

Refining the Measurement of Women's Autonomy: An International Application of a Multi-dimensional Construct

Rina Agarwala, *Princeton University*

Scott M. Lynch, *Princeton University*

Abstract

Women's autonomy has long been a central concern for researchers examining the social position of women in developing countries. However, little emphasis has been placed on the measurement of autonomy, despite its importance for assessing the validity of comparative research. In this research, we use confirmatory factor analyses to determine (1) whether items thought to measure autonomy in fact form a reliable measure of autonomy, (2) whether the relationship between multiple dimensions of autonomy are strong enough to justify a discussion of autonomy as a single underlying construct, and (3) whether comparative research on autonomy is possible between two countries (India and Pakistan). We find that our indicators capture four distinct dimensions of autonomy that are moderately related, and that, while the model structures replicate fairly well across the two countries we study, there are measurement differences that make comparative research challenging.

The concept of women's autonomy has been an important one in sociology and social demography for more than two decades (e.g., Connell 1987; Cubbins 1991; Ferree and Hall 1996; Kane and Sanchez 1994; Mason 1986). Early literature defined autonomy as "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) More recently, the definition of autonomy has been broadened to include "the ability to influence and control one's personal environment" (Safilios-Rothschild 1982) or "the capacity to obtain information and make decisions about one's private concerns and those of one's intimates." (Dyson and Moore 1983). These definitions assert a single construct that captures the multifaceted ability to gain control over the circumstances of one's life. Among women, attaining such control is viewed as a key to improving their living conditions. An in-depth examination of this construct, therefore, can bolster our efforts to alleviate gender inequalities.

In this paper, we investigate the empirical utility of the "autonomy" construct by examining (1) whether items thought – and commonly used – to measure several aspects of autonomy in fact form a reliable, unified scale and (2) assuming they do, whether the scale can be used successfully in comparative research. We use the term autonomy to reflect "the extent to which women exert control over their own lives within the families in which they live, at a given point in time." (Jejeebhoy 2000: 205) In other words, we define autonomy in its static, individual form, rather than as a group process. As Jejeebhoy (2000) argues, various terms, including "status," "autonomy" and "empowerment," have been used over time to capture some element of gender equality in the household and community. "Status" was often

We wish to thank Alaka Basu and five anonymous readers for their extremely valuable comments on earlier drafts. Direct correspondence to Rina Agarwala, Department of Sociology and Office of Population Research, 284 Wallace Hall, Princeton University, Princeton, NJ 08544. E-mail: agarwala@princeton.edu.

confused with “prestige” or “esteem” in the eyes of men, while “empowerment” referred to a more dynamic process of challenging existing power relations and gaining greater control over sources of power. Unlike autonomy, the term “empowerment” emphasizes the collective aspect of power, in addition to the individual one.¹ While the collective aspect of power is no doubt paramount to gender equality, especially in certain cultural contexts, it is the individual aspect of power (captured through the term “autonomy”) that appears more frequently in the quantitative literature on gender and power. This paper, therefore, fits within the context of the existing literature on autonomy.

Background

Autonomy has been equated with mechanisms used to alter rigid gender stratification, a central concept in sociological discourse (Connell 1987; Ferree and Hall 1996). Throughout the 1970s, modernists attempted to narrow sexual inequalities by increasing women’s choices in education, employment and reproduction. In response to persisting gender inequalities, despite such attempts to assimilate women into modern life, a new generation of scholars pointed to the vital role that women’s power or autonomy plays in accessing and maximizing choices in the first place (Blumberg 1984; Fernandez -Kelly 1994; Nussbaum 2000).

In the mid-1980s, scholars began to analyze the implications of autonomy through in-depth empirical research in developing countries. Employing quantitative analyses of field-level data, they showed that increased female autonomy was correlated with reduced fertility, improved child nutrition and education, and improved standards of living for women (Balk 1994; Basu 1992; Basu and Basu 1991; Dyson and Moore 1983). Based on such findings, the United Nations asserted in 1995 that women’s autonomy is essential to human dignity and must be considered a basic human right (UNDP 1995). Since then, several new studies have further examined the important determinants and effects of women’s autonomy in different contexts. (Balk 1997; Hashemi, Schuler and Riley 1996; Hehui 1995; Jejeebhoy and Sathar 2001; Kritz and Makinwa-Adebusoye 1999; Morgan and Niraula 1995).

To date, however, there has been little evidence to show if and how these existing findings can be compared across the different contexts. Moreover, few have seriously considered how autonomy should be measured, within and across context. This is surprising given both the wealth of literature underscoring why autonomy is an important concept to study and the rich data available to analyze women’s autonomy in different countries. Precise measures of theoretical constructs, such as autonomy, can help provide empirical evidence for their existence and are essential to ensuring robust work on the roots and consequences of the constructs. In turn, such work can improve our understanding of the dynamics of gender stratification as well as our policy prescriptions to narrow gender gaps throughout the world.

Indirect Proxy Measures of Autonomy

Since Mason (1986) first highlighted the conceptual and methodological issues that arose in the early studies on women’s autonomy, a consensus on how we *cannot* measure autonomy has emerged. Autonomy cannot be measured, as it often was, using a single observable characteristic, such as women’s education or labor force participation rates, as an approximate indicator of autonomy. These one-dimensional proxies, or indirect measures, are highly imperfect and have grave policy implications, especially when used to analyze the predictors and effects of autonomy (Balk 1994; Jejeebhoy 1991; Vlassoff 1994).

First, proxies for autonomy are extremely context dependent, which makes comparative research on autonomy difficult. For example, while women's education may be highly correlated with women's increased autonomy in one setting, it may be completely uncorrelated in another.² Second, proxy measures do not provide ample evidence for how well they capture the construct of autonomy. Without an understanding of how various proxies are correlated with one another and to autonomy, it is difficult to compare the results from studies using different proxies. Third, proxies blur the channels through which autonomy works. For example, when examining the impact of autonomy on fertility using education as a proxy for autonomy, we cannot be certain whether a decrease in fertility from increased education is due to the direct effects of education, or to education's effect on autonomy, which in turn affects fertility. Finally, proxies obscure which dimension of autonomy is being measured (Whyte 1978). Scholars have long argued that autonomy is comprised of multiple dimensions; each dimension is determined by, and predicts, different demographic and socioeconomic factors.

Direct Measures of Autonomy

Recently, scholars have turned from using indirect proxies to quantify autonomy, choosing instead 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 1995).

Studies using direct measures have addressed many of the inadequacies of the earlier indirect-measure approach. They have explicitly quantified the multi-dimensionality of autonomy, thereby clarifying the roots and consequences of each dimension. In addition, direct-measures have illuminated the channels through which economic and social factors (such as education and labor force participation) affect autonomy, rather than confounding the causes and effects of autonomy (Goetz and Sen Gupta 1996; Kritz and Makinwa-Adebusoye 1999; Mason 1997).

Some of the inadequacies of the earlier literature, however, remain unaddressed in the literature using direct measures. First, few have addressed the issue of context dependency; elements that may genuinely reflect autonomy in one context may be irrelevant or impossible in another (Davis and Robinson 1991; Entwisle, Henderson, Short, Bouma and Fengying 1995; Kenworthy and Malami 1999; Lobao and Brown 1998; Orloff 1993). A failure to address this issue has undermined the scope of comparative research on autonomy.

Second, recent research has not shown how reliably each direct measure, or observed item, reflects a particular dimension of autonomy. Most studies construct a dimension of autonomy as a simple summative index of a series of dichotomous, observable items. For example, the mobility dimension 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; Morgan and Niraula 1995). Few studies test the inter-item reliability of each index by measuring the correlations between the items within each dimension. In some cases, internal consistency of a dimension has been tested, but the measurement error within each item has not been considered (Balk

1994; Mason 1986). Of all research on autonomy, Mason's works are perhaps the most significant in paying considerable attention to the measurement of autonomy. For example, in reference to the domestic violence dimension used in one of her papers, Mason writes, "This scale [index] is relatively weak and is used here primarily for convenience." (Mason 1997: 3)

Finally, while direct measures have highlighted the important multidimensionality of autonomy, some of the literature has undermined the justification for conceptualizing autonomy as a single, empirical concept by analyzing the predictors or effects of each dimension of autonomy separately (Morgan and Niraula 1995; Vlassoff 1992). In such cases, scholars are effectively assuming that the correlations between the dimensions are zero, thereby weakening the power of their own conclusions regarding the net causes or impact of autonomy in general. Analyzing each dimension separately provides robust evidence only for the impact of each individual dimension of autonomy.

Among the studies that do analyze the strength of the relationship between the dimensions of autonomy, the relationships appear to be low to moderate. This raises important questions as to why we should study autonomy as a single measure at all. Balk (1994) interprets the low correlation coefficients between the dimensions she studied as evidence for the "successful measurement of largely distinct dimensions of women's status." (Balk 1994: 43) Other scholars, however, interpret the weak and inconsistent associations between some dimensions as evidence that those dimensions reflect aspects of autonomy that differ from those reflected by other dimensions that are more positively related (Jejeebhoy 2000; Mason and Smith 1999).

Alternatively, the weak correlations may be due to measurement methods. None of these studies accounted for measurement error within the observed items used to measure each dimension of autonomy. Nor did 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 (2000) and (2001) summed all six dimensions into a single index for autonomy, which forces all the dimensions to have equal weights and a correlation of one. Consideration of measurement error and differential weights could possibly produce higher correlations between the dimensions and provide evidence for considering autonomy as a more unified construct. Alternatively, it may shed greater light on why certain 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.

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); we analyze 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 that has tried to operationalize the multiple dimensions of autonomy.

The India survey includes 1,842 rural Indian women, covering 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.³ It also draws evenly between Hindus and Muslims and over samples castes that are numerically small. In order to ensure this ethnic mix, 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, ages 15-39, were randomly selected for interviews 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 of which affect 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), 10 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 10 sites.

It should be noted that the samples are not representative of the entire population of India or Pakistan, and our results cannot be generalized to the national level. However, the samples taken in India and Pakistan are relatively similar to one another and are representative of a population that the women's autonomy literature is concerned with, namely rural women in developing countries.

Methods

We use a 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. In this research, we begin with a confirmatory factor analytic approach in analyzing the Indian data, but then refine the measurement models based on substantive considerations and modification indices. We then replicate the analyses with the Pakistani data following a strictly confirmatory factor analytic approach. This approach allows us to assess whether, given a fixed "best" model for the Indian data, the model fits the Pakistani data. The choice of country on which we based our initial models was arbitrary. In the case that the Pakistani data do not fit the Indian model, starting with the Pakistani data would simply lead to a different best model. This fact should be kept in mind: We do not profess that our scale for autonomy is the right one – rather, we propose that it is right for the rural Indian context, and our goal is to determine whether this model is transferable to another culture.

We began with a set of 54 items thought to reflect autonomy. These items were chosen based on two criteria: (1) they were measured in our data, and (2) existing research has used them to represent autonomy. Because this paper aims to test existing measures of autonomy, we do not assert or test entirely new observable items that may represent autonomy. Therefore, we do not include items that meet the first criteria but not the second. Unfortunately, due to data limitations, we were also unable to include items that meet the second criteria and not the first.⁴ Based on the 54 selected items, we then constructed the following four dimensions of autonomy to test on our Indian data: autonomy from violence (7 items); autonomy in family decisions (16 items); autonomy in community involvement (15 items); and 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, according to the existing literature.

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

Table 1: 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?
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**
 Who in your family decides the following, and who...has the greatest say in this decision? What food to buy for family meals?
39. ...whether to purchase major goods for the household such as a TV?
 40. ...whether or not you should work outside the home?
 41. ...whether to purchase or sell animals?
 42. ...what gifts to give when relatives marry?
 43. ...whether to purchase or sell gold/silver jewelry?
 44. Do you and your husband ever talk alone with each other about what to spend money on?
 45. 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)?
 46. When you earn money, do you usually give all of it to your husband...?
 47. Regardless of who keeps the money you earn, do you usually have a say in how your earnings are used...?
 48. Do you have say in how the household's overall income is spent?
 49. Do you get any cash in hand to spend on household expenditures?
 50. If you wanted to buy yourself a dress, would you feel free to do it without consulting your husband?
 51. 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?
 52. If you wanted to buy a small gift for your parents or other family members, would you feel free to do it?
 53. Do you personally own any other property... (combined with: could you use [it] as you wish without...permission)?
 54.
-

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, in addition to 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 1986). 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 using the Indian data. These models were then revised based on inspection of the parameter estimates and modification indices. 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, 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) 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 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 Indian analyses using data from Pakistan in a multiple group analysis. Multiple group analysis allows the simultaneous estimation of models for two or more groups with the imposition of increasing numbers of constraints. In brief, if all measurement model parameters can be constrained to be equal across groups, then the groups can be said to be equivalent with respect to measurement (Bollen 1989). In this setting, such a finding would provide evidence that the measurement of autonomy may be independent of context.

We choose to use data from Pakistan to replicate our model for one primary reason – the existing evidence for the highly context-dependent nature of autonomy. A comparison across two relatively similar cultures (India and Pakistan) provides leverage with which to determine whether our autonomy measures are replicable independent of confounding factors that could arise from substantial differences in context. This is a conservative test of the viability of comparative research on autonomy and can only provide a first step in a longer analysis on the subject.

Results

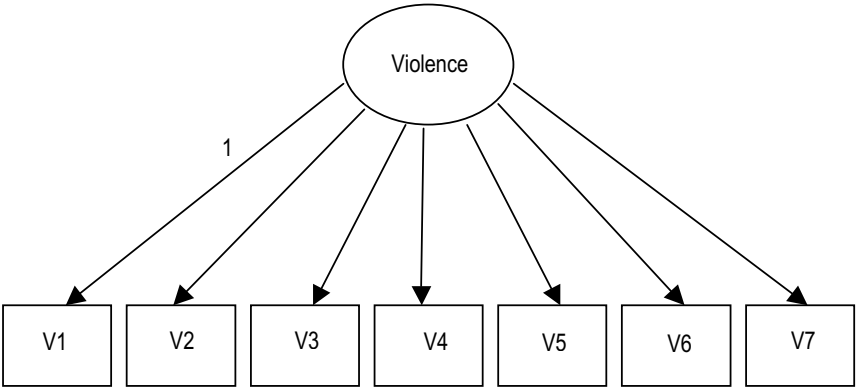
One-Dimensional Measurement Models

Violence

Figure 1 shows the initial model estimated for the dimension of autonomy from violence.⁵ As the figure indicates, all seven 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. Alternatively, there may be a substantively meaningful difference between experiences of violence and the perceived legitimacy of violence.

Figure 1a. Initial Measurement Model for Autonomy from 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 and 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 that reflects perceptions of the legitimacy of violence in the household and the other that reflects feared or actual violence. Figure 2 shows the structure of the final model.

Figure 1b. Final Measurement Model for Autonomy from Violence

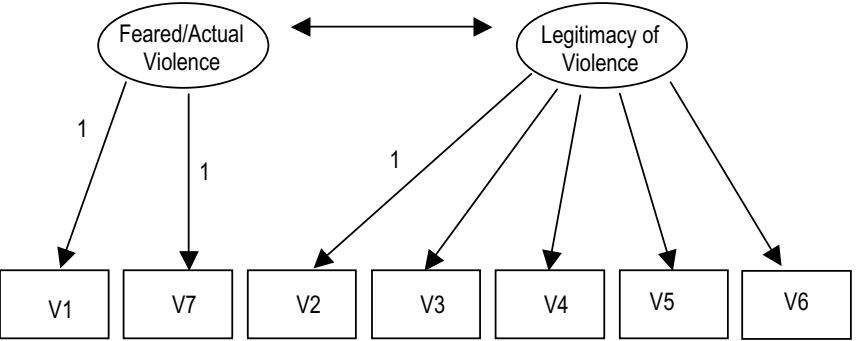


Table 2: Results of One-dimensional Factor Analyses

<i>Model</i>	$\chi^2(df)$	<i>RMSEA/IFI</i>	<i>Notes^a</i>
VIOLENCE			
One-factor model	361.49(14)***	.12	.99 item reliabilities low for v1 and v7 item reliabilities higher, loadings small for v2-v6 Factor correlation = .18
Wording effect added	179.09(13)***	.08	.99
Two-factor model	123.52(14)***	.07	1.00
FAMILY DECISIONS			
One-factor model (v13, v14)	2574.10(104)***	.11	.78 v22, v23 load poorly; MI suggest C(v13, v14)
Dropped v22 & v23; + error corr.	1353.17(76)***	.10	.87 MI suggest C(v16-v21)
Wording effect added	900.99(75)***	.08	.92 v9 loads poorly. MI suggest error (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 C(v33-v37)
Wording effect added	1008.24(89)***	.08	.91 MI suggest C(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

^aMI stands for "Modification Indices," C(.) stands for error correlations between (.). See Table 1 for variable names.

Family Decisions

As with violence, autonomy in making (non-economic) family 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, 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 3 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 (for items v33-v37). These items ask whether the respondent has engaged in several activities during 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 a 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 4 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 5 shows the final model for autonomy in household economics.

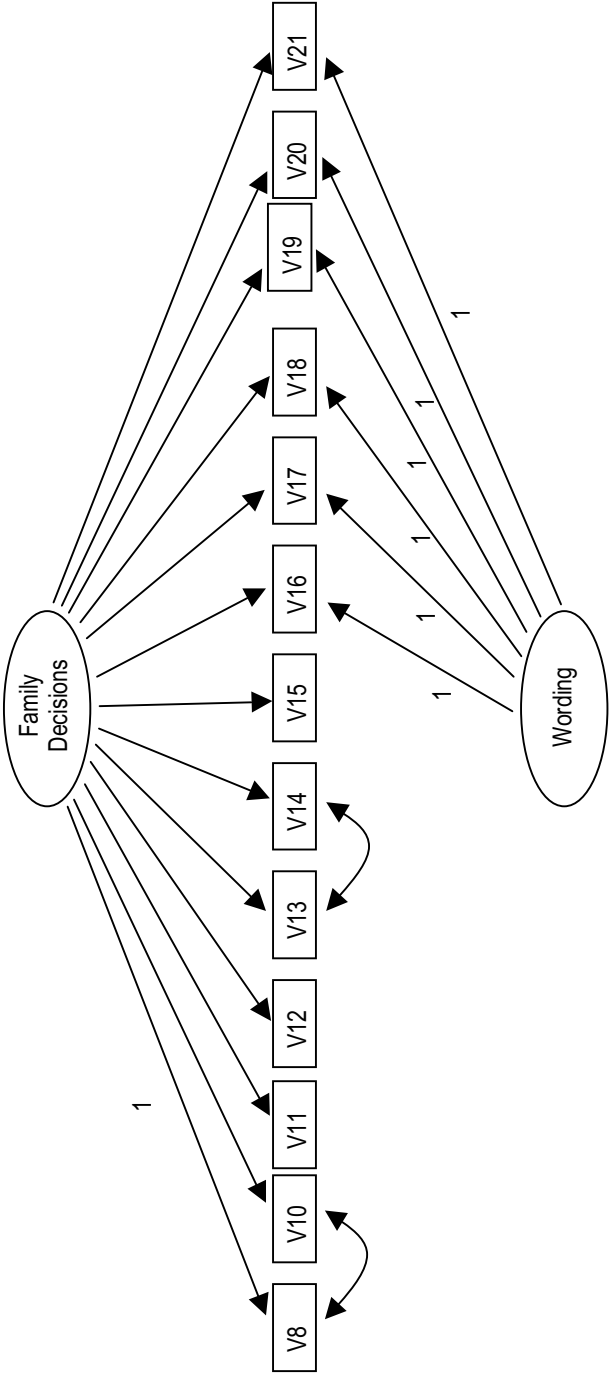


Figure 2. Final Measurement Model for Autonomy in Family Decisions

Figure 3. Final Measurement Model for Autonomy in Community Involvement

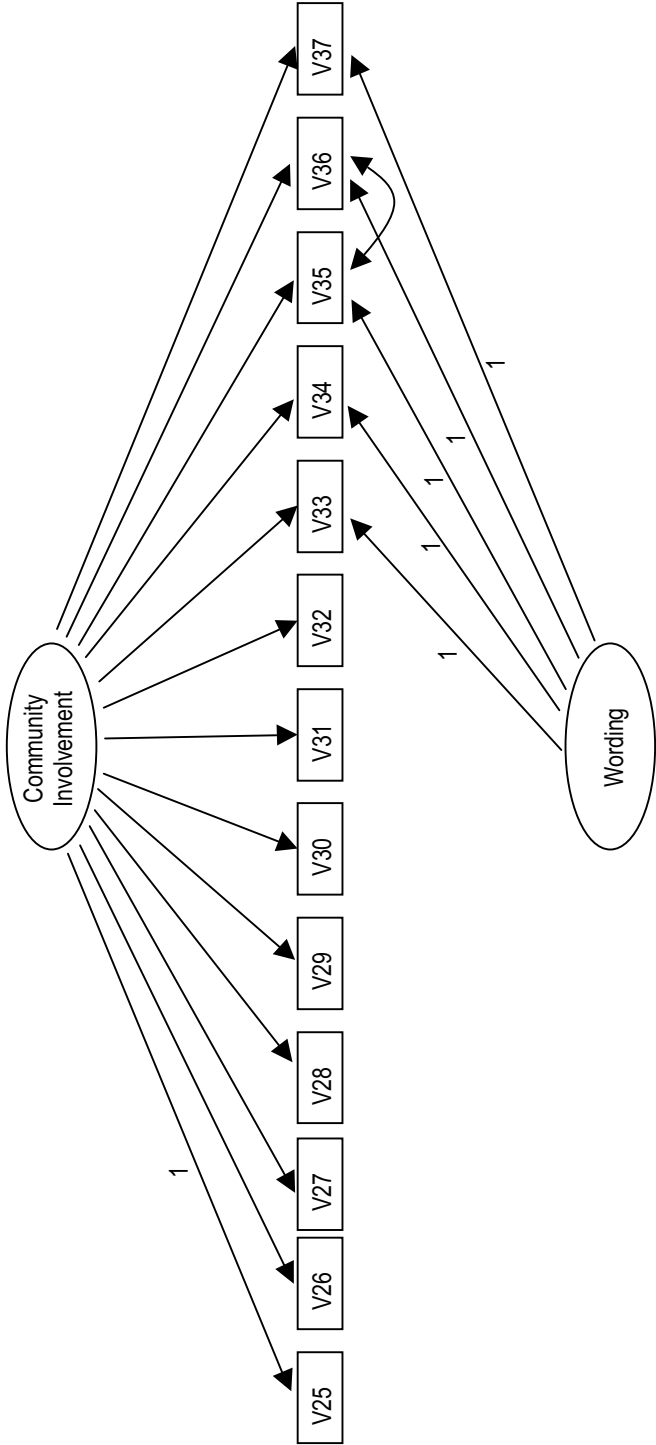
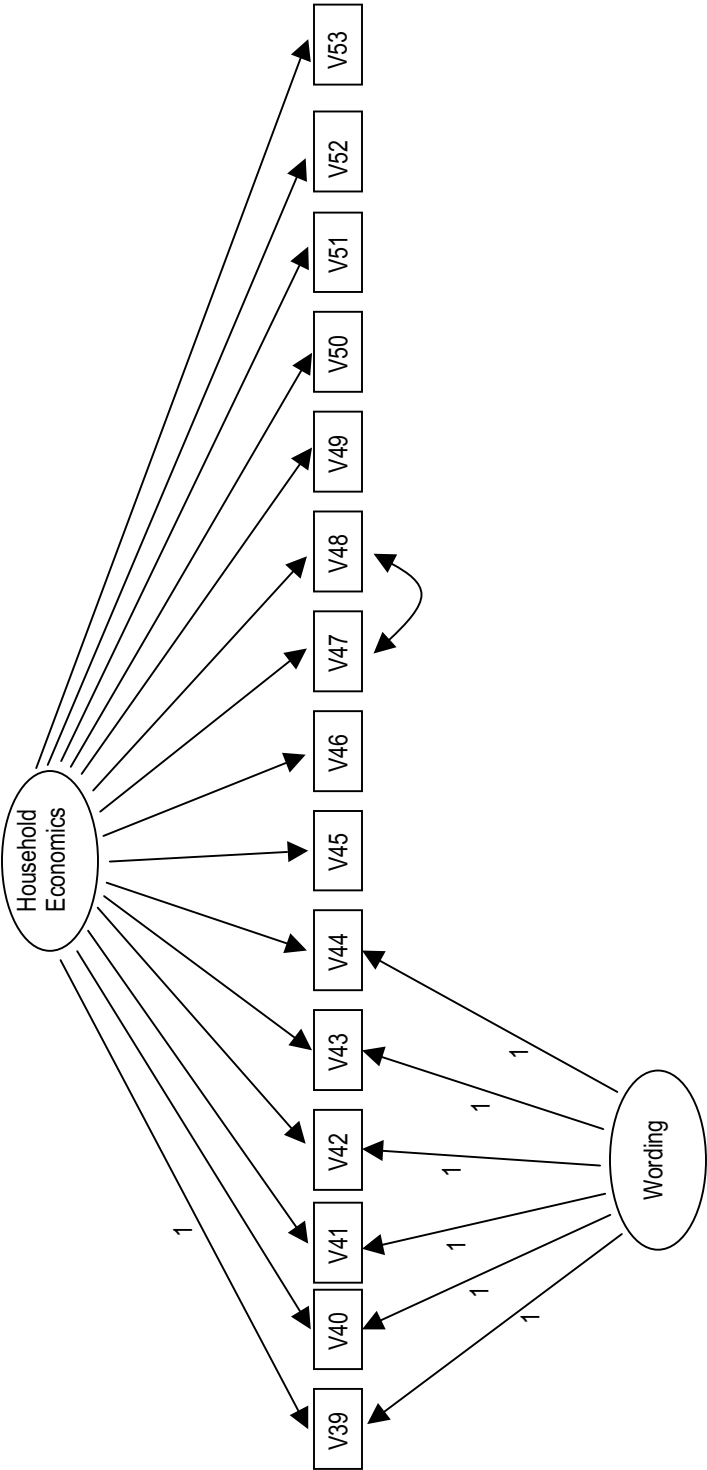


Figure 4. Final Measurement Model for Autonomy in Household Economic Decisions



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.

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

<i>Model</i>	<i>CFA χ^2(d.f.)</i>	<i>Summed Scale χ^2(d.f.)</i>	<i>Difference χ^2(d.f.)</i>
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)***

*** $p < .001$

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 the legitimacy 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 correlation 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.

Table 4: Correlations Between Dimensions of Autonomy

	<i>Violence 1</i>	<i>Violence 2</i>	<i>Family</i>	<i>Community</i>	<i>Economics</i>
Violence 1	.63***				
Violence 2	.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

*** $p < .001$

India-Pakistan Comparison

Finally, we replicated the results for the Indian data with virtually identical data from Pakistan. Several studies have argued for the important role that context plays in autonomy (Dyson and Moore 1983; Jejeebhoy 2000; Jejeebhoy and Sathar 2001; Mason 1986; Mason and Smith 1999). As stated earlier, a comparison across two relatively similar cultures (India and Pakistan) provides leverage with which to determine whether our autonomy measures are replicable independent of confounding factors due to major differences in context. In order to do the replication, we conducted two sets of analyses. First, we re-estimated the models using the Pakistani data alone. Second, we estimated a multiple group model that allowed 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 Pakistani 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 Pakistani data only. The results for the model for violence using the Pakistani data are comparable to the results using the data on India. The RMSEA is below .1, and the IFI is .99. The results for the model for family decisions are, in fact, better for the Pakistani data than the Indian 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 Pakistani 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 indicates comparable fit to the Pakistani data. In that model, the RMSEA was .058 (slightly better than the .08 for the Indian data), and the IFI was .98 (very slightly worse than the .99 for the Indian 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 the 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, is quite good.

Discussion and Conclusion

In this research, we have examined in detail the measurement of women's autonomy, an important theoretical construct in sociological and demographic literature on gender and development. In doing so, we hope to contribute to a deeper discussion on how best to model autonomy in future empirical research and ultimately improve our efforts to alleviate gender inequalities. The results of the analysis provide several interesting insights that can help further our understanding of the measurement of women's autonomy in developing countries. In particular, the results provide evidence to support and question some of the models and claims in the autonomy literature to date.

Table 5: Results of Replication of Factor Analyses of Indian Data with Data from Pakistan

<i>Dimension of Autonomy</i>	<i>Pakistani Data</i>	<i>Multiple Group Analyses (Pakistani and Indian Data)</i>		
		Unconstrained	Constrained	Variances Free
Violence	χ^2			
		200.41(28)***	227.76(42)***	217.27(32)***
		.065	.055	.063
Family Decisions	RMSEA	.99	.99	.99
	IFI			
Community Involvement	χ^2			
		957.92(124)***	1297.43(153)***	1244.04(136)***
		.068	.072	.075
Household Economics	RMSEA	.95	.93	.93
	IFI			
Violence	χ^2			
		1081.51(126)***	1485.34(154)***	1250.00(138)***
		.073	.078	.075
Family Decisions	RMSEA	.93	.90	.91
	IFI			
Community Involvement	χ^2			
		1603.56(176)***	1774.49(208)***	1712.97(190)***
		.075	.072	.075
Household Economics	RMSEA	.99	.99	.99
	IFI			

*** p < .001

First, we measured the robustness of four dimensions of autonomy: freedom from violence, participation in non-economic family decisions, community involvement, and participation in household economic decisions. Our measurement models for each dimension fit the data well, and the separate autonomy latent variables were intercorrelated but not perfectly so. These results lend concrete support to the existing literature arguing for the multidimensionality of autonomy (Balk 1994; Jejeebhoy and Sathar 2001; Mason 1986) by showing that autonomy items do indeed cluster into distinct and meaningful dimensions. Studies on autonomy should, therefore, continue to measure autonomy in its multidimensional form. With regard to violence, we found this aspect of autonomy should be further divided into two sub-dimensions: "feared and actual violence" and "views on the legitimacy of violence." This finding suggests that in India and Pakistan, a woman's views on the legitimacy of violence do not reflect the same aspect of autonomy as do her experiences of actual violence. Examining these two aspects separately may, therefore, yield more precise results in terms of the causes and effects of domestic violence against women. For example, a women's group may raise women's understanding that domestic violence is unacceptable, but it may not provide women with the social or physical resources needed to prevent it.

Second, we tested each dimension of autonomy against the most common method of measuring autonomy in the literature to date: summed scales (Hashemi, Schuler and Riley 1996; Jejeebhoy 2000; Morgan and Niraula 1995). We found that because summed scales do not account for measurement error and differential weighting of the items included in the scales, they provide a less appropriate measure for autonomy. Instead, autonomy research might benefit from using models that provide more robust measures of autonomy that allow for differential weight on observable items within a dimension and account for measurement error on each item. It is clear that measuring a theoretical construct through a series of observable items bears a host of potential complications due to measurement error. Nevertheless, such work is important to pursue. Accounting for the measurement error, as well as the differential weights on the observable items, is a relatively low cost way of strengthening the important research on women's autonomy. Given the rich data available on autonomy, future research could examine how the causes and effects of autonomy using this new measure of autonomy differ from studies using simple summed measures.

Third, we combined all four dimensions of autonomy into a single model to determine the extent to which the dimensions of autonomy were related. Here, again, our results lend concrete support for the argument that the dimensions of autonomy are, for the most part, moderately related (Balk 1994). This finding indicates that the various dimensions of autonomy can indeed be considered part of a single underlying construct, but also that they have distinct contributions to autonomy. Future research on women's autonomy should continue to study autonomy as a single, yet multidimensional, concept. Because each dimension can variously affect and be affected by different factors, interpretations on the causes and effects of autonomy as a whole would be facilitated by analyses that do not examine each dimension separately, but rather combine all dimensions into a single model. This single model, in turn, should not merely sum all autonomy indicators to produce a single measure of women's autonomy, but rather allow for the individual contributions of each dimension by accounting for differential weighting and measurement error.

An important exception to the dimensions being related, however, is the perceived legitimacy of violence sub-dimension. Because legitimacy 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. This finding presents important implications for future research on autonomy. Views on the legitimacy of violence may, in fact, reflect education or social norms in a community, which in turn can be a cause or effect of autonomy, rather than part of autonomy itself. In other words, our findings indicate that legitimacy of violence may

be measuring concepts that are unrelated to our understanding of autonomy as an indicator of control over one's life. Studies have often combined this dimension into the single construct of autonomy. Future research should examine the issue of violence in greater depth to better understand its relationship with other dimensions of autonomy. One way to do so could be to include alternative observable items on violence in future questionnaires. For example, answers to questions on one's capacity to respond to or prevent violence might be better related to other dimensions of autonomy than are the answers to questions on actual violence or perceptions of the legitimacy of violence.

We also found an important exception to the finding that while the dimensions (aside from the perceived legitimacy of violence sub-dimension) are part of a single underlying construct, each dimension has a distinct contribution to autonomy. The relationship between two dimensions, family decisions and household economic decisions, was very close to 1, suggesting that these two dimensions may be collapsed into a single latent construct or dimension and examined simultaneously. Again, much of the current literature examines these dimensions separately, thereby implying that they reflect different aspects of autonomy. This research, however, suggests that they reflect the same aspect. This finding implies that the dynamics involved in empowering women in household decision-making simultaneously affect a wide range of issues from reproduction and child education to family employment allocation. On a practical level, examining family decisions and household economic decisions as a single dimension can facilitate future empirical research on autonomy. More importantly, however, on a policy level, this finding indicates that efforts to empower women in less controversial household decisions, such as nursing a sick child or punishing a misbehaving child, may have positive effects on empowering women in more controversial household decisions, such as family planning or women's control over money. Future research should study the connections between women's autonomy in these various household decisions.

Finally, we replicated the Indian 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 Pakistani data only). On the other hand, the results are less clear in terms of whether the measurement of autonomy is comparable in terms of factor loadings and variances. These results are significant in that they suggest, contrary to our initial expectations, that even relatively small changes in context can affect the weights used in the measurement of autonomy and thus the concept of autonomy itself.

Our findings raise questions on the robustness of comparative research on autonomy (Jejeebhoy and Sathar 2001). At the least, they show that comparative research on autonomy must allow for differences in measurement errors and variances of latent variables measuring autonomy – no matter how small is the change in context. Again, this finding suggests that summed scales should not be used in studying autonomy, especially cross-culturally. In the current globalizing economy, where women are increasingly crossing borders for work, it is important to understand women's relative power across various contexts, and it is essential that measures of autonomy remain flexible enough to accommodate contextual changes. At the same time, future research may also test to what extent the model structure, or dimensionality of autonomy, changes with larger differences in context. Having different models with which to measure autonomy can make comparing autonomy levels across contexts difficult. Therefore, new constructs that can better accommodate contextual changes may need to be developed to measure the extent of power women have over their lives. In the long run, measurement concerns across context, particularly among sub-populations such as immigrants, may become less relevant as cultural practices become independent of geographic location.

While this study has provided a systematic study of the measurement of women's autonomy, it is not without limitations. First, the items used in measuring autonomy in this

research are specific to this survey. 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 needs to be paid to how autonomy is measured in future research. Such attention can help us continuously improve our survey questions so we can eventually capture the best observable indicators of autonomy. In addition, the findings in this study are subject to the samples in the data used.

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. Part of this test can be done using the remaining three countries in the SWAF data set. Beyond this, however, data limitations to date make such a test difficult. Specifically, autonomy is often measured with different indicators not only in different surveys, but even within the same survey applied in different countries. In addition, samples and sampling techniques must be comparable across data sets of different countries. This limitation is thus not so much a shortcoming of this research specifically, but a shortcoming of any comparative work on women's autonomy.

Notes

1. For more on this discussion, see Batliwala 1994; Dyson and Moore 1983; Jejeebhoy 2000.
2. For differing views on the relationship between education and autonomy. See Oropesa 1997; Sathar, Crook, Callum and Kazi 1988; Sathar and Kazi 1990.
3. Development indicators were based on income, percentage of roads surfaced and other economic criteria.
4. Due to the quality and depth of the survey, however, there were very few items in this category. Most of the items in this category dealt with political participation.
5. "Autonomy from violence" is the ability to avoid domestic violence against oneself. The observable items used in this model are those used in much of the existing literature and are thus available in the data set used for this study.

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 Bangladesh: A Social Demographic Analysis." *Population Studies* 51:153-72.
- Basu, Alaka Malwade. 1992. *Culture, the Status of Women, and Demographic Behaviour: Illustrated with the Case of India*. Clarendon Press & Oxford University Press.
- Basu, Alaka Malwade, and Kaushik Basu. 1991. "Women's Economic Roles and Child Survival: The Case of India." *Health Transition Review* 1:83-103.
- Batliwala, Srilatha. 1994. "The Meaning of Women's Empowerment: New Concepts from Action." in *Population Policies Reconsidered*. G. Sen, A. Germain and L. Chen, editors. Harvard University Press.

- Blumberg, R.L. 1984. "A General Theory of Gender Stratification." *Sociological Theory*. R. Collins, editor. Jossey-Bass.
- Blumberg, Rai Lesser. 1994. "Women's Work, Wealth, and Family Survival Strategy: The Impact of Guatemala's ALCOSA Agribusiness Project." *Women, the Family, and Policy*. E. Chow and C. Berheide, editors. State University of New York Press.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. Wiley.
- Connell, R.W. 1987. *Gender and Power*. Stanford University.
- Cubbins, Lisa A. 1991. "Women, Men, and the Division of Power: A Study of Gender Stratification in Kenya." *Social Forces* 69:1063-1083.
- Davis, Nancy J., and Robert V. Robinson. 1991. "Men's and Women's Consciousness of Gender Inequality: Austria, West Germany, Great Britain, and the United States." *American Sociological Review* 56:72-84.
- Dixon-Mueller, Ruth. 1978. *Rural Women at Work: Strategies for Development in South Asia*. Johns Hopkins Press.
- Dyson, Tim, and Mick Moore. 1983. "On Kinship Structure, Female Autonomy, and Demographic Behavior in India." *Population and Development Review* 9:35-60.
- Entwisle, Barbara, Gail E. Henderson, Susan E. Short, Jill Bouma and Zhai Fengying. 1995. "Gender and Family Business in Rural China." *American Sociological Review* 60:36-57.
- Fernandez -Kelly, Patricia M. 1994. "Broadening the Scope: Gender and the Study of International Development." Pp. 143-168. *Comparative National Development: Society and Economy in the New Global Order*. D. Kincaid and A. Portes, editors. University of North Carolina Press.
- Ferree, Myra Marx, and Elaine J. Hall. 1996. "Rethinking Stratification from a Feminist Perspective: Gender, Race, and Class in Mainstream Textbooks." *American Sociological Review* 61:929-50.
- Goetz, Anne Marie, and Rina Sen Gupta. 1996. "Who Takes the Credit? Gender, Power and Control over Loan Use in Rural Credit Programs in Bangladesh." *World Development* 24:45-63.
- Hashemi, Syed, Sidney Schuler and Ann Riley. 1996. "Rural Credit Programs and Women's Empowerment in Bangladesh." *World Development* 24:635-53.
- Hehui, Jin. 1995. "A Study of Rural Women's Decision-Making Power on Reproduction and Fertility." *Chinese Journal of Population Science* 7:241-57.
- Jejeebhoy, Shireen J. 1991. "Women's Status and Fertility: Successive Cross-sectional Evidence from Tamil Nadu, India 1970-1980." *Studies in Family Planning* 22:217-30.
- _____. 2000. "Women's Autonomy in Rural India: Its Dimensions, Determinants, and the Influence of Context." *Women's Empowerment and Demographic Processes: Moving Beyond Cairo*. H. Presser and G. Sen, editors. 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. 1986. *PRELIS: A Preprocessor for LISREL (2nd edition)*, Mooresville, Indiana: Scientific Software Inc.
- Kane, Emily W., and Laura Sanchez. 1994. "Family Status and Criticism of Gender Inequality at Home and at Work." *Social Forces* 72:1079-102.
- Kenworthy, Lane, and Melissa Malami. 1999. "Gender Inequality in Political Representation: A Worldwide Comparative Analysis." *Social Forces* 78:235-68.

- Kritz, Mary M., and Paulina Makinwa-Adebusoye. 1999. "Determinants of Women's Decision-Making Authority in Nigeria: The Ethnic Dimension." *Sociological Forum* 14:399-424.
- Lobao, Linda M., and Lawrence A. Brown. 1998. "Development Context, Regional Differences among Young Women, and Fertility: The Ecuadorean Amazon." *Social Forces* 76:819-49.
- 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." Honolulu, Hawaii: East-West Center.
- Mason, Karen, and Herbert Smith. 1999. "Female Autonomy and Fertility in Five Asian Countries." Paper presented at the Annual Meetings of the Population Association of America, New York, March 25-27.
- Morgan, S. Phillip, and Bhanu Niraula. 1995. "Gender Inequality and Fertility in Two Nepali Villages." *Population and Development Review* 21:541-61.
- Nussbaum, Martha. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge University Press.
- Orloff, Ann Shola. 1993. "Gender and the Social Rights of Citizenship: The Comparative Analysis of Gender Relations and Welfare States." *American Sociological Review* 58:303-328.
- Oropesa, R.S. 1997. "Development and Marital Power in Mexico." *Social Forces* 75:1291-1317.
- Safilios-Rothschild, Constantina. 1982. "Female Power, Autonomy and Demographic Change in the Third World." *Women's Role and Population Trends in the Third World*. R. Anker, M. Buvinic and N. Youssef, editors. Croom Helm.
- Sathar, Zeba, Nigel Crook, Christine Callum and Shahnaz Kazi. 1988. "Women's Status and Fertility Changes in Pakistan." *Population and Development Review* 14:415-432.
- Sathar, Zeba, and Shahnaz Kazi. 1990. "Women, Work and Reproduction in Karachi." *International Family Planning Perspectives* 16:66-69, 80.
- 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*. Oxford University Press.
- Vlassoff, Carol. 1994. "Gender Inequalities in Health in the Third World: Uncharted Ground." *Social Science and Medicine* 39:1249-59.
- _____. 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 Pre-industrial Societies*. Princeton University Press.