

Aldo Benini

## The Wealth of the Poor

*Simplifying living standards measurements with Rasch scales?*

### Summary

Evaluating poverty reduction requires repeated measures of the living standards of the poor. Traditionally, surveys collected data on household income or expenditure. These estimates require extensive data on each surveyed household. In recent years, measures derived from durable household goods and from access to services have been tried out as substitutes for income or expenditure-based ones.



**A monitoring officer of “RDRS Bangladesh” is interviewing a member of a neighborhood group affiliated with this NGO. The lady works with (and probably owns) a sewing machine, an item used in wealth indices assuming that its presence is reliably communicated and discriminates well between poorer and less poor households.**

However, while these wealth scales economize on data collection, their analysis has not usually been less complex. It calls for statistical expertise that local survey-implementing organizations often lack. A different technique, based on Rasch scales, has properties that make it is easy to apply and evaluate. It can make wealth measurement more accessible to local stakeholders.

The precondition is that the selection of household goods and services used in the scale must be pre-validated. I demonstrate the power of a simple 10-item scale with data from a large household survey in Bangladesh.

## Introduction

In household surveys concerned with poverty and its reduction, the measurement of living standards is a core component. Several international survey traditions have built up an impressive body of theory and methodology in this field. Best known are the Living Standards Measurement Surveys (LSMS) and the Demographic and Health Surveys (DHS). Their data collections from numerous countries are so rich and so similar in design that more and more comparative analyses have appeared. Most recently, Banerjee and Duflo (2007) have offered a very readable and instructive summary, *“The Economic Lives of the Poor”*.

These survey traditions have known steep learning curves also in their methodologies. One of the earlier debates was about income vs. consumption measures and their adequacy to capture chronic or transient poverty. Reliable measurement of either variable, however, was complex (Grosch and Glewwe 2000). In recent years, and more so after Filmer and Pritchett’s (2001) *“Estimating wealth effects without expenditure data”*, household wealth measures have come into favor as simple living standard proxies.

These wealth measures combine data on durable household goods and use of services. This data assumedly is easy, quick to collect and reliable because it necessitates few questions; and these elicit yes-no answers (“Do you have a radio?”) or a choice from among very few commonly known options (“What kind of toilet does your family use?”). The validity of such measures has been investigated with data from several countries (Rutstein and Johnson 2004).

The statistical treatment of this kind of data, however, is not straightforward. It is certainly beyond the analytical capacity of most local survey implementing organizations. Filmer and Pritchett and others have used ordinary principal component analysis. They have been criticized for this choice by Kolenikov and Angeles (2004). The method that these authors favor – polychoric principal components – is at the fingertips of even fewer analysts (they did make available an implementation in STATA 8).

The coexistence of easy data collection segments (as far as living standard data are concerned) and of difficult analysis to compute a final measure fits well with the typical highly stratified expertised system of development research and aid (see, with the example of the World Bank, Lera St. Clair 2004). This can be questioned in the context of rapid assessments and of participatory assessments. Is it possible to simplify further, both in data collection and in analysis, without trading off too much of the validity and precision that the wealth measure tradition has acquired?

Ideally, a wealth measure should be easy to assemble *and* analyze locally, such as by the monitoring and evaluation staff of development NGOs, by local research groups, or by citizen groups aided by them. Survey quality standards would require also that data on included items be reliable, and the compound measure be strongly correlated with already validated (possibly more complex) measures. The simplified measure should also be easy to calculate, perhaps with a help of a simple scoring rule.

## Strategies for a simpler index

A simpler wealth measure could be built following one or several of these strategies:

- Collecting fewer items
- Collecting items with fewer categories
- Simplifying the rules for combining the items into the final index

Each of these strategies has its downsides too. Reducing the number of questions asked about different wealth items may result in the loss of precision if items typically acquired within a certain stage of household asset growth are deleted, and in their lieu too many redundant ones retained. Some durable household goods come in distinct and (for living standard measurements) significant quality grades. For example, the difference between a sanitary latrine and an open latrine may be just as important as, or more important than, the difference between an open latrine and no latrine. Prematurely reducing the question (or the analytic use of the response) to a simple “Household has latrine – yes/no” alternative gives away valuable information. Inappropriate index formation may compromise the validity of the measure, by distorting the weights that different items should be assigned.

In order to avoid those pitfalls, simplified measures should be constructed from existing large living standards data sets and their household wealth measures already validated across several surveys. To this end, one may explore smaller sets of variables contained in the larger initial set and examine the loss of predictive power against the validating variable. I am doing this as an initial demonstration, using data from Bangladesh, hoping to later re-evaluate the simplified tool against other data, including household income data, collected elsewhere in the country.

In addition to reducing the number of items, two more simplifications can be used:

- Reducing the response categories to binary yes/no alternatives
- Using the count of all questions answered “Yes” as a preliminary household wealth index

The analytic vehicle for this is the Rasch scale, a statistical model that is much more popular in psychology and educational testing (Embretson and Hershberger 1998) than in poverty research (for two exceptions, however, see Winship and Jencks 2002: ; Fusco 2005). In simple words, the Rasch model follows the basic idea that when we ask a sufficiently large number of subjects a sufficiently large number of questions related to the same construct (e.g. mathematical intelligence, or purchasing power), then the response pattern will yield measures for the ability of the subjects as well as for the relative difficulty of each item.

This concept is intuitively applicable to the growth of household assets as poor people work their way out of poverty. For example, most, but not all, households will own one or several beds and be in a rainproof dwelling before they acquire a TV set. Of course, the additional difficulties in moving from step A to B and then from B to C may differ. In

a simple scenario, members of most households with annual incomes of USD 1,000 and more may no longer sleep on the floor, most of those making USD 2,000 and more may afford to rent homes with good roofs, but only those making USD 4,000 or more may have TV in their majority.

The Rasch scale presumes an underlying ability to produce the observed items. For example, the – observed or assumed – household income stream over several years is at the origin of the durable goods that the visiting data collector may observe (with his own eyes or by administering a questionnaire). In Rasch lingo, the underlying ability is known as the “latent trait” and is inferred from the pattern of response. The “Rasch score”, by contrast, is the simple count of items present, or, in the interview perspective, the number of item questions answered “Yes”.

The great attraction of the Rasch scale, in a participatory research perspective, is that the score and the latent trait are in a simple mathematical one-to-one relationship. All individuals with the same score have the same estimate of the latent trait. For example, if a survey probes for the presence of ten distinct durable goods, and Household A reports only these three: Bed, bicycle, refrigerator, while Household B owns a bed, telephone and TV only from among the same ten-item list, then A and B have the same score of 3 and are estimated to have the same ability to acquire and retain durable goods currently.

The consequence is obvious: Since the score captures all the relevant information regarding the latent trait (Hardouin 2007b: 23), it should be more than satisfactory for the purpose of providing ranks, which require only information on order and not magnitude. In other words, a count of the household goods present from the list is enough to assign the household to a wealth rank. Households can be assigned to poorer and less poor groups on the basis of the score alone although the substantive interpretation of such grades as “very poor”, “poor”, etc., “rich” depends on additional knowledge. Since calculating the Rasch score is a very simple operation, it can be done locally and immediately; and a rough wealth ranking or other uses of the score in survey analysis can proceed without outside statistical support. The latent trait and other more sophisticated statistics can be calculated later, or not at all, depending on needs and the availability of expertise.

However, score equivalence – household A and B in our example have the same underlying ability – holds only if the items included in the Rasch scale relate to the same dimension. Tests for item suitability too require statistical expertise. The objective therefore is to pre-validate a scale – a set of binary variables speaking to the same topic – using a recognized survey dataset and then apply it in subsequent surveys.

## **DHS Bangladesh 2004**

Almost two hundred Demographic and Health Surveys (DHS) have been carried out in several dozen countries for over twenty years, are nationally-representative household surveys with large sample sizes (usually between 5,000 and 30,000 households). (<http://www.measuredhs.com/>). Many of the datasets are available to interested

researchers by ORC Macro, a research firm that has played a key role in the development and expansion of these surveys.

In Bangladesh, six surveys were carried out between 1994 and 2004. The 2004 survey (NIPORT et al.) collected data on 10,500 households. For the purposes of this paper, we concern ourselves only with a very small segment of the data, chiefly those used in the wealth index construction (the background segment from the survey report is given on page 11 below).

Some of the variables used by DHS were not binary. I could not find the recoding rules by which DHS transformed them and passed them to the algorithm used to compute the index. Similarly, I could not locate a few of the variables used (e.g., “household has some servants”) in the data table. For the purposes of the Rasch analysis (using the “raschtest” procedure that Hardouin, op.cit., offered in STATA 8), I have dichotomized variables with multiple categories. These are the four with the attribute “Good” in the table below. For example, a household is coded as living in a dwelling with “good walls” if the primary wall material in the DHS codes was either “tin”, “finished”, or “cement / concrete / tiled”.

**Table 1: Durable household goods and their prevalence in DHS 2004 sample households**

Item	Urban	Rural	All
Good water	99%	97%	97%
Cot or bed	94%	91%	92%
Good roof	96%	90%	91%
Watch or clock	79%	62%	66%
Chair or bench	68%	64%	65%
Table	65%	61%	62%
Some land (other than homestead)	38%	56%	52%
Good walls	70%	47%	52%
Electricity	77%	30%	41%
Radio	37%	29%	30%
Almirah	47%	24%	29%
Bicycle	18%	26%	24%
Good toilet	48%	16%	23%
Television	49%	16%	23%
Sewing machine	13%	3%	5%
Telephone	16%	2%	5%
Motorcycle or scooter	4%	1%	2%

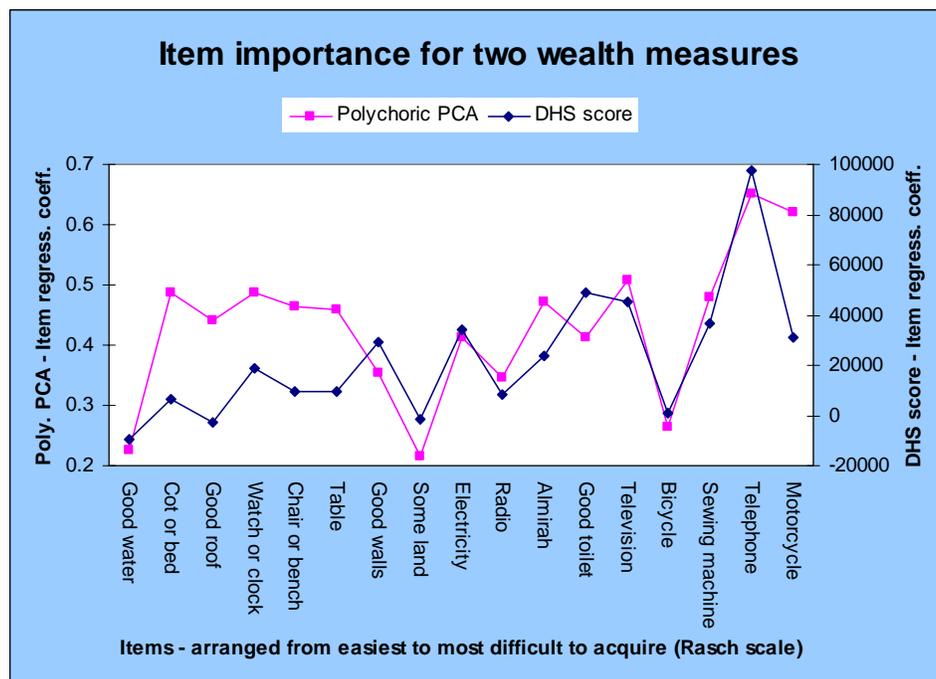
It is obvious that there are gaps between the prevalence levels of the 17 items. There is a wide one yawning between “good roof” and “watch or clock”. There is, although only for rural households, a steep drop from “good walls” to “electricity”. There is another gap separating the prevalence of TV from that of sewing machines. This is not ideal for scale construction, but generally survey designers have little control over the spacing unless the item frequencies are known from earlier statistics.

The DHS data table holds two wealth indicator variables, one being the factor score from an ordinary principal components analysis, the other the classification of households by quintiles. At this point, I need to quickly digress into some technicalities the results of which I will use for the simplification of the Rasch scale. Readers not interested in these arcane matters may skip to the diagram below.

As mentioned earlier, the use of ordinary PCA on binary and ordinal variables has been criticized by Kolenikov and Angeles (2004: , op.cit.). I have performed the alternative procedure that they recommend (polychoric PCA) on the 17 binary variables that I was able to locate or construct from the DHS data. This model produces three components with eigenvalues greater than one, but the first is so much stronger that it alone needs to concern us. It explains 47 percent of the variance. Unfortunately, I have not found documentation on the PCA that the DHS statisticians used and thus cannot compare the polychoric component statistics to the loadings that they obtained using the ordinary flavor.

Regardless, I then performed two regressions on the set of 17 variables, first of the polychoric PCA first-component score, then of the DHS score (which in the publicized data set had been rescaled). The results – regression coefficients that express the importance of each item for the wealth score<sup>1</sup> - are graphically displayed here:

**Figure 1: The relative importance of 17 items for two household wealth indices**



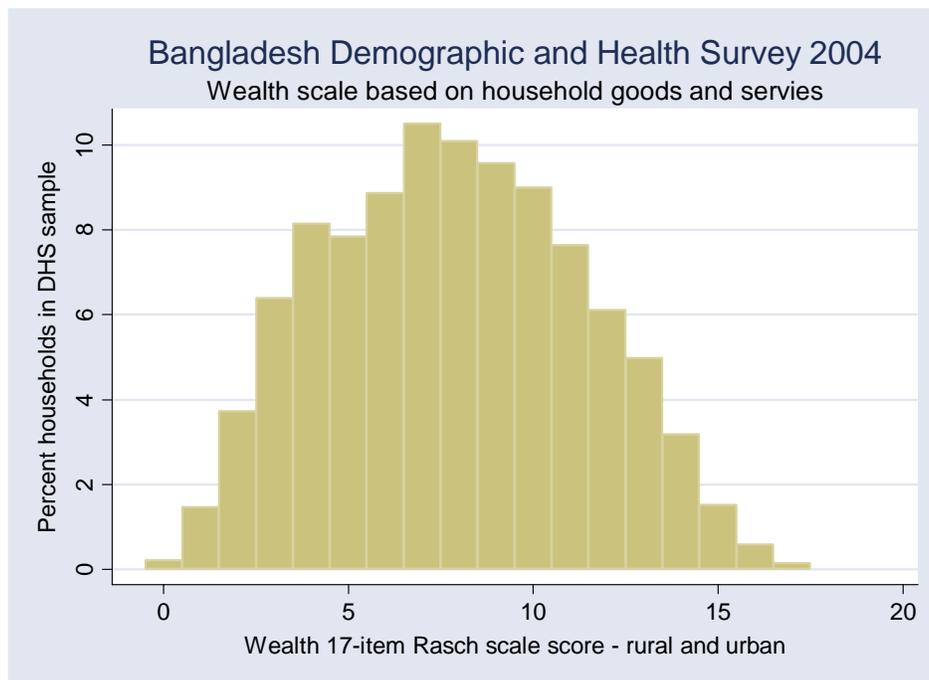
<sup>1</sup> The polychoric PCA factor score is nothing else but a linear combination of the 17 variables on which it is based, and therefore is completely explained by them. The same set accounts for 83 percent of the DHS score variance. The residuals are due to omitted variables (I did not have the whole set used by DHS), dichotomization of some ordinal ones, and the non-linear (normalization) rescaling by DHS.

The 17 items are arranged in the order of difficulty that the Rasch analysis revealed for their acquisition – from access to good water (which, in Bangladesh, is almost universal, not considering here the problem of arsenic contamination) to motorcycles, the possession of a slim minority of richer households. It is conspicuous that, starting from “household dwelling has good walls” the two line graphs follow each other fairly closely, with the exception of the coefficients of “motorcycles” at the far right. This is not the case of the left-side third of the items, which holds the items that poor people find easier to acquire. In this region, the blue line runs much lower than the red, with the exception of the first item, “access to good water”. Apparently, ordinary PCA downplays the importance of highly prevalent items. In other words, the DHS score is less sensitive to items that discriminate in the lower range of the wealth spectrum.

I let this graph, and the earlier frequency table, guide the selection of items that are to be eliminated from the Rasch scale item sets step by step. Candidates for exclusion are those items that were poorly associated with the first component (statistically speaking) or, if you like, with the wealth dimension (substantively speaking). Good water access is more the result of public policy in Bangladesh than of individual household wealth. Land ownership is important, but in a fast diversifying occupational structure, many households have advanced outside land ownership. Bicycles are not closely associated because of urban / rural differences and because households which owned some bicycles may have disposed of them when they could afford a motorcycle. This, of course, is speculation, just as when we conjecture that the effect of radio ownership is relatively weak because of upgrading to TV.

## Rasch analysis of 17 DHS 2004 wealth indicators

Table 2: Distribution of the Rasch score, all 17 item scale



The Rasch scale analysis worked out for 10,437 households that presented complete data in the 17 variables. The percentwise histogram of the 18 groups – from score 0 to score 17 – shows a somewhat normal distribution (graph on previous page).

The latent trait – the ability to acquire the items – is not defined at the extremes because the category of those without any items is a mix of households of indiscriminately low ability, and similarly the ability of the richest (who owned all 17 items) varies in ways that the data cannot determine. But the number of households in these score groups is very low (22 and 15), and for ranking purposes they could be assigned an artificial ability value, smaller than -4.58 for zero scorers, and larger than 4.99 for the top group.

**Table 3: Rasch score and ability parameters**

<b>Score</b>	<b>Ability parameter</b>	<b>Households</b>
0		22
1	-4.58979	154
2	-3.46174	391
3	-2.59867	669
4	-1.88293	854
5	-1.28611	822
6	-0.77095	929
7	-0.30516	1100
8	0.13334	1057
9	0.56028	1003
10	0.98918	943
11	1.43504	800
12	1.91753	640
13	2.4638	521
14	3.10915	333
15	3.90289	159
16	4.99471	62
17		15

At first sight, the ability increments from score value to score value are not too dramatically different for a good scale – about 0.4 in the center, and 1.1 at the extremes. The latent trait has good predictive power for the DHS factor score; based on a local regression that models a non-linear relationship, the ability parameters account for 74 percent of the latter’s variance.

Other tests later revealed that three items among the 17 were not suitable for a unidimensional scale<sup>2</sup>, but here we are chiefly concerned with simplification.

What would happen if we reduced the number of items in the Rasch scale, in order to save on data collection, or to trade off interview resources to other subjects?

<sup>2</sup> See the Mokken scale analysis in the appendix.

We conduct this experiment by retaining at first 10 of the 17 items, then only 5, the ones shown in gray in this table:

**Table 4: Items retained in the reduced Rasch scales**

Item	Prevalence in DHS sample	Initial 17-item scale	10-item scale	5-item scale
Good water	97%			
Cot or bed	92%			
Good roof	91%			
Watch or clock	66%			
Chair or bench	65%			
Table	62%			
Some land	52%			
Good walls	52%			
Electricity	41%			
Radio	30%			
Almirah	29%			
Bicycle	24%			
Good toilet	23%			
Television	23%			
Sewing machine	5%			
Telephone	5%			
Motorcycle	2%			

Elimination, as explained above, was guided chiefly by looking at the coefficients in the regression models and out of concern to maintain reasonable spacing across the prevalence or item difficulty spectrum.

The results are telling. With 10 items retained, the ability parameters of the Rasch scale exhaust 82 percent of the DHS score variance, slightly better than in the full 17-item scale. However, halving the number of items down to five reduces the R-square to a meager 0.58.

In other words, there is simplification and oversimplification. From these results – which are based on one data set only – it appears that dichotomization and use of the simple Rasch score – the count of items present in the household – were productive. Reducing the item set to roughly two third the number of original items did not result in a loss of predictive validity. Reduction by two thirds would be an oversimplification.

At the same time, a Rasch score is a simple measure, but no simple magic wand. Notably, some of the traditional poverty groupings, such as deciles or quintiles, cannot be exactly formed on the basis of the score. The 10-item score has a maximum of 11 unique values; the DHS wealth factor score, as a continuous variable, has 5,598 unique values realized in this sample! Groupings that try to approximate predefined ranges or n-tiles, such as the wealth quintiles of DHS, are therefore imprecise, as this cross-tabulation shows.

**Table 5: Crosstabulation of Rasch score groups and DHS wealth quintiles**

Rasch scores				DHS wealth index					
Group	Score range	N	Percent	Poorest	Poorer	Middle	Richer	Richest	Total
1	0-2	2,179	21%	74.1%	21.9%	3.3%	0.7%	0.0%	100%
2	3	1,867	18%	26.0%	49.5%	20.2%	3.8%	0.5%	100%
3	4	1,802	17%	2.1%	29.7%	50.1%	15.7%	2.4%	100%
4	5-6	2,424	23%	0.0%	2.1%	23.7%	53.5%	20.7%	100%
5	7-10	2,209	21%	0.0%	0.0%	0.1%	13.0%	86.9%	100%
Total		10,481	100%	20.4%	19.0%	18.4%	18.7%	23.6%	100%

More fundamentally, of course, neither wealth indices following the Filmer-Pritchett line nor Rasch scale are a viable substitute for income or expenditure data when detailed dynamic or composition analysis of household budgets is needed.

## Conclusion

With only 10 items – some of which require a yes/no-style question (“Is your home connected to the electricity net?”), others are derived from responses to a multiple choice question (“What kind of toilet does your family use?”) -, a measurement of household wealth seems feasible that reasonably approximates the values calculated with more extensive data and more sophisticated methods. The wealth score based on these 10 items is extremely simple – the count of the items present in the household.

The selection of these items, however, is the result of a re-analysis of a large data set that used items and methods tested in numerous other surveys. The reduced Rasch scale with 10 items will need to be tested in other surveys in Bangladesh, particularly such as estimate annual household incomes, against which the score can then be calibrated. It is likely that the NGO RDRS Bangladesh will collect data in 2007 as part of their panel surveys of program participants that will permit such a test.

If such a simple measure is found good enough an estimate of a household’s relative wealth position, both data collection and some analysis forms can be placed at a reach of actors closer to, and perhaps even with some representation from, the surveyed communities and households. One need not be as radical as to expect seamless integration with such participatory tools as wealth ranking done in grassroots meetings. Sample surveys and participatory assessments thrive in different, if sometimes interacting traditions. But even small steps in conferring enhanced analytic possibilities to local organizations seem desirable in themselves.

## **Appendix**

### ***The DHS wealth index***

The DHS Bangladesh 2004 report explains the wealth index as follows (NIPORT et al.: Chapter 2, 23):

#### **2.7 WEALTH INDEX**

In this report, an index of household economic status was created and used as a background characteristic with information on household ownership of assets and use of selected services.<sup>1</sup> The economic status index used here was developed and tested in a large number of countries in relation to inequities in household income, use of health services, and health outcomes (Gwatkin et al., 2000). It is an indicator of the level of wealth that is consistent with expenditure and income measures (Rutstein, 1999). The wealth index was constructed using principal components analysis (Rutstein and Johnson, 2004). Asset information was collected with the *2004 BDHS Household Questionnaire* and covered information on household ownership of a number of consumer items, ranging from a television to a bicycle, as well as dwelling characteristics, such as source of drinking water, sanitation facilities, and type of material used for flooring. Each asset was assigned a weight (factor score) generated through principal components analysis, and the resulting asset scores were standardized in relation to a normal distribution with a mean of zero and standard deviation of one (Gwatkin et al., 2000). Each household was then assigned a score for each asset, and the scores were summed for each household; individuals were ranked according to the total score of the household in which they resided. The sample was then divided into quintiles from one (lowest) to five (highest). A single asset index was developed for the whole sample; indexes were not prepared for urban and rural populations.

Fn 1:

Variables include ownership of items listed in Table 2.9, except homestead; household characteristics in Table 2.8, except food consumption; and whether a household has at least one domestic worker.

### ***Selected statistical output***

#### **Sequence of procedures**

In partial analogy with the wealth index creation, I performed a principal components analysis on a set of 17 of the DHS-designated household wealth variables. As the objective was to test a Rasch model, polytomous variables (4 of the 17) were dichotomized. In the line of Kolenikov et al. (op.cit.), I used polychoric PCA. For later item reduction, I regressed the DHA wealth index factor score (which had been normalized) and the polychoric PCA first component score on the 17 binaries. The results were visualized in the importance diagram on page 6.

The polychoric PCA confirmed a dominant first component. In the Rasch context, however, the unidimensionality of the item set can be tested through a Mokken scale procedure (Mokken 1971). Hardouin (2007a) gives the rationale; he wrote the STATA “msp” procedure, which I used; I also used STATA’s generic “mokken”, which produces a more didactic output telling the user “how good” his scale is.

Subsequently, I performed the Rasch analysis, calculating item difficulty parameters, score and the latent trait, using Hardouin’s “raschtest”. The quality of the Rasch scale as a suitable substitute for the (ordinary PCA-derived) DHS wealth index was evaluated by regressing the index on the latent trait (the subjects’ ability variable). Because of omitted variables (I did not have the whole set used by DHS), dichotomization of some ordinal ones, and the non-linear (normalization) rescaling by DHS, I used a local regression procedure (“mrunning” in STATA; see Royston and Cox 2005).

Polychoric PCA, Rasch scale analysis, and regressions of the DHA wealth factor score on the Rasch latent trait were performed for 17-, 10 and 5-item sets. Mokken scale analysis was applied to the 17- and 10-item sets. The item reductions, however, were guided chiefly by the item frequencies (concern for regular spacing from rare to almost universal items) and by the effects of the items on the polychoric PCA first component score (concern for unidimensionality). I preferred polychoric PCA quantities to the Mokken scale statistics because the former accepted sample weights.

## Items and item dichotomization

variable name	storage type	display format	value label	variable label
hv206	byte	%8.0g	hv206	has electricity
hv201	byte	%8.0g	hv201	source of drinking water
hv205	byte	%8.0g	hv205	type of toilet facility
hv215	byte	%8.0g	hv215	main roof material
hv214	byte	%8.0g	hv214	main wall material
sh31b	byte	%8.0g	sh31b	almirah (wardrobe)
sh31c	byte	%8.0g	sh31c	table
sh31d	byte	%8.0g	sh31d	chair/bench
sh31e	byte	%8.0g	sh31e	watch or clock
sh31f	byte	%8.0g	sh31f	cot or bed
hv207	byte	%8.0g	hv207	has radio
hv208	byte	%8.0g	hv208	has television
hv210	byte	%8.0g	hv210	has bicycle
hv211	byte	%8.0g	hv211	has motorcycle/scooter
sh31k	byte	%8.0g	sh31k	sewing machine
hv221	byte	%8.0g	hv221	has telephone
sh42	byte	%8.0g	sh42	owns any land

I also used, chiefly for descriptive purposes, the urban vs. rural variable and, for frequencies and polychoric PCA, the DHS-calculated sample weights. “raschtest” does not accept weights.

Four of the 17 indicators were polytomous, with implied partial ordinality. I created dichotomous variables:

goodwater	byte	%8.0g	Household has access to good water (based on hv201)
goodtoilet	byte	%8.0g	Household has good toilet (based on hv205)
goodroof	byte	%8.0g	Household dwelling has good roof (based on hv215)
goodwalls	byte	%8.0g	Household dwelling has good walls (based on hv214)

by recoding the superior variants as 1. For example,

source of drinking water (hv201)	Household has access to good water		Total
	0	1	
-----	-----	-----	-----
pipel inside dwelling	0	667	667
pipel outside dwellin	0	222	222
tubewell	0	8,724	8,724
shallow tubewell	0	40	40
deep tubewell	0	469	469
surface well/other we	96	0	96
pond/tank/lake	228	0	228
river/stream	52	0	52
other	1	0	1
-----	-----	-----	-----
Total	377	10,122	10,499

## Polychoric PCA

Here I exemplify with the eigenvalues / proportion of variation explained as well as the scoring coefficients tables for the 17-item analysis (the STATA procedure name is “polychoricpca”):

Principal component analysis

k	Eigenvalues	Proportion explained	Cum. explained
1	8.072373	0.474845	0.474845
2	1.727664	0.101627	0.576473
3	1.387052	0.081591	0.658064
4	0.964059	0.056709	0.714773

[Output for subsequent component not shown here]

Instead of the scoring coefficients that the procedure routinely reports, a table of the (polyserial) correlation coefficients between the first three component scores and the items is more instructive. Coefficients greater than 0.50 are grayed.

	w17poly1	w17poly2	w17poly3
hv206	0.74	-0.35	0.12
goodwater	0.31	-0.10	0.66
goodtoilet	0.69	-0.35	-0.16
goodroof	0.71	-0.06	0.69
goodwalls	0.63	-0.23	0.44

sh31b	0.81	-0.13	-0.17
sh31c	0.85	0.63	0.04
sh31d	0.86	0.61	0.04
sh31e	0.85	0.34	-0.16
sh31f	0.84	0.58	0.57
hv207	0.61	0.23	-0.27
hv208	0.86	-0.26	-0.09
hv210	0.47	0.77	0.01
hv211	0.83	-0.18	-0.53
sh31k	0.67	-0.20	-0.31
hv221	0.92	-0.55	-0.37
sh42	0.41	0.65	-0.03

## Regression of the wealth index factor score on the 17 items using DHS sample weights as analytic weights

Source	SS	df	MS	
Model	7.1385e+13	17	4.1991e+12	Number of obs = 10474
Residual	1.4738e+13	10456	1.4095e+09	F( 17, 10456) = 2979.07
				Prob > F = 0.0000
				R-squared = 0.8289
				Adj R-squared = 0.8286
				Root MSE = 37544
Total	8.6124e+13	10473	8.2234e+09	

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
hv271					
hv206	34223.45	932.8695	36.69	0.000	32394.85 36052.05
goodwater	-9788.413	2198.143	-4.45	0.000	-14097.19 -5479.632
goodtoilet	49080.79	1003.552	48.91	0.000	47113.64 51047.94
goodroof	-2958.541	1415.441	-2.09	0.037	-5733.076 -184.0051
goodwalls	29355.22	828.6753	35.42	0.000	27730.86 30979.59
sh31b	23888.9	988.7181	24.16	0.000	21950.82 25826.97
sh31c	9686.583	1172.37	8.26	0.000	7388.513 11984.65
sh31d	9510.444	1180.642	8.06	0.000	7196.16 11824.73
sh31e	18666.26	965.3201	19.34	0.000	16774.04 20558.47
sh31f	6309.938	1450.328	4.35	0.000	3467.018 9152.859
hv207	8151.408	879.5718	9.27	0.000	6427.279 9875.537
hv208	45058.7	1138.455	39.58	0.000	42827.11 47290.29
hv210	794.2746	921.3191	0.86	0.389	-1011.687 2600.236
hv211	31079.53	2860.529	10.86	0.000	25472.35 36686.72
sh31k	36721.6	1721.776	21.33	0.000	33346.59 40096.61
hv221	97608.97	1979.51	49.31	0.000	93728.75 101489.2
sh42	-1377.319	781.2863	-1.76	0.078	-2908.789 154.1515
_cons	-97948.03	2500.287	-39.17	0.000	-102849.1 -93046.99

## Mokken scale analysis

For this step, I reproduce output for two item sets, the original 17, and then after the first reduction to 10. The motivation is to demonstrate the lack of unidimensionality in the 17 items, and the better quality of the 10-item scale. For its easier intuitive output, I use the output from STATA's generic "mokken" procedure, not from Hardouin's "msp":

### 17 items

Variable	Obs	%Pos	Hi	z(H)	Label
----------	-----	------	----	------	-------

Variable	Obs	%Pos	Hi	z(H)	Label
hv211	10499	0.023	0.649	50.576	has motorcycle/scooter
hv221	10495	0.065	0.645	80.166	has telephone
sh31k	10498	0.065	0.479	59.581	sewing machine
hv210	10498	0.238	0.271	58.253	has bicycle
hv208	10497	0.264	0.520	117.561	has television
goodtoile	10500	0.274	0.422	96.422	Household has good toilet (based on hv205)
sh31b	10495	0.323	0.483	114.715	almirah (wardrobe)
hv207	10498	0.324	0.355	84.278	has radio
hv206	10498	0.445	0.442	105.812	has electricity
sh42	10499	0.502	0.217	51.774	owns any land
goodwalls	10500	0.549	0.382	89.838	Household dwelling has good walls (based on hv214)
sh31c	10497	0.636	0.550	122.467	table
sh31d	10497	0.660	0.556	121.029	chair/bench
sh31e	10494	0.686	0.558	116.797	watch or clock
goodroof	10500	0.915	0.522	63.103	Household dwelling has good roof (based on hv215)
sh31f	10497	0.916	0.620	74.816	cot or bed
goodwater	10500	0.964	0.265	21.859	Household has access to good water (based on hv201)
Mokken H			0.443	251.606	

3 items have a scalability coefficient below the cutpoint 0.300  
 Consider dropping hv210 sh42 goodwater

According to Mokken (1971:185),  $0.40 \leq H < 0.50$  is a medium scale

## 10 items

Variable	Obs	%Pos	Hi	z(H)	Label
hv211	10499	0.023	0.756	43.878	has motorcycle/scooter
hv221	10495	0.065	0.854	74.490	has telephone
hv208	10497	0.264	0.647	106.630	has television
goodtoile	10500	0.274	0.539	90.128	Household has good toilet (based on hv205)
sh31b	10495	0.323	0.578	99.026	almirah (wardrobe)
hv206	10498	0.445	0.594	100.803	has electricity
goodwalls	10500	0.549	0.507	81.521	Household dwelling has good walls (based on hv214)
sh31e	10494	0.686	0.618	85.362	watch or clock
goodroof	10500	0.915	0.587	53.191	Household dwelling has good roof (based on hv215)
sh31f	10497	0.916	0.592	53.533	cot or bed
Mokken H			0.595	180.543	

According to Mokken (1971:185),  $H \geq 0.50$  is a strong scale

## Rasch scale analysis

As with the Mokken scale analysis, the 17-item and 10-item output is given here:

## 17 items

At first, the item difficulty parameters were calculated:

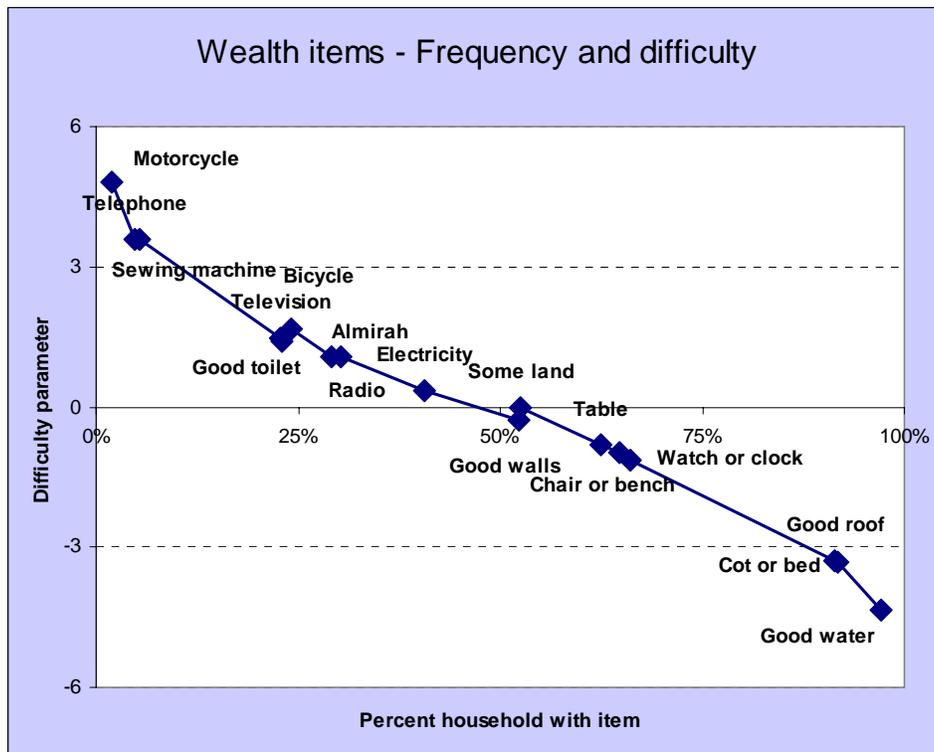
Number of items: 17  
 Number of groups: 18 (16 of them are used to compute the statistics of test)  
 Number of individuals: 10437 (26 individuals removed for missing values)

Items	Difficulty parameters	std Err.	Q1	d.f.	p-value
hv206	0.33496	0.03350	94.845	15	0.00000
goodwater	-4.35633	0.06364	643.561	15	0.00000
goodtoilet	1.40484	0.03534	143.508	15	0.00000
goodroof	-3.30460	0.04807	33.659	15	0.00380
goodwalls	-0.27661	0.03357	232.564	15	0.00000
sh31b	1.07416	0.03449	200.246	15	0.00000
sh31c	-0.80990	0.03428	339.543	15	0.00000
sh31d	-0.96661	0.03461	327.958	15	0.00000
sh31e	-1.13475	0.03502	282.330	15	0.00000
sh31f	-3.31068	0.04814	92.630	15	0.00000
hv207	1.07047	0.03448	170.367	15	0.00000
hv208	1.47127	0.03555	408.267	15	0.00000
hv210	1.66139	0.03621	1219.257	15	0.00000
hv211	4.80193	0.07504	89.727	15	0.00000
sh31k	3.57411	0.05102	30.620	15	0.00987
hv221	3.57785	0.05107	315.220	15	0.00000
* sh42	0.00000	.	2421.319	15	0.00000
Van den Wollenberg test			6631.175	240	0.00000
Andersen LR test		Z1=	5345.482	240	0.00000

\*: The difficulty parameter of this item had been fixed to 0

The following graph shows the relationship between item frequency and item difficulty. The relationship is nearly, but not perfectly, monotonic.

**Figure 2: Relationship between item prevalence and Rasch difficulty parameters**



Next, for each score (except the extremes) the corresponding latent trait value (“ability parameter”) is calculated:

Group	Score	Ability parameters	std Err.	Freq.	ll
0	0			22	
1	1	-4.58979	0.09577	154	-143.8800
2	2	-3.46174	0.04938	391	-727.4900
3	3	-2.59867	0.03425	669	-1658.4676
4	4	-1.88293	0.02760	854	-2895.5703
5	5	-1.28611	0.02586	822	-3577.2727
6	6	-0.77095	0.02287	929	-4475.6793
7	7	-0.30516	0.02021	1100	-5503.4046
8	8	0.13334	0.02018	1057	-5293.7825
9	9	0.56028	0.02060	1003	-5093.3197
10	10	0.98918	0.02147	943	-4677.0798
11	11	1.43504	0.02399	800	-3676.5383
12	12	1.91753	0.02821	640	-2700.1090
13	13	2.46380	0.03363	521	-1919.8165
14	14	3.10915	0.04614	333	-996.4988
15	15	3.90289	0.07512	159	-414.7521
16	16	4.99471	0.15007	62	-97.9978
17	17			15	

The loss at the extremes (household with no estimated latent trait) is relatively small for 17 items but increases for scales with fewer items.

## 10 items

Number of items: 10  
 Number of groups: 11 (9 of them are used to compute the statistics of test)  
 Number of individuals: 10168 (19 individuals removed for missing values)

Items	Difficulty parameters	std Err.	Q1	d.f.	p-value
hv206	-4.33710	0.07357	40.009	8	0.00000
goodroof	-8.20506	0.08507	28.715	8	0.00036
goodwalls	-5.02108	0.07410	339.563	8	0.00000
goodtoilet	-3.04135	0.07308	99.517	8	0.00000
sh31b	-3.45475	0.07319	65.621	8	0.00000
sh31e	-5.93232	0.07541	69.460	8	0.00000
sh31f	-8.20986	0.08510	11.403	8	0.17990
hv208	-2.95729	0.07306	191.453	8	0.00000
hv211	1.59056	0.10118	48.001	8	0.00000
* hv221	0.00000	.	43.463	8	0.00000
Van den Wollenberg test			843.483	72	0.00000
Andersen LR test			Z1= 810.997	72	0.00000

\*: The difficulty parameter of this item had been fixed to 0

Group	Score	Ability parameters	std Err.	Freq.	ll
0	0			194	
1	1	-8.47735	0.05119	617	-650.8844
2	2	-7.06118	0.03080	1368	-1618.9730
3	3	-5.87994	0.02381	1867	-3003.1370
4	4	-4.91743	0.02217	1802	-3622.9077
5	5	-4.07302	0.02455	1358	-3315.5773
6	6	-3.24657	0.02833	1066	-2495.5144
7	7	-2.29909	0.03393	946	-1760.9575
8	8	-0.93381	0.04854	713	-571.2878
9	9	0.93453	0.06909	431	-210.9568
10	10			119	

## Validation of the three Rasch scales

by local regression of the DHS wealth factor score on the latent trait. To check what influence the loss of cases due to missing latent traits at the extremes of the Rasch score (household with zero or with all items present) had the association between the Rasch scale and the DHS index, I calculated also the rank order correlation between the index and the Rasch *score* (which is available for *all* cases with complete item information). The “spearman” procedure does not allow weights, though.

Items in scale	Obs. estimated Rasch trait	w.. latent	R-sq (on latent (sample-weighted))	(on trait)	Spearman's (w. score) (unweighted)
17		10,437		0.74	0.91
10		10,168		0.82	0.92
5		9,304		0.58	0.86

## References

- Banerjee, A. V. and E. Duflo (2007). "The Economic Lives of the Poor." Journal of Economic Perspectives 21(1): 141-167.
- Embretson, S. E. and S. L. Hershberger, Eds. (1998). The new rules of measurement : what every psychologist and educator should know. Mahwah, N.J., L. Erlbaum Associates.
- Filmer, D. and L. Pritchett (2001). "Estimating wealth effects without expenditure data – or tears: an application to educational enrollment in states of India." Demography 38: 115-32.
- Fusco, A. (2005). La Contribution des Analyses Multidimensionnelles à la Compréhension et à la Mesure du Concept de Pauvreté : Application Empirique au Panel Communautaire des Ménages [The Contribution of Multidimensional Analysis to the Understanding and Measurement of the Concept of Poverty: Evidence from the ECHP]. Nice, France, University of Nice Sophia Antipolis, School of Law, Political Science, Economics and Business, Research Center for Macro-Economics and International Finance. Ph.D. thesis.
- Grosh, M. and P. Glewwe, Eds. (2000). Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 years of the Living Standards Measurement Study. 3 volumes. Washington DC, The World Bank.
- Hardouin, J.-B. (2007a). "Non parametric Item Response Theory with SAS and Stata." Journal of Statistical Software (under review).
- Hardouin, J.-B. (2007b). "Rasch analysis: Estimation and tests with raschtest." The STATA Journal 7(1): 22-44.
- Kolenikov, S. and G. Angeles. (2004). "The Use of Discrete Data in Principal Component Analysis With Applications to Socio-Economic Indices. CPC/MEASURE Working paper No. WP-04-85." Retrieved 21 January 2005, from <https://www.cpc.unc.edu/measure/publications/pdf/wp-04-85.pdf>.
- Lera St. Clair, A. (2004). Global Knowledge, Global Politics: the World Bank as a Transnational Expertised Bureaucracy. Transnational Knowledge Elites panel at

- the Fifth Pan-European International Relations Conference: Constructing World Orders, September 9-11, 2004, The Hague.
- Mokken, R. J. (1971). A Theory and Procedure of Scale Analysis: With Applications in Political Research The Hague, Mouton
- NIPORT et al. Bangladesh Demographic and Health Survey 2004. Dhaka, Bangladesh and Calverton, Maryland [USA], National Institute of Population Research and Training, Mitra and Associates, and ORC Macro.
- Royston, P. and N. J. Cox (2005). "A multivariable scatterplot smoother." Stata Journal 5(3): 405-412.
- Rutstein, S. O. and K. Johnson (2004). The DHS wealth index [DHS Comparative Reports No. 6]. Calverton, Maryland, USA, ORC Macro.
- Winship, S. and C. Jencks (2002). Changes in Food Security after Welfare Reform: Can We Identify a Policy Effect? [Report prepared for the US Department of Agriculture's Economic Research Service]. Boston, Department of Sociology and John F. Kennedy School of Government, Harvard University.

## **Acknowledgement**

I am grateful for the access that the data archive staff at ORC Macro, a research firm conducting DHS, has given me to the Bangladesh survey data. However, any errors in using this data or interpreting its context are my own responsibility.

Professor Jean-Benoît Hardouin, Department of Biomathematics and Biostatistics, University of Nantes, Nantes, France is the author of the STATA procedures "msp" (Mokken scale analysis) and "raschtest", which I used in this study. He provided valuable technical advice and also made the thesis of Alessio Fusco (op.cit.) available.

## **Author contact**

abenini [ at ] starpower [ dot ] net

Draft 5 April 2007

Please do not quote or cite without prior permission.