floyd 2006; McBride and Mills 1993; O'Brien 2005; Sanderson and Sanders-Thompson 2002). The first dimension of *accessibility* signifies being physically available and present as a parent. This encompasses supervisory care or activities that require a less intensive degree of interaction, such as cooking in the kitchen while the child plays nearby. Thus, it is taken to refer to a secondary activity without one-to-one engagement in which the father is present and available for childcare if required either by the child or a partner. *Engagement*, or interaction, Lamb argues, represents the one-to-one interaction time with the child, described by Lamb as the most intensive component of involvement because it is entirely child-centred. Examples include: feeding the child and playing or helping with homework. Lamb maintains that this does not include 'multi-tasking' activities e.g. doing the cooking whilst helping the child with homework as this involves a less intensive degree of interaction.[3] According to Lamb, *responsibility* is the third and most important component of involvement but is difficult to quantify because the planning, worry and thought that goes into being responsible often occurs when the parent is doing something else. Responsibility for the child's welfare and care is different from that of being able and willing to help out when convenient or when needed and mobilised by the 'main carer'. Responsibility involves *knowing* in detail what is needed and *ensuring* the particular aspects of childcare that are required are provided by anticipating, planning and arranging provision. For example, knowing when the child needs to go to the doctor, making the appointment and ensuring the child gets to it;

[3] We argue that this is too strong a definition of engagement; it seems uncontroversial that engaging with a child can be achieved alongside other activities, particularly when the child is very young. As long as the 'engaging' activities constitute direct interactions with the child the possibility that they are done alongside other tasks cannot be ruled out. For example, feeding the baby may be done alongside having a conversation with the mother or feeding other children in the household.

making arrangements for childcare and ensuring the child has clothes to wear and food to eat. One way a father might be responsible is to take responsibility for housework in order to maintain a clean and safe standard of living for the child; while also relieving the other parent (i.e. the mother) of these tasks so that she can concentrate on other activities such as looking after the child.[4] Indeed, Dermott (2008) argues that after the birth of a child, housework becomes an 'acknowledged task' because chores suddenly attain a child-specific dimension to them (Dermott 2008: 53)[5]. For example, doing the laundry will most likely include the child's laundry and cleaning the house may be considered more important when the wellbeing of the child who lives there is taken into account. More generally, housework in households containing children can be viewed as implicit childcare and as indicators of Lambs responsibility dimension. Therefore, housework contributions made by fathers will be considered an important aspect of paternal involvement here.

3. Measuring paternal involvement in the Millennium Cohort Study

There are several different ways in which to capture paternal involvement empirically. For example, analysing parents' perceptions and accounts of their roles, relying on only mothers' or only fathers' reports of paternal contributions, comparing the amount of time men and women put into childcare and so on. The study here primarily concentrates on the fathers' accounts of their childcare practices when children are very young but the analyses rely on the mothers' reports to measure fathers' contributions to housework. This reliance on a combination of mothers' and fathers' reports for different variables is driven by the structure of the questionnaire design but doing this also allows us to gain a different perspective on fathers' roles..

Selecting variables

The first step was to select MCS variables that were considered to represent some underlying aspect of one of the two dimensions of involvement. The questionnaire for the sweep one survey is

[4] Of course, this can work in reverse where fathers primarily look after the children while the other parent does the housework.

[5] Even though her study of twenty-five fathers revealed that they did not consider housework it to be part of fathering (irrespective of whether they were actually engaged in housework or not).

divided into eleven modules according to topic[6] with modules 2 (fathers involvement with baby), 8 (self-completion questions) and 9 (employment and education) identified as containing the most suitable variables for measuring paternal involvement in childcare and housework[7].

Twenty sweep one variables were initially selected from three modules on the basis that they represented some underlying aspect of one of the three dimensions of involvement, however, several data issues came to light, which required thirteen variables to be dropped. This resulted in a total of seven variables identified as both captured reliably and appearing to measure some underlying aspect of paternal involvement (see appendix for frequency tables and correlation matrix). Table 1 lists the twenty variables initially selected and highlights the final seven that were retained.

[6] Modules were: 1) Non-resident parents; 2) Father's involvement with the baby; 3) Pregnancy, labour and delivery; 4) Baby's health and development; 5) Childcare; 6) Grandparents and friends; 7) Parent's health; 8) A self completion section (questions include attitudes to marriage, parenting, work and psychological assessments); 9) Employment and education; 10) Housing and the local area; 11) Interests and time with the baby.

[7] Although module 4 contained questions on child health and development, there were no measures of who takes most responsibility for the child's health, which would have been a suitable indicator for paternal responsibility. Instead, questions focus on the physical attributes of children to indicate their health such as their height and weight for instance. Module 5 contained questions on childcare, but there were no direct measures of how often fathers performed certain childcare tasks as was the case in module 2. In the 'childcare' module 5, questions mainly focused on formal childcare arrangements and were only asked to the mother (i.e. main respondent). Although module 11 'time with baby' may appear relevant, there were only two questions that measured fathers' attitudes towards whether they felt they had enough time with their baby as opposed to how they actually used this time.

Table 1: Sweep one variables that potentially measure paternal involvement at age nine months

VARIABLE	RESPONDENT Partner	Main
Total hours of work per week (inc overtime) [TOHR] / [WKHR]*	✓	✓
Usual hours of work per week (exc overtime) [WOHR]	✓	×
Frequency works evenings [EVEW]	✓	✓
Frequency works weekends [WKWE]	✓	✓
Frequency works away overnight [WKAW]	✓	✓
Frequency works at night [NGTW]	✓	✓
Frequency spends time with friends [FRTI]	✓	✓
Frequency go out as a couple [COLT]	✓	✓
Frequency looks after baby on own [LOAF]	✓	×
Frequency feeds baby [OFFE]	✓	×
Frequency changes baby's nappies [NACH]	✓	×
Frequency gets up in night for baby [GETU]	✓	×
Mostly around and generally looks after child [GECA]	×	✓
Mostly responsible for changing nappies [CHNA]	×	✓
Mostly responsible for feeding baby [REFE]	×	✓
Mostly responsible for looking after the baby when ill [LKIL]	×	✓
Mostly responsible for getting up at night for baby [GEUP]	×	✓
Who takes most responsibility for cooking main meal [COOK]	×	✓
Who takes most responsibility for laundry and ironing [LAUN]	×	✓
Who takes most responsibility for cleaning the home [CLEA]	×	✓

*[TOHR] = Partner variable, [WKHR] = Main variable

The work hour variables ([TOHR] / [WKHR]) were considered relevant. All other variables under this construct i.e. those that measured atypical patterns of paid work - [EVEW], [WKWE], [WKAW] [NGTW] - were discarded because although they appear to have face validity they are ambiguous with respect to overall involvement. For example, regular evening work would appear to suggest low involvement but evening work might be done on top of shorter working hours, or in order to fit round a partner working during the day, which, could lead to greater paternal involvement during the daytime. A measure of a father's work hours provides some indication of the total amount of time the father is at work and away from his child regardless of when this occurs.

Variables measuring 'frequency spends time with friends' [FRTI] and 'frequency goes out as a couple' [COLT] under the accessibility construct were also dropped. The frequency fathers go out as a couple without the child was dropped because little information about paternal involvement could be gained from using this variable: over half of respondents went out (without their child) rarely and about a third went out only occasionally[8]. Moreover, it is typical for couples to go out only for short periods of time such as for a few hours in the evening; this is unlikely to have a significant impact on a father's involvement. Similarly, when fathers spend time with their friends, this is most likely to be for a few hours at a time and will therefore have little impact on fathers' contributions to their children's care. Indeed, [FRTI] measures how *often* fathers go out with friends and there is no indication of how long fathers spend with friends making the measure ambiguous for this purpose. Furthermore, spending time with friends does not necessarily preclude the baby who may be 'in tow' during such activities (e.g. see Sayer, Bianchi and Robinson 2004).

All of the other sweep one partner variables were retained as they measure core childcare and domestic work tasks. All fathers from the 'main' childcare variables are imputed into the partner variables because activities can be matched up with one exception: 'frequency fathers look after children when they are ill' [IKIL] is a unique measure because there is no equivalent partner variable. Furthermore, looking after a child when he/she is ill does not constitute a *core* activity because it does not require an ongoing daily commitment from either parent; involvement is completely dependent on how often the child gets ill, which will be on intermittent occasions (unless the child in question suffers from a chronic illness, which is fairly rare and cannot be measured in the MCS[9]). This variable was consequently dropped. The three variables measuring domestic work were retained and are solely reliant on the reports of the main respondent, most of whom are mothers. Generally, mothers tend to take the most responsibility for household labour (Singleton and Maher 2004; Crompton 2006), which is the rationale behind the survey only choosing the main respondent (almost always the mothers) to report domestic work (Calderwood et al 2005).

[8] 57.6% went out as a couple less than once a month or 'hardly ever/never' and 31.4% went out as a couple once a month or more but less than weekly

[9] This scenario also applies to [FATHGETUP] as involvement is dependent on how often the baby wakes up at night which could be once, twice or more. There were also a small proportion of fathers with babies who never woke up at night (13.1%, n= 1,545). This variable was retained however; the implications of this decision are discussed in detail in Norman (2010).

An initial exploratory principal component analysis using the remaining eight variables produced three factors. However, factor three effectively consisted of a single variable - fathers work hours. Although in principle this is an appropriate proxy measure of accessibility, it cannot be used alone since further analyses on a single-item component in factor analysis would be equivalent to analysing the conditions that influence fathers' work hours, which is not the focus here.

In summary, the final seven variables that are used broadly measure the frequency fathers spend engaging in implicit or explicit childcare. So the measure of involvement derived here is based on *how often* fathers make contributions to these tasks. This is a key indicator of paternal involvement, further supported by Sayer, Bianchi and Robinson (2004) who argue that there is a cultural expectation for fathers to spend time with children; thus involved fathering generally means a greater time commitment to childcare. Measuring the frequency a father performs certain childcare tasks will indicate whether he has a high or low level of involvement.

4. An evaluation of three data reduction methods

Factor analysis is one of the main methods of data reduction and works by exposing patterns of relationships between variables, identifying highly intercorrelated variables and reducing a large number of variables to a smaller number of factors that can be used in subsequent analyses. Thus, underlying latent variables, which are not otherwise directly measurable, are identified through these techniques with parsimony achieved because the greatest amount of common variance is accounted for in a correlation matrix with the smallest number of explanatory components. There are variations of factor analysis however, with implications attached to choosing one method over the other.

PAF and PCA are two branches of EFA and will be used to explore the set of variables selected to measure involvement. However, the way these techniques classify data differ due to the ways in which they estimate variance among the variables. PAF is used to generate latent variables (or factors), which account for the covariances among a larger set of observed variables. PCA is used to simply reduce a number of variables into a smaller, more manageable set, although it can still be used to explore how data is structured. PCA components (which are linear combinations of observed variables) account for total variances whereas factors from PAF account for common variance out of a total variance; (i.e. common variance is separated from unique variance) (Albright and Park 2008). Both PCA and PAF will be used to explore the structure of the seven variables. The use of two exploratory tools provides a more thorough initial analysis, prepares the data for the final confirmatory factor analysis and highlights the implications of choosing one technique over the other.

The third method is Confirmatory Factor Analysis (CFA), which is a different technique used to test hypotheses about a predicted factor structure; including evaluating the goodness-of-fit of a model produced using EFA methods (Albright and Park 2008). It relies on a method of estimation called Maximum Likelihood to evaluate factor solutions so is often used in the later stages of metric development to refine and improve solutions produced by EFA. CFA is often used to assess whether a proposed factor structure adequately fits the data better than alternative structures (Floyd and Widaman 1995: 293).. Thus for this study, it will be used to confirm whether the two dimensional solution is better than a one-dimensional one (i.e. all variables load onto the same dimension).

Using all three data classification techniques serves to refine the latent measure of involvement as much as possible so that it is accurate, reliable and fits the data in the most appropriate way. Thus, PCA, PAF and CFA will be used to develop a latent variable of paternal involvement. The solutions produced by EFA and CFA often differ due to their different criteria for successful exploratory and confirmatory solutions. The purpose of EFA is to retain factors that account for significant amounts of variance in the data, while CFA assesses goodness-of-fit based on the variance remaining after the factors are taken into account. Therefore, whilst EFA can identify factors that account for significant variance in the data, CFA will highlight whether any significant additional variance remains. This means that the different techniques used by CFA and EFA are sensitive to different features of the data and can often produce different factor solutions (Floyd and Widaman 1995). Therefore, the process of using all three techniques will highlight the differences between each analysis and the implications of using each one interchangeably.

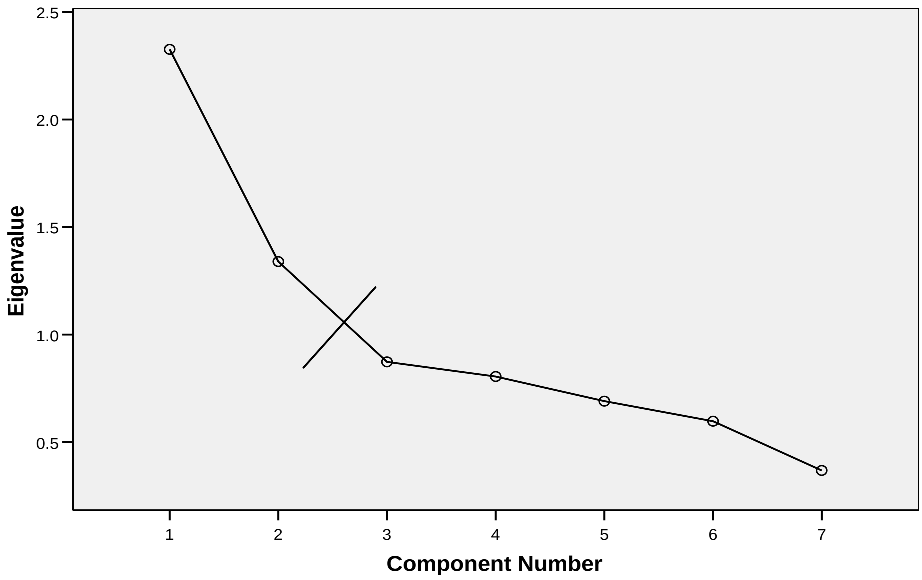
One of the most critical decisions in EFA is the number of factors to retain. Various methods have been proposed and we have selected two given they have been widely commended as the most reliable (e.g. see Hayton et al 2004; Franklin et al 1995). The first is a scree plot, which is a scatterplot of eigenvalues plotted against their ranks in terms of magnitude. This procedure proposes retaining as many factors as there are eigenvalues that fall below the last substantial drop on a scree plot. In addition to this, we also use a second method - Parallel Analysis (PA) - to further confirm the scree plot results. This procedure involves extracting eigenvalues from random datasets that parallel the actual dataset in terms of the number of cases and variables. Eigenvalues from the original dataset and the random datasets are then compared and the factors/components that are retained are those obtained from real data with eigenvalues greater than those generated by chance through PA (O'Connor 2000). Thus, factors which do not account for more variance than the parallel factors obtained for random numbers are not of interest because meaningful components extracted from actual data should have larger eigenvalues than parallel ones obtained from random data (Hayton et al 2004: 194). The results from the PA and scree plot are also shown in the next section with the results from PCA, PAF and CFA also presented respectively.

5. Paternal involvement at age nine months: results of the Principal Components Analysis

Seven variables were entered into PCA thus seven linear components were identified before extraction with eigenvalues initially set to 0. Missing cases were excluded in a listwise fashion meaning the sample size was reduced to 10,112[10].

We use both parallel analysis (PA) and a scree plot to verify whether two factors should be retained. The scree plot is shown in Figure 1 and indicates that two components should be retained since two components have been extracted before the last substantial drop shown by the line that cuts across the graph.

Figure 1: The Scree plot of all eigenvalues produced by PCA



Source: MCS 1, (n=10,112)

Next, PA was run using PCA extraction methods using random normal data generation. 10,112 cases were used in this analysis with seven variables to generate 1000 datasets with mean eigenvalues extracted for the 95th percentile. Results are presented in Table 2.

[10] The analysis (i.e. PCA extraction of three components) was also run with cases excluded pairwise to explore the impact of missing values. This had minimal effect on results: the same components were produced with very similar component loadings.

Table 2: Parallel Analysis results showing the raw data Eigenvalues and the Mean and Percentile Random Data Eigenvalues.

Root	Raw Data	Means	Percentile random data Eigenvalues
1.000000	2.341676	1.037124	1.050333
2.000000	1.337098	1.022103	1.031698
3.000000	.874949	1.010539	1.018293
4.000000	.804445	.999923	1.006524
5.000000	.680607	.989694	.996703
6.000000	.595251	.977425	.985779
7.000000	.365973	.963192	.974400

Source: MCS 1, (n=10,112)

Table 2 confirms the results of the scree plot: there are only two components with Eigenvalues greater than those which occurred by chance in a randomly generated dataset that has an equal number of cases and variables as the raw data. Thus, component 3 and above should be discarded.

Following confirmation by PA, PCA was run again and the Eigenvalues threshold was set to 1 in order to retain the first two components only. The components were obliquely rotated using direct oblimin and set with a delta of 0 in order to discriminate between the components and obtain a neater structure (Field 2005: 634; Abdi 2003). Rotating components obliquely is necessary in order to allow the factors to correlate, which is reasonable because, as discussed in chapter two, the dimensions of involvement are conceptually related.

Two components were extracted by the PCA and the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was used to gauge the stability of the component solution. This measures the ratio of the squared correlation between variables to the squared partial correlation between variables. It takes a value between 0 and 1 where a value of 0 indicates that the sum of the partial correlations is large relative to the sum of correlations, which suggests there is “diffusion” in the pattern of correlations and PCA is inappropriate for the analysis. A value between 0.5 and 1 indicates the pattern of correlations is relatively compact suggesting PCA yields distinct and reliable components (Field 2005). The KMO shows a value of 0.7, which suggests the components are reasonably stable (Field 2009: 647). Table 3 shows a list of the communalities.

Table 3: Communalities after extracting two components

Variable	Initial	Extraction
Frequency fathers look after the baby alone [BABY]	1.000	.435
Frequency fathers feed the baby [FEED]	1.000	.710
Frequency fathers get up in the night [GETUP]	1.000	.314
Frequency fathers change the baby's nappy [NAPPY]	1.000	.667
Who mostly cooks main meal [COOKING]	1.000	.286
Who mostly cleans the home[CLEANING]	1.000	.646
Who mostly does the laundry and ironing [LAUNDRY]	1.000	.621

The communality of a variable represents the variance that the particular variable shares with the extracted components so can be interpreted as measuring the *reliability of the indicator* (Floyd and Widaman 1995). Each communality is the squared multiple correlation for the variable as dependent, using the factors as predictors. Because PCA assumes all variance is common, or shared, all variables begin with a communality of 1.0 (i.e. 100% of the variance) and components are continually extracted until all of the variance is accounted for. For this analysis, eigenvalues are set to over 1.0, which means components are extracted only if they account for more variance than that which is captured by one single variable. After extraction, the communalities of all the variables were over 0, which suggests all the variables had some degree of common variance.

Although three of the communalities (i.e. those associated with [BABY], [GETUP] and [COOKING]) are fairly low, they do contribute to well-defined components. The remaining variables have relatively high communalities at over 0.5, which suggests the extracted components capture a high proportion of the variability in the variables measuring the frequency fathers feed the baby, change nappies, clean the house and do the laundry. The total variance (common and unique) captured by the two extracted components in the variable of 'involvement' was just over half (52.5%).

The components were rotated obliquely using direct oblimin because they initially had an ambiguous component structure. The delta score was increased from 0 to 0.2, to allow for a greater correlation between components but this resulted in components with only slightly higher correlations (0.282) and a structure that was relatively unchanged. Increasing the delta score even further caused the factor structure to disintegrate (where the factors to become less distinguishable). Therefore, the delta score was not increased further and a level of 0 was used.

Table 4 shows the pattern matrix of the rotated solution. Table 5 shows the structure matrix of the rotated solution, which takes into account the relationship between the factors.

Table 4: The pattern matrix of the components extracted by the PCA

Variable	Component	
	1	2
Frequency fathers look after the baby alone [BABY]	.659	.002
Frequency fathers feed the baby [FEED]	.837	.021
Frequency fathers get up in the night for the baby [GETUP]	.570	-.050
Frequency fathers change the baby's nappy [NAPPY]	.800	.064
Who mostly cooks main meal [COOKING]	.126	.492
Who mostly cleans the home [CLEANING]	-.063	.816
Who mostly does the laundry and ironing [LAUNDRY]	-.063	.800

Source: MCS1, (n=10,112)

Table 5: The structure matrix of the components extracted by the PCA

Variable	Component 1	Component 2
Frequency fathers look after the baby alone [BABY]	.660	.152
Frequency fathers feed the baby [FEED]	.842	.211
Frequency fathers get up in the night for the baby [GETUP]	.558	.079
Frequency fathers change the baby's nappy [NAPPY]	.814	.245
Who mostly cooks main meal [COOKING]	.237	.520
Who mostly cleans the home [CLEANING]	.123	.802
Who mostly does the laundry and ironing [LAUNDRY]	.119	.786

Source: MCS1, (n=10,112)

The pattern matrix (Table 4) clearly shows that two distinct components have been extracted. Face analysis suggest that factor 1 corresponds to 'engagement' (activities which are direct childcare) and factor 2 corresponds to responsibility 'responsibility' (non engaged activities that nevertheless contribute to the child's care). The structure matrix (Table 5) reiterates this structure. There is a slight, albeit fairly weak, correlation between factors nevertheless (.228). Some correlation is expected since both domestic work and childcare are related to 'responsibility' and such tasks are often done simultaneously e.g. doing the washing up whilst chatting to the child for example although the PCA results indicate engagement and responsibility are two dimensions of involvement that are not closely related.

The pattern matrix shows that feeding and nappy changing account for a greater proportion of the variance in fathers' engagement (component one) (70% and 64% respectively) than looking after the baby alone and getting up at night (42% and 32% respectively). Cleaning and laundry account for a greater proportion of the variance in paternal responsibility (component two) (66% and 64% respectively) than cooking main meals (25%).[11] The smaller loading that 'cooking' has on the second component of 'responsibility' (compared to the other domestic work variables) indicates

[11] The squared factor loading for an indicator variable is the percent of variance in that variable explained by the component; this is analogous to Pearson R^2.

that there is a significant proportion of variability unexplained by the factors. This suggests that the reasons that men cook or not may be unrelated to childcare responsibilities.

6. Paternal involvement at age nine months: results of the Principal Axis Factoring Analysis

The seven variables were entered into Principal Axis Factoring (PAF), which uses a similar strategy to PCA but seeks the least number of factors. It only accounts for the common variance (correlation) among a set of variables.

Table 6 shows the communalities that were produced from the seven variables (with missing cases excluded listwise and eigenvalues set to over 1 since this rule was previously confirmed by Parallel Analysis (see Table 2). This time, all communalities do not start from 1.0 (as in PCA) because PAF uses the correlation matrix in which the diagonal elements are iteratively derived estimates of the communalities (i.e. the R^2 of a variable using all factors as predictors) (Garson 2009).

Table 6: Communalities of the variables after extracting two components

Variable	Initial	Extraction
Frequency fathers look after the baby alone [BABY]	.190	.238
Frequency fathers feed the baby [FEED]	.449	.691
Frequency fathers get up in the night [GETUP]	.113	.142
Frequency fathers change the baby's nappy [NAPPY]	.424	.568
Who mostly cooks main meal [COOKING]	.079	.109
Who mostly cleans the home[CLEANING]	.186	.466
Who mostly does the laundry and ironing [LAUNDRY]	.177	.347

Source: MCS1, (n=10,112)

Communalities are generally lower compared to those produced by PCA (in Table 3) because they exclude unique variance; however, the variables with the lowest communalities are the same i.e. [BABY], [GETUP] and [COOKING] although they are even lower here suggesting the factors capture a particularly low level of variability in these tasks.

Table 7 shows that the total variance explained by the two factors was 36.6%, which is lower than the variation accounted for by the components extracted in the PCA (52.5%).

Table 7: Total variance explained by the two factors extracted in the PAF

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cum %	Total	% of Variance	Cum %	Total
1	2.342	33.453	33.453	1.817	25.963	25.963	1.753
2	1.337	19.101	52.554	.746	10.651	36.614	1.093
3	.875	12.499	65.053				
4	.804	11.492	76.545				
5	.681	9.723	86.268				
6	.595	8.504	94.772				
7	.366	5.228	100.000				

Source: MCS1, (n=10,112)

Table 8 shows the original factor matrix, which was then rotated in the same way as the PCA (direct oblimin with a delta of 0) in order to clarify the factor structure. The rotated pattern and structure matrix are shown in Table 9 and 10 respectively.

Table 8: The factor matrix of the factors extracted by PAF

Variable	Factor	
	1	2
Frequency fathers look after the baby alone [BABY]	.477	-.104
Frequency fathers feed the baby [FEED]	.794	-.246
Frequency fathers get up in the night [GETUP]	.366	-.092
Frequency fathers change the baby's nappy [NAPPY]	.733	-.175
Who mostly cooks main meal [COOKING]	.276	.182
Who mostly cleans the home[CLEANING]	.339	.593
Who mostly does the laundry and ironing [LAUNDRY]	.311	.501

Source: MCS1, (n=10,112)

Table 9: The pattern matrix of the factors extracted by PAF

Variable	Factor	
	1	2
Frequency fathers look after the baby alone [BABY]	**.482**	.017
Frequency fathers feed the baby [FEED]	**.847**	-.047
Frequency fathers get up in the night [GETUP]	**.377**	.001
Frequency fathers change the baby's nappy [NAPPY]	**.750**	.011
Who mostly cooks main meal [COOKING]	.132	**.261**
Who mostly cleans the home[CLEANING]	-.062	**.702**
Who mostly does the laundry and ironing [LAUNDRY]	-.031	**.600**

Source: MCS1, (n=10,112)

Table 10: The structure matrix of the components extracted by the PAF

Variable	**Factor**	
	1	**2**
Frequency fathers look after the baby alone [BABY]	**.488**	.186
Frequency fathers feed the baby [FEED]	**.830**	.250
Frequency fathers get up in the night [GETUP]	**.377**	.133
Frequency fathers change the baby's nappy [NAPPY]	**.754**	.274
Who mostly cooks main meal [COOKING]	.223	**.307**
Who mostly cleans the home[CLEANING]	.183	**.680**
Who mostly does the laundry and ironing [LAUNDRY]	.179	**.589**

Source: MCS1, (n=10,112)

Results from the PAF are comparable to results from the PCA: two distinct factors are extracted with childcare variables loaded onto factor one and domestic work variables loaded onto factor two. In the pattern matrix (Table 9), variables have slightly lower loadings on the factors compared to the PCA (except for [FEED]) but are roughly similar. Again, the structure matrix (Table 10) echoes this quite clearly with all variables loading relatively highly onto the same factor (as shown by the highlighted loadings) except for [COOKING], which once again, has a lower loading of .261. Thus, even when unique variance is excluded in PAF, loadings appear to be relatively similar to PCA suggesting that for this particular analysis, distinguishing between common and unique sources of variance has little impact on the results.

There is a stronger correlation between factors (0.350) than there is between PCA components (0.228) although this difference is only moderate. Increasing the delta to 0.2 increases the correlation of the factors to .415, suggesting that engagement and responsibility are moderately related.[12]

To summarise, the PAF extracted two latent involvement variables that were similar to the components extracted by the PCA: factor one represented engagement and factor two represented responsibility. Both exploratory techniques have produced similar results, which indicate that the seven variables from the MCS can be combined into two linear components or factors that

[12] Increasing the delta score still further caused the factor structure disintegrate.

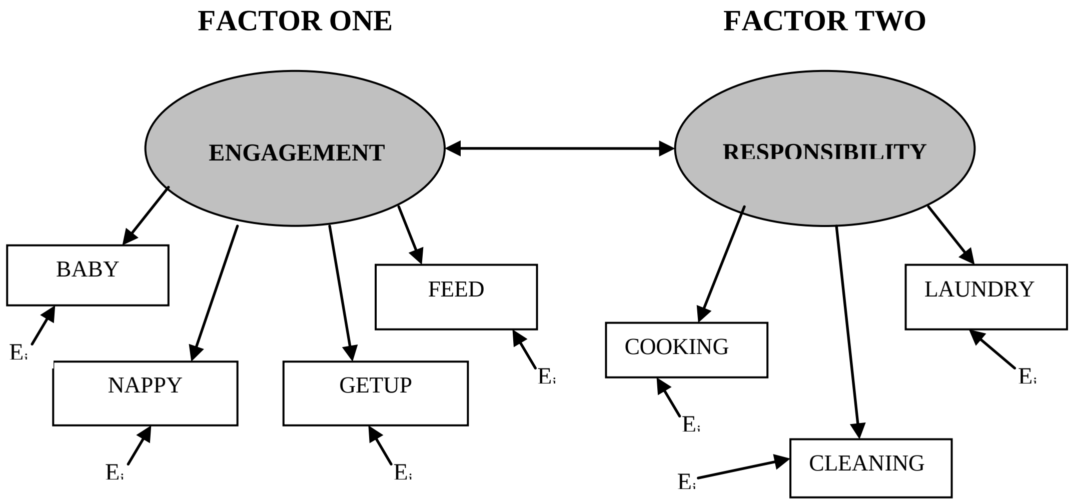
represent two dimensions of involvement. Using either method would derive the same latent variables although PAF produced factors that had a slightly higher correlation compared to PCA.

7. Paternal involvement at age nine months: results of the Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) is used here to *confirm* the two factor structure extracted from the variables in the exploratory analyses, as well as the theory that involvement can be classified into the dimensions of engagement and responsibility. This will be achieved by evaluating the 'goodness of fit' indices produced by CFA. The 'goodness-of-fit' tests are based on the variance remaining after the factors are taken into account, that is, while the PAF identified factors that accounted for significant variance in the data, the CFA refines this by highlighting whether any significant additional variance remains (Floyd and Widaman 1995).

The CFA model for this study specifies a priori measurement model with two factors: the first factor will have four item loadings - [BABY], [GETUP], [FNAPPY], [FEED] - and the second factor will have three item loadings: [COOKING], [CLEANING] and [LAUNDRY]. The factors should have a moderate correlation as shown by both exploratory factor analyses. As in the other analyses, missing cases were treated with a listwise deletion. The measurement model is summarised in Figure 2.

Figure 2: The measurement model: engagement and responsibility. Single headed arrows indicate direct effects; double headed arrows indicate unmeasured covariance.



E = error

An assumption of CFA is that variables have multivariate normal distributions. If variables violate this assumption then the chi-square statistic, which is regarded the main goodness-of-fit index, can be biased towards a Type I error (i.e. rejecting a model that should not be rejected) while standard errors can also become moderately or severely deflated. Multivariate normality is, however, rare for categorical indicators. An alternative is to use a polychoric correlation matrix in conjunction with weighted least squares estimates. This estimates the parameters of sub-interval variables in order to overcome the lack of multivariate normality in categorical data (Garson 2009; DiStefano and Hess 2005). Thus, the model is estimated using the means and variances adjusted weighted least squares (WLSMV) option rather than the simpler Weighted Least Squares (WLS).[13] WLSMV has also been recommended over the WLS method by Muthén, du Toit, and Spisic (1997) because the residuals tend to be closer to zero. DiStefano and Hess (2005: 236) state that the WLSMV estimation technique "rescales" categorical, non-normal data through a scaling factor that adjusts

[13] Yu (2002) explains that the former uses the diagonal weight matrix with robust standard errors and a mean and variance adjusted chi-square test statistic whereas the latter uses the full weight matrix to compute standard errors and chi-square.

for non-normality. This results in a chi-square fit statistic that is adjusted for both mean and variance anomalies within the data and more robust standard error terms.

For the two-factor model to be successfully confirmed, the measurement model must be significantly different from the *null model* in which the covariances in the covariance matrix for the latent variables are all assumed to be zero. There are various measures of fit that can be used in CFA; these usually require a null model against which the specified model is compared.

Goodness-of-fit indices indicate whether a model can be successfully confirmed while ‘difference tests’ indicate whether the proposed two factor model (in Figure 2) fits the data better than a one factor model in which all variables load onto the same factor. Various indices of goodness-of-fit exist (see Chau and Hocevar 1995 for a review). One of the main indices is the chi-square test for goodness of fit which provides the difference between expected and observed covariance matrices.. Thus, a *significant* chi-square indicates *lack of satisfactory fit* because the model’s covariance structure is significantly different from the observed covariance matrix; in these cases the model should be rejected (Garson 2009). However, large sample sizes and non-normal data tend to bias the chi-square test in favour of model rejection so whenever large samples are used - such as the sample for this analysis - the chi-square is likely to have a significant p-value. Therefore, a negative model chi-square finding may be discounted for large samples (e.g. 200 or more cases), providing other model fit measures support it, because the chi-square can be misleading (Garson 2009; DiStefano and Hess 2005).

Since the sample size for this study exceeds 200 cases, three alternative goodness-of-fit tests will be used. These other tests compare the given model with a null model, or one with no structure, which is usually the independence model where chi-square is at its maximum (i.e. the worst case model). These tests are referred to as *relative fit* and are fairly independent of sample size, therefore deemed more reliable than the chi-square test (see DiStefano and Hess 2005). The first test is the *Bentler Comparative Fit Index (CFI),* which compares the existing model fit with a null or independence model where the latent variables are assumed to be uncorrelated. In other words, the CFI compares the covariance matrix predicted by the model to the *observed* covariance matrix to assess the percent of lack of fit, which is accounted for by going from the null to the specified model. The CFI ranges from 0 to 1 where a value closest to 1 indicates a very good fit of the data. By convention, the CFI should have a value greater than 0.9 for the model to be successful as this would indicate 90% of the covariation in the data can be reproduced by the given model (Garson 2009).

The second test, similar to CFI, is the *Tucker-Lewis Index (TLI)* (also known as the non-normed fit index). Again, this reflects the proportion by which the specified model improves fit compared to the null model. Thus, a value of .50 means that the model improves fit by 50% compared to the null model. Again, a value closest to 1 indicates a good fit with a cut off of about >=.95 as confirmation of good model fit. Generally, values below .90 indicate a need for model respecification. Thus CFI and TLI will be used to assess whether the model summarised in Figure 2 is a good fit of the data since the chi-square is expected to be significant due to the large sample size. Both have been commended as acceptable indicators of model fit (Chau and Hocevar 1995; Garson 2009). The Root Mean Square Error of Approximation (RMSEA) is the third test of good fit and relates to the residual in the model. This evaluates whether the fit between model and data is "close" and takes a value between 0 and 1 with a smaller value indicating better model fit. Acceptable model fit is indicated by a value of 0.06 or less according to Chau and Hocevar (1995)[14]. RMSEA is a popular method of fit in CFA partly because it does not require comparison with a null model. It is also less biased by sample size (although it can overestimate goodness of fit for very small samples). Garson (2009) argues that RMSEA is useful because it corrects for model complexity as shown by the fact that degrees of freedom is its denominator. Thus, the chi-square will be noted but successful model fit will be determined by the CFI, TLI and RMSEA.

[14] RMSEA = sqrt (CS/(n*df)) where CS is the chi-square, n is the sample size, and df is the degrees of freedom.

Tables 11 and 12 present the results of the CFA model. Table 11 shows a summary of the chi-square for the baseline model (when only the constant and no variables are in the equation) along with a summary of the chi-square and other model fit indices (CFI, TLI and RMSEA) for the independent model (when all variables are included). Table 12 shows the standardised parameter estimates of the variables (equivalent to the factor loadings in the PCA and PAF).

Table 11: Results of goodness-of-fit tests from the CFA

	Chi-Square	**df**	**CFI**	**TLI**	**RMSEA**
Baseline model	23569.4**	13	--	--	--
Independent model	252.5**	10	0.990	0.987	0.045

**p=<0.001 Source: MCS 1 (n=10,112)

Table 12: Standardised parameter estimates from the CFA

Variable	**Estimate**	**Std. Error**	**Est. Std. Error**[1]
Factor 1			
FATHBABY	0.509**	0.008	66.961
FATHFEED	0.850**	0.006	137.881
FATHGETUP	0.403**	0.010	41.325
FATHNAPPY	0.817**	0.006	126.167
Factor 2			
COOKING	0.511**	0.015	35.140
CLEANING	0.735**	0.015	49.372
LAUNDRY	0.766**	0.016	48.593
Correlation of factors = 0.390			

**p=<0.001 Source: MCS 1 (n=10,112)

[1] Estimated Standard Error i.e. the estimate of the sample mean's standard error, i.e. s / n, where n is the sample size and s is the estimate of the population standard deviation.

In Table 11, the first fit-measure is the chi-square. The larger the difference in the chi-square between the baseline and independent model, the more the independent variables can be seen to contribute to the model by more than just a random amount. There appears to be a large difference between the chi-square statistic for the baseline and the chi square statistic for the independent model, which suggests the variables make a significant improvement to the model. Although the chi-square is significant (p=<0.001), this will be discounted here for reasons already discussed. The CFI, TLI and RMSEA show that the two-factor model is a good fit to the data. The CFI is well over the 0.9 cut-off and indicates that 99% of the covariation in the data can be reproduced by the given model. Similarly, the TLI is over the cut-off of 0.95, also confirming the model is a good fit of the data. The RMSEA is also less than the recommended cut-off of 0.06 providing further confirmation of good model fit.

In Table 12, the parameter estimates of the variables are listed along with their standard errors. Each estimate is statistically significant at the 1% level. Standardised parameter estimates are used since they constitute transformations of unstandardised estimates so that their measures are the same and can therefore be used for informal comparisons of parameters. Standardised estimates correspond to effect-size estimates (Suhr 2006).

According to Fortnell, Tellis and Zinkhan (1982)[15], parameter estimates of .70 or more are considered acceptable because the amount of variation shared with a latent construct is greater than the error variance. There are also guidelines for parameter estimates suggested by Comrey and Lee (cited in Distefano and Hess 2005) whereby loadings greater than .70 are considered excellent, .63 are very good, .55 are good, .45 are fair and .32 are poor. Although these are in relation to exploratory factor analysis, Distefano and Hess have suggested these criteria can be used as an approximate guide for CFA.

According to these guidelines, all variables have “very good” loadings with the exception of three variables, which can be considered “fair to good”. These are [BABY] [COOKING] and [GETUP] where the latter variable has the lowest loading at 0.4. This corresponds to the results of the PCA and PAF, where communalities (and loadings for [COOKING] and [GETUP]) were lowest for these variables. This suggests there is minimal difference between the solutions produced from the

[15] cited in Hulland et al (1996)

exploratory and confirmatory factor analyses. None of the variables have "poor" loadings so are all considered to be useful here.

The final analysis concerns the difference test or *chi-square difference testing of measurement invariance*, which establishes whether the specified two factor model fits the data better than a one factor model. Here the original (multifactor) model is compared to one that is constrained by forcing all the correlations among the factors to be 1.0. If the constrained model is not significantly worse than the unconstrained model, then a one-factor model fits the data just as well as a multi-factor model and so is to be preferred on the basis of parsimony (Garson 2009).

The 'difference test' works by comparing model "H0" and model "H1" where H0 is *nested* within H1 (the less restricted model). The chi-square value and degrees of freedom of the less restrictive model (H1) is subtracted from the chi-square value and degrees of freedom of the nested, more restrictive model (H0). The chi-square difference value is then compared to the chi-square value in a chi-square table using the difference in degrees of freedom between the more restrictive and less restrictive models. A significant chi-square difference value indicates that constraining the parameters of the nested model significantly worsens the fit of the model whereas a non-significant chi-square difference value indicates that constraining the parameters of the nested model does not significantly worsen fit. Thus, the difference test value must be *significant* for the two-factor model to be confirmed as a better fit of the data (Muthén and Muthén 2007).

The difference test value is statistically significant at the 1% level indicating that the two factor model fits the data significantly better than a one factor model.[16] This confirms that the results of both PCA and PAF are good representation of the latent structure in these data.

[16] Muthén (2007) points out that the chi-square and degrees of freedom are adjusted to obtain a correct p-value in the difference testing. Therefore it is *only the p-value that is meaningful* here; the chi-square is not a true indicator of the model so will not be reported.

8 Summary and conclusions

This paper has described the process of deriving latent variables, of *paternal involvement*, for a sample of employed fathers within the MCS in order to represent their involvement with their children aged nine months. Two dimensions of involvement – corresponding to two elements of Lamb's conceptualisation engagement and responsibility – were extracted from childcare and domestic work variables. PCA and PAF were run using seven variables in order to explore the data and CFA was run afterwards to confirm the structure. Table 13 provides a summary of the component/factor loadings from each model (taken from the pattern matrix in the exploratory analyses).

Table 13: A summary of model results from the PCA, PAF and CFA

Variable	**Loading***					
	PCA		**PAF**		**CFA**	
	C1	C2	F1	F2	F1	F2
BABY	**.659**	.002	**.482**	.017	**.509**	--
FEED	**.837**	.021	**.847**	-.047	**.850**	--
GETUP	**.570**	-.050	**.377**	.001	**.403**	--
NAPPY	**.800**	.064	**.750**	.011	**.817**	--
COOKING	.126	**.492**	.132	**.261**	--	**.511**
CLEANING	-.063	**.816**	-.062	**.702**	--	**.735**
LAUNDRY	-.063	**.800**	-.031	**.600**	--	**.766**
Correlation (C1/F1 & C2/F2)	**.228**		**.350**		**.390**	

C1/2 = Component 1/2, F1/2 = Factor 1/2.

**Pattern matrix Loadings for the PCA and PAF*

Table 13 confirms that the three data classification procedures produced similar results: two factors were extracted from the exploratory PCA and PAF and these were further confirmed by the CFA. Two variables have consistently low loadings in all three analyses: 'getting up at night' and 'cooking main meals'. This indicates that these two activities account for the least amount of variation in the factors of engagement and responsibility as extracted. Clearly some of the variability in these variables is due to factors not captured within the two factor model suggesting that the generic factors of engagement and responsibility may not present a complete picture of paternal involvement. Nevertheless, the two variables to load moderately onto the factors and overall the factors account for over 50% of the variation in the observed variables.

Each procedure also shows a small correlation between components/factors of between 0.2 and 0.4 suggesting there is a weak to moderate association between engagement and responsibility. When only common variance is accounted for by the factors (in the PAF), the correlation between them is higher. Increasing the delta coefficient to 0.2 to allow for greater correlation between factors increases the association between factors even further but this is not by a substantial amount. Increasing the delta score still further caused the factor structure to disintegrate. Therefore, a delta of 0 was used. Since the PAF and CFA are judged to be more robust than PCA, it can be concluded that engagement and responsibility are moderately correlated at around 0.4.

The purpose of using all three techniques, aside from developing the most accurate and robust measure, was to explore the differences within each technique and the effect that these differences would have on results and further analyses that use the latent variables. This exercise revealed that there is unlikely to be any practical difference in which of the PCA components or the PAF factors are used as variables in tertiary analyses (the correlation between the scores on the PCA components and PAF factors were 0.97 for component/factor one and 0.98 for component/ factor two). However, these conditions did not appear to cause any major differences to the component/factor structure that was produced by all three methods suggesting Floyd and Widaman's thesis does not apply in this case.

This paper has demonstrated that PCA can provide a sound initial insight into the structure of data, which can be further refined through PAF and CFA when the final goal is to derive a latent measure. The use of all three techniques has resulted in a more robust and accurate analysis of the seven dependent variables resulting in two, moderately correlated latent variables, which could be used in further analyses to measure the conditions to model paternal involvement. The method employed in this paper involves a triangulation of techniques to produce latent variables of involvement to cross-check their validity and reliability. Here, the same two factors have been produced in three different ways suggesting that they effectively summarise two dimensions of paternal involvement when children are aged nine months. This also challenges the criticisms directed mainly to the technique of PCA, which has often been labelled an inferior and more biased method compared to PAF. Here PAF and CFA produce the same result as PCA with the only differences being slightly higher factor loadings and slightly lower correlations between components for the PCA.

Appendix

Table 1: Frequency fathers contribute to the childcare tasks: looking after the baby, feeding the baby, changing the baby's nappy and getting up at night for the baby

Frequency	Look after baby on own	Feed baby	Change nappy	Get up at night
More than once a day	15.5	24.3	35.5	7.0
Once a day	15.1	28.4	19.7	8.1
A few times a week	28.6	27.6	22.0	15.7
Once or twice a week	24.2	10.2	7.3	14.5
Less than once a week	12.6	5.7	7.4	17.3
Never	3.8	3.7	8.0	24.2
Total	**99.9**	**99.9**	**99.9**	**86.8**
Baby never wakes up	n/a	n/a	n/a	13.1
Missing: Not applicable/refusal	0.1	0.1	0.1	0.1
TOTAL	**100.0**	**100.0**	**100.0**	**100.0**

Source: MCS 1, aged nine months (n=11,767)

Table 2: Frequency fathers contribute to the housework tasks: cooking the main meal, cleaning the house and doing the laundry and ironing

Frequency	Cooking	Cleaning	Laundry
Father does most	10.9	2.0	2.2
Shares equally with partner	21.8	20.6	13.9
Mother/other does most	66.5	76.6	83.2
Total	**99.2**	**99.2**	**99.2**
Missing: Not applicable/refusal	0.8	0.8	0.8
TOTAL	**100.0**	**100.0**	**100.0**

Source: MCS 1, aged nine months (n=11,767)

Table 3: Spearman correlation matrix of the seven variables measuring paternal involvement

	Baby	Feed	Nappy	Get up	Cooking	Cleaning	Laundry
Baby	1.000						
Feed	0.403	1.000					
Nappy	0.357	0.645	1.000				
Get up	0.192	0.296	0.285	1.000			
Cooking	0.117	0.172	0.207	0.073	1.000		
Cleaning	0.104	0.136	0.151	0.065	0.232	1.000	
Laundry	0.093	0.131	0.151	0.078	0.199	0.406	1.000

Source: MCS 1, aged nine months (n=11,767)

References

Abdi, H. (2003). *Factor Rotations. Encyclopedia for research methods for the social sciences*. A. Lewis-Beck, A. Bryman and T. Futing. Thousand Oaks (CA), Sage.

Albright, J. J., & Park, H. M. (2008). *Confirmatory Factor Analysis using Amos, LISREL, Mplus, SAS/STAT CALIS*. Technical Working Paper. The University Information Technology Service (UITS), Center for Statistical and Mathematical Computing, Indiana University.

Cabrera, N., C. Tamis-LeMonda, et al. (2000). *Fatherhood in the Twenty-First Century*. Child Development 71(1): 127-136.

Chau, H. and D. Hocevar (1995). *The effects of number of measured variables on goodness-of-fit in confirmatory factor analysis*. Paper presented at the annual conference of the American Educational Research Association: San Francisco.

Coltrane, S. and R. Parke (1998). *Reinventing Fatherhood: Toward an Historical Understanding of Continuity and Change in Men's Family Lives*. Philadelphia, National Center on Fathers and Families

Dermott, E. (2003). *'The Intimate Father': Defining paternal involvement*. Sociological Research Online 8(4)

Dermott, E. (2008). Intimate Fatherhood: A sociological analysis. London, Routledge.

DiStefano, C. and B. Hess (2005). *Using Confirmatory Factor Analysis for Construct Validation: An Empirical Review.* Journal of Psychoeducational Assessment 23: 225-241.

Doucet, A. (2006). *'Estrogen-filled worlds': fathers as primary caregivers and embodiment.* The Sociological Review 54(4): 696-716.

Fabrigar, L., D. Wegener, et al. (1999). *Evaluating the Use of Exploratory Factor Analysis in Psychological Research.* Psychological Methods 4(3): 272-299.

Field, A. (2005). Discovering Statistics using SPSS (2nd edition). London, Sage.

Field, A. (2009). Discovering Statistics using SPSS (3rd edition). London, Sage.

Floyd, F. and K. Widaman (1995). *Factor Analysis in the Development and Refinement of Clinical Assessment Instruments*. Psychological Assessment 7(3): 286-299.

Franklin, S. B., D. J. Gibson, et al. (1995). *Parallel analysis: a method for determining significant principal components*. J. Veg. Sci. 6: 99-106.

Garson, G. D. (2009). *Factor Analysis*. http://faculty.chass.ncsu.edu/garson/PA765/factor.htm (Accessed 10 April 2010)

Hayton, J. C., D. G. Allen, et al. (2004). *Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis*. Methodological Resources 7(2): 191-205.

Hulland, J. Chow, Y. and Lam, S. (1996) *Use of causal models in marketing research: A review,* International Journal of Research in Marketing Volume 13, Issue 2, April 1996, Pages 181–197

Lamb, M. (ed) (1986). *The Father's Role: Applied Perspectives*. New York, John Wiley & Sons.

Mcbride, B. and G. Mills (1993). *A Comparison of Mother and Father Involvement With Their Preschool Age Children*. Early Childhood Research Quarterly 8: 457-477.

Mikelson, K. (2008). *He Said, She Said: Comparing Mother and Father Reports of Involvement.* The Journal of Marriage and Family 70: 613-624.

Morman, M. and K. Floyd (2006). *Good Fathering: Father and Son Perceptions of What It Means to Be a Good Father*. Fathering 4(2): 113-136.

Muthén, B., S. H. C. du Toit, et al. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished technical report. Available at: http://www.statmodel.com/papers.shtml.

Muthén, L. K. and B. O. Muthén (2007). *Mplus User's Guide*, 5th edition. Los Angeles, Muthén & Muthén.

O'Brien, M. (2005). *Shared caring: bringing fathers into the frame.* Working Paper Series No. 18, Manchester: Equal Opportunities Commission.

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(3): 396-402.

Preacher, K. and R. MacCallum (2003). *Repairing Tom Swift's Electric Factor Analysis Machine*. Understanding Statistics 2(1): 13-43.

Sanderson, S. and V. Sanders-Thompson (2002). *Factors Associated With Perceived Paternal Involvement in Childrearing*. Sex Roles 46(3/4): 99-111.

Sayer, L., S. Bianchi, et al. (2004). *Are Parents Investing Less Time in Children? Trends in Mothers' and Fathers' Time with Children*. American Journal of Sociology 110(1): 1-43.

Suhr, D. (2006). *The Basics of Structural Equation Modeling.* http://www.lexjansen.com/wuss/2006/tutorials/TUT-Suhr.pdf (Accessed 5 May 2010).

Tinsley, H. and D. Tinsley (1987). *Uses of Factor Analysis in Counseling Psychology Research*. Journal of Counseling Psychology 34(4): 414-424.

Williams, S. (2008). *What is Fatherhood: Searching for the reflexive father.* Sociology 42(3): 487-502.