The Service Delivery Underperformance Index: A Multidimensional Approach to Measuring the Inadequacies in Service Delivery

By Melanie Allwine and James Foster*

A new approach to the measurement of service delivery is introduced. The Service Delivery Underperformance Index (SDUI) adapts the Alkire and Foster (2011) methodology used for poverty measurement to measure the underperformance, or multiple inadequacies, in service delivery. The index satisfies numerous properties. It focuses on underperformance, satisfies dimensional monotonicity, subgroup decomposability, and decomposability by dimensions and indicators. Dimensions and indicators that could be used to populate the index are discussed, referencing past work that has shown the negative impact of the poor delivery of services on health and education of individuals. It is demonstrated how the Alkire and Foster (2011) methodology is applied to calculate the index using facilities as the unit of analysis. Significance and robustness of the resulting rankings are discussed.

1. Introduction

More than half of the eight Millennium Development Goals (MDGs) are directly related to improving health and education. While funding for healthcare and education has increased since the announcement of the MDGs, outcome variables in many low and middle income countries have stagnated (World Bank; International Monetary Fund (2006)). Clearly, increased health funding has been insufficient for countries to meet these MDGs. This observation suggests that an increased emphasis should be placed on the delivery of healthcare and education services, in particular on measuring and improving its performance.

This paper introduces a new approach to the measurement of service delivery that measures the underperformance, or multiple inadequacies, in service delivery. Though many research papers have discussed how inadequacies in service delivery impact outcomes, such as health and education, no work has been done on creating a single measure of poor performance in the delivery of services.

Using Sen's theory of development (Sen (1999)) and Foster and Handy (2008*a*)'s theory of external capabilities, each facility is shown to give the surrounding population additional capabilities or freedoms through its adequate provision of service delivery. The poor delivery of services then deprives individuals of capabilities. If

* George Washington University, Correspodance author's email: mallwine@umw.edu

development is the expansion of peoples freedoms, then we should focus on those individuals that experience multiple deprivations of freedom.

The SDUI adapts the Alkire and Foster (2011) methodology used for poverty measurement to allow the index to satisfy numerous properties. There is a focus on underperformance and those dimensions facilities are inadequate in. It satisfies dimensional monotonicity, which implies that the breadth of capabilities that people are not receiving from local facilities matters. Subgroup decomposability (so that if service delivery improves in one group and is unchanged in all other groups, service delivery will necessarily improve for both groups in total) and decomposability by dimension and indicator (so that policy can identify those dimensions and indicators that are making the largest contribution to the poor delivery of services). It will also allow for the inclusion of ordinal variables.

Indicators that could be used to populate the index are introduced. These indicators have been used in previous research to reflect the underperformance of service delivery in the health and education sector and in many cases shown to negatively impact health and education outcomes. These indicators are categorized into three core dimensions: the ineffective use of resources/financing, the ineffective use of infrastructure, and the ineffective use of labor.

It is shown how theAlkire and Foster (2011) methodology is applied to calculate the index using facilities as the unit of analysis. An in-depth explanation is given on how to apply adequacy cutoffs (the threshold below which a facility is considered inadequate in an indicator of service delivery) and the underperformance cutoff (the number of inadequacies required to classify a facility as poorly delivering services) to a data matrix of inadequacies for all facilities and calculate the SDUI. No discussion is given as to how to choose each of the cutoffs as this is left for application.

Lastly, it is discussed how to determine the statistical significance and robustness of the rankings determined by the SDUI. This includes using Spearman's Rho, Kendall's Tau, and Pearson's rank correlation coefficients to test the robustness of the rankings to changes in the cutoffs and weights.

The layout of the rest of this paper is as follows. Section 2 gives a literature review. Section 3 presents the theoretical need for a measure of underperforming service delivery. Section 4 shows how the SDUI is computed. Section 5 discusses the properties satisfied by the SDUI. Section 6 discusses the choice of weights on indicators and dimensions. Section 7 discusses statistical significance and robustness of the rankings to changes in the adequacy cutoffs, weights, and underperformance cutoff. Section 8 concludes.

2. Literature Review

There are numerous measures of the quality of health and education services. For instance, the OECD Health Care Quality Indicators Project uses indicators of mortality and screening rates for specific illnesses to compare the quality of healthcare across countries. The National Center for Education Statistics recom-

mends that 12 indicators of education quality be used to measure the quality of eduction in schools. These include: the academic skills, knowledge, experience and professional development of teachers. Focused and rigorous curriculum and technology in the classroom as well as class size. Disciplinary environment, high academic expectations/goals, leadership, and faculty. The quality of healthcare has also be measured depending on the purpose of the measure of quality. For instance, Bruce (1990) measures the quality of care for family planning. Nicolucci, Greenfield and Mattke (2006) measures the quality of care for diabetes.

A majority of measurements of quality found in research are used in regressions to determine the impact of quality on the dependent variable of interest. For instance, Hong, Montana and Mishra (2006) construct an index of the quality of family planning services at the facility level in order to analyze how quality of family planning services impact IUD use in Egypt. Peabody, Gertler and Leibowitz (1998) analyze the relationship between infant birth weight and measures of the quality of care. Gertler, Patrinos and Rubio-Codina (2008) show that parents direct involvement in the management of schools resulted in lower number of students failing a grade and lower number of students repeating a grade. Fertig and Schmidt (2002) finds that an index of poor basic school conditions which aggregates shortages of teachers and lack of instructional materials has a significantly negative impact on reading performance from the PISA dataset. Suryadarma and Rogers (2006) analyzes the effect of school and teacher quality on fourth graders math and dictation tests. They find evidence that teacher absenteeism has a significant negative effect on math tests, while students at schools that had recent staff meetings had higher scores on both math and dictation tests.

Note that measures of quality vary depending on the indicators used. Some include measures of basic inputs available and utilization rates of services. This paper focuses on those indicators of quality that reflect the poor performance, or underperformance, of the delivery of services. For instance, as indicators of quality of healthcare Hong, Montana and Mishra (2006) include the supervision of healthcare providers and updating of family planning registers to construct an index of the quality of family planning services at the facility level. Suryadarma and Rogers (2006) use teacher absenteeism as a measure of quality. Bjorkman and Svensson (2009) use cleanliness of the facility as an indicator of the quality of healthcare provision when analyzing the impact of increasing community involvement on the quality of healthcare.

Measures that specifically focus on the delivery of services, not on quality of services overall, are relatively new. Lewis and Pettersson (2009*b*) and Lewis and Pettersson (2009*a*) compile lists of indicators of healthcare and education delivery and propose a "dashboard" or "menu" approach for the measurement of service delivery. In this approach, the interested party chooses the indicators they would like to use (or those where data is available) and measures each from potentially separate data sets to come up with a series of numbers which determine the overall performance of service delivery. Bold et al. (2010) suggests indicators of

the delivery of services, including indicators of healthcare providers effort and knowledge and basic input availability for the measurement of service delivery. The authors also suggest that these indicators could be aggregated by normalizing each to a number between 0 and 1, take the arithmetic mean of each indicator, and then take the arithmetic mean of all of the indicators. This is similar to the aggregation approach used in the construction of the Human Development Index (HDI).

These measures of service delivery differ from the measure introduced here as they do not focus on underperformance. The theoretical argument below shows that according to Sen's theory of development from Sen (1999) and Foster and Handy (2008*a*)'s definition of external capabilities, each facility is considered to give the surrounding population additional capabilities or freedoms through its adequate provision of service delivery. The poor delivery of services then deprives individuals of capabilities. If development is the expansion of peoples capabilities, then we should focus on those facilities that deprive people of multiple capabilities.

Bold et al. (2010) requires that indicators all be cardinal. However, many indicators of service delivery, such as the performance of doctors and teachers, are not cardinal. Surveys of healthcare and education delivery that collect cardinal data are also very expensive and suffer from high rates of non-response. Lewis and Pettersson (2009*b*) and Lewis and Pettersson (2009*a*) do not aggregate indicators, which may lead to difficulties in making comparisons of the delivery of services across different groups and over time.

3. Theory

The methodologies presented here are directly related to Sen's capability approach. This approach measures a person's well-being by the capabilities, or freedoms, available to him or her to lead the kind of life he or she values. Therefore, development is measured as an expansion of peoples capabilities, or freedoms. (Sen (1999)) A key freedom that Sen discusses is the freedom of social opportunity. "Social opportunities refer to the arrangements that society makes for education, health care and so on, which influence the individual's substantive freedom to live better. These facilities are important not only for the conduct of private lives (such as living a healthy life and avoiding preventable morbidity and premature mortality) but also for more effective participation in economic and political activities. (Sen (1999), 39) Social opportunities for health and education are provided by facilities, and therefore the facility can be thought of as providing capabilities to the population in the area that the facility serves.

Foster and Handy (2008*b*) defines external capabilities as the capabilities that are conferred to individuals through their "direct connection or relationship with another person" (Foster and Handy (2008*b*),1). The prime example being the additional capabilities provided by literate family members to illiterate ones, discussed in Basu and Foster (1998). In this example, a family member who is illiterate experiences an additional capability through his or her relationship with

a literate family member. In the case of service delivery, the facility provides additional capabilities to individuals in the surrounding area. A high quality healthcare facility in the area allows a person to live a healthier life by providing adequate services such as checkups and vaccinations by quality personnel and the appropriate treatment for illnesses. To measure the capabilities being provided to the individual by the facility we are forced to measure the quality of services delivered to the population that the facility serves as a proxy of the capabilities the individual receives from these services.

If development is an expansion of a person's capabilities, then poverty is a lack of capabilities. This is the approach taken in the measurement of poverty in the Alkire and Foster (2011) methodology and its application by Alkire and Santos (2011) in the Multidimensional Poverty Index. Given that good quality facilities provide capabilities to individuals, poor quality facilities deprive people of these capabilities. Therefore, we should focus on those facilities that are depriving people of capabilities, in other words the worst performing facilities.

4. Measurement of Service Delivery

1. Dimensions of Service Delivery

As discussed in the literature review, there are many indicators that have been used in previous research that represent the underperformance of service delivery in health and education. Amin, Das and Goldstein (2008), defines service delivery as the effectiveness with which inputs are utilized to improve outcomes. Using this definition, the indicators used in previous research can be represented by the effective use of:

1. Resources and Financing
2. Infrastructure
3. Providers

These dimensions are similar to the classifications of indicators used by Bold et al. (2010) and Lewis and Pettersson (2009*b*) and Lewis and Pettersson (2009*a*). Resources and financing are combined because the way facilities receive resources depends on the country. In some cases, the government issues resources (such as medical supplies) directly to the facility, while in other cases the facility is given the financing to purchase their own resources.

These are simply suggested dimensions. One of the benefits of using the Alkire and Foster (2011) methodology is that the government can choose the dimensions (as well as indicators to represent these dimensions) to populate the index, depending on the structure of the sector. Following the suggestions of Alkire and Santos (2011), dimensions could be chosen based off of group participants discussions of the value of capabilities, enduring consensus, or theory.

A summarizing framework of the indicators that have been used in previous research to reflect underperforming service delivery are listed in Table 2.1 and some examples of these indicators from both the health and education sector are

given below the table.

Table 1—Dimensions and Indicators that can be used to populate the SDUI

Dimension	Indicators
Financing and Resources:	- Financial accountability of the state to the sector. - Supervision and monitoring of financing and resources (including budget and supply leakages and payroll irregularities[1]).
Infrastructure:	- Adherence to quality assurance activities - Maintaining the physical condition of the facility. -Patient satisfaction ratings.
Providers:	- Provider absenteeism - Provider performance - Provider education and training - Corruption (such as job purchasing[2] and informal payments) - Patient satisfaction ratings.

The state is accountable to its people for efficiently financing the health and education sectors. The WHO recommends that governments designate at least 8% of their budget expenditure to the health sector (National Institute of Statistics Rwanda, Ministry of Health Rwanda, and Macro International Inc. (2008)). UNESCO's EFA Global Monitoring Report suggests that at least 15 to 20% of countries' national budgets be spend on education. Gauthier and Wane (2009) find that only 1% of non-recurrent budget allocated to regional health services actually reaches local service providers, and that after taking into account leakages of funding, expenditure on health has a positive and significant relationship with the number of healthcare consultations in Chad. The leakage of drug supplies (including the presence of ghost patients) and mismanagement of revenues has been found to have a negative impact on healthcare delivery, as presented in McPake et al. (1999).

Quality assurance activities are activities that are performed at the facility level to maintain the quality of services delivered. In the health sector they may include supervisory check lists of equipment, medications, healthcare providers (such as observation checklists), supervision of healthcare providers, updating of records, facility reviews of mortality, and audits of medical records. These are some of the quality assurance activities listed in the DHS Service Provision Assessment survey. Gage and Zomahoun (2012) use data from health facility and household surveys in five states in Nigeria to examine the association of family planning service delivery and contraceptive outcomes. They find that there is a strong positive association of the knowledge of contraceptive methods and the

[1] the difference between payroll roster and number of actual providers working

[2] portion of providers claiming illegal payments were made to influence managerial decisions- including hiring and assignments

use of quality assurance systems by family planning health facilities in the local government area (LGA).

Hong, Montana and Mishra (2006) using the 2003 Egypt Interim Demographic and Health Survey (EIDHS) include the supervision of healthcare providers and updating of family planning registers to construct an index of the quality of family planning services at the facility level. Authors find that IUD use among women was significantly positively associated with quality of family planning services, independent of characteristics of the facility and women surveyed. This held for women that received their IUD from a public facility, but not from a private facility.

Quality assurance activities in education may include approved curriculum, development of explicit learning outcomes, regular evaluation of providers and recorded feedback, records of the monitoring of progress and achievements of students, and records of staff meetings. Suryadarma and Rogers (2006) use recent staff meetings as an indicator of quality of schools and find evidence that schools that had recent staff meetings had higher scores on both math and dictation tests for fourth grade students.

Maintaining the physical condition of the facility, for instance whether sharps have been put back into sharps containers and if surfaces are clean are imperative to the usefulness of the facility to improve the health of patients. Cleanliness of the facility is one of the indicators used by Bjorkman and Svensson (2009) to represent quality of healthcare provision. It is also used in Basinga et al. (2010) as a measure of quality of healthcare centers.

In the healthcare sector, provider performance has been measured by direct observation of providers and vignettes, where the interviewer acts as a patient and provides a very brief description of symptoms and the provider is evaluated based off of their response. Das and Hammer (2005) and Leonard and Masatu (2010) use vignettes to measure healthcare provider performance. Das and Hammer (2005) find that, among the top 20% of providers, there was still a more than 50% chance that providers would harm patients with their lack of knowledge in the case of viral diarrhea and an over 25% chance for patients with preeclampsia. Direct observations can also be used of providers observations with patients. Bjorkman and Svensson (2009), Peabody, Gertler and Leibowitz (1998), and Gage and Zomahoun (2012) use direct observations of healthcare providers. Bjorkman and Svensson (2009) use the quality of information provided to patients, the appropriate use of equipment, and the appropriate provision of supplements and vaccines for children to represent the quality of healthcare provision. Peabody, Gertler and Leibowitz (1998) use a measure of the completeness of clinical examinations as a measure of quality and find that women who had access to facilities that did a more complete clinical examination, had infants who weighed, on average, 128 g more than infants born in areas with lower quality facilities. Gage and Zomahoun (2012) use the quality of provider-client interaction in family planning in the LGA as another measure of quality and found it was positively associated

with current use of a modern method of contraception as well as the odds of currently using a modern method for men and women.

Provider performance in the education sector can be measured by teachers adherence to protocols, such as whether the teacher is actually teaching when in school. Chaudhury et al. (2006) find that though one-quarter of government primary school teachers were absent from school, only half of the teachers present were actually teaching when enumerators arrived at the schools.

Provider education is commonly used as a measure of provider quality. Hong, Montana and Mishra (2006) use both education levels and availability of training to construct an index of the quality of family planning services at the facility level and analyzed the relationship between the index and family planning outcome variables. Existence of training for providers to keep up with medical knowledge is important for the upkeep of the education providers have. Gage and Zomahoun (2012) use health worker training in family planning services as an indicator of quality and find it had a positive association with the odds of both lifetime use and current use of a modern method.

The education level of providers of education is a typical measure of teacher quality measured by reported teachers education completion as well as certification. It has been used in almost every paper discussing the quality of teachers. A sample of this research is Suryadarma and Rogers (2006), Das et al. (2005), and Kremer et al. (2005). It can also be measured by teachers scores on basic exams.

Provider absenteeism is a key indicator that has been used to measure the delivery of services by numerous researchers. Chaudhury et al. (2006) look at healthcare and education providers absence across the countries and find high amounts of absenteeism. Bjorkman and Svensson (2009) use healthcare provider absence as an indicator of quality of healthcare provision. Suryadarma and Rogers (2006) find evidence that teacher absenteeism has a significant negative effect on math test scores for fourth grade students. Kremer et al. (2005) finds that teacher absence is negatively associated with student attendance, as well as test scores for grade 4 children. Noting that it is likely that these results are underestimated due to measurement error of teacher absence and drop outs that may occur in schools with poorer quality teachers. Chaudhury et al. (2004) finds that teacher absence predicts lower scores in english, but not math in Bangladesh. Das et al. (2005) find a significant negative impact of teacher absence on english and math scores using panel data in Zambia.

One must be careful about using patients satisfaction ratings, as Aldana, Piechulek and Al-Sabir (2001) find that patient satisfaction ratings may not reflect quality, and that patient satisfaction ratings may have more to do with the cultural background of people than actual quality.

Survey data that has been collected to capture this data is usually measured using facility level surveys. This can be see in DHS's SPA survey and in some LSMS surveys that have facility level data. There are usually provider surveys included in the facility level surveys as well as patient surveys. Another option, if data was

available, would be to use patient level surveys of the indicators. Patients could answer questions about whether infrastructure was being maintained, whether providers were available on a given date, and corruption practices of providers. In some way this gets closer to Sen's theory of development as expanding the capabilities of individuals.

Though some of these indicators are causal, such that if provider are uneducated, they are likely to provide worse services, some are complements while others are substitutes, such as medication supplies and providers, this is not the direct concern of this measure. The measure specifically looks at indicators that are important for service delivery.

An increasing concern in the formation of indices is whether an indicator is necessary if the correlation between itself and another indicator is strong. Though this is a valid concern, it has been discredited in the poverty realm by economists such as Amartya Sen and Angus Deaton:

"Indeed, precisely because income deprivations and capability deprivations often have considerable correlational linkages, it is important to avoid being mesmerized into thinking that taking note of the former would somehow tell us enough about the latter. The connections are not that tight, and the departures are often much more important from a policy point of view than the limited concurrence of the two sets of variables." (Sen (1999), 20)

This also holds for healthcare delivery, as some of these indicators are likely to be correlated with one another, but each indicator is required to define overall healthcare delivery within a country, and it is likely that the relationship between any two indicators is not constant across countries or even within countries.

2. Properties that a measure of underperforming service delivery should satisfy

The aggregation approach used to aggregate the indicators above will fully determine the properties of the index. For instance, using an aggregation approach such as is used in Bold et al. (2010) would imply that the measure will not take into account the breadth of inadequacies at the facility, it will require cardinal data, and it will not focus on underperformance. In deciding on the aggregation approach it is important to determine which properties one would like the index to satisfy.

To measure the underperformance of service delivery, the aggregation approach should imply that the index has a focus on underperformance and those dimensions/indicators facilities are inadequate in. This is in accordance with our theoretical base using Sen's theory of development and Foster and Handy (2008*a*)'s definition of external capabilities that each facility is considered to give potential patients in the area additional capabilities or freedoms. If development is considered the expansion of peoples freedoms, then to increase development we should focus on those individuals that are being deprived of capabilities.

Dimensional monotonicity should be satisfied, such that if a underperforming facility becomes inadequate in another indicator or dimension then the index

should increase in value and therefore represent worse delivery of services. This property implies that the breadth of capabilities that people are deprived of by facilities matters. A facility where providers don't show up and supplies are unavailable is worse than a facility that only has providers that are absent. It also states that a facility with multiple inadequacies has a more negative impact on individuals than numerous facilities with only one inadequacy.

It should also satisfy subgroup decomposability, such that if service delivery improves in one group and is unchanged in all other groups, service delivery will necessarily improve for both groups in total. It should be decomposable by dimension and indicator so that policy can identify those dimensions and indicators that are making the largest contribution to the poor delivery of services. It should also allow for ordinal variables as many indicators of service delivery are likely to be ordinal and not cardinal.

All of these qualities are compatible with the Alkire and Foster (2011) aggregation approach used for poverty measurement. The next section adapts this methodology to the calculation of service delivery at the facility level to calculate the SDUI.

Note that a framework of simply regressing the indicators of healthcare delivery on health outcomes does not satisfy the above properties. Though a regression approach is useful in explaining and predicting health outcomes, this index is attempting to measure the poor delivery of services so that governments are aware of what is going on in a country. Solely by measuring the poor delivery of services, the hope is to draw attention and keep the state accountable to its people.

Note that there are also many outcome variables that are impacted by healthcare delivery, such as productivity, perceived well-being, under-5 mortality, life expectancy, and many others. This makes it very difficult to use a regression framework to create a measure of the quality of healthcare services. One would potentially have to create a health index to regress indicators of healthcare delivery on. This is the approach taken by the World Health Organization (2000) in the health sector. However it is likely that each indicator of service delivery will impact each health outcome differently, which would not be taken into account using this approach.

3. Computation of the Index

To aggregate these indicators into an index, the Alkire and Foster (2011) methodology is used. First, the indicators chosen must be measured and applied to the facility level.For the purpose of this index, the facility will be the level at which each of the indicators will be represented. This is because, as mentioned above, facilities are considered to be giving individuals external capabilities. We are concerned with those individuals who are deprived of these capabilities. However, the indicators above are not only measured at the facility level. Any indicator that is not measured at the facility level needs to be

transformed to a facility level indicator.

Financial accountability of the state to the sector is measured at the state level. Depending on the financial structure of the sector, this could mean that each facility within the state receives the same value for this indicator. This would be the case where all facilities rely on financing from the government. However, if private facilities are not reliant upon the government for financing, or only receive a small portion of funding from financing, then the indicator could be evaluated separately for public and private facilities. Though measuring an indicator for the facility at the state level conveys less about differences in service delivery within a country then indicators measured at the facility and provider level, it gives us critical information on differences in service delivery across countries. A similar approach is taken for poverty analyses using the MPI. Village level variables reflecting infrastructure, such as the availability of roads to individuals, have been discussed in the MPI.

Supervision and monitoring of financing and resources, adherence to quality assurance activities and maintaining the physical condition of the facility are measured at the facility level. Provider performance, absenteeism, education and training, and corruption of the provider are all measured at the provider level. The provider level indicators are then aggregated to the facility level. For instance, taking the share of providers with adequate performance or the share of providers without the basic education requirements being met. Patient satisfaction ratings are measured at the patient level and then must be aggregated to the facility level, again using using share of patients who are satisfied.

These indicators will then be represented in a matrix of facilities achievements. Using similar notation as Alkire and Santos (2011), let $y = [y_{ij}]$ denote the $n \times d$ matrix of achievements for facility i in dimension j, where each value within the matrix (y_{ij}) represents an achievement y for facility i in dimension j and each row of the matrix represents the achievements of a facility i in all of the dimensions.

Next, a weighting scheme must be chosen for the dimensions such that each dimensions receives a weight w_j such that all weights sum up to the number of dimensions ($\sum_{j=1}^{d}(w_j) = d$).

Now we must identify those facilities that are underperforming (i.e. those facilities that have multiple inadequacies), as these are the facilities we want the index to capture and monitor. First, an adequacy threshold must be chosen for each dimension, call this z_j. z_j is chosen such that if the value of the dimension for a facility y_{ij} falls below the threshold z_j, a facility is considered inadequate in this dimension. Achievements of each facility in each of the dimensions are evaluated according to the threshold, and a matrix of inadequacies ($g = [g_{ij}]$) is generated replacing each element of y (y_{ij}) with a value of 0 if $y_{ij} < z_j$ and 1 if 0 if $y_{ij} \geq z_j$.

From the matrix g a matrix of inadequacy counts (i.e. the weighted number of inadequacies a facility experiences) is calculated by calculating a weighted summation of each of the rows in matrix g. In mathematical notation, vector

$c = [c_i]$ where $c_i = \sum_{j=1}^{d}(w_j \times g_{ij})$ and c is therefore a $n \times 1$ vector representing the weighted number of inadequacies each facility experiences.

To identify facilities that are underperforming, a underperformance threshold K must be chosen to represent the weighted number of dimensions a facility must be inadequate in to be considered underperforming. Generate a vector ρ that replaces each value in vector c in the following way. Let $\rho_i = 0$ if $c_i < K$ and let $\rho_i = 1$ if $c_i \geq K$. This vector ρ identifies which facilities are classified as underperforming.

Construct a second matrix $g(K)$ called the censored inadequacy matrix that sets all values of g to 0 when a facility is not classified as underperforming. In mathematical terms $g_{ij}(K) = 0$ if $\rho_i = 0$ else $g_{ij}(k) = g_{ij}$. From the censored inadequacy matrix a censored inadequacy counts vector can be calculated as simply the sum of each of the rows of $g(K)$ or $c(K) = [c(K)_i]$ where $c(K)_i = \sum_{j=1}^{d}(w_j \times g(K)_{ij})$.

There are numerous ways to calculate the the SDUI from this information. First, one can simply calculate the average of the censored inadequacy matrix $g(K)$, meaning that $SDUI = \mu(g(K)) = (1/n)(1/D)\sum_{i=1}^{n}\sum_{j=1}^{d}(g(K)_{ij})$. This is the same as simply multiplying the share of facilities classified as underperforming by the average share of inadequacies that underperforming facilities experience. The share of facilities that are classified as underperforming can simply be calculated as the average of the vector ρ. This is called the facility count ratio, or FCR. The average share of inadequacies that a underperforming facility experiences is calculated by simply dividing each entry of vector $c(K)$ by the number of dimensions, d, and taking the average of the vector for all facilities classified as underperforming. In mathematical terms $A = (1/d)(1/q)\sum_{i=1}^{n}(c(K)_{ij})$. It is clear from this approach that the SDUI takes into account not only the incidence of underperforming facilities, but also the intensity of inadequacies within underperforming facilities.

5. Properties of the SDUI

The Alkire and Foster (2011) methodology satisfies numerous properties that the authors show to be important for poverty measurement. These same properties will hold for the SDUI. In a similar fashion as Alkire and Foster (2011), assume there are two subgroups (x and y) of the population of facilities where there are n(x) number of facilities of subgroup x and n(y) number of facilities of subgroup y. One could assume there are two types of facilities (x and y).

Subgroup decomposability of the SDUI implies that not only can the index be decomposed to find the SDUI for any grouping, but also this grouping when weighted by its population share can be added to yield the original SDUI. This can be seen below:

$SDUI(x, y) = SDUI(x) * n(x)/n(x, y) + SDUI(y) * n(y)/n(x, y)$

This implies that if service delivery improves in one subgroup and is unchanged in all other subgroups, service delivery will necessarily improve for the whole

country. In this way it is able to address inequalities across regions, countries, managerial types, and facility types in order to target those individuals with the smallest capability sets and therefore those areas with the least development.

Decomposability of the SDUI by each dimension and indicator implies that one can determine the contribution of each of the indicators and dimensions to the SDUI. For the jth dimension of the SDUI this can be written as follows:

$Contr_j = (1/d)(1/n)(1/SDUI) \sum_{i=1}^{n} g(K)_{.i}$

Where $g(K)_{.i}$ is simply the jth column vector of the censored inadequacy matrix $g(K)$. This is necessary for policy where many of the indicators of service delivery can be manipulated by policymakers to make improvements to the delivery of services.

As mentioned in the theory section, the breadth of the capabilities that people are not receiving matters. A facility where providers don't show up and supplies are unavailable is worse than when one facility only has providers that don't show up and another only has supplies that are unavailable. Evidence of this can be seen in the results of McPake et al. (1999) where it was found that health facilities that had multiple inadequacies in service delivery demonstrated a distinct pattern of low utilization rates in Uganda. Therefore implying that the people themselves realize that multiply inadequate facilities are more harmful. It is also clear that this will not only have an impact on utilization, but also on health outcomes both indirectly through low utilization rates and directly. The literature review lists a sample of the many papers that have documented the negative relationship between these indicators and outcome variables resulting from service delivery.

The SDUI takes into account this breadth of capabilities through the property of dimensional monotonicity such that if a facility becomes inadequate in an additional dimension, then the index will increase in value, reflecting a higher amount of underperformance. This is satisfied through A.

As mentioned in the theory section, we want our measure to focus on underperformance, as people who are not receiving multiple capabilities are being deprived of more freedoms. The SDUI has a focus on underperformance such that if any changes in the delivery of services by facilities not deemed underperforming has no impact on the index. This is a valid property assuming that governments will want to focus solely on those facilities that are underperforming, or multiply inadequate. In the same argument, the SDUI has a focus on only those dimensions that facilities are inadequate in. Such that if an underperforming facility improves in a dimension that is not inadequate, the index remains the same. We care about the capabilities people are deprived of that are provided by the facility.

The SDUI will also reflect inequality amongst the underperforming facilities. Imagine two groups A and B whereby B is simply the average of the indicators amongst underperforming facilities of A. Then B will necessarily have a lower SDUI than A reflecting a better delivery of services. This is very important for a measure of service delivery, as we do not want some individuals to have a large number of capabilities that they are deprived of. This also implies that

one cannot simply improve the measure by taking the best providers from the worst performing facilities and moving them to facilities that are just classified as underperforming and improve the service delivery, as this is not an improvement of service delivery. Those individuals served by original facility are even more deprived then before.

The SDUI is easily understood. This is extremely important in a measure of service delivery so that it can be applied easily by countries. First, the Alkire and Foster (2011) methodology is a relatively simple counting methodology. Second, the methodology has been adapted in numerous countries at this point to measure poverty and therefore it will be easy for policymakers to adapt the same technology for the measurement of service delivery.

It allows for ordinal variables. However, if cardinal data is available, then the SDUI satisfies the property of monotonicity, such that if a facility becomes worse in any of its inadequate dimensions, then the index will necessarily increase in value, reflecting a higher amount of underperformance. Note however that it is unlikely that cardinal data will be available for all dimensions.

The SDUI satisfies replication invariance and therefore the index can be compared across different population sizes. The SDUI will necessarily fall between the value of 0 and 1 (called normalization property). It also satisfies symmetry such that each facility receives the same emphasis. Symmetry is an important property in poverty measurement, the idea being that each person is valuable and should be valued the same. However, there is an argument that in measuring service delivery, maybe some facilities if serving a larger portion of the population should receive a higher weight. This can be taken into account by adding weights to each facility based off of their population share.

6. Choosing Weights

Weights in the Alkire and Foster (2011) framework can be assigned on dimensions, indicators within dimensions, and on facilities themselves. In the poverty realm, it is clear that each person should receive the same weight as we care about all human beings. In this same argument, if we care about all people the same then we should potentially give a higher weight to facilities that serve a larger number of people.

When setting weights on indicators and dimensions we must realize that these are not weights in the usual sense that a composite indicator employs them, instead here each indicator has been dichotomized to a 0,1 variable demonstrating whether the facility is inadequate in the respective indicator. When the facility is inadequate in the indicator, it is this number that represents the relative value of that inadequacy relative to the value in another indicator. Since the terminology weights has become common, the same terminology is used here.

Also when setting weights we must go back to the theoretical base above. We are measuring the capabilities that people are being deprived of by facilities. Therefore, similar to the discussion of weights by Alkire and Santos (2011), weights

on deprivations are a political decision. Alkire and Santos (2011) can use norms used in other research such as the MDGs and HDI, there are no such norms for service delivery, For this reason when applying the SDUI to data it is suggested that the researcher starts with an assumption of equal weights on dimensions and equal weights on indicators within dimensions, and then run robustness checks.

The argument that one can simply run regressions of the service delivery indicators on outcome measures to determine weights is not valid. First of all, because there are numerous outcome variables that will be impacted by the delivery of healthcare and education services, including mortality rates, life expectancy rates, productivity, well being, and many more. Running regressions using each of these dependent variables will give many different weights. Second, the weights in the Alkire and Foster (2011) framework are normative and each weight reflects the relative value of that inadequacy relative to the value in another indicator when a facility is underperforming. This cannot be reflected in a simple regression framework.

7. Choosing Underperformance Threshold K

The underperformance cutoff K is the sum of weighted indicators in which a facility must be classified as inadequate in order to be considered underperforming. Therefore, as the underperformance cutoff increases the classification becomes stricter and therefore the number of facilities considered underperforming will fall while the intensity or breadth of the inadequacies experienced by these underperforming facilities will increase. In accordance with the work on the MPI in Alkire and Santos (2011), K is a policy variable that dictates how many multiple inadequacies a facility must have to be classified as underperforming. When a political decision has not been made on the value of K, one can assume a plausible value for K and then run robustness checks.

8. Statistical Significance and Robustness Analysis

To test statistical significance of rankings, one can simply run the usual t-test for each of the rankings of interest. To test the robustness of the results to a change in the underperformance cutoff (K), one can check the ordering implied by the index for plausible values. Dominance of the resulting ordering is implied by a lower value of the index for all values of K (Foster and Shorrocks (1988)). However this can be a strict requirement and if a comparison is being made across more than 2 groups then another test of robustness may be needed.

To test the robustness of the underperformance cutoff (K), weights, and adequacy cutoffs (z_j) one can compare robustness of rankings to different specifications by computing rank correlation coefficients using Spearman's Rho, Kendall's Tau, and Pearson. Pearson's rank correlation coefficient simply calculates the percentage of pair-wise comparisons that are robust when different specifications are used. Kendall's Tau compares each pair within the group to determine if the

ranking of the pair is concordant (implying that the ranking is the same as the original ranking) or discordant (implying that the ranking does not hold under the new specification). Kendall Tau's rank correlation coefficient is then simply the number of concordant pairwise ranks (C) minus the number of discordant pairwise ranks (D) divided by the total number of concordant pairwise ranks ($\tau = (C - D)/(C + D)$). Spearman's rank correlation is computed by ranking all groups under the original and new specification and for each country calculating the difference in the two ranks (r_i) for each group i. Then Spearman's Rho is simply $\rho = 1 - (1/n(n^2 - 1))6 \sum_{i=1}^{n} r_i^2$.

9. Conclusion

This paper defined a new methodology for measuring the underperformance, or multiple inadequacies, in service delivery. Though previous research has discussed the negative impact inadequacies in service delivery on the health and education of individuals, no measure has been created to reflect the poor delivery of services.

Using Sen's theory of development and Foster and Handy (2008*a*)'s definition of external capabilities, each facility is considered to give potential patients in the area capabilities or freedoms. Because development is considered the expansion of people's freedoms, we should focus on those individuals with multiple deprivations of freedoms.

Indicators that could be used to populate the index were discussed. These indicators have been used in previous research to reflect the poor delivery of health and education services. These indicators are categorized into three core dimensions: the ineffective use of resources/financing, the ineffective use of infrastructure, and the ineffective use of labor. It was shown how the Alkire and Foster (2011) methodology is applied to calculate the SDUI using facilities as the unit of analysis. An in-depth explanation is given on how to apply adequacy cutoffs (the threshold below which a facility is considered inadequate in an indicator of service delivery) and the underperformance cutoff (the number of inadequacies required to classify a facility as underperforming in service delivery) to a data matrix of inadequacies for all facilities and calculate the SDUI.

It is shown that the SDUI satisfies numerous properties. It concentrates on underperformance and those dimensions in which facilities are inadequate. Takes into account the breadth of inadequacies that a facility satisfies through the property of dimensional monotonicity. Decomposability by subgroup, such that if service delivery improves in one group and is unchanged in all other groups, service delivery will necessarily improve for both groups in total. Also satisfies decomposability by dimension and indicator so that policy can identify those dimensions and indicators that are making the largest contribution to the poor delivery of services. Allows for the use of ordinal variables. Statistical significance and robustness tests of rankings to changes in weights and underperformance thresholds were discussed.

REFERENCES

Aldana, Jorge Mendoza, Helga Piechulek, and Ahmed Al-Sabir. 2001. "Client satisfaction and quality of health care in rural Bangaldesh." *Bulletin of the World Health Organization*, 79(6): 512–517.

Alkire, Sabina, and James Foster. 2011. "Counting and multidimensional poverty measurement." *Journal of Public Economics*, 95.7: 476–487.

Alkire, Sabina, and Maria Emma Santos. 2011. "Acute multidimensional poverty: A new index for developing countries."

Amin, Samia, Jishnu Das, and Markus Goldstein. 2008. *Are you being served?: new tools for measuring service delivery.* The International Bank for Reconstruction and Development/ The World Bank.

Basinga, Paulin, Paul J Gertler, Agnes Binagwaho, Agnes LB Soucat, Jennifer R Sturdy, and Christel MJ Vermeersch. 2010. "Paying primary health care centers for performance in Rwanda." World Bank Policy Research Working Paper Series 5190.

Basu, Kaushik, and James E. Foster. 1998. "On Measuring Literacy." *The Economic Journal*, 108(451): 1733–1749.

Bjorkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on COmmunity-Based Monsitoring in Uganda." *Quarterly Journal of Economics*, 124(2): 735–769.

Bold, Tessa, Bernard Gauthier, Jacob Svensson, and Waly Wane. 2010. " Delivering Service Indicators in Education and Health in Africa." World Bank Policy Research Working Paper 5327.

Bruce, Judith. 1990. "Fundamental Elements of the Quality of Care: A Simple Framework." *Studies in Family Planning*, 21(2): 61–91.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Mularidharan, and Halsey Rogers. 2004. "Roll Call: Teacher Absence in Bangladesh." *World Bank Working Paper*.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. "Missing in action: teacher and health worker absence in developing countries." *The Journal of Economic Perspectives*, 20(1): 91–116.

Das, Jishnu, and Jeffrey Hammer. 2005. "Which Doctor? Combining Vignettes and Item-Response to Measure Doctor Quality." *Journal of Development Economics*, 78(2): 348–383.

Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan. 2005. "Teacher Shocks and Student Learning: Evidence from Zambia." *World Bank Working Paper.*

Fertig, Michael, and Christoph M. Schmidt. 2002. "The Role of Background Factors for Reading LIteracy: Straight National Scores in the PISA 2000 Study." *IZA Discussion Paper No. 545.*

Foster, James E., and Christopher Handy. 2008*a*. "External Capabilities." *Oxford Poverty and Human Development Initiative, OPHI Working Paper Series.*

Foster, James E., and Christopher Handy. 2008*b*. "External Capabilities." *OPHI Working Paper Series.*

Foster, James Eric, and Anthony F. Shorrocks. 1988. "Poverty orderings and welfare dominance." *Social choice and welfare*, 5(2-3): 179–198.

Gage, Anastasia J, and Delayo Zomahoun. 2012. "Influence of the Service Delivery Environment on Family Planning Outcomes in Nigeria."

Gauthier, Bernard, and Waly Wane. 2009. "Leakage of Public Resources in the Health Sector: An Empirical Investigation of Chad." *Journal of African Economies*, 18(1): 52–83.

Gertler, Paul, Harry Patrinos, and Marta Rubio-Codina. 2008. "Empowering parents to improve education: Evidence from rural mexico." *World Bank Working Paper Series Paper 3935.*

Hong, Rathavuth, Livia Montana, and Vinod Mishra. 2006. "Family planning services quality as a determinant of use of IUD in Egypt." *BMC Health Services Research*, 6(79).

Kremer, Michael, Nazmul Chaudhury, F. Halsey Rogers, Karthik Muralidharan, and Jeffrey Hammer. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association*, 3(2/3): 658–667.

Leonard, Kenneth, and Melkiory Masatu. 2010. "Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance." *Journal of Development Economics*, 93(2): 226–234.

Lewis, Maureen, and Gunilla Pettersson. 2009*a*. "Governance in Education: Raising Performance." World Bank Human Development Network Working Paper.

Lewis, Maureen, and Gunilla Pettersson. 2009*b*. "Governance in Health Care Delivery: Raising Performance." World Bank Policy Research Working Paper 5074.

McPake, Barbara, Delius Asiimwe, Francis Mwesigye, Mathias Ofumbi, Lisbeth Ortenblad, Pieter Streefland, and Asaph Turinde. 1999. "Informal economic activities of public health workers in Uganda: implications for quality and accessibility of care." *Social Science and Medicine*, 49(7): 849–865.

National Institute of Statistics Rwanda, Ministry of Health Rwanda, and Macro International Inc. 2008. "Rwanda Service Provision Assessment Survey 2007."

Nicolucci, Antonio, Sheldon Greenfield, and Soeren Mattke. 2006. "Selecting Indicators for the quality of diabetes care at the health systems level in OECD countries." *International Journal for Quality in Health Care*, 1: 26–30.

Peabody, John W., Paul J Gertler, and Arleen Leibowitz. 1998. "The policy implications of better structure and process on birth outcomes in Jamaica." *Health Policy*, 43: 1–13.

Sen, Amartya. 1999. *Development as Freedom.* New York: Anchor Books.

Suryadarma, Daniel, and F. Halsey Rogers. 2006. "Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia." *Education Economics*, 14.4: 401–429.

World Bank; International Monetary Fund. 2006. "Global Monitoring Report: Millennium Development Goals - Strengthening Mutual Accountability, Aid, Trade, and Governance." Washington, DC: World Bank.

World Health Organization. 2000. *World Health Report 2000: Health Systems: Improving Performance.* World Health Organization.