Refining the Measurement of Women's Autonomy

Rina Agarwala
Woodrow Wilson School of Public Policy and Office of Population Research
Princeton University

Scott M. Lynch
Department of Sociology and Office of Population Research
Princeton University

Rina Agarwala
284 Wallace Hall
Princeton University
Princeton, NJ 08544
(609) 258-5088

agarwala@princeton.edu

Scott Lynch
114 Wallace Hall
Princeton University
Princeton, NJ 08544
(609) 258-7255

slynch@Princeton.edu

Refining the Measurement of Women's Autonomy

Women's autonomy has long been a central concern for researchers examining the social position of women in developing countries. To date, however, most research has placed little emphasis on the measurement of autonomy; instead, research has relied on simple summed autonomy indexes. In this research, we examine the measurement properties of women's autonomy by explicitly considering measurement error and the differential importance of indicators that are thought to reflect autonomy. We use confirmatory factor analyses to determine a) which indicators, among a variety, produce the most appropriate measures of different dimensions of autonomy; b) whether a factor analytic approach offers a better fit to the data than the summed indexes that are commonly used; and c) whether the measurement of autonomy is consistent across two developing countries (India and Pakistan). We find that our indicators adequately capture four distinct dimensions of autonomy, that summed indexes lose considerable information relative to factor analytic measures, and that, while the model structures replicates fairly well across two countries, there are measurement differences that may make comparative research more difficult than expected.

Refining the Measurement of Women's Autonomy

The concept of women's autonomy has been an important one in social demography and sociology for at least the last two decades (e.g., (Mason 1986)[1]. Autonomy has been variously defined as "the ability to influence and control one's personal environment" (Safilios-Rothschild 1982); "the capacity to obtain information and make decisions about one's private concerns and those of one's intimates (Dyson and Moore 1983a); and "the degree of access to and control over material and social resources within the family, in the community, and in the society at large" (Dixon-Mueller 1978). These definitions highlight the important implications that women's autonomy can have on women and their families and thus point to autonomy as an important concept to study. Empirically, increased female autonomy has been shown to be correlated with reduced fertility, improved child nutrition and education, and improved standards of living for women (Balk 1994; Basu 1992; Dyson and Moore 1983b). Based on several such empirical studies conducted on the implications of women's autonomy during the 1980s and early 1990s, the United Nations asserted in 1995 that women's autonomy should be considered a basic human right and essential to human dignity (UNDP 1995).

[1] Various terms, including "status", "autonomy", "empowerment", have been used over time to capture some element of gender equality in the household and community. While "status" implied a more static state and was often confused with "prestige" or "esteem" in the eyes of men, "autonomy" and "empowerment" referred to a more dynamic process of challenging existing power relations and gaining greater control over sources of power. (see Batliwala, Srilatha. 1994. "The Meaning of Women's Empowerment: New Concepts from Action"." in *Population Policies Reconsidered*, edited by G. Sen, A. Germain, and L. Chen. Boston: Harvard University Press, Dyson, Tim and Mick Moore. 1983a. "On Kinship Structure, Female Autonomy, and Demographic Behavior in India." *Population and Development Review* 9:35-60, Mason, Karen. 1986. "The Status of Women: Conceptual and Methodological Issues in Demographic Studies." *Social Forces* 1:284-300.) In this research, we use the term "autonomy".

While several studies have highlighted the important determinants and effects of women's autonomy (Balk 1994; Balk 1997; Blumberg 1994; Hashemi, Schuler, and Riley 1996; Jejeebhoy 2000; Jejeebhoy and Sathar 2001; Morgan and Niraula 1995a), few have seriously considered how autonomy should be measured. Most research has decomposed the concept of autonomy into several dimensions (such as mobility, control over income, power within the household etc.); each dimension is measured separately with simple scale variables or indexes comprised of the sum of several, often dichotomous, observable items (Balk 1994; Balk 1997; Hashemi, Schuler, and Riley 1996; Jejeebhoy 2000; Morgan and Niraula 1995a). Such an approach, while useful in highlighting the complexity and multidimensionality of autonomy, undermines the conceptualization of autonomy as a single, overarching concept.

In an attempt to retain the concept of autonomy as a single concept (yet multidimensional at the same time), some studies have added a composite index of "total autonomy" to their analyses of each individual dimension of autonomy. The composite index of "total autonomy" is then comprised of simply the sum of the dimensions (Hashemi, Schuler, and Riley 1996; Jejeebhoy 2000; Jejeebhoy and Sathar 2001). This approach suffers as well, because, although it recognizes the unifying construct of autonomy, it ignores that the separate dimensions of autonomy may not be perfectly related. Regardless of which approach is used (i.e. measuring autonomy as separate dimensions and measuring autonomy as a sum of individual dimensions), both suffer from (1) ignoring the differential importance of the dimensions and the differential importance of the items *within* a single dimension and (2) from ignoring measurement error in the observed items that make up each dimension.

In this research, we examine the measurement properties of women's autonomy items by explicitly considering measurement error and the differential importance of the observable items or indicators that are thought to reflect the larger autonomy construct. We use confirmatory factor analyses to determine a) which observable items, among a variety, produce the most reliable measures of different dimensions of autonomy; b) whether a factor analytic approach offers a better fit to the data than the summed indexes that are commonly used (at both the level of individual dimensions and the level of total autonomy); and c) whether the measurement of autonomy is consistent across two developing countries (India and Pakistan). To answer the first question, we use confirmatory factor analysis on data from India to test and refine a multidimensional measure of autonomy. To answer the second question, we treat summed indexes (within and among autonomy dimensions) as a special case of our more general measurement model of autonomy as a single construct and measure the loss-of-fit incurred by ignoring measurement error in the observed items and differential weighting of autonomy indicators. To answer the third question, we replicate our model using data from Pakistan.

BACKGROUND

Indirect Proxy Measures of Autonomy

Since Karen Mason's seminal research in 1986 first highlighted the conceptual and methodological issues in social demographic studies on women's autonomy (Mason 1986), a consensus on how we *cannot* measure autonomy has at least emerged. Autonomy cannot be measured, as it often was, using simple, unidimensional proxies or

indirect measures, such as women's education or labor force participation rates. These measures attempted to use a single observable characteristic as an approximate indicator of autonomy. For example, a high education level was said to indicate a high degree of autonomy, and a low education level was said to indicate a low degree of autonomy.

Such indirect measures are highly imperfect and have grave policy implications, however, especially when used to analyze the predictors and effects of autonomy (Balk 1994; Jejeebhoy 1991; Vlassof 1994). First, proxies for autonomy are extremely context dependent. For example, while women's education may be highly correlated with women's increased autonomy under one setting, it might be completely uncorrelated in another. Although many scholars argue that women's increased education is correlated with women's increased autonomy, Oropesa (1997) and Sather et al. (1998) argue that in Pakistan, educational institutions merely reinforce dominant cultures of gender inequality and, therefore, do not necessarily translate into increased women's autonomy as it might elsewhere. Using education level as a proxy for autonomy in such settings could give erroneous results on predictors and effects of autonomy.

Second, proxy measures do not provide ample evidence for how well the measure is captures the construct of interest—in this case, autonomy. This is particularly problematic when different studies use different proxies to measure autonomy. The various proxies may be differentially correlated to autonomy and differentially correlated to one another. For example, Dyson and Moore use nearness to kin as a proxy for women's autonomy, because they find that nearness to kin is a significant determinant of women's sense of security and power (Dyson and Moore 1983a). Sather et al, on the other hand, use women's education, work participation, and age at marriage as three

different proxies for autonomy (Sathar, Crook, Callum, and Kazi 1988). It is unclear whether one of these proxies is a better or worse measure of autonomy than another one. If all these proxies are equally fit measures of autonomy, then they should also be highly correlated with one another. However, studies have not been able to show that nearness of kin is highly correlated with women's education level or work participation. The apparent lack of correlation or relation between the different proxy measures of autonomy make it difficult to compare the strength of various measures and compare the results from studies using different proxies for autonomy. This, in turn, makes it difficult to draw conclusions on autonomy per se.

Third, single proxy measures of autonomy blur the channels through which autonomy works. In other words, it is unclear whether the measured effect of autonomy is due to the direct effects of the proxy variable or the indirect effects of autonomy. For example, when using education as a proxy for autonomy to study the effects of women's autonomy on fertility, we cannot be certain whether a decrease in fertility is due to education for reasons unrelated to autonomy or due to education's effect on autonomy, which in turn affects fertility. Moreover, even if it the latter is true, it is unclear how education is related to autonomy. For example, does education provide women with increased opportunities for mobility, which in turn increases their autonomy, or does the material learned through education teach women how to be more autonomous, or both? This ambiguity from studies using proxy measures to understand autonomy makes policy translations of empirical findings difficult.

Finally, and perhaps most importantly, single item proxy measures of autonomy fail to capture the multidimensionality of autonomy. Empirical studies have long argued

that autonomy is a complex construct comprised of multiple dimensions (Whyte 1978). Mason classified these dimensions into three broad categories: power, prestige, and wealth (Mason 1986). Different dimensions may be determined by, and may predict, different demographic and socioeconomic factors. Using single item proxies to measure autonomy, however, obscures which dimension of autonomy is actually being measured. For example, in South Asia, women's labor force participation (one common proxy used to measure autonomy) has been found to be correlated with increased power in the household, but decreased social prestige (Jejeebhoy and Sathar 2001; Mason and Smith 1999; Menon-Sen and Kumar 2001; Sathar, Crook, Callum, and Kazi 1988). In West Africa, on the other hand, scholars have noted that women's labor force participation increases their economic power, but has no effect on their social or legal power (Safilios-Rothschild 1980). Therefore, while the proxy might indeed be reflecting autonomy, it sheds little light on which *aspect or dimension* of autonomy it is capturing. As noted above, the proxy might measure one dimension of autonomy in one context and an entirely different dimension in another.

On the whole, indirect measures of autonomy soon appeared to be inadequate in understanding the details of women's autonomy, its predictors and its effects.

Direct Measures of Autonomy

Recently, scholars have turned from using indirect proxies to quantify autonomy to using more direct measures. These direct measures consist of a combination of observable items or indicators that are categorized into different dimensions of autonomy, such as access to and control over resources, participation in economic and child-related

decisions, self-esteem, mobility, freedom from domestic violence, and political awareness and participation, (Balk 1994; Balk 1997; Blumberg 1994; Hashemi, Schuler, and Riley 1996; Jejeebhoy 2000; Jejeebhoy and Sathar 2001; Morgan and Niraula 1995a).

Studies using a direct-measure approach have addressed many of the inadequacies of the earlier indirect-measure approach. Most important, they have explicitly quantified the mutli-dimensionality of autonomy. We can now better understand which factors affect or are affected by which specific dimensions of autonomy. In addition, direct-measure studies have illuminated the channels through which economic and social factors (such as education and labor force participation) affect autonomy, rather than confounding 'causes' and 'effects' of autonomy—a point that was missing from the early literature on autonomy. Studies using the direct-measure approach have thus made substantial contributions to our knowledge of the factors that influence, and are influenced by, autonomy.

Some of the inadequacies of the indirect-measure approach, however, remain unaddressed in the direct-measure approach literature on autonomy. First, few have addressed the issue of context dependency. While the earlier literature was criticized for using indirect proxies that are context dependent, the current literature fails to analyze the context dependency of direct measures. For example, elements that may be empowering in one context may be irrelevant or impossible in another. One way to address this issue is to allow the weights (e.g. factor loading) of the various direct measures (or observable items) to vary by context. In this way, we can better approach a measure of autonomy

that uses direct measures and is flexible enough to accommodate different cultural contexts.

Second, research using the direct-measure approach has still not shown how reliably direct measures reflect different dimensions of autonomy. While the indirect-measure literature has failed to show how well one proxy measures autonomy as compared to another, direct-measures fail to highlight the differential importance of the various direct measures or observed items used to measure a single dimension of autonomy. Most direct-measure studies construct each dimension of autonomy as a simple summative index of a series of items attained through dichotomous questions. For example, mobility is often measured by asking women whether they are able to go alone to a series of places, such as the market, the health center, the next village etc. Those who answer "yes" receive 1 point, those who answer "no" receive 0 points. At the end of the series of questions, a respondent's answers are aggregated to equal her score for mobility ranging from 0 to the total number of questions asked on mobility. This procedure is then repeated for each dimension (Balk 1994; Hashemi, Schuler, and Riley 1996; Jejeebhoy 2000).

Few studies test the inter-item reliability of each index by measuring the correlations between the items within each dimension. Among the few cases in which the internal consistency of a dimension is tested, the measurement error within each item is not considered. Balk (1994), for example, uses data from rural Bangladesh to study the effects of women's mobility, decision-making, and gender attitudes (three dimensions of autonomy) on their fertility. She uses Cronbach's alpha coefficients to measure the reliability or internal consistency of each of her dimensions and finds that her mobility

measure is not particularly internally consistent. However, she does not model the measurement error within the items used to measure mobility, and thus her results may be biased downward as the mobility dimension may be comprised of items with very high measurement error.

Of all research on autonomy, Mason's works are perhaps the most significant in paying considerable attention to the measurement of autonomy. Mason (1996), using data from a five-country study, analyzes the determinants of women's economic decision-making power in the family—one of the several dimensions of autonomy often studied. Using principal components analysis, she finds that only 6 of the 8 questions thought to reflect this dimension of autonomy have strong factor loadings. Therefore, her subsequent analysis of women's economic decision-making power uses an index created by summing only those six items. In a subsequent paper, Mason analyzes two additional dimensions of autonomy, mobility and threat of violence by their husbands. The mobility index consists of seven items regarding whether the respondent could go alone to the following places: the market, a health center, fields outside the village, community center, the home of a relative or friend, a fair, a temple, or to the next village. Although Mason writes that the items are strongly interrelated according to principal components analysis, she discusses another analysis using the Rasch Model that suggests not all seven items scale equally well in all five countries studied. The index on domestic violence consists of two questions (1) whether the husband beats the respondent, and (2) whether the respondent is afraid to disagree with her husband. Mason writes, "this scale [index] is relatively weak and is used here primarily for convenience" (Mason 1997), p. 3). Like

Balk, Mason tests the strength of each index but does not consider the measurement error on each item within in the dimensional index.

Finally, while the direct-measure approach has indeed highlighted the important multidimensionality of autonomy, some of the literature has undermined the justification for analyzing autonomy as a single, empirical concept by analyzing the predictors or effects of each dimension of autonomy separately (Morgan and Niraula 1995b; Vlassoff 1992). In such cases, scholars are effectively forcing the correlations between the dimensions of autonomy to be zero. Although these scholars begin and conclude with arguments regarding autonomy in general, their analyses do not test the strength of the relationships between the dimensions of autonomy and thus provide no empirical evidence for autonomy as a single, overarching concept. Vlassoff (1992), for example, uses longitudinal data to study the associations and change over time between women's autonomy and their fertility in rural India. Her study analyzes the effect of three separate dimensions of autonomy (control over resources, decision-making power, and mobility) with no analysis of the relationship between the three dimensions. Control over resources and mobility are found to be significantly associated with desired fertility, while decision-making power is not. Based on these findings, Vlassoff writes, "The main conclusion to be derived from these findings is that the decline in fertility goals…could not be attributed to changes in women's status [or autonomy]". By not combining her three dimensions of autonomy, however, the power of her conclusion regarding the net effect of autonomy is weakened.

Morgan and Niraula (1995) also study the association between women's autonomy and fertility, using data from rural Nepal. They, too, divide autonomy into

three dimensions: women's control over resources, decision-making power, and mobility, with each dimension measured by a series of questions. However, the relationships or correlations between these dimensions are not tested (nor are the inter-item reliability of the questions tested). Toward the end of the paper, Morgan and Niraula write that they "construct a simple index of individual autonomy". The index on autonomy, however, is comprised of only three of the eight questions under one dimension of autonomy: mobility. The other two dimensions, along with the other 5 sub-questions under mobility, are left out of the scale for individual autonomy. Unlike Vlassoff, they find that increased women's autonomy reduces fertility. Once again, however, by not analyzing the strength of the correlations between dimensions and justifying the final selection of dimensions, the conclusions regarding the effect of autonomy as a single concept are undermined.

Among the few studies that do analyze the strength of the relationship between the dimensions of autonomy, the relationships appear to be rather low to moderate. Balk (1994), for example, presents an inter-index correlation matrix for total autonomy, where the relationships between each dimension of autonomy studied are positive and significant. Balk notes that the correlation coefficients between the dimensions are low (none exceed 0.30), providing evidence for the "successful measurement of largely distinct dimensions of women's status" (Balk 1994), p. 43). Jejeebhoy uses the same data set studied in this research to analyze the determinants of women's autonomy in rural India. Using partial correlation coefficients, controlling for state and religion, she analyzes the strength of the associations between six dimensions of women's autonomy (economic decision-making, child-related decision-making, mobility, freedom from

threat, access to resources, and control over resources). While she finds that the associations are generally positive and significant, 95 out of the 105 coefficients presented are less than .25—implying a rather moderate association overall. Most notably, Jejeebhoy finds that the associations with freedom from threat and control over resources were weak and inconsistent—implying that these measures may reflect an aspect of autonomy entirely distinct from the other four dimensions (economic and child-related decision-making, access to resources, and mobility) (Jejeebhoy 2000).

In both cases, the authors do not discuss why, if the dimensions of autonomy are so weakly correlated, we should study autonomy as a single measure at all. Moreover, neither study takes out the measurement error within the items used to measure each dimension of autonomy. Nor do they allow the observed items within each dimension or the dimensions themselves to have differential weights. For example, in creating a measure for total autonomy, Jejeebhoy sums all six dimensions into a single index for autonomy. By doing so, she forces all the dimensions to have equal weights and a correlation of one. Instead, allowing for measurement error and differential weights could possibly produce higher correlations between the different dimensions of autonomy.

Hashimi et al. (1996) use a slight variation on the summative index for each dimension by making each dimension dichotomous. Respondents are given a score of 1 for being "empowered" on a certain dimension if they score in the 25^{th} to 30^{th} percentile on all the questions reflecting that dimension, and a 0 for "unempowered" on that dimension if they score below the 25^{th} percentile. In an attempt to analyze overall "autonomy", they create a composite indicator in which respondents are classified as

"empowered" if they score a 1 on 5 or more of the 8 empowerment dimensions measured. However, there is little justification given for the chosen dimensions or for the cut-off points, and a considerable amount of measurement error is introduced by the dichotomization of the items.

Mason and Smith (1999), using the same data set across all five available countries, also find a weak correlation between threat of violence and the other autonomy dimensions studied (economic and fertility decision-making and mobility). However, neither study looks at measurement error in the observed measures. Consideration of measurement error may shed greater light on why these dimensions are not as strongly correlated to the other dimensions, and to what extent we can continue to analyze the weakly associated dimensions as a part of an overall construct of autonomy.

In sum, the literature to date has typically a) ignored measurement error in the observable items thought to capture the different dimensions of autonomy, b) failed to consider the differential importance of items within a dimension, and c) either failed to consider the relationship between dimensions of autonomy or ignored the differential importance of the different dimensions by creating a simple summed index of various dimensions. We argue here that greater attention must be paid to these issues concerning the measurement of autonomy if the concept is to be useful in studying the status of women in developing countries and if comparative work is to be at all possible.

DATA

This paper uses data from the Survey on the Status of Women and Fertility (SWAF), a survey conducted in 1993-94 on women and their husbands in five countries

in Asia (Smith, Ghuman, Lee, and Oppenheim Mason 2000). This study analyzes the data from two of the five survey countries, namely India and Pakistan. The survey was specifically designed to measure women's autonomy and its relationship to reproductive behavior. It is one of the first surveys to try and operationalize the multiple dimensions of autonomy.

The India survey includes 1,842 rural Indian women. It covers both North and South India, and both Hindus and Muslims. The survey was conducted in two states--Uttar Pradesh (UP) in North India and Tamil Nadu (TN) in South India. Traditionally, UP has had lower indicators of gender equality than TN. The survey samples evenly from two more developed sub-districts and two less developed sub-districts in each state[2]. It also draws evenly between Hindus and Muslims and over samples castes that are numerically small. In order to ensure this ethnic mix in the sample, villages were merged into clusters of 1,000-2,000 households. One cluster from each of the sub-districts was chosen at random and houses in each chosen cluster were put on house lists, which constituted the sampling frame. Approximately 800 currently married women, aged 15-39, were randomly selected for interview in each of the four sites. Husbands who were present were also interviewed.

The Pakistan survey includes 1,050 rural Pakistani women. All interviews were conducted in the province of Punjab, which covers 52 percent of Pakistan's population and 56 percent of its geographical area. The province was divided into three agro-ecological zones that were developed by the Pakistan Agricultural Research Council. The three zones, North Barani Belt, the Central Zone, and the Southern Zone, represent varying degrees of development, cultural and linguistic traditions, and feudal regimes, all

[2] Development indicators were based on income, percentage of roads surfaced, and other economic criteria

of which impact gender equality indicators. The North Barani Belt reflects higher indicators of gender quality than the highly feudal Southern Zone. Based on Pakistan's Federal Bureau of Statistics' master list of rural Primary Sampling Units (PSUs), ten sites were randomly selected from the three zones. Prior to sampling, the PSU list was restricted to sites with a population size between 2,500 and 4,999 in 1991. After conducting a household census in each site, a sampling fraction was adopted to ensure a minimum of 100 interviews in each of the ten sites.

METHODS

We use a modified confirmatory factor analysis approach to examine the measurement of autonomy (Bollen 1989). Confirmatory factor analysis, unlike exploratory factor analysis, places *a priori* structure on the data and allows the explicit testing of competing hypotheses regarding the measurement properties of indicators thought to reflect a theoretical construct. We began with a set of 54 items thought to reflect autonomy; all 54 items have been used to represent autonomy in previous research. We then revised our preliminary models based on inspection of the parameter estimates and modification indices. We constructed the following four dimensions of autonomy for testing: a) autonomy from violence (7 items); b) autonomy in family decisions (16 items); c) autonomy in community involvement (15 items); and d) autonomy in household economics (16 items). Table 1 presents a listing of all 54 items, along with the dimensions of autonomy they are thought to reflect.

All items were coded as ordinal. Estimators for confirmatory factor analyses require covariance or correlation matrices as input for estimation. These matrices are

generally computed using Pearson product-moment covariances or correlations. However, these measures are inappropriate for ordinal data. Given that the variables measuring autonomy in this study (as well as most others) are measured at the ordinal level, a more appropriate approach is to estimate polychoric correlations between the variables, and to use these resulting matrices as input into the structural equation modeling software. Polychoric correlations are a measure of the correlation between the latent continuous (and normally distributed) variables thought to underlie the crudely measured observed variables. Methodological studies have shown that, aside from being a more theoretically appropriate measure of the association between ordinal variables, polychoric correlations correct for the negative bias that using Pearson correlations produces, making the results of structural equation analyses stronger (Joreskog and Sorbom 1988). Thus, for these analyses, matrices of polychoric correlations were used in estimation, and the weighted least squares estimator was used to estimate all measurement models.

The analyses proceeded in four steps. (1) We first estimated simple one-factor models for each dimension of autonomy. These models were then revised. In some cases, significant similarity in wording of items within each dimension required the inclusion of either correlations between errors of similar items or the inclusion of an item-level random effect to eliminate the effect of wording. In other cases, beyond wording differences, the initial models indicated that the items reflected different sub-dimensions of autonomy. In those cases, we revised the model to consist of more than one substantive latent factor. In some cases, certain items did not load well on any factor. These items were deleted, leaving us with a revised, final set of factors. (2) In the next

step, we tested each of these final models against one roughly equivalent to the standard approach to measuring autonomy: using summed scales. (3) After the dimensions of autonomy were investigated, we then combined all dimensions into a single model in order to determine the extent to which the dimensions of autonomy were related. Finally, (4) we replicated the analyses using data from Pakistan in a multiple group analyses forcing model parameters to be identical across both sets of data in an effort to determine the extent to which the concept of autonomy may be useful for comparative research.

RESULTS

One-Dimensional Measurement Models

Violence

Figure 1a shows the initial model estimated for the dimension of autonomy from violence. As the figure indicates, all 7 violence indicators were initially assumed to reflect a common factor. Table 2 shows that the first model estimated had an excellent fit, as assessed by the Incremental Fit Index (IFI=.99), but had a significant chi-square (361.49, 14 d.f., $p<.001$) and an RMSEA greater than .1. By these measures, the model does not fit the data well (see Bollen, 1989: IFIs should be above .9 for an acceptable fit and above .95 for an excellent fit; chi squares should be nonsignificant; and RMSEAs should be below .05 for an excellent fit and below .1 for acceptable fit). An examination of the model parameters revealed that items v1 and v7 loaded poorly on the factor. The poor loading could be due to either measurement or substantive differences between these items and the other items. Items v2-v6 ask whether the respondent feels a husband would be "justified in beating his wife" under various circumstances, while items v1 and v7 ask

about actual fear and experience of violence. Wording differences may account for the poor loading of v1 and v7, or it may be a substantively meaningful difference between perceptions of violence and the perceived legitimacy of violence. In our second model, we added another latent variable representing a wording effect. The loadings for this effect were all constrained to be 1, with the variance of the latent variable free, making the latent variable essentially a random effect. This model fit the data better, with a much smaller chi square, an RMSEA below .1, and an IFI just under 1. In a third model, we removed the latent variable for wording a created a separate latent variable. This model fit the data better than the previous model and was retained as the final model. Thus, autonomy from violence is represented by two factors: one reflecting perceptions of the legitimacy of violence in the household; the other reflecting perceptions of actual violence. Figure 1b shows the structure of the final model.

Family Decisions

As with violence, autonomy in making (non-economic) household decisions was first modeled as a single latent variable with all 16 items loading on it. This initial model did not fit the data well, with a large chi square, and RMSEA>.1 and an IFI of .78 (see Table 2 for results of the following sequence of models). The loadings for v22 and v23 were very low (as were their reliabilities—the explained variance for these indicators), and modification indices indicated that adding a correlation between the errors of v13 and v14 would significantly improve the model's fit. This error correlation is reasonable, given that both items concern the schooling of children and have very similar wording. In the second model, items v22 and v23 were eliminated, and the error correlation was

added. This model fit the data considerably better, but the RMSEA and IFI still suggested the model needed improvement. Modification indices suggested that the model would be strengthened by the addition of error correlations between several of the items v16-v21. As with several of the violence items, these items are all similarly worded, asking whether the respondent feels uncomfortable speaking in front of various people. In our third model, we thus added a latent variable to capture a wording effect. The RMSEA for this model was acceptable, as was the IFI. Modification indices suggested that v9 loaded very poorly (as it had in the previous models), and that an error correlation between v8 and v10 would improve model fit considerably. Thus, in our final model we dropped v9 and added the suggested error correlation. The RMSEA for this model was no better than that of the previous model, but the IFI improved slightly (from .92 to .93). Finally, we attempted a two-factor model in lieu of the wording effect, but that model did not have a better fit than the wording effect model. Figure 2 shows the final model for autonomy in family decisions.

Community Involvement

The base model for autonomy in community involvement had a reasonable RMSEA (.08), but an unacceptable IFI (.89). As with the previous set of analyses, a number of error correlations were suggested by the modification indices (between items v33-v37). These items ask whether the respondent has engaged in several activities in the previous week. Thus, in the second model, we added a latent variable to capture this wording effect. This model fit the data better, with the IFI now indicating acceptable fit (.91). Modification indices suggested the addition of an error correlation between v35

and v36, and in the next model we added the correlation. The results of that model indicated very good fit of the model to the data, with the RMSEA dropping to .06 and the IFI increasing to .95. In this model, items v24 and v38 continued to evidence poor loadings and reliabilities. In a final model, we eliminated these variables. The results of the final model indicated excellent fit: the RMSEA was .05, and the IFI was .97. As before, we attempted one final model with a substantive latent factor included rather than a wording effect, but those results were not better than those of model 4. Figure 3 shows the final model for autonomy in community involvement.

Household Economics

The initial one-factor model for autonomy in household economic decisions fit the data very well, with an RMSEA of .10 and an IFI of .98 (see Table 2). However, the modification indices suggested that the inclusion of an error correlation between items v47 and v48 would significantly improve the model's fit. In the second model, we added this error correlation with a slight improvement in the RMSEA and IFI. Modification indices suggested the addition of error correlations between items v39-v44, items which have very similar wording. Thus, in the next model, we included a wording effect for these items. This model fit the data only slightly better, based on the RMSEA. v54 continued to load poorly in this model, so in a final model, we dropped the variable. This model fit the data only slightly better (but not seen after rounding of the RMSEA). Figure 4 shows the final model for autonomy in household economics.

Comparison to Summed Scales

Our next step in the analyses was to compare these final four models to summed scales to determine if summed scales produce a significant lack of fit of the models to the data. In these models, all factor loadings were constrained to 1, wording effects were removed, and all measurement errors and error covariances were set to 0. Table 3 shows the results of these models. In all cases, after setting the factor loadings equal to 1 for all variables and constraining the measurement errors to be 0, the model chi squares (as well as the RMSEAs and IFIs, not reported in table) showed a significant loss of fit.

Four-Factor Combined Model

In the next step in the analyses, we combined all four (five, counting the second substantive violence scale) autonomy factors into a single factor analysis model in an effort to determine the extent of the relationship between the latent factors. Table 4 shows the correlations between the substantive factors. These results reveal several interesting patterns. First, the correlations between the second violence factor (perceptions of violence) and the other factors are either very weak (between the first violence factor and the second) or statistically 0 (between the second violence factor and the other latent factors). Second, the correlations between the family decisions factor and the household economic decisions factor is very close to 1. Third, the remaining correlations reveal moderate relationships between all other factors.

India-Pakistan Comparison

We replicated the results for the Indian data with virtually identical data from Pakistan. In order to conduct the replication, we conducted two sets of analyses. First, we re-estimated the models using the Pakistan data alone. Second, we estimated a multiple group model first allowing the parameters for the Indian and Pakistani data to be freely estimated across groups. Next, we constrained all parameters to be equal across the two datasets. Finally, we constrained all the factor loadings to be equal across the datasets but allowed the error variances and variances of the latent variables to be freely estimated across groups. In these analyses, we had to examine each dimension of autonomy independently, because the sample size for the Pakistan data was too small to estimate the complete asymptotic covariance matrix (used in WLS estimation using polychoric and polyserial correlations) for all the variables.

Table 5 presents the results of these two comparisons. The first column of the table presents the results of the models for the Pakistan data only. The results for the model for violence are comparable to the results for the data for India. The RMSEA is below .1, and the IFI is .99. The results for the model for household decisions are, in fact, better for the Pakistan data than the India data. The RMSEA indicates excellent fit at .045, and the IFI indicates the same at .98. In contrast, the model for community involvement does not appear to fit as well to the Pakistan data. The RMSEA was .102 (compared to .05 in the Indian data), and the IFI was .82 (compared to .97 in the Indian data). Finally, the model for household economic decisions indicated comparable fit to the Pakistan data. In that model, the RMSEA was .058 (slightly better than the .08 for the India data), and the IFI was .98 (very slightly worse than the .99 for the India data).

The remaining columns in the table show the results from multiple group analyses of the Indian and Pakistani data together. The first of the remaining columns shows the results of estimating the parameters freely across the countries; the second column shows the results of constraining all parameters to be equal across groups; and the third column shows the results of constraining only the factor loadings. For the sake of brevity, we do not discuss all of these results. In brief, the results of chi-square difference tests (not shown in table, but easily found by subtracting the unconstrained chi square from the constrained chi square) show that a significant loss of fit results if parameters are forced to be equal. Although the loss of fit is considerably less when all the variances and covariances of the latent variables—as well as the error variances—are freely estimated across data sets, difference chi square tests continue to show a significant loss of fit. However, the overall fit of the models, based on the RMSEAs and IFIs, show the models fit quite well.

DISCUSSION AND CONCLUSION

In this research, we have examined in detail the measurement of women's autonomy, an important theoretical construct in demographic literature on development. In doing so, we hope to contribute to a deeper discussion on how best to model autonomy in future empirical research. The results of the analysis provide several interesting insights that can help further our understanding of women's autonomy in developing countries.

First, building on Mason's argument that autonomy is a multidimensional concept, we measured the robustness of four dimensions of autonomy: freedom from

violence, participation in household economic decisions, participation in non-economic family decisions, and community involvement. The results of this analysis show that autonomy items do indeed cluster into distinct and meaningful dimensions insofar as the measurement models fit the data well, and the separate autonomy latent variables were intercorrelated but not perfectly so. With regard to violence, we found this aspect of autonomy should be further divided into actual violence and views on the legitimacy of violence. Nevertheless, our results lend concrete support to Mason's argument for the multidimensionality of autonomy.

Second, we tested each dimension of autonomy against models that reflect the most common approach to measuring autonomy: summed scales. We found that, although such scales have been the most common method of measuring autonomy in the literature to date, measurement error and differential weighting of the items included in such scales make such scales poor measures of autonomy. These results indicate that researchers should not use simple summed scales in measuring autonomy.

Third, we combined all four dimensions of autonomy into a single model to determine the extent to which the dimensions of autonomy were related. Our results showed that the dimensions of autonomy are, for the most part, moderately related. This finding indicates that the various dimensions of autonomy can be considered as being related to a single underlying construct, but also that they have distinct contributions to autonomy. An important exception, however, is the perception of violence dimension. Since the perception of violence was weakly correlated with the experience of violence as well as the other dimensions, we argue that it should not be used in measures of autonomy. These findings suggest that autonomy indicators should not be summed to

produce a single measure of women's autonomy, but rather that each dimension should be examined separately. The only exception may be regarding the family decisions and household economic decisions constructs, given that the correlation between these latent constructs was very close to 1.

Finally, we replicated the India analyses using data from Pakistan to examine the extent to which autonomy may be useful for comparative research. These results indicate that the model *structure* for autonomy is replicable across countries (based on the acceptable fit of the models of the Pakistan data only). On the other hand, the results are somewhat less clear in terms of whether the measurement of autonomy is comparable in terms of factor loadings and variances. For example, the community involvement dimension did not seem to fit as well in the Pakistan data as in the India data. Ultimately, the results suggest that comparative research using measures of autonomy should, at a minimum, allow for differences in measurement errors and variances of latent variables measuring autonomy. Once again, this finding suggests that summed scales should not be used in studying autonomy, especially cross-culturally.

While this study has provided the most systematic study of the measurement of women's autonomy to date, it is not without limitations. A key limitation is that the items used in measuring autonomy in this research are survey-specific. That is, other surveys use different measures and may therefore reach other conclusions regarding the extent of the relationship between dimensions of autonomy. Nonetheless, we feel that greater attention at least needs to be paid to how autonomy is measured in future research.

A second limitation to these analyses is that we used data for only two countries: India and Pakistan. A more substantial test of the comparability of measurement of autonomy across developing countries is needed. However, data limitations to date preclude such a test. Specifically, autonomy is often measured with different indicators not only in different surveys, but also even within the same survey applied in different countries. This limitation is thus not so much a shortcoming of this research specifically, but is a shortcoming of any comparative work on women's autonomy.

REFERENCES

Balk, Deborah. 1994. "Individual and Community Aspects of Women's Status and Fertility in Rural Bangladesh." *Population Studies* 48:21-45.

—. 1997. "Defying Gender Norms in Rural Bagladesh: A Social Demographic Analysis." *Population Studies* 51:153-172.

Basu, Alaka Malwade. 1992. *Culture, the status of women, and demographic behaviour : illustrated with the case of India*. Oxford & New York: Clarendon Press & Oxford University Press.

Batliwala, Srilatha. 1994. "The Meaning of Women's Empowerment: New Concepts from Action"." in *Population Policies Reconsidered*, edited by G. Sen, A. Germain, and L. Chen. Boston: Harvard University Press.

Blumberg, Rai Lesser. 1994. "Women's Work, Wealth, and Fmaily Survival Strategy: The Impact of Guatemala's ALCOSA Agribusiness Project." in *Women, The Family, and Policy*, edited by E. Chow and C. Berheide. Albany: State University of New York Press.

Bollen, Kenneth A. 1989. *Structural equations with latent variables*. New York: Wiley.

Dixon-Mueller, Ruth. 1978. *Rural women at work : strategies for development in South Asia*. Baltimore: Published for Resources for the Future by the Johns Hopkins Press.

Dyson, Tim and Mick Moore. 1983a. "On Kinship Structure, Female Autonomy, and Demographic Behavior in India." *Population and Development Review* 9:35-60.

Hashemi, Syed, Sidney Schuler, and Ann Riley. 1996. "Rural Credit Programs and Women's Empowerment in Bangladesh." *World Development* 24:635-653.

Jejeebhoy, Shireen J. 1991. "Women's Status and Fetility: Successive Cross-sectional Evidence from Tamil Nadu, India 1970-1980." *Studies in Family Planning* 22:217-230.

—. 2000. "Women's Autonomy in Rural India: Its Dimensions, Determinants, and the Influence of Context." in *Women's Empowerment and Demographic Processes: Moving Beyond Cairo*, edited by H. Presser and G. Sen. New York: Oxford University Press.

Jejeebhoy, Shireen J. and Zeba Sathar. 2001. "Women's Autonomy in India and Pakistan: The Influence of Religion and Region." *Population and Development Review* 27:687-712.

Joreskog, Karl G. and Dag Sorbom. 1988. "PRELIS: A Preprocessor for LISREL (2nd edition), Scientific Software."

Mason, Karen. 1986. "The Status of Women: Conceptual and Methodological Issues in Demographic Studies." *Social Forces* 1:284-300.

—. 1997. "How Family Position Influences Married Women's Autonomy and Power in Five Asian Countries." East-West Center, Honolulu.

Mason, Karen and Herbert Smith. 1999. "Female Autonomy and Fertility in Five Asian Countries."

Menon-Sen, Kalyani and AK Shiva Kumar. 2001. *Women in India: How Free? How Equal?* New Delhi: UNDP, India.

Morgan, S. Phillip and Bhanu Niraula. 1995a. "Gender Inequality and Fertility in Two Nepali Villages." *Population and Development Review* 21:541-561.

Safilios-Rothschild, Constantina. 1980. "A class and sex stratification theortetical model and its relevance for fertility trends in the developing world." Pp. 189-202 in *Determinants of Fertility Trends: Theories Re-Examined*, edited by C. Holn and R. Machensen. Liege: Ordina Editions.

—. 1982. "Female Power, Autonomy and Demogrpahic Change in the Third World." in *Women's Role and Population Trends in the Third World*, edited by R. Anker, M. Buvinic, and N. Youssef. London & Canberra: Croom Helm.

Sathar, Zeba, Nigel Crook, Christine Callum, and Shahnaz Kazi. 1988. "Women's Status and Fertility Change in Pakistan." *Population and Development Review* 14:415-432.

Smith, Herbert L., Sharon Ghuman, Helen J. Lee, and , and Karen Oppenheim Mason. 2000. "Status of Women and Fertility." Pp. Machine-readable data file.

UNDP. 1995. *Human Development Report*. NY, NY: Oxford University Press.

Vlassof, Carol. 1994. "Gender Inequalities in Health in the Third World: Uncharterd Ground." *Social Science and Medicine* 39:1249-1259.

Vlassoff, Carol. 1992. "Progress and Stagnation: Changes in Fertility and Women's Position in an Indian Village." *Population Studies* 46:195-212.

Whyte, Martin King. 1978. *The Status of Women in Preindustrial Societies*. Princeton: Princeton University Press.

Figure 1a. Initial Measurement Model for Autonomy from Violence.

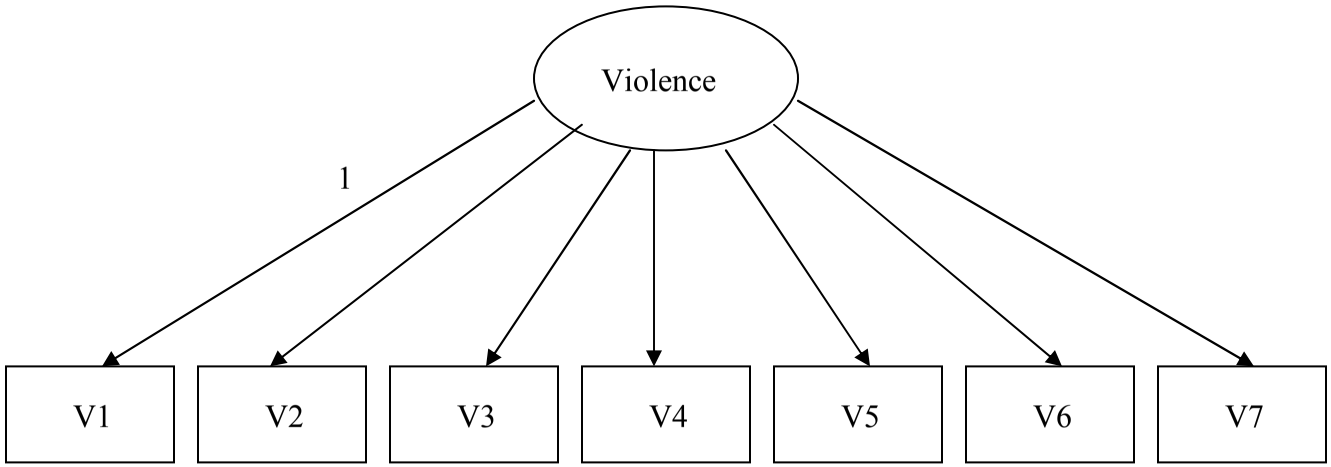

Figure 1b. Final Measurement Model for Autonomy from Violence.

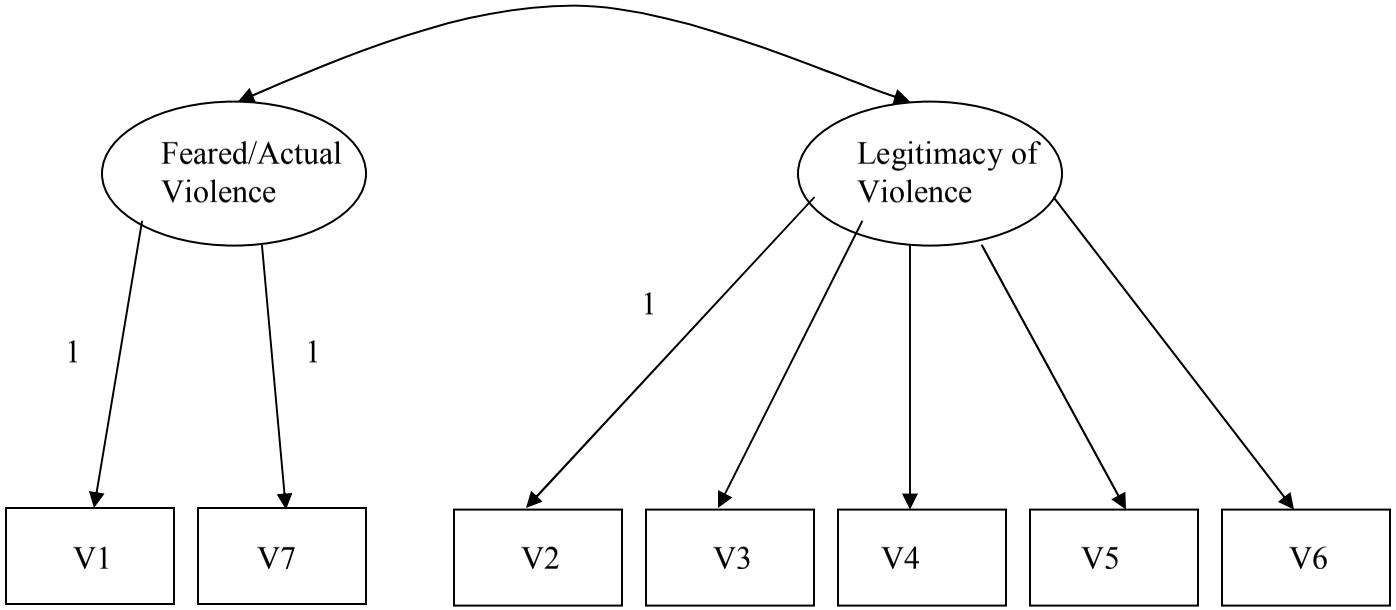

Figure 2. Final Measurement Model for Autonomy in Family Decisions.

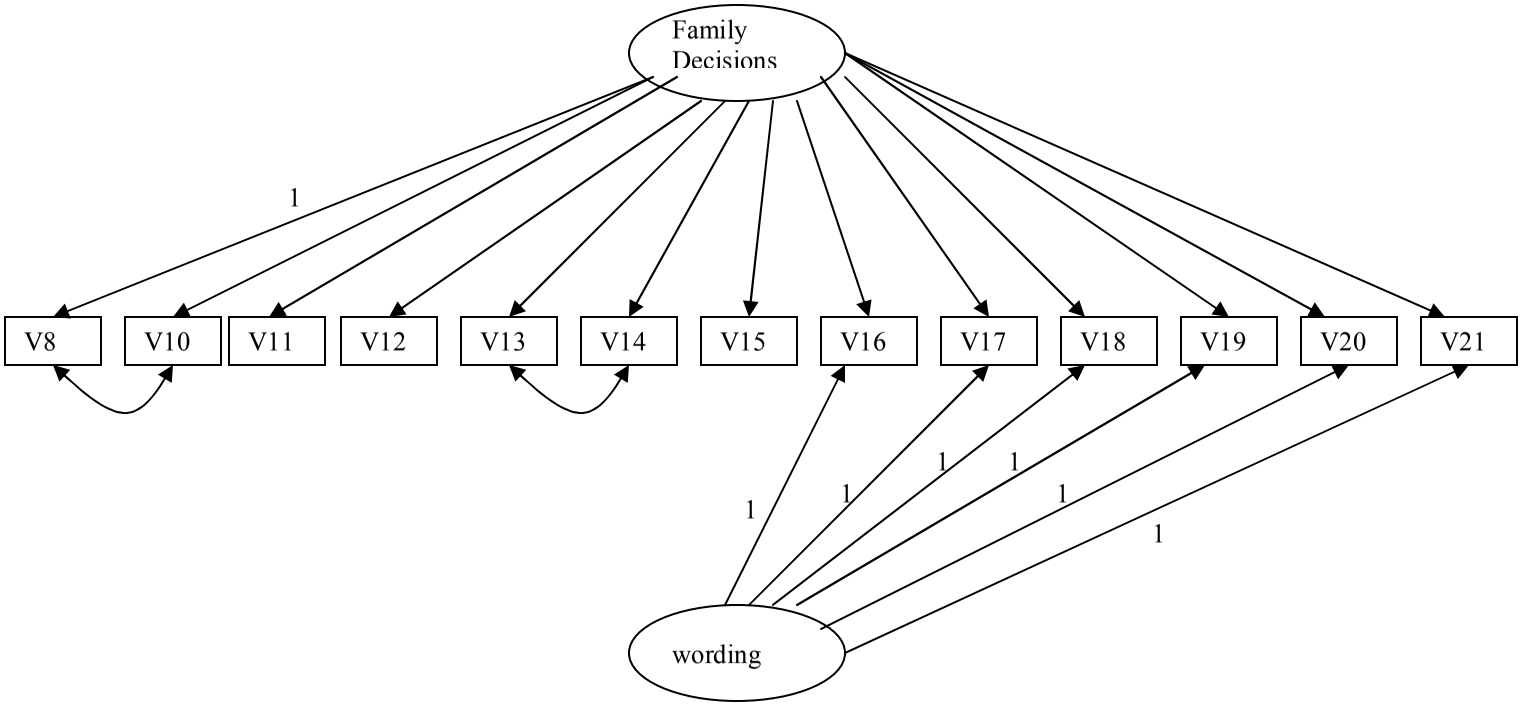

Figure 3. Final Measurement Model for Autonomy in Community Involvement.

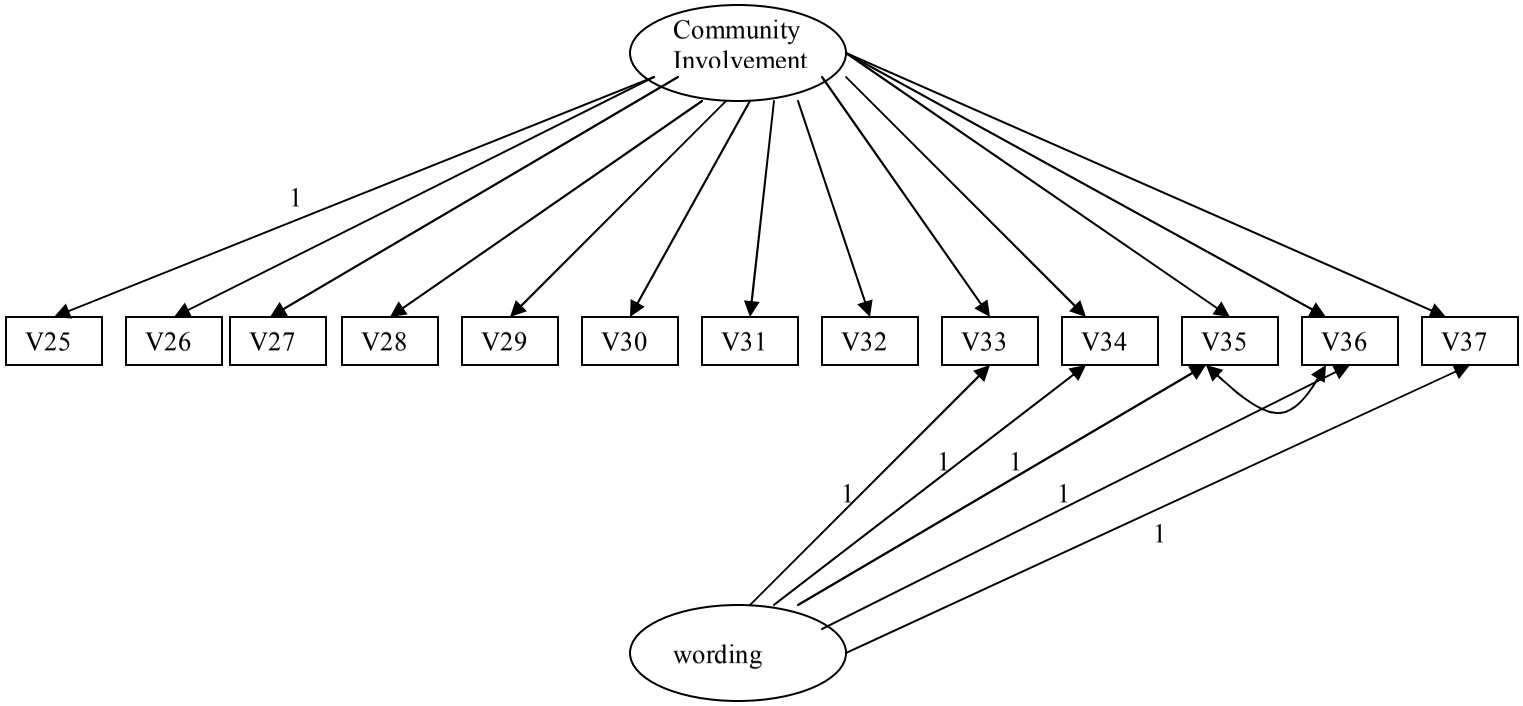

Figure 4. Final Measurement Model for Autonomy in Household Economic Decisions.

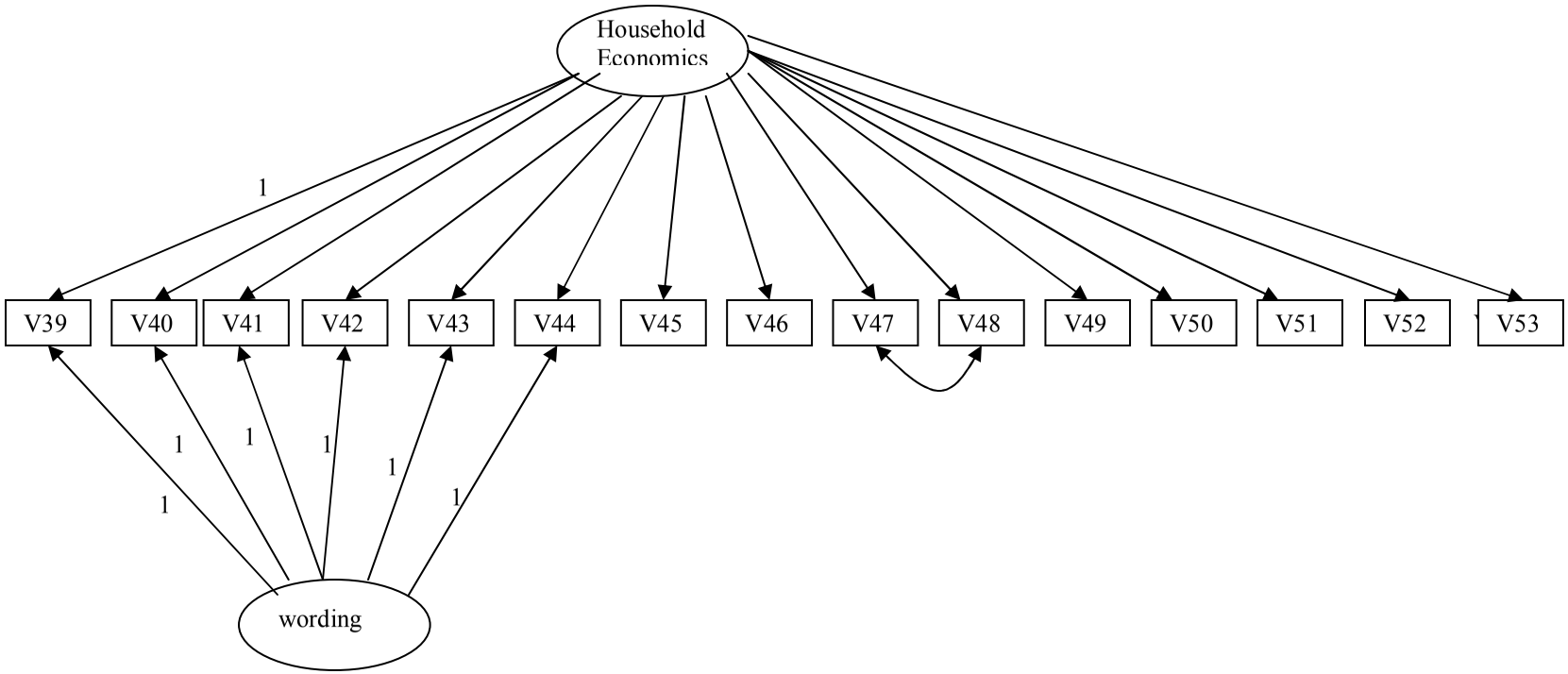

Table 1. Autonomy Indicators and Dimensions of Autonomy Used in the Analyses

Violence

1. Are you afraid to disagree with your husband because he will be angry with you and how often does this happen?
2. Would a husband be justified in beating his wife if she was disrespectful to his parents or other senior members of his family?
3. …if she neglected household chores?
4. …if she was disobedient or did not follow his orders?
5. …if she was a drunkard or drug addict?
6. …if she beat the children frequently?
7. Does your husband ever hit or beat you, and does this happen fairly regularly or only rarely?

Family Decisions

8. Who in your family decides the following, and who…has the greatest say in this decision: What food to prepare for family meals
9. …how many children to have.
10. …inviting guests to your home
11. …whether to punish children for misbehaving
12. …what to do if a child falls sick
13. …how much schooling to give your children
14. …what kind of school to send the children to
15. …to whom to marry your children
16. Do you feel uncomfortable speaking/giving an opinion in the presence of the following people: husband?
17. …Father-in-law?
18. …Mother-in-law?
19. …Elder brothers-in-law?
20. …Elder sisters-in-law?
21. …Outside men?
22. Have you ever discussed how many children to have?
23. Have you ever discussed whether to use birth control?

Table 1, cont'd

Community Involvement

24. Do you have to ask your husband or senior family member for permission to go to: any place outside your compound
25. …the local market
26. …the local health center
27. …fields outside the village
28. …a community center in the village
29. …home of relatives or friends in the village
30. …a nearby fair
31. …a nearby shrine
32. …the next village
33. …In the past week have you read a newspaper
34. …listened to the radio
35. …watched television
36. …watched a movie
37. …did you go to the cinema
38. Do you and your husband ever talk alone with each other about what is happening in the community?

Household Economic Decisions

39. Who in your family decides the following, and who…has the greatest say in this decision? What food to buy for family meals
40. …whether to purchase major goods for the household such as a TV
41. …whether or not you should work outside the home
42. …whether to purchase or sell animals
43. …what gifts to give when relatives marry
44. …whether to purchase or sell gold/silver jewelry
45. Do you and your husband ever talk alone with each other about what to spend money on?
46. If your husband were unable to support you, would you be able to support yourself and your children (combined with: Is there someone else you could rely on for support)?
47. When you earn money, do you usually give all of it to your husband…?
48. Regardless of who keeps the money you earn, do you usually have a say in how your earnings are used…?
49. Do you have say in how the household's overall income is spent?
50. Do you get any cash in hand to spend on household expenditures?

51. If you wanted to buy yourself a dress, would you feel free to do it without consulting your husband?
52. If you wanted to buy yourself a small item of jewelry, such as a pair of earrings or bangle, would you feel free to do it?
53. If you wanted to buy a small gift for your parents or other family members, would you feel free to do it?
54. Do you personally own any other property…(combined with: could you use [it] as you wish without…permission)?

Table 2. Results of Unidimensional Factor Analyses

Model	χ^2(df)	RMSEA	IFI	Notes[a]
VIOLENCE				
One-factor model	361.49(14)***	.12	.99	item reliabilities low for v1 and v7
Wording effect added	179.09(13)***	.08	.99	item reliabilities higher, loadings small for v2-v6
Two-factor model	123.52(14)***	.07	1.0	Factor correlation=.18
FAMILY DECISIONS				
One-factor model	2574.10(104)***	.11	.78	item loadings low for v22, v23; MI suggest error correlations for (v13,v14)
Dropped v22 & v23; + error corr.	1353.17(76)***	.10	.87	MI suggest error correlations for v16-v21
Wording effect added	900.99(75)***	.08	.92	v9 loads poorly. MI suggest error correlation for (v8,v10)
Dropped v9; added error correlation	767.62(62)***	.08	.93	two factor model attempted but not better
COMMUNITY INVOLVEMENT				
One-factor model	1210.09(90)***	.08	.89	MI suggest error correlations for items v33-v37
Wording effect added	1008.24(89)***	.08	.91	MI suggest error correlations for (v35,v36)
Error correlation added	590.50(88)***	.06	.95	v24 and v38 load poorly
v24 and v38 dropped	338.35(63)***	.05	.97	two factor model attempted but not better
HOUSEHOLD ECONOMIC DECISIONS				
One-factor model	2073.31(104)***	.10	.98	MI suggest error correlation for (v47,v48)
Added error correlation	1645.31(103)***	.09	.99	MI suggest correlations for v39-v44
Added wording effect	1353.38(102)***	.08	.99	v54 continues to load poorly
Dropped v54	1209.86(88)***	.08	.99	two factor model attempted but not better

[a] MI stands for "Modification Indices." See Table 1 for variable names.

Table 3. Comparison of Confirmatory Factor Analysis Results to Summed Scales

Model	CFA χ^2(d.f.)	Summed Scale χ^2(d.f.)	Difference χ^2(d.f.)
Violence	123.52(14)***	9933.96(25)***	9810.44(11)***
Household Decisions	767.62(62)***	4562.47(90)***	3794.85(28)***
Community Involvement	338.35(63)***	3521.63(90)***	3183.28(27)***
Household Economics	1209.86(88)***	5929.58(119)***	4719.72(31)***

*** p<.001

Table 4. Correlations Between Dimensions of Autonomy

	Violence1	Violence2	Family	Community	Economics
Violence1	.63***				
Violence2	.16***	.87***			
Family	.49***	.07***	.52		
Community	.35***	0	.67***	.95	
Economics	.47***	0	.98***	.66***	.60

note: Variances on diagonal; correlations off diagonal; model fit: chi square=14,886.48(1064)***; RMSEA=.08; IFI=.98
*** p<.001

Table 4. Correlations Between Dimensions of Autonomy

	Violence1	Violence2	Family	Community	Economics
Violence1	.63***				
Violence2	.16***	.87***			
Family	.49***	.07***	.52		
Community	.35***	0	.67***	.95	
Economics	.47***	0	.98***	.66***	.60

note: Variances on diagonal; correlations off diagonal; model fit: chi square=14,886.48(1064)***; RMSEA=.08; IFI=.98
*** p<.001