Measuring Subjective Economic Empowerment
Christopher Woodruff

Extensive evidence shows that participation in paid work increases both income and control over decisions at the individual, household and perhaps community levels. Increased control over household and personal decisions is one of the most important aspects of empowerment – the feeling of having control of one's life. It is also one of the most commonly used measures of economic empowerment. For example, researchers ask whether women have a say in important household decisions such as purchase of assets, or schooling of children, and whether they are allowed to move about on their own, and if they feel safe in doing so. In each of these realms, ample evidence demonstrates that paid work is causally associated with empowerment.

But control of one's life is more than the sum of these decisions. Even as paid work increases control in decision-making, it may also decrease the feeling of control over time. Participation in the labor market increases the demands on the individual's time. Women, especially women with young children, typically retain a large share of the responsibility for managing the household even after they enter paid work. Long work hours combined with demands at home may increase stress and the feeling of losing control over one's life.

In this note, I make a case for expanding our view of empowerment to capture issues related to control over time. The discussion is organized in four sections. In the first, I outline the case for expanding our measures of economic empowerment, and in the second section I discuss measures that might allow us to do so. In the third, I discuss empirical techniques typically employed now to capture these measures, and what data on those measures say. In the final section, I discuss the challenges we face with existing measures, and I offer some ideas for overcoming those challenges.

Time as a fundamental object of control

Suppose we wish to evaluate the impact of a childcare intervention that increases the number of childcare centers in areas where jobs are available for women. What outcomes should we measure? One obvious outcome is women's participation in paid work. We might be interested in shifts into paid work, in shifts from self-employment to wage work, and in shifts across sectors within self-employment. In terms of hours of work, we might think that the biggest effects would come from the extensive margin – the availability of childcare decreases the cost to women of entering labor markets – though possibly there would be some effect on the intensive margin – perhaps women self-employed at the start of the program would work longer hours. A movement into paid work, an increase in hours, or shifts to higher-paying occupations should all increase female empowerment as traditionally measured.

But there is another effect that could be equally important. The childcare centers might provide women who were *already working* with a much better and more secure alternative source of care for their young children. Even without any shift in occupation, the childcare centers might lead to a reduction in stress, and an increase in mental bandwidth for these women. Indeed, in many contexts where women work in spite of good alternatives for

childcare, women in this category may be more numerous than those affected at the margins of labor market participation. These women would not necessarily see an increase in income or an increase in household decision-making. But they would experience more control over their time from the increased reliability of formal (rather than informal) childcare arrangements, and a tangible gain in well-being from reduced worries about their children's welfare.

While the traditional measures of empowerment are clearly relevant here, I propose that we cannot ignore some measure of well-being. This is especially true in situations where wage work is common. The reason for this is that an important part of the issue here comes from the lumpiness of work. Compared with either rural agricultural work or self-employment, non-agricultural wage work often comes with earnings that are both higher and more stable. But stable wage work most often also comes with a loss of control over the number of hours worked. Women employed in the garment sector in Bangladesh earn higher wages than almost all relevant alternatives; at the same time, they have no choice but to work six 10-hour days or more per week. Their only available choice is whether or not to take the job. There is no option to work part time, and little flexibility to miss work when circumstances in the household require their attention.

Flexibility is especially important to women who are managing household responsibilities – a point underscored by the reasons self-employed women give as the rationale for entering a particular sector. As noted in Woodruff (2014), from a list of 10-12 motivations for choosing a particular sector in which to start a business, the most common reason given by samples of women in Sri Lanka and Ghana is that the chosen sector gives "more flexibility to look after children or other family members than other sectors or activities."[1] As the gap increases between more lucrative wage labor opportunities and self-employment earnings, we should expect women to move into wage work. But we should also recognize that this movement comes with a loss of flexibility, a loss of control over time.

Measuring Stress and Well-being

While we most commonly measure income or consumption, we are actually concerned with the more abstract idea of "utility." Perhaps the most straightforward proxy for utility is suggested by the now-considerable literature focused on measures of "happiness" or "well-being." Concerns have been raised because happiness and well-being scales are derived from self-reports. Hence – particularly in the case of evaluating interventions – we should worry about their reporting bias, an issue I address in the next section[2]. We discuss here three ways of measuring (positive) well-being or happiness, and then we examine measures related to stress and control over time.

Well-being is most commonly measured using surveys. The first approach directly elicits the extent of the feeling itself, through one of three survey questions. The first question assesses happiness on a scale:

[1] The wording in Ghana was: "Because of the need to balance family obligations and work."

[2] There has been some recent work measuring happiness through brain scans, but I ignore that here as impractical in the relevant settings.

Taking all things together, would you say you are: 1) Very happy…4)Not happy at all.

The second asks for a rating of life satisfaction:

All things considered, how satisfied are you with your life as a whole these days? (scale 1-10)

Finally, the most involved of these direct approaches is the Cantril Self-Anchoring Striving Scale. In this approach, respondents are shown a ladder with 10 rungs, and are then asked to imagine that the top rung represents "the best possible life for you" and the bottom rung "the worst possible life for you." Respondents are then asked to place themselves on the ladder. They may also be asked to imagine their life five years hence, and to give their expected position on the ladder at that time.

An alternative survey approach measures affect, asking respondents, for example, how many times they have smiled or laughed in the past 24 hours. In a slightly more involved version, respondents are shown a list of 20 words (e.g., interested, irritable, guilty, alert) and asked to indicate to what extent (1= "not at all" … 5= "extremely") they feel that way in the present, or perhaps felt that way in the past week. Negative feelings are reverse scored, and the responses are summed. Deaton (2008) argues that life satisfaction is not the same as happiness. Life satisfaction asks respondents "…to make an overall *evaluation* of their lives" (p. 55, emphasis in original), while happiness measures affect. Using data from the Gallup World Poll[3], Graham and Chattopadhyay (2013) provide support for Deaton's view, showing that there is a crucial difference between the measures that focus on immediate feelings (happiness and affect) and the Cantril Scale. Reponses based on the life satisfaction questions are more likely to reflect views over a longer period of time, and to be less affected by immediate attitudes. Earlier work using data from the World Values Survey showed that, across countries, happiness increases with per capita income at lower levels of income, but then levels off. There is no correlation among high-income counties. Using the Gallup data, Deaton confirms the earlier pattern at lower income levels, but overturns the earlier conclusions for high-income countries. He shows that life satisfaction is increases with income per capita at all levels of income.

A complementary approach to measuring overall well-being is to focus on indicators of stress. In this vein, two very commonly used surveys are the the Generalized Anxiety Disorders (GAD-7) and the Patient Health Questionnaire (PHQ2/9). These simple seven- and two-question instruments are shown in the appendix.[4]

Perhaps we can use surveys of the same type to unpack channels through which effects occur. To take our childcare example, we might imagine that access to childcare removes

---

[3] See http://www.gallup.com/178667/gallup-world-poll-work.aspx . The Gallup World Pool surveys representative samples of adults in 160 countries.

[4] The long version of the PHQ includes 9 questions. See: http://www.phqscreeners.com/sites/g/files/g10016261/f/201412/PHQ-9_English.pdf.

specific concerns, which in turn reduces levels of anxiety and stress. The GAD-7 and PHQ2/9 involve questions of feelings at a high level. One possible way to gain insights into these issues is to use the Daily Hassles Scale, which asks respondents to what extent over the preceding two weeks they have experienced very specific hassles – being without electricity, heavy traffic, loss of sleep, etc. The appendix shows one version of a daily hassles instrument, but the questions can be targeted to specific issues likely to be sources of stress for the relevant population can help to uncover specific causes. Kanner et al (1981) list 117 questions and Holm and Holroyd (1992) 64 questions measuring hassles in several different realms of life. In the context of research and policy, the key here is to identify actionable sources of stress.

The term well-being itself seems an all-encompassing one, and optimists might believe that it provides measure that captures everything we need to know. But stress presents a complex issue. Stress is thought of as being a measure in the sphere of mental health, but if we focus on *outcomes* of stress, we might also want to include measures of physical health. The GAD-7 and PHQ2 scales do attempt to incorporate some measure of physical discomfort, but arguably fall short of a full incorporation of physical well-being. But as with mental well-being, measuring physical well-being is not straightforward, as there are widely recognized concerns with self-reported measures, and physical measures are costly and often intrusive.

What do we learn about Well-being and Gender?

There are reasons to believe that, in a development context, measuring well-being is particularly important for women. Graham and Chattopadhyay (2013) use data from the Cantril scale questions included in the Gallup World Poll to examine differences between female and male well-being gaps. They find that in the full, unweighted, sample, women are happier than men at a global level. However, when they group the 180 countries in the survey by GDP per capita, women's positive well-being gap disappears in low-income countries. Indeed, women are significantly less happy in Sub Saharan Africa.

The finding that women are happier at a global level is reversed when we look at data on stress rather than positive well-being. The medical and psychological literature has documented that women are twice as likely to suffer from chronic worry, anxiety and depression (Pilgrim et al. 2011, Patel et al. 2014). This has been attributed in part to the larger number and distressing nature of the daily hassles they face, such as threats to personal safety, and having to juggle work and household responsibilities.

There is surprisingly little evidence that variation in stress levels among working populations is associated with differences in well-being. In one study finding some connection, Steinisch et al (2014) measure stress levels of garment workers in Bangladesh through cortisol collected in hair samples. They find that stress is higher among workers with *better* prospects for promotion to management. This finding is consistent with some previous research in higher-income countries (e.g., Boyce and Oswald 2012). However, prospects for promotion is the only one of a dozen measures of work-related pressures Steinisch et al collect which is significantly related to measured cortisol levels.

I am not aware of large-scale cross-country surveys revealing gender differences in stress levels in low-income countries. At a much more micro level, I recently participated in a survey of households in a community near garment factories in Gazipur, just north of Dhaka. We asked respondents about access to services, community safety, and labor force activities; we also collected basic demographic information. We then gave participants the GAD-7 survey and PHQ-2 survey. We standardize and sum each of the responses to the GAD and PHQ questions, and then regress the resulting index of stress against a variety of reported conditions.[5] The data reveal very logical and reasonable patterns. Overall, women report stress levels almost 0.3 standard deviations higher than men. Some factors – low levels of education – are associated with higher stress among both males and females. Other factors are gender-specific. For women, higher stress is correlated with having pre-school children, being the head of the household, and lacking a savings account. Among males, stress is higher for those not working, those commuting more than 20 minutes to work, and those who report being sick in the past two weeks. Having pre-school children is not correlated with stress levels among men. We selected the particular neighborhood for study due to the presence of several garment factories in the area. Two-thirds of those in the sample who report working are employed in the garment sector. We find no association between garment work and stress levels. Indeed, if anything, stress is reduced by having someone else in the household employed in the garment sector, though this positive effect is not significant in all specifications. Nevertheless, this suggests an important channel of positive influence for the garment sector, and one that might be difficult to measure through other means.

The Gazipur data indicate that the GAD7 and PHQ2 measure stress in a reasonable way. Considerable variance emerges with these measures, and that variance shows sensible patterns. The data leave us with the impression that these measures can provide a substantial amount of information.

Issues with Measuring Well-being and Stress

An obvious concern with the measures of well-being and stress that we have discussed is that they come from self-reports. This leads to several potential issues. First, critics note that two individuals in the same circumstances, but with different outlooks on life, might give very different responses to questions about stress or well-being. That is, the criteria that respondents use for judgment may vary. We should particularly worry that the life outlooks affecting the responses might vary in ways that correlate with the characteristics of the respondents. Second, where respondents have little at stake, certain types of respondents – for example, those who are more conscientious – may exert more effort in thinking about the most valid response. In some settings, raising the stakes for responses – for example, paying for correct answers to a cognitive test – might alleviate this problem. But in other contexts, raising the stakes may not be possible, or may be counter-productive. High-stakes environments create a different set of potential problems, because

[5] Standardizing and combining the responses is not the normal practice in the psychology literature, but is more in keeping with recent approaches in economics. There may be more objection to combining the PHQ and GAD indices into a single measure, as the two are intended to measure different attitudes. However, they are, in our data, at least, very highly correlated.

respondents may know what the "correct" answer is and give that regardless of their feelings. An employer screening asking applicants whether they are honest is unlikely to receive honest responses. Many of the survey questions regarding stress or well-being are administered in low-stakes settings. But we should be concerned with responses in the context of program evaluation, where respondents in the treatment group may feel incentives to overstate changes in wellbeing.

With regard to measures of stress, one potential solution to these issues created by self reports is to measure biomarkers, and particularly, cortisol through either blood, saliva, hair or fingernails. Cortisol is a marker for stress, and when accurately measured, provides an accurate measure of stress levels. Saliva and blood provide a measure of current stress levels; hair, a timeline over the previous period up to six months (Kirschbaum et al. 2009); and fingernails, a measure from perhaps four to five month previous (Izawa et al. 2015). In most studies, however, cortisol measurement will not be feasible for both budget and logistical reasons. Each of the methods is fairly expensive, and they are either intrusive (especially blood and hair, where a lock the diameter of a pencil needs to be cut off right at the scalp); of questionable reliability (especially saliva, where the level is affected by recent meals, the time of day, and other factors aside from stress); or untested (fingernails, which have been used less often).

How well do self-reported measures of stress correlate with biological markers? Unfortunately, not as well as we might wish. Where the two measures have been collected in the same samples, the correlation between them is often weak and, in some cases, contradictory. Karlson et al. (2012) conduct a systematic review of 27 studies of work-related stress that measured both salivary cortisol and self-reported stress. The 27 studies contain 185 separate analyses. Only 42 of the 185 analyses show a significant correlation between the self-reported responses and measured cortisol levels, and in 13 of these, the correlation is unexpectedly negative. The low correlation could reflect the well-known difficulty of properly measuring biological markers, though, somewhat troublingly, Karlson et al. find that positive correlations are more likely in studies they rate as being of lower quality. They discuss several explanations for the low correlations. These include adherence to proper protocols for cortisol collection, a factor likely to be particularly challenging in low-income country settings, and failure to control for confounding factors. Karlson et al. also note that work-related stress is often intermittent, and the sampled populations are often relatively homogeneous in the level of stress they face, for example, because they are all from a single firm and occupation. Finally, the stress levels in most work / life situations may not be severe. In that context, participants may become habituated to the levels of stress, lessening the cortisol response to stressful situations. These suggest that self-reported stress levels, reflecting stress over longer periods of time, may be more relevant indicators.

Certainly more work is warranted in this area. The comparative advantage of social scientists in this regard is in thinking about ways to improve and validate survey questions. One area that has faced similar issues and is perhaps more advanced relates to the measurement of the effect of non-cognitive skills on education, labor market outcomes and other outcomes in adult life. This literature faces similar measurement issues in key

variables – the Big 5 personality diagnostic, for example. This literature also faces issues of concerns about bias in responses in difficult context, for example, when students are asked to rate teachers according to some criteria.

Where questions ask for a strength of opinion – how strongly does the respondent agree with a given statement – the concern arises that respondents will use different relative scales to anchor their response. One solution to this is to anchor the response through other questions. Kyllonen (2016) proposes a series of "anchoring vignettes." For example, education researchers want to know how effective a teacher is from the perspective of the student. Simply asking the student to rate the teacher is problematic both because each student will use a different scale – and that scale may be correlated with outcomes or student characteristics that offer potential insights. The anchoring vignettes would describe characteristics of very good and very bad teachers, and ask the students to rate the hypothetical teachers. Their responses reveal information about the scale the students use to rate their own teachers, and, hence, can be used to calibrate those responses across students.

A second technique suggested by Kyllonen is to structure questions so that respondents are required to choose between alternatives rather than rating each alternative separately. These "forced choice" questions help in situations where respondents are asked a series of agree/disagree questions, and have a bias to agree with everything, or to fake responses. For example, respondents in work settings are likely to report that their supervisor is kind, generous, and helpful even when s/he is not, because they may fear their supervisors will learn of their responses. Giving respondents a short list of characteristics and asking them to choose the single trait that is most strongly true allows them to provide information without being seen to criticize their supervisor.

One concrete idea in the spirit of the forced choice approach would offer respondents a choice to participate in one of several lotteries. For example, one lottery might involve providing transportation to and from work; one, prepared meals; and one, cash. The choice the respondent makes with regard to which lottery she prefers to enter would be informative with regard to the cause and monetary value of stress. Where the chosen lottery is actually carried out and winners given the prize, respondent will have stronger incentive to give a thoughtful and honest response, similar to higher-stakes testing. In any case, we are likely to be stuck in a world of self-reported responses for many important issues that offer potential to understand these issues; thus, clever ways to elicit the most valid responses possible will be beneficial.

A final challenge in analyzing well-being and stress issues stems from a time lag. Negative effects of stress in the present period may affect life satisfaction in the opposite way in a later period. Work now may cause stress, but it may also lead to larger investments in children, and accumulation of assets that allow for more flexibility and financial security at a later period. Time will be needed to collect enough panel data with any of these measures to be able to gauge these long-term effects. The difficulty of accounting for longer-term effects should not dissuade us from using measures of well-being; all measures we use in

an economic or social realm are imperfect. Nevertheless, the issue of lagged effects should be kept in mind.

Concluding thoughts and other issues

Empowerment measures control over one's life. The most common measures we employ reflect control through decision-making and freedom of movement. But *control over time* should also be a point of focus. Overwork and daily stresses reduce well-being, and may also reduce mental bandwidth, leading to errors in decision-making. (See, for example, Mani et al., 2013.) Hence, stress may not only diminish well-being directly, but indirectly through poor decision-making as well.

As it stands now, the most practical and best measures of control over time are survey questions. An important issue is that we don't yet have work which strongly validates these measures. So, while I believe the combination of the Cantril Scale question on life satisfaction and the PHQ2 / GAD-7 are a reasonable set of measures to employ, we should also build a research agenda around the validation of those measures. Somewhat surprisingly and worryingly, the limited literature comparing the responses to GAD/ PHQ with cortisol measures does not show strong correlations. More work is needed to understand why this is the case. But we also need to take lessons from the non-cognitive skills literature, and work to validate measures through more sophisticated survey methodologies – structured vignettes and forced choice questions, for example. We should encourage work in this area whose aim is to validate survey data and survey methods. Finally, we need to understand more about the underlying causes of stress. These are very likely to vary with the specific context. On this, of particular importance is uncovering causes which are actionable through policy.

References:

Boyce, Christopher and Andrew Oswald, 2012. "Do people become healthier after being promoted?," *Health Economics,* Vol. 596, 580—596.

Deaton, Angus, 2008, "Income, Health, and Well-being Around the World: Evidence from the Gallup World Poll," *Journal of Economic Perspectives*, Vol 22(2), pp. 53-72.

Graham and Chattopadhyay (2013) *International Journal of Happiness and Development* Vol 1(2), pp. 212-232.

Holm, Jeffrey and Kenneth Holroyd, 1992. "The Daily Hassles Scale (Revised): Does it Measure Stress or Symptoms?" *Behavioral Assessment*, Vol.14, pp. 465-82.

Izawa, S., L. Miki, M. Tsuchiya, T. Mitani, T. Midorikawa, T. Fuchu, T. Komatsu, and F.Togo, 2015. "Cortisol level measurements in fingernails as a retrospective index of hormone production," Psychneuroendocrinology, 54(1), pp. 24-30.

Kanner, Allen, James Coyne Catherine Schaefer, and Richard Lazarus, 1981. "Comparison of Two Modes of Stress Measurement: Daily Hassles and Uplifts versus Major Life Events." *Journal of Behavioral Medicine*, Vol. 4(1), pp. 1-39.

Karlson , Björn, Petra Lindfors, Roberto Riva, Christin Mellner, Töres Theorell and Ulf Lundberg, 2012, "Psychosocial Work Stressors and Salivary Cortisol," in Margareta Kristenson, Peter Garvin, Ulf Lundberg, eds., *The Role of Saliva Cortisol Measurement in Health and Disease,* pp. 43-66

Kirschbaum, C, A Tietze, N. Skiluda, and L. Dettenborn, 2009. "Hair as a retrospective calendar of cortisol production-Increased cortisol incorporation into hair in the third trimester of pregnancy," Psychneuroendocrinology, 34(1), pp. 32-37.

Kyllonen, Patrick C., 2016, "Designing Tests to Measure Personal Attributes and Non-Cognitive Skills", in Lane, Suzanne; Raymond, Mark R.; Haladyna, Thomas M. (eds.) Handbook of Test Development, Second Edition. New York: Routledge, 2016, p190-211.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, 2013. "Poverty Impedes Cognitive Function," *Science*, 341(6149), pp. 976-980.

Patel, V, H. Minas, A Cohen and MJ Prince, 2014. *Global Mental Health: Principals and Practice*. New York: Oxford University Press.

Steinisch, Maria, Rita Yusuf, Jian Li, Tobias Stalder, Jos A. Bosch, Omar Rahman, Christian Strumpell, Hasan Ashraf, Joachim E. Fischer and Adrian Loerbroks, 2014, "Work stress and hair cortisol levels among workers in a Bangladeshi ready-made garment factory — Results from a cross-sectional study," Psychoneuroendocrinology Vol. **50**, pp. 20-27.

Ussher, Jan M, 2011, "Gender Matters: Differences in Depression Between Women and Men, " in Pilgrim, D, A. Rogers, and B Pescosoudo, Eds, The SAFE Handbook of Mental Health and Illness. London: Sage Publications.

Appendix: GAD7 and PHQ2 survey questions.

Section 3. Anxiety (GAD7)					
Enumerator read: "Over the LAST TWO WEEKS, how often have you been bothered by the following problems?"		Not at all (0 days)	Sometimes 1-5 days	More than half days (6-10 days)	Nearly every day (11-14 days)
3.1.	Feeling nervous, anxious, or on edge	0	1	2	3
3.2.	Not being able to stop or control worrying	0	1	2	3
3.3.	Worrying too much about different things	0	1	2	3
3.4.	Trouble relaxing	0	1	2	3
3.5.	Being so restless that it is hard to sit still	0	1	2	3
3.6.	Becoming easily annoyed or irritable	0	1	2	3
3.7.	Feeling afraid, as if something awful might happen	0	1	2	3
Section 4. PHQ2					
Enumerator read: "Over the LAST TWO WEEKS, how often have you been bothered by the following problems?"		Not at all (0 days)	Sometimes 1-5 days	More than half days (6-10 days)	Nearly every day (11-14 days)
4.1.	Little interest or pleasure in doing things	0	1	2	3
4.2.	Feeling down, depressed, or hopeless	0	1	2	3

A Daily Hassles Scale

	Enumerator read: Over the LAST TWO WEEKS, how often have you experienced the following situations?"	Not at all (0 days)	Sometimes (1-5 days)	More than half days (6-10 days)	Nearly every day (11-14 days)	RTA -88 DNK -99
2.1.	You didn't have enough money for own food, hygiene, housing, transportation and health	0	1	2	3	
2.2.	You needed a credit, loan from informal lenders or savings and you couldn't get it	0	1	2	3	
2.3.	Felt overwhelmed with home duties and maintenance	0	1	2	3	
2.4.	You felt disturbed by the amount of noise in your community or at the factory	0	1	2	3	
2.5.	You felt disturbed by the amount of garbage (waste) accumulated in your community	0	1	2	3	
2.6.	You didn't have access to clean water and good sanitation (at home)	0	1	2	3	
2.7.	You didn't get enough sleep (quantity)	0	1	2	3	
2.8.	You didn't rest well or relax at your home (quality of rest at home)	0	1	2	3	
2.9.	Some women experience situations when walking in the streets. I am going to read some examples: Males staring persistently or winking at them, calling them to come close, singing or whistling, making sounds, making gestures, grab their hand or other parts of their bodies, tickling….How often did any of these happened to you in the last two weeks?	0	1	2	3	
2.10.	You felt afraid when walking on the street (as if someone was going to assault or rob you)	0	1	2	3	
2.11.	Someone close to you (for example, a relative) said things that made you feel bad about yourself	0	1	2	3	
2.12.	You didn't have enough money for children's food, education or care	0	1	2	3	
2.13.	Your husband spent too much money on things that are not to pay for the family needs	0	1	2	3	
2.14.	You didn't have enough people to talk to or help you	0	1	2	3	
2.15.	You had unpleasant feelings of weakness or dizziness	0	1	2	3	
2.16.	Body aches (headaches, stomachache etc.)	0	1	2	3	
2.17.	Palpitations	0	1	2	3	
2.18.	Chinta rog (worry illness)	0	1	2	3	
2.19.	Has anyone in your community being robbed or physically assaulted (in any way)	0	1	2	3	