Benchmarking Public Policy

Methodological Insights from Measurement of School Based Management

Suhas D. Parandekar

The World Bank
East Asia and the Pacific Region
Human Development Department
June 2014

Abstract

This working paper presents a benchmarking analysis of School Based Management (SBM) using empirical data from the Philippines. School based management is widely used as a policy tool in many countries that seek to improve the quality of service delivery through decentralization. School based management typically takes many years to have an impact on educational outcomes, but policy makers need to know sooner how well the policy is being implemented. The paper extends the well-known Rasch methodology from the literature on student achievement, including the Programme for International Student Assessment, to the measurement of the implementation of school based management by computing a Rasch measure of the implementation of school based management. To test whether the resulting benchmarked measure is plausible and has practical

policy value, the measure is tested for correlations with standardized measures of personality and political skills of school principals, developed in the psychology and political science literatures. The paper will be useful for readers interested in studying school based management as well as those interested more generally in the methodology of benchmarking implementation of public policy where the ultimate results are subject to long implementation periods. The methodology presented in this paper can be applied to enhance the rigor of the ongoing Systems Approach for Better Education Results (SABER) exercise to benchmark educational policies in various domains. That exercise is set to become one of the flagship policy analytical tools being developed by the World Bank and partner agencies.

This paper is a product of the Human Development Department, East Asia and the Pacific Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at sparandekar@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Benchmarking Public Policy: Methodological Insights from Measurement of School Based Management

Suhas D. Parandekar

Key Words: Public Policy; Measurement; Benchmarking; School Based Management; SABER; Philippines; Rasch measurement.

JEL Classification: I25, I28, Z18

Acknowledgements

Thanks are due to the Department of Education (DepED) of the Government of Philippines who approached the World Bank to conduct a study on the implementation of School Based Management (SBM) in the Philippines. Thanks are due to AusAID for financing this study through the AusAID World Bank Philippines Development Trust Fund No. 071200 and the Basic Education Public Expenditure Review Trust Fund No. 099307. This Working Paper is one of a series of analytical pieces prepared as part of the overall study and the paper itself does not purport to represent the view of the World Bank or of AusAID.

The overall study was conducted by a World Bank and AusAID team under the leadership and guidance of Motoo Konishi, WB Country Director, Philippines; Xiaoqing Yu, WB Sector Director, EASHD; Octavia Borthwick, Minister-Counsellor, AusAID Manila; and Luis Benveniste, WB Sector Manager, EASHE. The WB team was led by Suhas Parandekar, Senior Education Economist and included Futoshi Yamauchi, Senior Economist, and Lynnette Perez, Senior Education Specialist. The AusAID team was led by Ken Vine, Senior Education Adviser and included Lea Neri, Senior Program Officer; Hazel Aniceto, Portfolio Manager, Education; Teresita Felipe, Education Specialist; and Quintin Atienza, Senior Program Officer.

The WB-AusAID team was supported by a team of consultants led by Vandana Sipahimalani-Rao and included Brian Gozun, Jason Alinsunurin, Rouselle Lavado, and Maria Alma Pineda. The TNS Philippines survey team did an excellent job of administering the school survey in three divisions and collating and cleaning the survey data. The team benefitted from excellent support of the World Bank Manila Office and the AusAID Manila Office. Kristine San Juan-Ante and Corinne V. Bernaldez, EACPF, provided excellent administrative and logistical support from Manila and Chandra Chakravarthi, Anna Coronado and Maya Razat provided administrative support and Takiko Koyama provided research assistance from World Bank headquarters.

Thanks are due to Futoshi Yamauchi, Harry Patrinos and Luis Benveniste for providing very useful comments and suggestion on earlier versions of this paper. Any errors and remaining defects are the responsibility of the author, who may be contacted at sparandekar@worldbank.org.

A. Introduction

Increasingly strident calls for results and accountability from public spending are an accepted fact. At the same time, it is acknowledged that many results related to public policy interventions require a long time. Public policy often seeks to bring about ambitious and fundamental behavioral change from diverse groups of large numbers of people. Examples include efforts to streamline government, to reduce corruption, to enhance an economy's innovative capacity, and to enable people to be healthier. Long and uncertain implementation periods typify complex endeavors such as educational reform, which involve students, teachers, parents, school administrators, local, regional and national government officials who interact in multiple ways. While waiting for potential *results* far away in the future, public policy practitioners and researchers try to measure reform implementation *processes* and organizational *structures* that are eventually believed to lead to the desired social outcomes. Since it is quite difficult, if not impossible, to define absolute measures of abstractly constructed performance measures, the default is to use relative measures – where the performance of units is compared to one another, *benchmarked* to some actual or conceptual ideal.

Most measures are presented in the form of a multi-dimensioned index. Typically, the dimensions are broken down into sub-dimensions and variables that are constructed by experts in the field. Generally, the choice of dimensions and variables as well as the choice of weights is arbitrary – very rarely can one find any rigorously constructed theory of measurement or empirical evidence that explains why the elements of the index are what they are.¹ Sometimes there is an implicit weight of one for factors or dimensions and sometimes there are arbitrarily imposed explicit weights. The construction of these indices raise a series of questions - the most basic one of which concerns *construct validity* – does the index indeed measure what it purports to measure? Then there is the question of *inter-temporal* and *inter-personal invariance* or *reliability* – meaning would one be measuring the same thing when one conducts the measurement after some time or if some other person or group of persons conducts the measurement? A method general enough to cover different kinds of performance measurement, but simple enough to understand and compute would be really useful.

This paper examines the issue of benchmarking for the case of a popular and extensively studied policy reform known as School Based Management (SBM). SBM is a widely researched policy mechanism that can bring the benefits of decentralization to public service delivery in schools. It is widely acknowledged that SBM measures take a long time before they can have an economically significant impact on outcomes such as improved student learning. A corollary is that the level of implementation of SBM can serve as an intermediate outcome that can be used to determine the extent

¹ Well-known examples of expert opinion based benchmarking indices are the “Human Development Index” produced by UNDP and the “Doing Business” index produced by the World Bank. The World Bank’s “Systems Approach for Better Education Results” or SABER initiative extends the index construction method to benchmark education policy across countries. SABER covers a range of topics for policy benchmarking – from early childhood education to workforce development.

to which SBM is being successfully applied. Policy feedback loops can be set up to improve the systemic performance regarding SBM, perhaps through better information sources, resource flows and accountability measures.

The paper seeks to examine the SBM issue in depth with an empirical analysis of SBM implementation in the Philippines. The paper uncovers interesting insights into the implementation of SBM, the leadership of school principals and the relation of both to school performance. These insights can be useful to policy makers and researchers interested in SBM. At the same time, the paper seeks to illustrate the functioning of Rasch modeling, a popular technique used to measure student academic achievement. Rasch modeling is extended to the issue of measurement of SBM, and the methodological lessons uncovered here could be of relevance to many other contexts of benchmarking public performance.

The paper is organized as follows. Section B provides background regarding the four different streams that are combined in this paper: (i) School Based Management – in general and with specific reference to the Philippines; (ii) the Big Five Inventory (BFI) of personality traits from the applied psychology literature that is applied in this paper to Philippine principals; (iii) measurement of leadership abilities using a measure called the Political Skill Inventory (PSI); and (iv) Rasch analysis. This is followed by a detailed description of the empirical data used in the paper in Section C. The main analytical results that form the core of the paper are presented in the next two sections: Section D explores the benchmarking of SBM across the chosen sample of SBM schools; Section E presents findings relating SBM data with data on principal personality and political skills. Finally Section F provides some concluding observations and ideas about future directions for research.

B. Background

Our search for a methodologically sound technique to benchmark policy implementation using the example of SBM and principal leadership brings together four diverse strands of literature. This section provides an overview of the key concepts in the literature regarding school based management, personality traits, political skills and Rasch analysis. The background information is not intended to be comprehensive and the reader interested in more detailed treatment will find some useful references in this section to pursue further enquiry.

1. School Based Management (SBM) – Autonomy with Accountability

What constitutes SBM? School autonomy and local accountability are the underlying features defining school based management. Autonomy means the ability for schools to be empowered to take decisions. Accountability usually stems in SBM from a higher level of participation from the local community, usually through a school committee which includes school level administrators, parents and teachers as members. In an influential recent book (Bruns, Filmer and Patrinos, 2011), the authors describe possible areas of decision making decentralized to the school level. The list includes topics ranging from monitoring of student performance to the hiring and firing of teachers. Decisions can cover various pedagogical and administrative tasks. The school may be able to take decisions regarding the curriculum and the monitoring of teacher performance and the school may be able to allocate its own budget, and in some cases, even hire and fire teachers. Variations in the areas of decision making delegated to the school, and the degree of autonomy provided for those decisions are two of the inputs into a typology of school based management. Another detailed examination of SBM (Barrera-Osorio et. al., 2009) describes the different types of controlling arrangements at the school level, depending on the devolution of authority to the principal, the teachers and the parents.

Impact of SBM: Decentralization of decision making to the school level is expected to lead to an improvement in the performance of individual schools and consequently the entire education because it clarifies and simplifies governance arrangements for service delivery. “By giving a voice and decision-making power to local stakeholders who know more about local needs than central policy makers do, it is argued that SBM will improve education outcomes...” (Bruns, Filmer and Patrinos, 2011; p. 16). The authors explain in detail the so called ‘SBM results chain’ between implementation of SBM and school performance. The increased participation of local stakeholders is expected to lead to greater transparency and effectiveness in the use of resources at the school level. There is an ‘increased understanding of the rules of the game.’ With better planning and performance measurement and monitoring to go with resources, the school would likely have more open and welcoming environments for all the actors. The services delivered of the school would then be of a higher quality, resulting in improvement in educational indicators such as lower repetition and drop-out and better test scores. All this takes time, and the authors cite studies that indicate 8 to 10 years to see appreciable results in student achievement. A stream of literature somewhat parallel to the service

delivery set of arguments comes from educational researchers seeking to explain what makes an effective school. In addition to the work on SBM done by economists, educational researchers have also studied SBM (Briggs and Wohlstetter, 2003). The authors arrive at a list of elements indicated in the literature to be characteristics of successful SBM schools: a vision focused on teaching and learning; use of decision making authority to bring about meaningful changes in teaching and learning; development of teachers' knowledge and attitudes towards a learning community; the distribution of power across stakeholder and shared leadership, and mechanisms for collecting and communicating information about school performance.

Cross-national benchmarking of SBM policies: A recent benchmarking study provides a cross-country comparison of SBM implementation along a set of dimensions.² A simple version of the benchmarking framework is reproduced in Table 1 below – it is being applied in many educational policy areas as part of the SABER framework. Scoring for the study used a more complicated variant that included sub-dimensions under each of five dimensions, with numerical values of 1, 2, and 3 attached to the Low, Medium and High classifications. The numerical values are treated as cardinal numbers and totaled – a further classification of the totals leads to a four-fold classification of SBM in a country as being “Latent”, “Emerging”, “Established” and “Mature” along each policy dimension and also an overall dimension. The actual scoring is done by experts who presumably know the policy in each country though it is not clear how comparability is established across different experts rating different countries.

Table 1: Autonomy and Accountability at the School Level			
Managerial Factor	Strength		
	LOW	MEDIUM	HIGH
(A) Teacher and Personnel Management	Centralized Hiring and Firing	Regional Hiring and centralized firing	Local hiring and firing
(B) Budget Planning and Approval	Centralized budget based on payroll plus an allotment for materials and utilities	Decentralized budget with regional variations. Budget based on payroll and equity considerations	Decentralized at school level. Budget approved by the school council and funds transferred directly to the school
(C) Teacher Assessment	None	Routine evaluations, no direct accountability	Schools conduct routine evaluations that provide teachers and schools with incentives to perform better
(D) Student Assessment	None or based on local tests	Periodic standardized testing but results not made public	Routine standardized testing; results made public

Source: Arcia et al., 2011

² The SABER framework – see <http://www.worldbank.org/education/saber>. The “School Autonomy and Accountability” policy domain describes the detailed indicators for each of the five dimensions.

SBM benchmarking instrument in the Philippines: The Department of Education (DepED) in the Philippines introduced a manual (DepED, 2009) that began to be used since 2010 to calibrate the level of SBM implementation in schools in Philippines. SBM was being rolled out nationally in the Philippines since 2006 and the tool was devised as a way to measure the impact of the policy, for which large amounts of resources were being spent by DepED on SBM training and provision of school managed grants. Similar to the SABER framework, three levels – standard, progressive and mature, were applied to a hierarchical set of items across themes such as accountability and resources – the instrument has a rather comprehensive list of 457 items.³ A crucial motivation for this paper is that the DepED instrument, like the SABER framework mentioned before, is a list of items conceptually linked to SBM implementation – but so far lacking a formal method to check the instrument for validity and reliability.

We next look at two fields where measurement of a difficult and subjective concept has been addressed by developing tools and testing them across many contexts – if a similar effort is made for SBM related research, it would generate much greater confidence about the validity of the findings and policy conclusions from such research.

2. Big Five Inventory (BFI) of Personality Traits

A widely cited seminal piece in the literature (Goldberg, 1990) is credited with introducing the “Big Five” terminology. Goldberg describes in detail the process of categorization to understand better the description of traits from thousands of adjectives. This included an early effort to narrow down a list of 1,431 descriptive adjectives to 75 categories, still too numerous to use effectively in research. Goldberg applied the technique of factor analysis to further narrow down those 75 categories to five factors – he presents evidence to show how the result of five factors is robust to alternative choices to determine the underlying factors. Goldberg further tested alternative methods starting from a cluster of synonyms and testing with a different group of college students the task of describing personalities of individuals – and uncovered the same underlying structure of five factors. Other authors have worked on variants of the Goldberg methodologies and come up with the same result of five factors. One branch of the literature has explored the applicability of the five factor model across different cultures. Findings regarding the five factor model from 56 countries have been reported (Schmitt, et al., 2007).

The most commonly accepted variant of the five factor model is termed as the Big Five Inventory (BFI), an inventory of 44 items that survey respondents rate on a five point scale (John, Naumann and Soto, 2008). Respondents are asked to rate on a Likert scale “Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.” Factor analysis carried out under multiple respondent samples and alternative inventory specifications has resulted in an agreed index measure of five factors

³ DepED has since revised the methodology of the SBM assessment tool.

explained in Table 2. The BFI is reproduced in Annex Table 1, together with the variable combinations that yield the scores for the five factors in Table 2. The table also explains the meaning of each of five factors – extraversion, agreeableness, conscientiousness, neuroticism and openness by adjectives that describe low and high scores on a particular factor.

Table 2: Big Five Factors		
	Low Scorers	High Scorers
Extraversion	Loner; Quiet; Passive; Reserved	Joiner; Talkative; Active; Affectionate
Agreeableness	Suspicious; Critical; Ruthless; Irritable	Trusting; Lenient; Soft-hearted; Good-natured
Conscientiousness	Negligent; Lazy; Disorganized; Late	Conscientious; Hard-working; Well-organized; Punctual
Neuroticism	Calm; Even-tempered; Comfortable; Unemotional	Worried; Temperamental; Self-conscious; Emotional
Openness	Down-to-earth; Uncreative; Conventional; Uncurious	Imaginative; Creative; Original; Curious

Source: Coon and Mitterer, 2010

Personality traits as measured by BFI have shown to be of widespread applicability and BFI has been used to gain insight in various applied fields including occupational stress (Bakker et al., 2002), job performance (Hurtz and Donovan, 2000), entrepreneurship (Zhao et al., 2010; Caliendo et al., 2011), learning styles, and academic achievement (Nofle and Robins, 2007). BFI has been applied in studies of leadership and meta-analysis of the findings from over a hundred studies has been conducted (Judge, et al., 2002; Bono and Judge, 2004). This paper may be the first published application of BFI to School Based Management (SBM), though at least one previous study has explored the relationship between participative management and personality traits following the five factor model (Benoliel and Somech, 2010).

3. Political Skills Inventory (PSI)

Pfeffer (1981) is credited with being the first to use the term ‘political skill’ and to popularize the idea through a series of books that all organizations are inherently political in nature. In spite of the negative value supposedly attached to “playing politics” in avowedly apolitical organizations like profit-making corporations, Pfeffer persuasively argues from the perspective of more than two decades of research, that politics is an important determinant of personal and organizational success. ‘Politics’ here is taken to mean the securing and deployment of information and resources to meet specific goals. Essentially, to be political is to be influential over the behavior of other individuals, with or without realization by the other individuals that they are being influenced. There is clearly a negative value judgment attached to the deceitful manipulation of others for hidden selfish goals, but the exercise of influence more generally is a crucial element of leadership. Ferris, et al., 2005, define political skills as “the ability to effectively understand others at work, and to use such knowledge to influence others to act in ways that enhance one’s personal and/or organizational objectives”. The same authors declare

“we see political skill as independent from general mental ability and related to personality traits.” Their measure of political skills is outlined here.

Political Skill is constituted of four dimensions for which there is substantial evidence of content validity of the construct as established mainly through factor analysis across different samples. Ferris et al. describe the four dimensions as follows: (i) *Social astuteness*: “Individuals possessing political skill are astute observers of others and are keenly attuned to diverse social situations. They comprehend social interactions and accurately interpret their behavior, as well as that of others, in social settings”; (ii) *Interpersonal Influence*: “Politically skilled individuals have a subtle and convincing personal style that exerts a powerful influence on those around them. Individuals high on interpersonal influence ... are capable of appropriately adapting and calibrating their behavior to each situation in order to elicit particular response from others”; (iii) *Networking Ability*: “Individuals with strong political skill are adept at developing and using diverse networks of people...easily develop friendships and build strong, beneficial alliances and coalitions; and (iv) *Apparent Sincerity*: Politically skilled individuals ... are or appear to be, honest, open, and forthright... focuses on the perceived intentions of the behavior exhibited. The PSI is an inventory of 18 items that map on to the four dimensions of political skills. They were chosen from a larger list of 40 items after factor analysis and subsequent reform over different survey samples (Ferris et al., 2005).

While not yet as large as the literature on BFI on personality traits which was developed much earlier, there have also been a number of studies that seek to apply and extend the PSI to other countries such as Russia and China (Lvina et al., 2009; Shi and Chen, 2012). A number of studies have sought to relate political skills to performance in varied occupations. A meta-analysis of 35 such studies reports the interesting finding that “as the interpersonal and social requirements of the occupations increased, so did the strength of the positive relationship between political skill and task performance ratings” (Bing et al., 2011). An interesting set of studies combines PSI and BFI to examine their impact on leadership performance and one study explores the interaction between PSI and the personality traits of agreeability and conscientiousness (Blickle, et al., 2008). Interestingly, the authors find low performance for the combination of high agreeableness and low political skills. Theoretical reasons for possible mediation of personality traits through political skills on leadership behavior and performance have also been presented (Phipps and Prieto, 2011). How teacher leadership in schools is affected by political skills has also been investigated (Brosky, 2011). The combination of personality traits and political skills clearly holds the promise of interesting insights regarding leadership provided by school principals.

4. Rasch Modeling

Rasch models belong to the family of Item Response Theory (IRT) models. Rasch modeling became very popular in the 1980s as a method of assessing student achievement or proficiency from standardized testing. The method is applicable generally to any setting where measurement is not

possible directly because what is being measured is abstract, like student ability. Many book length treatments regarding Rasch modeling are available (Andrich, 1988; Bond and Fox, 2007). Rasch modeling for student testing is quite simple to understand. The objective is to uncover the value of the latent variable of student *ability* by testing students with a battery of questions of differing level of *difficulty*. The basic idea is that many students can answer easy questions, and as questions become more difficult, only students with a commensurate higher level of ability can answer them. At the extreme end of the difficulty scale, only a few of the highest ability students can answer the most difficult questions. A rough measure of ability of a student in a test is given by the percentage of correct responses by a student. Similarly, a measure of the difficulty of a particular question can be by examining the percentage of students who answered the question correctly. It is this simple mathematical symmetry between student ability and question difficulty that is at the core of the Rasch model.

The Rasch model predicts the log-odds of the probability that a person, v , of ability level θ_v can correctly answer an item i , of difficulty level β_i , as indicated by the following equation:

$$P(X_{vi} = 1|\theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)} \quad \dots[1]$$

Equation [1] is intuitively very easy to understand, probability is modeled in the log-odds form because the logarithmic transformation stretches out the middle of the difficulty/ability distribution or compresses the ends of it to enable scaled comparison. For example, to move from 90% to 95% of correct answers should be more difficult than moving from 50% to 55% of correct answers, assuming that difficulty is monotonically distributed, so that someone who answers a question of a given level of difficulty is more likely to answer a question of lower difficulty.

Rasch model for polytomous choice: The Rasch model can be extended from a dichotomous (yes or no – correct or incorrect) response to a graded or polytomous response, such as in a 5 point Likert scale, ranging through (1 disagree completely 2 disagree somewhat 3 neither agree nor disagree 4 agree somewhat and 5 agree completely). Instead of ability of students, the model is generalized to the location of a person or other unit of analysis on a latent trait similar to ability in the sense that it cannot be directly measured. In this paper, instead of the ability level of a student, we seek to model the level of SBM implementation in a school. In place of a battery of tests of differing level of ability, we administer a series of items to school principals regarding the implementation of SBM at their school, and ask them to provide a rating on a Likert scale about how much they agree about the item in reference to their school. Formally, the Rating Scale Model (RSM) is as follows:

$$P(X_{vi} = h|\theta_v, \beta_i, \omega_h) = \frac{\exp[h(\theta_v - \beta_i) + \omega_h]}{\sum_{l=0}^m \exp[l(\theta_v - \beta_i) + \omega_l]} \quad \dots[2]$$

In this model there are h graded response categories each category is an observed threshold of an underlying continuous variable that is latent or unobserved. In the model of Equation [2], in addition to the person parameters θ_v and the item parameters, β_i , there are $m+1$ category parameters $\omega_0, \omega_1, \dots, \omega_m$. Probabilities are modeled successively and cumulatively – meaning the probability of grading a ‘2’ is modeled against the probability of not having graded a ‘1’; the probability of modeling a ‘3’ is modeled against the probability of not having graded a ‘2’ and so on. Maximum Likelihood estimation is used to generate estimates of the parameters. The data matrix of persons and items is ordered by raw score and information from the order and value of the raw scores is sufficient to generate estimates of the model parameters, subject to some key assumptions.

Uni-dimensionality: Rasch modeling is based on the assumption of the existence of a single dimension that is to be measured – for example the ability level of students or the level of SBM implementation. In formal terms, we can say that equations [1] and [2] above are characterized completely – the probability of correct answer or of rating ‘3’ instead of ‘2’ on a Likert scale depends solely on the modeled parameters.

Local conditional independence: This is a corollary of uni-dimensionality. Conditional on the single dimension θ , the probability of a correct answer or a particular rating on one item is independent of the probability of a correct answer or a rating on another item. If there were a correlation between two items, it would mean that there is another dimension other than θ that connects the two items.

Sufficiency: The raw composite score (the number of times an item has been solved in the case of the dichotomous model or the sum of the category score times the category number) is all that is required to identify the item parameter β . The converse holds for the subject parameter, θ . That is, it does not matter which items have got which score for the subject and it does not matter which subject scored a particular item.

Monotonicity: This assumption holds that the response probability values increase with θ . In the case of student assessment, a high ability student who correctly answers an item of a given level of difficulty will have a higher probability of answering a question of a lower level of difficulty. If you get the difficult question right, you are unlikely to give a wrong answer to a much easier question.

C. Data

Two datasets are used in this paper – a census level panel dataset of all public schools in the Philippines, and a survey sample of 150 schools that was chosen from the census data.

BEIS-NAT matched panel database (2005-2010): The dataset contains approximately 38,000 public elementary schools and 7,200 public high schools. The dataset was generated by merging school level Basic Education Information System (BEIS) administrative data with National Achievement Test (NAT) data that includes school level mean scores on standardized student tests for grades 5 and 9. Data was matched across school years to generate a panel dataset. A separate set of working papers (Yamauchi, 2012; Yamauchi and Parandekar, 2013) present analysis of the BEIS-NAT data to examine the impact of SBM and to analyze patterns of inequality regarding results and resources. A detailed field survey of 150 schools was also conducted to investigate SBM related issues.

Three division field survey: The three divisions of Pangasinan II, Bohol and Surigao Del Sur were chosen as the sample divisions for the field survey. These were chosen to represent the three island groups of Luzon, Visayas and Mindanao respectively. The divisions were also selected because they represent varied experiences with SBM implementation. Surigao Del Sur schools have a longer experience with school level planning and management as part of the Third Elementary Education Project (TEEP) which started in Surigao Del Sur, along with in 22 other divisions, in 2002. Bohol has had the benefit of experience from the AusAID financed Strengthening Implementation of Visayas Education (STRIVE) which introduced school level planning and SBM related data management systems in the division. Pangasinan II, on the other hand, has not had the experience of any externally aided project, and SBM tools such as School Improvement Plans were introduced in most schools as late as 2009/2010.

Stratified random sampling: All schools in the three divisions were distributed across strata formed by three key variables and to select randomly within the strata. The variables were the level of implementation of SBM in a school as determined by an SBM index; the NAT score performance across the past 6 years computed as a NAT index and the school size (enrollment). The SBM index and NAT index was used to generate a matrix of schools were randomly selected from each of the four corner cells and the middle cell of the matrix. The purpose of this stratification was to ensure *a priori* that the sample had a sufficient variation in SBM implementation and NAT score performance and that there was adequate representation of schools of different sizes.

Respondent groups: Within each sampled school, detailed interviews were carried out with the school principal, the chairperson of the school governing council (usually a parent), and a teacher randomly selected from the roster of teachers in the school. The same set of questions regarding SBM was implemented to the three respondent groups.

SBM Rating Inventory (SRI): This paper presents a first approach to develop a rigorous benchmarking measure for SBM – the method presented can be applied across different contexts to further refine and arrive at a refined inventory that meets the criteria of reliability and validity. The SBM inventory used here contains 38 items. The 38 cards were shuffled and the respondent was asked to rate the variable mentioned on the card for *importance* to school quality on a 7 point Likert scale from 1 – extremely unimportant to 7 ‘extremely unimportant’. After completing the rating on *importance* for all items, the whole exercise was repeated with reshuffled cards, except the respondents were now asked to provide the rating for each card on a 10 point scale regarding the *implementation* of each factor in the school from 1 ‘extremely poorly implemented’ to 10 ‘extremely well implemented.’⁴

TABLE 3: SBM Conceptual Dimensions	
Variable Name	Conceptual dimension and variables within each dimension
	PARTICIPATION
PAR_1	- Closer integration of school with local community
PAR_2	- Parent participation in purchases made with the school budget
PAR_3	- Participation from alumni association and barangay or local NGOs in organizing events and celebrations
PAR_4	- Parent participation in organizing events and celebrations
PAR_5	- Parents help to make sure that children study at home and come to school on time
	RESOURCES
RES_1	- Funds raised by the parents to supplement government budget
RES_2	- School has adequate financial resources to purchase educational material
RES_3	- Adequate resources to pay for MOOE other than educational material
RES_4	- Higher level of school budget
	TEACHING PRACTICES
TCH_1	- Teachers are able to get the training they need to upgrade their skills
TCH_2	- Teachers are encouraged to innovate new teaching methods
TCH_3	- Teachers have a say in pedagogy related decisions at school
TCH_4	- Teachers provide timely and periodic feedback to parents about their child’s school performance
TCH_5	- Teachers have special meetings amongst themselves to discuss pedagogical issues
	AUTONOMY
AUT_1	- School can upgrade building or facilities as required
AUT_2	- School undertakes own school-based procurement for school needs
AUT_3	- Support from local government with financial matters
AUT_4	- Parent participation in decision about allocation of school budget
AUT_5	- Parent participation in evaluation of teacher performance
AUT_6	- School is able to secure additional number of teachers when required
	SCHOOL IMPROVEMENT PROCESS
SIP_1	- More effective school administration
SIP_2	- Goals in the School Improvement Plan are realistic and attainable
SIP_3	- School records are kept in perfect order

⁴ We borrow this technique from the vast literature on customer satisfaction from the discipline of marketing management. In this literature, consumers rate products on the importance of a set of product features or service quality dimensions and they also rate their satisfaction or their perception of performance of the product on those features or dimensions (Martilla and James, 1977).

SIP_4	- Students have extra-curricular opportunities like sports, music and dance
SIP_5	- School grounds and toilets are always clean
PRINCIPAL LEADERSHIP	
PRI_1	- School head provides pedagogical feedback to teachers
PRI_2	- School head has harmonious relationships with local authorities
STUDENT LEARNING	
LRG_1	- Deeper attention to student learning and other student outcomes
LRG_2	- All stakeholders are aware of NAT results
LRG_3	- Teachers hold preparatory classes to help students do well in the NAT
LRG_4	- Teachers maintain portfolios of each students achievements
LRG_5	- School does well in regional competitions like Math Olympics
GOVERNANCE	
GOV_1	- More DepED control or oversight of school functioning
GOV_2	- DepED division and district supervisors visit school frequently
GOV_3	- School receives technical assistance as needed from district and division officers
EQUITY AND INCLUSIVENESS	
EQI_1	- Care better for students with learning difficulties or special needs
EQI_2	- Remedial classes are held for weaker students
EQI_3	- School feeding program ensures that no children are hungry

Reducing Response Bias: All instruments were pre-tested and refined before being deployed in the field. The survey was implemented by an experienced survey research firm with a rigorous protocol in place for training of field investigators and for quality control on the process of data collection and compilation. The same field interviewer conducted the interviews with the different respondents at a school in order to avoid introducing bias due to differential interviewer effects. The survey was conducted over a 3 week period at the beginning of March 2012, before schools began their final examination period and were then released for summer break – it was thus a neutral time period for the school. For purpose of recording the ratings the back of each card described in Table 3 had a specific code that was generated so as to be value neutral – ‘KSE4’, ‘LTM8’, and ‘SFK7’ and so on. The nine classifying dimensions are mentioned in Table 3 for explanation, the cards provided to respondents did not provide any classification information. The shuffling of cards before each interview and the coding system used ensured that each respondent would get a different order of presentation of questions – it was important to eliminate position order bias for the items and also to help the respondents think independently about importance and implementation.

Figure 1 depicts an analysis of the importance and implementation ratings by school principal (standardized across ratings for each individual, variable names are defined in Table 3). It is useful to examine the preponderance of *Participation* and *Autonomy* variables in Quadrant III in Figure 1 - indicating low importance and low implementation - a finding that indicates that perhaps the principals are not so convinced about the core SBM message. However, *School Implementation Plan* variables appear to be clustered in Quadrant I, which also includes a number of variables related to *Teaching* and *Learning* practices, which bodes well. A visual analysis as in Figure 1 may reveal interesting hypotheses to test further and is relatively simple to understand. Rasch measurement provides a more rigorous way to analyze the same data on importance and implementation.

across all items. ‘Difficulty’ here means it is difficult to agree on the implementation or the importance ratings. Table 4 indicates that principals found it very difficult to agree the statement that RES_2 “School finances adequate for educational material” is implemented well. Just like only the students of highest ability can correctly answer the most difficult test questions, only the high SBM level schools would be the ones with high ratings on difficult items. In the right hand panel of Table 4, participation and autonomy variables appear at the top, with positive behavior of the principals and teachers appearing at the bottom – meaning these are the easiest variable to agree with.

Table 4: SBM Ratings by Principals: Rasch Analysis of 38 items			
IMPLEMENTATION	$\hat{\beta}$	IMPORTANCE	$\hat{\beta}$
RES_2 School finances adequate for educational material	1.01	AUT_5 Parental participation in teacher performance evaluation	1.78
TCH_5 Teachers have special pedagogy related meetings*	0.88	PAR_2 Parent participation in purchasing decisions	1.59
LRG_5 School does well in regional scholastic competitions*	0.85	PAR_3 Participation from local community associations*	1.14
TCH_1 Teachers are able to get training they need*	0.65	AUT_4 Parental participation in school budget allocation	1.12
AUT_1 School can upgrade buildings or facility as needed*	0.64	GOV_1 More central control/oversight of school functioning*	0.97
RES_4 Higher level of school budget*	0.62	LRG_5 School does well in regional scholastic competitions*	0.93
PAR_2 Parent participation in purchasing decisions*	0.61	GOV_2 Central authorities visit school frequently*	0.85
PAR_3 Participation from local community associations*	0.55	RES_1 Funds raised by parents to supplement govt. budget*	0.7
TCH_4 Teachers provide regular feedback to parents*	0.51	SIP_4 Extra-curricular activities like sports and music*	0.67
RES_3 School finances adequate for other operational exp*	0.34	PAR_4 Participation from Parents in event organization*	0.59
GOV_2 Central authorities visit school frequently*	0.32	AUT_1 School can upgrade buildings or facility as needed*	0.44
TCH_3 Teachers have a say in pedagogical decisions	0.31	GOV_3 School receives technical assistance from central level*	0.32
EQI_1 Better care for students with special needs/difficulties	0.23	AUT_6 School can get additional teachers when required*	0.21
PAR_5 Parent responsibility for child effort/attendance*	0.22	TCH_5 Teachers have special pedagogy related meetings*	0.19
GOV_1 More central control/oversight of school functioning*	0.16	EQI_1 Better care for students with special needs/difficulties*	0.17
EQI_3 School feeding programs ensure no children are hungry	0.15	TCH_3 Teachers have a say in pedagogical decisions*	0.14
RES_1 Funds raised by parents to supplement govt. budget	0.09	EQI_3 School feeding programs *	0.14
GOV_3 School receives technical assistance from central level*	0.09	AUT_3 Financial support from local government*	0.07
PAR_4 Participation from Parents in event organization	-0.03	AUT_2 School undertakes school based procurement*	0.05
AUT_5 Parental participation in teacher performance eval.	-0.03	PAR_1 Closer integration with school community*	0.05
SIP_3 School records are kept in perfect order*	-0.09	RES_3 School finances adequate for other operational exp*	0.03
LRG_2 All stakeholders are aware of NAT results*	-0.11	LRG_2 All stakeholders are aware of NAT results*	-0.08
TCH_2 Teachers are encouraged to use innovative methods	-0.19	RES_4 Higher level of school budget	-0.1
SIP_5 School grounds and toilets are always clean*	-0.21	PRI_1 School head provides pedagogical feedback to teachers*	-0.16
PRI_2 School head harmonious relations with local authorities*	-0.23	RES_2 School finances adequate for educational material	-0.18
AUT_4 Parental participation in school budget allocation*	-0.27	SIP_5 School grounds and toilets are always clean*	-0.51
AUT_3 Financial support from local government*	-0.35	LRG_4 Teachers have portfolios of student's achievements*	-0.51
AUT_2 School undertakes school based procurement*	-0.39	SIP_3 School records are kept in perfect order*	-0.73
SIP_1 More effective school administration*	-0.39	LRG_3 Teachers hold preparatory classes for NAT performance	-0.79
PAR_1 Closer integration with school community*	-0.41	EQI_2 Remedial classes are held for weaker students*	-0.81
LRG_4 Teachers have portfolios of student's achievements*	-0.47	SIP_1 More effective school administration	-0.93
AUT_6 School can get additional teachers when required*	-0.48	SIP_2 Goals in the School Improvement Plan are realistic*	-0.93
EQI_2 Remedial classes are held for weaker students*	-0.57	PAR_5 Parent responsibility for child effort/attendance*	-0.96
LRG_1 Deeper attention to student learning and other results*	-0.58	TCH_2 Teachers are encouraged to use innovative methods*	-0.99
SIP_4 Extra-curricular activities like sports and music*	-0.62	TCH_1 Teachers are able to get training they need*	-1.02
PRI_1 School head provides pedagogical feedback to teachers*	-0.83	TCH_4 Teachers provide regular feedback to parents*	-1.08
LRG_3 Teachers hold preparatory NAT classes *	-0.88	PRI_2 School head harmonious local relations *	-1.14
SIP_2 Goals in the School Improvement Plan are realistic*	-1.12	LRG_1 Deeper attention to student learning and other results*	-1.24

Table 4 difficulty measures are useful to intuitively understand the construction and the usefulness of Rasch measures of SBM. Some of the items Table 4 have been marked with an asterisk – based on the model infit mean square number generated from the Rasch analysis. This number is the ratio of the actual variance in ratings as compared to the Rasch model predicted variance in ratings, with a

weighting scheme that assigns low weights to outliers (see Smith, et al., 2008). The literature recommends that items that have IN.MSQ greater than 1.5 be rejected as being too noisy (Linacre, 2013). In this paper, to be more conservative, we remove items with IN.MSQ greater than 1.20 – 8 items are eliminated from the implementation model and 7 from the importance model. We then recomputed the ‘person measure’ or the $\hat{\theta}$ values that we interpret as an SBM measure.

This SBM measure is methodologically superior to simply taking the numerical mean of the Likert scores. An indication of this methodological superiority can be obtained by comparing the Rasch SBM measure with the simple arithmetical measure of the total of the Likert scores - adding together the Likert scores across the items for each principal with an implied equal weight of 1 for each item. The Rasch method recognizes that not all items in a rating scale inventory have equal weight and discriminating power. Figure 2 below shows schools lined up in a sorting order or ‘parade of dwarves’ from the highest total Likert measures school to the lowest total measure. However, each blue bar represents the school’s Rasch SBM measure. If the differential weighting (according to ‘difficulty’ of items) would have made no difference, Figure 2 would have been smooth with no spikes and the rank order would have been preserved. Figure 2 thus shows how Rasch analysis incorporates the valuable information that all items are not the same.

Rasch analysis helps us to answer a very simple but important and often overlooked question – does there actually exist something called “SBM level” that can be used as a benchmark to compare across schools and to monitor performance change over time – validity and reliability. While reliability testing would require a repeated application, we have adequate data to test for validity. The method is to test for uni-dimensionality of the Rasch index.

Table 5 present four tests that shed light on the validity of Rasch measures, which are applied to the six applications of Rasch measures of SBM available in our study. Column (A) indicates that for each of the six applications, not more than 2 of the 38 items exceeded the threshold of 1.5 for *IN.MNSQ* or ‘in-fit mean square error’ which indicates a sign of problematic model misfit (violation of Rasch assumptions). Column (B) shows that the *item separation* values fall below the minimum cut-off of 3 in 3 of the six applications – which means that item separation could have been better if the items had been better designed. Column (C) indicates that *item reliability* falls below 0.9 in the case of 1 out of the 6 applications. Column (D) in Table 5 shows the result from the principal component analysis (PCA) of the residuals from the Rasch analysis (observed score – expected score). If the uni-dimensional Rasch model is valid, the residuals would not show a structure indicative of more dimensions in the data. A heuristic proposed by Linacre, 2004 is that the first eigenvalue from PCA of the residuals should not exceed 4 – Column (D) shows that five of the six measures meet Linacre’s criteria.

Table 5: Tests of validity of Rasch SBM Measure						
Group Number	Respondent Group	SBM question asked	(A) # of items with IN.MSQ > 1.5	(B) Item Separation	(C) Item Reliability	(D) Residuals vector 1 st eigenvalue
1	Principal	Implementation	0	6.65	0.98	2.9
2		Importance	1	2.50	0.86	2.5
3	SGC Head	Implementation	0	3.27	0.91	3.6
4		Importance	2	2.93	0.90	2.7
5	Teacher	Implementation	1	5.25	0.97	4.3
6		Importance	1	2.80	0.96	2.8

Without resorting to the characteristic matrix decomposition of the residuals, a simple scatter plot of the residuals against base values of the Rasch index also indicated the absence of any structure in the data. The four measures shown in Table 5 give us reasonable confidence regarding the existence of a single dimension that we can call “SBM level” in our sample of schools. Whereas the traditional method using arithmetic score computations simply assumes the existence of an “SBM” level, the use of Rasch methodology allows us to test for the validity in a rigorous way.

Analysis item-by-item: There is a straightforward extension of the ‘person-item’ map familiar to readers from the application of Rasch model to student assessment. In our case the ‘person-item’

map of Figure 3 indicates the comparison between SBM implementation levels at schools and item 'difficulty'.

The low level of overlap between the ability and difficulty distributions in Figure 3 indicates the need for improvement in the original instrument design. For the Rasch analysis to work well there should be a judicious mix of difficult and easy items that together are able to discriminate between schools of different SBM levels. Ideally, Figure 3 would have shown a symmetric or mirrored set of curves of ability and difficulty. A bulk of the items would be of moderate level of difficulty for most people; a few of the items would be very difficult and get high ratings only from some schools; and a few of the items would be so easy that very few schools would not rate it highly. In the actual case, the item distribution is shifted toward the bottom – overall the items are too easy, and there were not adequate number of items in the instrument to which only few respondents gave high ratings. Future research should attempt to redress this shortcoming through preliminary Rasch analysis of pilot test data and examine the profiles of each item.

Item characteristic curves: The simplest way to investigate properties of an item is to plot an empirical ‘item characteristic curve’ (ICC) and compare it with a theoretical ICC. In the simplest case, a person of average ability would answer an item of median difficulty correctly 50% of the time (tend to agree more in the case of principals surveyed here). An item would have graded level of difficulty – so that some very difficult items, a person of average ability would answer correctly only 20% of the time, while the person at the 20th percentile ability rank would be able to answer the question correctly 80% of the time. In a basic theoretical model, we can say that schools with an average level of SBM would get a score of 5 – this is equivalent to getting the answer right 50% if the time. Schools with higher level of SBM would get higher rating – say 7 or 8. We examine the empirical item characteristic curve (or ICC) for some items in Figure 4.

The ICC plots provide diagnostic information about the difficulty distribution of items which can be used to generate an ‘item bank’ for SBM in the same way as there are question banks for student assessment. Figure 4 (a) takes the two items at the extreme ends of the difficulty distribution and plots their ICCs. The solid curve without markers is the theoretical symmetric distribution, centered on a mean of zero being a rating of 5 and following an S shape to capture a normal ability distribution. The blue line shows the ICC for RES_2 ‘School has adequate financial resources to purchase educational material’ and it can be seen that it extends lower along the theoretical S-curve, to a nearly -1 level of difficulty. The red curve showing PRI_2 ‘School head has harmonious relationships with local authorities’ is only seen at the right side, it virtually has no discriminating power – it appears that all school principals in the Philippines have great relationships ! Figure 4 (b) shows the ICCs for all ‘autonomy’ items – this indicates a jumble of items with ICCs crossing one another and all concentrated on the right hand side. To the extent possible, it would be useful to see such a family of curves running parallel, with a left to right displacement showing easier items.

This section has shown that rigorous analysis of SBM implementation is possible – rather than relying on arbitrary indices, indices can be constructed which meet objective criteria regarding validity and reliability. However, in order for SBM related literature to adopt the suggested measurement technique, more empirical studies in different contexts need to be carried out and the process of item construction build successively on lessons learnt from earlier studies. The measurement of personality traits and political skills are two examples of fields where a large literature has emerged, and we turn next to the application of BFI and PSI to the same sample of 150 schools from the Philippines.

E. Principal Personality and Political Skills and SBM

This section presents empirical evidence using the SBM measure generated from the previous section to test hypotheses regarding the personality types and political skills of the principals. As outlined in Section B, the BFI and PSI instruments have become accepted in the literature. A similar research agenda is possible with regard to SBM as well which would enable for checking the presence or absence of the relationship between SBM and the other constructs. This section sketches the application in context of the sample of 150 schools in the Philippines. The issue of impact of SBM on school performance requires a much larger longitudinal sample size and a dynamic context that allows testing against a counter-factual – presented in another working paper related to this study (Yamauchi, 2012).

The hypotheses here are drawn from the huge literature on leadership and personality. In this paper, we posit that the principals of schools which have higher level of SBM implementation are effective in exercising leadership of this important reform effort and we can derive some insights from empirical analysis. In this section, there is no attempt being made to draw a causal link from the correlation – it is possible that principals possessing certain traits were early adopters of SBM and the correlations merely represent a self-selection of principals. It should be noted that at the time of the survey data collection, all the schools in the Philippines had already been converted to adopt the SBM model, so any self-selection would have occurred only with respect to the timing of the SBM in a school.

H1: Extraversion is positively associated with SBM implementation. Extraversion is the tendency to be outgoing and possess positive energy that is used in social situations to exercise assertiveness. Extraverted individuals tend to be expressive and talk a lot. This is the one BFI trait that is fairly consistent in the literature as a positive marker for leadership.

H2: Agreeableness and SBM implementation is not be correlated. Agreeable individuals are described as ‘nice’- they are compliant and passive and they like to get along with others. On the positive side, agreeable individuals display empathy and in a collegial school atmosphere it is possible that trusting and nurturing would work in a principal’s favor. On the negative side, agreeableness tends to be a follower rather than a leader’s trait – a leader needs to set a direction and bring about change.

H3: Conscientiousness and SBM implementation are positively correlated. The literature exploring BFI and managerial performance including entrepreneurship and BFI and academic achievement indicates a high correlation with conscientiousness. Conscientious individuals are hard-working and persevering in their effort for results. For implementation of SBM, even though stated claims from the principals could be a biased positive self-image for some individuals, hard-working principals would probably be more commonly found in high SBM schools.

H4: Neuroticism and SBM implementation are negatively correlated. Neurotic individuals are low on self-esteem and self-confidence. Neuroticism is associated with individuals who display a nervous disposition and react adversely to stress and crisis situations. Since unforeseen problems are the order of the day in many school contexts, principals who score high on neuroticism would probably not be able to exercise the motivation and leadership required for a school to get to a high SBM level.

H5: Openness and SBM implementation are positively correlated. Openness implies a curiosity to learn new facts, to be able to change one's mind when required and to identify patterns and insights that indicate creativity. SBM implementation requires innovation and thinking of alternatives and solutions on the spot, which would be helped if principals possess this trait.

H6: Political skills, including social astuteness, networking ability, interpersonal ability and apparent sincerity are all associated positively with SBM implementation. SBM implementation requires the principal to carry out a large and complex set of tasks in addition to the regular administration of the school. Political skills are required for identifying and harnessing resources, motivating stakeholders and solving management and pedagogical problems.

To explore the hypotheses mentioned here, we ranked the schools in order of the SBM measure, and classified the schools into 3 groups of high SBM, medium SBM and low SBM – we focus attention on 50 schools in the “high” SBM group and 51 schools in the “low” SBM group. The BFI and PSI scores are calculated according to definition of the two measures described in the annex. The BFI and PSI scores are further standardized across all individuals in the sample and expressed in standard deviation units. A score of 0.41 on “extraversion” for a person would mean that given the mean of zero for all individuals on extraversion, the subject individual received a score higher on extraversion as compared to the sample mean, by 0.41 standard deviation units. In this way the differentials can be compared. Table 6 shows the mean values of the constructs across the two groups. The average score for Extraversion was 0.09 for the high SBM group, and it was -0.16 for the low SBM group – this means a difference of 0.25 standard deviations across the groups – confirming H1 above.

Table 6: Personality Traits and Political Skills by SBM Performance groups (measured in standard deviation units by person)		
BFI or PSI construct	High SBM Performance	Low SBM Performance
Extraversion	0.09	-0.16
Agreeableness	0.10	-0.27
Conscientiousness	0.31	-0.31
Neuroticism	-0.10	0.16
Openness	0.28	-0.30
Social Astuteness	0.51	-0.49
Interpersonal Influence	0.17	-0.20
Networking Ability	0.30	-0.38
Apparent Sincerity	0.34	-0.51

Four of the five hypotheses regarding BFI are supported by the data in Table 6. As agreeableness is seen to be positively linked with SBM implementation, H2 is not supported. It is probable that unlike the cases studied in most of the leadership and BFI literature, in the case of SBM, it does help for principals to be agreeable. Principals who described themselves as trusting, cooperative, considerate and less apt to quarrel and find fault, perhaps have greater persuasive power in their relationships with community. More research would be required to understand the issue in depth, but it is possible that agreeableness as a personality trait is associated with more democratic behavior by school principals, and that such democratic behavior in turns elicits superior school team performance that is inherent in the SBM rating as constructed here. It is interesting to note that the most important differentiator in BFI traits are conscientiousness and that extraversion is the least important. The finding that openness is a strong discriminator is interesting.

Separately from the analysis presented in Table 6, we analyzed the correlation between the BFI measures and the amount of funding per student available to schools. Agreeableness, openness and extraversion are positively related to schools obtaining higher level of funds per student. On the other hand school heads who rated themselves relatively higher on the neuroticism scale, thus describing themselves as nervous, moody, tense, worriers who don't handle stress well and get depressed easily were in schools which obtained fewer per student funds. While these results cannot confirm causality, they do indicate that personality types of school heads can have a significant bearing on resources available to the school. Higher personnel funds per student can be obtained because of more number of teachers or better qualified teachers of a combination of these two factors. It is not clear why more "open" school heads would receive these higher resources, perhaps due to better communication skills but further research would be needed to understand the channels of causality.

Table 6 shows that all four dimensions of political skills are important for implementation of SBM. The most important of the PSI dimensions are social astuteness and apparent sincerity. Ferris et al., 2005 define individuals with high social astuteness as those who are "astute observers of others and are keenly attuned to diverse social situations. They comprehend social interactions and accurately interpret their behavior, as well as that of others, in social settings." Apparent Sincerity is exhibited by politically skilled individuals who "are or appear to be, honest, open, and forthright." Apparent sincerity "focuses on the perceived intentions of the behavior exhibited." Apparent sincerity is a very important political skill that is sometimes underrated by individuals who need it most. It is very important even for the most genuine and honest and hardworking individuals to be perceived as honest – if followers do not believe a leader, even if a leader is genuine, his or her efforts may not be fruitful. The word 'apparent' here is instrumental – though it can be taken to be somewhat cynical, because leaders may appear to be sincere who are not really so. If the political skill literature is to be believed, what matters is the perception of sincerity, more than merely the reality of it.

F. Conclusions

The objective of this paper is to provide an application of Rasch modeling to the context of measurement of a policy - School Based Management or SBM. It is widely accepted in the literature that SBM requires a number of years to eventually find appreciable impact in terms of outcomes such as student achievement. However, many stakeholders are interested in knowing the progress of the implementation of the policy itself – is SBM being implemented well in the schools? Two instruments were mentioned in the background section of this paper – the SABER framework for SBM introduced by the World Bank as part of a multi-country benchmarking study for SBM, and an assessment of SBM implementation in the Philippines. It was proposed that by applying Rasch modeling it is possible to bring a semblance of scientific validity and reliability to measurement of SBM.

The paper applied the Rasch modeling to an instrument applied in a sample survey of schools in the Philippines. The instrument is far from perfect, but the Rasch modeling shows that it is feasible to develop an instrument that is not completely subjective. Developing good quality instruments is an iterative and collaborative process, as shown by the examples of the BFI and PSI instruments. While BFI and PSI have used factor analysis to uncover structure in the data, the purpose is similar to this paper's use of Rasch analysis – to come up with a valid and reliable instrument that can be applied in differing contexts. The presence of interesting conclusions from the analysis presented in this paper indicates that indeed the resulting SBM measure can be used in a conceptually sound manner. The first conclusion of this paper is that further development of SABER instruments for benchmarking educational policy in SBM and other areas might use Rasch modeling to help improve the selection of items. Rasch analysis can be applied to help revise and reformulate an initial list of items in an iterative process so that eventually a good 'person-item' overlap can be developed for a rigorously defined measure.

For the superiority of a formal method such as Rasch modeling to be accepted, more SBM studies would need to deploy similar survey instruments including a common list of items like the BFI-44 or PSI-18. While there are country specific issues that may require unique items to be developed for an SBM study in a country, even such questions can be standardized from a template, and there will always be a relevance of generic measures of accountability and participation. Rasch analysis applied from an exploratory stage and used in subsequent iterations by different authors in varying contexts can lead to the development of an instrument that will have construct validity and reliability. A second conclusion of the paper is an exhortation for SBM researchers to use the Rasch method to work towards the creation of a common index of an SBM-n (e.g. SBM-38), similar to the BFI-44 or PSI-18 measures. Even beyond SBM, the Rasch method can be applied to other areas of benchmarking public policy where measurement is bedeviled by subjectivity. The development of such instruments will help to develop greater confidence in policy analysis and conclusion regarding policies where the eventual results take 8 to 10 years to develop. It is hoped that other researchers might be sufficiently engaged by this paper to explore the application of Rasch analysis to policy benchmarking.

Annex Table 1: Big Five Inventory

Rating 1 “mostly like me” to 5 “mostly not like me” in describing characteristics about self.

	Rating		Rating
1. Is talkative	----	23. Tends to be lazy	----
2. Tends to find fault with others	----	24. Is emotionally stable, not easily upset	----
3. Does a thorough job	----	25. Is inventive	----
4. Is depressed, blue	----	26. Has an assertive personality	----
5. Is original, comes up with new ideas	----	27. Can be cold and aloof	----
6. Is reserved	----	28. Perseveres until the task is finished	----
7. Is helpful and unselfish with others	----	29. Can be moody	----
8. Can be somewhat careless	----	30. Values artistic, aesthetic experiences	----
9. Is relaxed, handles stress well	----	31. Is sometimes shy, inhibited	----
10. Is curious about many different things	----	32. Is considerate and kind to almost everyone	----
11. Is full of energy	----	33. Does things efficiently	----
12. Starts quarrels with others	----	34. Remains calm in tense situations	----
13. Is a reliable worker	----	35. Prefers work that is routine	----
14. Can be tense	----	36. Is outgoing, sociable	----
15. Is ingenious, a deep thinker	----	37. Is sometimes rude to others	----
16. Generates a lot of enthusiasm	----	38. Makes plans and follows through with them	----
17. Has a forgiving nature	----	39. Gets nervous easily	----
18. Tends to be disorganized	----	40. Likes to reflect, play with ideas	----
19. Worries a lot	----	41. Has few artistic interests	----
20. Has an active imagination	----	42. Likes to cooperate with others	----
21. Tends to be quiet	----	43. Is easily distracted	----
22. Is generally trusting	----	44. Is sophisticated in art, music, or literature	----

BFI scale scoring (“R” denotes reverse-scored items)

Extraversion: 1, 6R, 11, 16, 21R, 26, 31R, 36

Agreeableness: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42

Conscientiousness: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R

Neuroticism: 4, 9R, 14, 19, 24R, 29, 34R, 39

Openness: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

Source: John, Naumann and Soto, 2008

Annex Table 2: Political Skills Inventory

Rating 1 Strongly Agree to 7 Strongly Disagree

	Dimension of political skill		Dimension of political skill
1. I always seem to instinctively know the right thing to say or do to influence others	NETWORKING ABILITY	10. I spend a lot of time and effort work networking with others	NETWORKING ABILITY
2. I have good intuition or ‘savvy’ about how to present myself to others	INTER-PERSONAL INFLUENCE	11. At work, I know a lot of important people and am well connected	NETWORKING ABILITY
3. I am particularly good at sensing the motivations and hidden agendas of others	INTER-PERSONAL INFLUENCE	12. I am good at using my connections and networks to make things happen at work	APPARENT SINCERITY
4. I pay close attention to people’s facial expressions	INTER-PERSONAL INFLUENCE INTER-PERSONAL INFLUENCE	13. I have developed a large network of colleagues and associates at work who I can call on for support when I really need to get things done	APPARENT SINCERITY
5. I understand people very well	SOCIAL ASTUTENESS	14. I spend a lot of time at work developing connections with others	APPARENT SINCERITY
6. It is easy for me to develop good rapport with most people	NETWORKING ABILITY	15. I am good at building relationships with influential people at work	NETWORKING ABILITY
7. I am able to make most people feel comfortable and at ease around me	SOCIAL ASTUTENESS	16. It is important that people believe I am sincere in what I say and do	SOCIAL ASTUTENESS
8. I am able to communicate easily and effectively with others	APPARENT SINCERITY	17. I try to show a genuine interest in other people	SOCIAL ASTUTENESS
9. I am good at getting people to like me	NETWORKING ABILITY	18. When communicating with others, I try to be genuine in what I say and do	SOCIAL ASTUTENESS

Scores for each dimension are the arithmetic means of ratings for each variable that is included in the dimension.

Source: Ferris, et al, 2005

References

- Andrich, David, 1988: *Rasch Models for Measurement*, Newbury Park: Sage Publications
- Arcia, Gustavo, K. Macdonald, H. Patrinos and E. Porta, 2011: *School Autonomy and Accountability*. Washington, D.C.: The World Bank.
- Bakker, Arnold B., K. I. Van der Zee, K.A. Lewig, and M. F. Dollard, 2006: "The Relationship between the Big Five Personality Factors and Burnout: A Study among Volunteer Counselors." *Journal of Social Psychology*, 146 (1):31-50.
- Barrera-Osorio, Felipe, Tazeen Fasih, and Harry Anthony Patrinos, 2009: *Decentralized Decision-Making in Schools: The Theory and Evidence on School Based Management*. Washington, D.C.: The World Bank.
- Benoliel, Pascale and A. Somech, 2010: "Who benefits from participative management?" *Journal of Educational Administration*, 48 (3):285-308.
- Bono, Joyce E., and Timothy A. Judge, 2004: "Personality and Transformational and Transactional Leadership: A Meta-Analysis." *Journal of Applied Psychology*, 89 (5):901-910.
- Bing, M., K. H. Davison, I. Minor, M. Novicevic and D. Frink, 2011: "The prediction of task and contextual performance by political skill: a meta-analysis and moderator test." *Journal of Vocational Behavior*, 79, 563-577.
- Blickle, Gerhard, J. A. Meurs, I. Zettler, J. Solga, D.Noethen, J. Kramer, and G. R. Ferris, 2008. "Personality, Political Skill and Job Performance." *Journal of Vocational Behavior* 72, pp. 377-387.
- Bond, Trevor G. and C. M. Fox, 2007: *Fundamental Measurement in the Human Sciences*. 2nd Edition. Toledo: University of Toledo.
- Briggs, Kerri L. and Priscilla Wohlstetter, 2003: "Key Elements of a Successful School-Based Management Strategy", *School Effectiveness and School Improvement*, 14(3):351-372.
- Brosky, Donald, 2011: "Micro-politics in the School: Teacher Leaders' use of Political Skill and Influence Tactics." *International Journal of Educational Leadership Preparation*, 6(1)ISSN 2155-9635.
- Bruns, B., D. Filmer, and H. Patrinos, 2011: *Making Schools Work: New Evidence on Accountability Reforms*. Washington, D.C.: The World Bank.
- Caliendo, Marco, F. Fossen and A. Kritikos, 2011: "Personality Characteristics and the Decision to Become and Stay Self-Employed." *IZA Discussion Paper Series*, Number 5566.

- Department of Education, 2009: *A Manual of the Assessment of School-Based Management Practices*. Manila: Department of Education.
- Coon, Deniss and J. O. Mitterer, 2010: *Introduction to Psychology: Gateways to Mind and Behavior*. 12th edition. Belmont: Wadsworth Cengage.
- Ferris, Gerald R., D. C. Treadway, R. W. Kolodinsky, W. A. Hochwater, C. J. Kacmar, C. Douglas, and D. D. Frink, 2005: "Development and Validation of the Political Skill Inventory." *Journal of Management*, 31(1):126-152.
- Goldberg, Lewis R., 1990: "An Alternative Description of Personality: The Big-Five Factor Structure." *Journal of Personality and Social Psychology*, 59(6):1216-1229.
- Hurtz, Gregory M., and J.J. Donovan, 2000: "Personality and Job Performance: The Big Five Revisited." *Journal of Applied Psychology*, 85(6):869-879.
- John, Oliver P., L. P. Naumann and C. J. Soto, 2008: "Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurement, and Conceptual Issues." *Handbook of Personality: Theory and Research*, ed. John, R. W. Robins, & L. A. Pervin, pp. 114-158. New York, NY: Guilford Press.
- Judge, Timothy A., J.E. Bono, R. Ilies and MW. Gerhardt, 2002: "Personality and Leadership: A Qualitative and Quantitative Review". *Journal of Applied Psychology*, 87(4):765-780.
- Linacre, J. M., 2013: *Winsteps® Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com
- Lvina, Elena, G. Johns and T. Bobrova, 2009: "Cross-Cultural Generalizability of the Political Skill Construct: A Validation of the PSI in Russian." *Working Paper, Concordia University*.
- Martilla, John A. and J. C. James, 1977: "Importance-Performance Analysis." *Journal of Marketing* 41(1):77-79.
- Noftle, Erik E., and R.W. Robins, "Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores". *Journal of Personality and Social Psychology*, 93(1):116-130.
- Pfeffer, Jeffrey, 1981: *Power in Organizations*. Boston: Pitman.
- Phipps, Simone T.A., and L.C. Prieto, 2011: "The Influence of Personality Factors on Transformational Leadership: Exploring the Moderating Role of Political Skill." *International Journal of Leadership Studies*, 6(3):ISSN1554-3145.
- Schmitt, David P., J. Allik, R.R. McCrae, and V. Benet-Martínez, 2007: "The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations." *Journal of Cross-Cultural Psychology*, 38 (2):173-212.

- Shi, J and Z. Chen, 2012: “Psychometric properties of the political skill inventory”. *Psychological Reports*, 110(1):233-246.
- Sjöberg, Sofia, 2008: “What do we know about traits predicting leader emergence and leader effectiveness?”. *Working Paper, Stockholm University*.
- Yamauchi, Futoshi, 2012: “An alternative estimate of SBM impacts on students’ achievements: Evidence from the Philippines.” Forthcoming, *Journal of Development Effectiveness*.
- Yamauchi, Futoshi and S. Parandekar, 2013: “School Resources and Performance Inequality: Evidence from the Philippines”, Working Paper, The World Bank, 2013.
- Zhao, Hao, S.E. Seibert and G.T. Lumpkin, 2010: “The Relationship of Personality to Entrepreneurial Intentions and Performance: A Meta-Analytic Review.” *Journal of Management*, 36: 381-404,.