Summer 2014

Assessment of contributing factors to the reduction of diarrhea in rural communities of Para, Brazil

Lee Voth-Gaeddert

ASSESSMENT OF CONTRIBUTING FACTORS TO THE REDUCTION OF DIARRHEA IN RURAL COMMUNITIES OF PARA, BRAZIL

by

LEE VOTH-GAEDDERT

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN ENVIRONMENTAL ENGINEERING

2014

Approved by

Daniel B. Oerther, Advisor
Robert Holmes Jr.
V.A. Samaranayake

PUBLICATION THESIS OPTION

This thesis has been prepared in the style utilized by the Journal of Water Science and Technology and Humanitarian Technology. Pages 20-44 and 56-73 will be submitted to Water Science and Technology and Pages 45-55 will be submitted to Humanitarian Technology. Appendices A, B, C and D have been added for purposes normal to thesis writing.

ABSTRACT

In developing communities the occurrence of diarrhea has been reported at elevated levels as compared to those communities in more developed regions. Diarrheal diseases were linked to over one million deaths in 2012 throughout the world. While multiple pathways are present for the transmission of diarrheal diseases, water has been the focus for many aid organizations. Point-of-use (POU) water treatment methods are a common tool used by aid organizations in efforts to provide potable water. The CAWST biosand filter is a POU tool that has shown removal effectiveness of pathogenic microorganisms ranging from 90-99%. However, minimal literature was found that reported on the effectiveness of the filter within the larger body of the complex system found in all communities. Therefore a hypothesis was derived to confirm that the intervention of a CAWST biosand filter is the most significant factor in the reduction of the diarrheal health burden within households in developing regions. Communities located along the Amazon River in Para, Brazil were selected for study. Structural Equation Modeling (SEM) was utilized to aid in representing the complex set of relationships within the communities. The Mahalanobis-Taguchi Strategy (MTS) was also used to confirm variable significance in the SEM model. Results show that while the biosand filter does aid in the reduction of diarrheal occurrences it is not the most significant factor. Results varied on which factor influenced diarrheal occurrences the greatest but consistently included education, economic status, and sanitation. Further, results from the MTS analysis reported education as the largest factor influencing household health. Continued work is needed for further understanding of these factors and their relationships to diarrhea reduction.

PREFACE

Within the following document a novel approach to understanding impacts of interventions designed to alleviate poverty in developing regions is presented. The approach or methodology is taken from the disciplines of social science and econometrics. The paper is also presented in such a way that it allows the reader to observe the progression of the research through the four manuscripts written over the course of the past fourteen months. For these reasons a preface was established to prepare the reader for the layout of the thesis, the terminology to be expected within the thesis, and the nomenclature found in the thesis.

As the reader will see in the table of contents, the publication thesis option was utilized to incorporate the four manuscripts produced throughout this project. The first section of this thesis gives an introduction to the topic and presents a narrative of the researchers personal journey to completing this thesis. In the introduction a significant amount of time is given to establishing a needed understanding for the rest of the subsequent papers. The breadth to which this project covered was significant and therefore warranted a lengthy introduction. The introduction was also crucial for bringing together all of the different sub topics found throughout the following four manuscripts, as each manuscript only covered a particular amount of material. The narrative offers a brief journey of the authors understanding towards the project. Manuscript one through four were written over the past year. Manuscripts one, two and four were written to an audience in the International Water Association and manuscript three was written for a humanitarian aid conference. Understanding who the target audience was is important for a full understanding by the reader. Finally, the second section of the thesis offers a conclusion to all of these manuscripts and brings the readers focus back to the original hypothesis.

The terminology found within this document is derived from several different origins which may confuse the reader. While written from an Environmental Engineering point of reference, the methodology used within the research stems from the fields of statistics, social science and information systems. However, it should also be noted that while the usage of Structural Equation Modeling and Mahalanobis-Taguchi Strategy were a focal point in this research, this thesis was not written to withstand the deepest scrutiny within the statistical field. A balance between scientific enterprise and statistical procedures needs to be accepted by the reader.

Terms such as latent and indicator variables, causal influences, direct and indirect effects, path analysis, and indicator significance are common throughout the body of literature. These terms are associated with the statistical tools used and an understanding of these terms is needed. The methodology utilized within this research, while statistically robust, is commonly used in the social science and econometric disciplines. Terms such as multidimensional poverty, socio-economic status, household education level and cognitive ability are also found throughout this thesis and help to capture the breadth of the project. These terms and others are defined within the nomenclature section and it is recommended that the reader take time to familiarize themselves with these terms.

ACKNOWLEDGEMENTS

I am grateful to Dr. Oerther for allowing me to be a part of his research team and for supporting me in my academic as well as personal life. I am fortunate to be able to call him my friend as well as mentor.

I would also like to acknowledge John A. and Susan Mathis for providing much of the funding to conduct our research. Without your generous support none of this would have been possible.

I am grateful to Dr. Samaranayake and Dr. Holmes for being on my committee and pushing me to be a better student and person both through my research and in my classes.

A number of people have been crucial at certain points throughout this project; some of those people include Gil Serique, Samuel Vasconcelos, Dr. Elizabeth Cudney, and Adebolaji Jobi-Taiwo.

Finally, I am always thankful for a very supportive family both immediate and distant. Without the support of Ken, Deb, Jesse and others I could never have accomplished what I have.

TABLE OF CONTENTS

Page

PUBLICATION THESIS OPTION iii

ABSTRACT iv

PREFACE v

ACKNOWLEDGEMENTS vii

LIST OF ILLUSTRATIONS xi

LIST OF TABLES xii

NOMENCLATURE xiii

SECTION

1. INTRODUCTION 1
   - 1.1. THE PROBLEM 1
   - 1.2. METHODOLOGY 3
     - 1.2.1. Structural Equation Modeling 3
     - 1.2.2. Two-Step Approach 4
     - 1.2.3. Step One - Measurement Model 5
       - 1.2.3.1. Latent variable 5
       - 1.2.3.2. Graphical model 9
       - 1.2.3.3. Model fit indices 9
     - 1.2.4. Step-Two Structural/Full Model 10
       - 1.2.4.1. Path analysis 10
       - 1.2.4.2. Identification 11
     - 1.2.5. Software 11
     - 1.2.6. Multicollinearity 12
     - 1.2.7. Confidence Bounds 12
     - 1.2.8. Data Source 12
     - 1.2.9. Mahalanobis-Taguchi Strategy 12
   - 1.3. BACKGROUND 13
     - 1.3.1. CAWST Biosand Filter 13
     - 1.3.2. Locations 15
2. NARRATIVE OF LEARNING 17
   - 2.1. MY PATH TO COLLEGE 17
   - 2.2. UNDERGRADUATE STUDIES 17
   - 2.3. GRADUATE STUDIES 18

PAPER
I. UTILIZING STUCTRUAL EQUATION MODELING TO CORRELATE BIOSAND FILTER PERFORMANCE AND OCCURANCE OF DIARRHEA IN ENSEADO DO ARITAPERA, PARA, BRAZI 20
Abstract 20
Introduction 20
Methodology 22
Results 24
Discussion and Conclusion 30
References 31
II. UTILIZING STRUCTURAL EQUATION MODELING AS AN EVALUATION TOOL FOR CRITICAL PARAMETERS OF THE BIOSAND FILTER IN A PILOT STUDY IN PARA, BRAZIL 33
Abstract 33
Introduction 33
Methodology 34
Results and Discussion 38
Conclusions 42
References 42
III. UTILIZING STRUCTURAL EQUATION MODELING IN THE DEVELOPMENT OF A STANDARDIZED INTERVENTION ASSESSMENT TOOL 45
Abstract 45
1. Introduction 45
2. Measuring Multidimensional Poverty 47
3. Structural Equation Modeling 49
4. Current Work 51
5. Conclusion 53
Acknowledgements 53
References 53
IV. ANALYZING INDICATORS OF MULTIDIMENSIONAL POVERTY FOR STRUCTURAL EQUATION MODELING USING MAHALANOBIS-TAGUCHI STRATEGY 56
Abstract 56
Introduction 56
Methodology 59
Mahalanobis-Taguchi Strategy 59

Stage I: Construction Of A Measurement Scale. 60

Stage II: Assessment Of The Measurement Scale................................ 61

Stage III: Identify The Useful Variables (Developing Stage). 61

Stage IV: Future Diagnosis With Useful Variables. 62

Structural Equation Modeling ……………………………………..…………….. 63

Results and Discussion 65

Pros And Cons Of SEM And MTS …………………………...……………….. 65

MTS Analysis On DHS Survey Data From Guatemala ……………………… 68

Conclusion 70

References.. 71

SECTION

3. CONCLUSION 74

APPENDICES 78

A. SURVEY OF OWNERS OF CAWST BIOSAND WATER FILTERS....... 78

B. DHS HOUSEHOLD SURVEY (GUATEMALA - SPANISH) 85

C. TOP 50 SIGNIFICANT INDICATORS FOR HEALTH IN DHS SURVEY USING MTS ANALYSIS 92

D. EXPLANATION OF VARIABLES 94

BIBLIOGRAPHY 98

VITA 100

LIST OF ILLUSTRATIONS

Page

Figure 1.1. Divelbiss et al. 2013 SEM full model used in Guatemala................................ 4

Figure 1.2. Reflective and Formative Measures (Coltman et al., 2008)............................. 6

Figure 1.3. Representation of shared variance among indicators (x_1-x_3) 8

Figure 1.4. A description of what is incorporated in the CAWST biosand filter found in the CAWST manual ……………………………..…………………… 14

Paper I

Figure 1. Divelbiss et al. 2013 SEM model used in Guatemala 24

Figure 2. Altered SEM model for use in Brazil 26

Figure 3. Fitted model that shows numeric (standardized) direct effects between variables using bootstrapped data... 27

Paper II

Figure 1. Measurement Model with parameter estimates, definitions in Table 1............. 36

Figure 2. Full Model with standardized parameter estimates 42

Paper III

Figure 1. Graphical display of a full SEM model with latent and observable.................. 51

Paper IV

Figure 1. Full Model with parameter estimates. 64

Figure 2. SEM latent variable described by three indicator variables 66

Figure 3. MTS visual simplified method of finding significant indicator variables......... 66

Figure 4. Results are reported in a signal-to-noise ratio (dB) where positive blue bars indicate significant variable.. 69

LIST OF TABLES

Page

Table 1.1 Supporting hypotheses identified from Divelbiss' study and literature reviews from Brazil ... 3

Table 1.2. A Framework For Assessing Reflective and Formative Models: Theoretical and Empirical Considerations (Coltman et al., 2008) ... 7

Table 1.3. Rules of path coefficients, applicable for both standardized (given) and unstandardized (Grace, 2006)... 11

Paper I

Table 1. Results from physical MPLUS7 model. ... 29

Table 2. Correlation matrix from raw data. ... 29

Paper II

Table 1. List of variables and their type ... 38

Paper IV

Table 1. List of variables and their type ... 65

Table 2. Indicators from MTS Analysis ... 70

Table 3. Indicators from (Divelbiss et al. 2013) ... 70

NOMENCLATURE

Acronym	**Description**
SEM	Structural Equation Modeling – a statistical procedure used to estimate causal relations based on a set of researcher specified hypotheses
MTS	Mahalanobis-Taguchi Strategy – a pattern recognition technique
SES	Socio Economic Status – the level of 'wealth' within the household
HEL	Household Education Level – the cognitive abilities of a household
DHB	Diarrheal Health Burden – issues relating to diarrhea within the household

Definitions

Latent Variable – a variable that cannot be measured directly but is hypothesized to exist

Indicator Variable – observed variable that helps to describe the latent variable

Causal Influence – the direction of effect that two variables share

Path Analysis – multivariate technique used to analyze multiple relationships simultaneously

Indicator Significance – the importance of an indicator variable in representing the concept described by the latent variable

Multidimensional Poverty – the complex set of parameters that influence poverty

Theoretical Approach – concepts derived from literature and personal experience

Empirical Approach – concepts derived from statistical based evidence

1. INTRODUCTION

1.1. THE PROBLEM

In 2012 6.6 million children under the age of five died from preventable diseases. Diarrhea is the second leading cause of death for children under five (World Health Organization, 2013). Diarrheal diseases can be transmitted through a variety of ways, these are referred to as the fecal to oral pathways (Ben-Joseph, 2013). These pathways include fluids, fingers, fields, flies and foods. One major pathway identified by the World Health Organization (WHO) is fluids, more specifically water (Bartram, 2008). A large portion of literature has been devoted to showing the relationship between the diarrheal health burden (DHB) of an impoverished household and the water that it consumes (Black, Morris, & Bryce, 2003; Fewtrell et al., 2005; Ozkan et al., 2007). The United Nations (UN), through the WHO and Millennium Development Goals (MDGs), has stated that clean water is the most important step in reducing diarrhea and sickness within developing regions (WHO Media Centre, 2013). This statement is manifested through the actions of many non-governmental organizations; NGOs (Ahmed & Svennerholm, 2009).

With this notion so heavily engrained in the culture of development aid, verification of this statement is crucial. However, while being a widely accepted notion, within academic literature support for this hypothesis is lacking. This is not to say that there aren't reports showing individual variables that offer a large reduction in diarrheal occurrences, but few reports holistically analyze all potential influences simultaneously regarding diarrhea. The third manuscript within this thesis offers a broader view in the area of implementation assessments. It is critical for the advancement in aid effectiveness to fully understand the complex environment in which it serves. As reflected in recent changes to organizational policies (Martindale, 2013; Peace Corps, 2012) a shift in focus is occurring to incorporate feedback loops to help understand these complex environments. Unfortunately, there is a significant gap within the literature in this area.

A recent study implemented by Divelbiss et al. (Divelbiss, Boccelli, Succop, & Oerther, 2013) in Guatemala assessed the impact of point-of-use (POU) water filtration interventions on local communities utilizing the statistical tool structural equation modeling (SEM). The POU Divelbiss observed in his study was the Centre for

Affordable Water and Sanitation Technology (CAWST) biosand filter which is used to improve individual household's water quality. The holistic approach coupled with the use of SEM offered a robust novel analysis technique for impoverished communities. This same novel approach was utilized in this study for local communities located along the Amazon River in Para, Brazil that rely heavily on aforementioned CAWST biosand filters. The central hypothesis of this study was to confirm that the CAWST biosand filter is the most significant factor in the reduction of diarrheal occurrences in developing communities.

To aid in the evaluation of this hypothesis, a complex set of relationships surrounding this hypothesis were also manifested into supporting hypotheses. These relationships were identified from Divelbiss' study as well as a literature review regarding Brazil. Household education level (HEL), socio-economic status (SES), additional water treatment, improved water source, adequate sanitation and proper water storage were all identified as variables with potential to impact diarrheal occurrences within the household. HEL, SES, additional water treatment, and improved water source were additionally identified for potential to impact the operation and maintenance of the biosand filter. These eleven different relationships can be seen in Table 1.1. except for the relationship between an improved water source and the filter operation and maintenance. Explanations of these variables can be found in the Nomenclature as well as Appendix D.

To assess the complex set of hypotheses, the experiment was carried out through the following tasks;

Task 1: Take the existing SEM model and the accompanying survey from Divelbiss et al. and modify for location and cultural differences in Brazil (see appendix A)

Task 2: Collect data through the implementation of household surveys and analyze data with SEM model, make adjustments if necessary

Task 3: Iterate until tests of model fit are passed, analyze results

Task 4: Use Mahalanobis-Taguchi Strategy (MTS) and data collected by the Demographic and Health Survey (DHS) Program to confirm the correct significant indicator variables were used in Divelbiss' study in Guatemala (see appendix B)

Tasks 1, 2 and 3 are all based on the methodology of SEM. Task 4 is based on the methodology of MTS. Together these tasks allowed for the experimentation of the primary hypothesis as well as the supporting hypotheses.

Table 1.1 Supporting hypotheses identified from Divelbiss' study and literature reviews from Brazil

Increased household educational level has a negative effect on the severity of diarrhea burden
Increased socio-economic status has a negative effect on the severity of diarrhea burden
Poor hygiene practices have a positive effect on the severity of diarrhea burden
Additional water treatment beyond the filter has a negative effect on the severity of diarrhea burden
Access to an improved water source has negative effect on the severity of diarrhea burden
Access to adequate sanitation has a negative effect on the severity of diarrhea burden
Proper water storage has a negative effect on the severity of diarrhea burden
Increased socio-economic status has a positive effect on filter operation and maintenance
Better personal hygiene practices have a positive effect on filter operation and maintenance
Additional water treatment has a positive effect on filter operation and maintenance
Increased household educational level has a positive effect on filter operation and maintenance

1.2. METHODOLOGY

1.2.1. Structural Equation Modeling. SEM is a statistical procedure used to estimate causal relations based on a set of researcher specified hypotheses. The relationships depicted in the hypotheses offered above (Table 1.1) are manifested in the complex multivariable model such as that shown in Figure 1.1. Within this model are two main types of variables, observed variables (variables that can be directly tested), represented by the boxes, and the latent or hidden variables (variables that cannot be directly tested for), represented by the ovals. The arrows between the variables represent the direction of the hypothesized casual relationships. Measurement error variables are represented by circles with an 'e'. Some relationships within Figure 1.1. may give the reader pause, such as the relationships between the latent variables (ovals) and the observed indicator variables (ie for SES; floor material, wall material, and population density). These relationships are part of what make SEM a robust statistical tool.

To begin, a model is hypothesizes based on literature, the researchers field experience, local practitioners, and other resources. Once this is established a survey is created to gather data privy to the model. Within this study a household survey was created to gather information on the factors hypothesized to affect diarrheal occurrences.

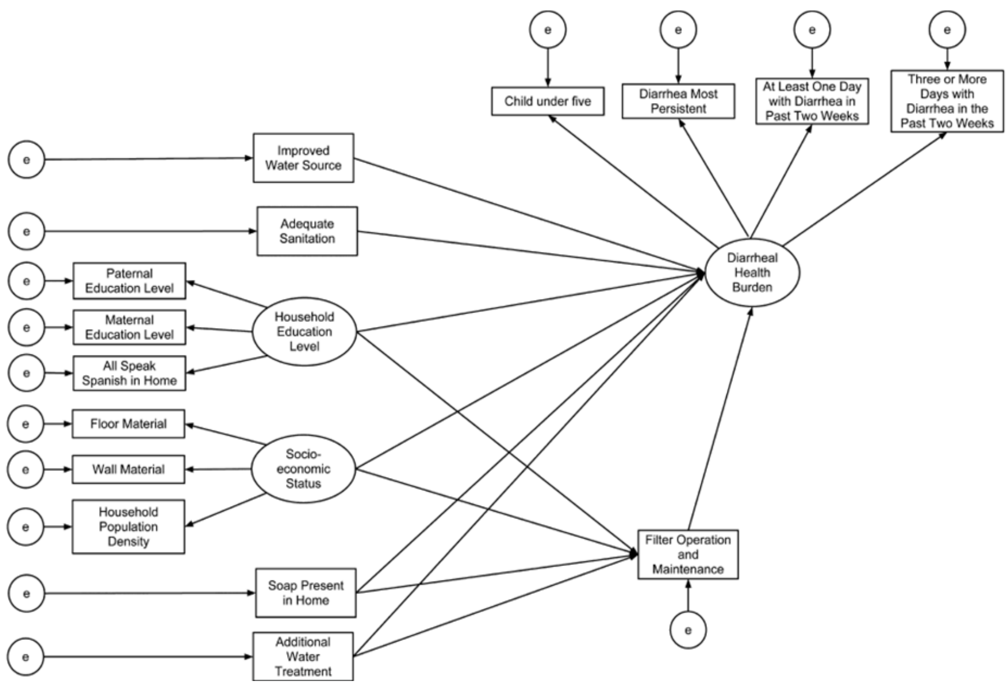


Figure 1.1. Divelbiss et al. 2013 SEM full model used in Guatemala

1.2.2. Two-Step Approach. SEM offers a wide variety of uses within different research fields. However, to use SEM as an exploratory technique (as this thesis does), a two-step approach is employed including: (step 1) the analysis of acceptable latent variables (this is referred to as the measurement model) and (step 2) then the analysis of fit indices of the model as a whole (a combination of the measurement model and the structural model as seen in Figure 1.1.). These steps, individually, can be seen as a Confirmatory Factor Analysis (CFA) approach where in one hypothesized model is compared to a data driven model using the covariance matrices of both for analysis. To assess the adequacy in similarity between the models, fit tests are utilized (see 'Model Fit Indices'). If the two models (hypothesized and data driven) are not adequately similar then adjustments are made. Once a model is altered to improve fit, this process becomes exploratory, requiring new data to be collected and tested against the altered model. Because repeated steps are involved, as fit is evaluated, failure occurs and iteration drives the model towards a better representation of the relationships present within the environment.

1.2.3. Step One - Measurement Model. Once a full model is hypothesized and data is collected to populate it, the latent variables are then analyzed in what is referred to as the measurement model. The latent variables are what make SEM unique, but also warrant much attention when used.

1.2.3.1. Latent variables. Among the complex set of relationships within a developing community several variables offer a challenge to represent. Within the literature the reader can find much agreement in the representation of poor water practices (ie gathering, source, treatment, storage, etc.) and poor sanitation (ie direct relationship to the five 'F' diagram). These concepts or variables can be represented empirically without much discussion. For example, a researcher can empirically test that if water is not treated properly before consumption, diarrheal occurrences will increase (Quick et al., 2002). This idea of proper treatment of water ('additional water treatment') can be found as an independent observable variable within the SEM model (see Figure 1.1.). However, the concepts within the literature that were not consistent in definition were that of education level (ie cognitive ability), socio economic status, and health. Among these concepts, definitions and assessment techniques varied widely. For this reason latent variables were used to represent these concepts in the SEM model.

A latent variable is one that is hypothesized to exist, but that has not been measured directly (Grace 2006). As one may see from Figure 1.1. the latent variables (ovals) are regressed on the supporting indicator variables (boxes). These indicator variables can also be thought of as manifestations of the concept that is the latent variable. The term 'indicator' is used to identify the fact that it is not an exact representation of the concept of interest. With this type of regression pattern the latent variable is referred to as a reflective latent variable, as it is reflected within its indicator variables. Conversely, a formative latent variable has the indicator variables regressing on it. Figure 1.2. offers a visual comparison between the two types of latent variables. Table 1.2. provides a side by side comparison in regards to theoretical and empirical considerations and Figure 1.3. offers a visual aid to the concept of representation in a reflective latent variable. The difference lies in the direction of causality. For the Effect Model the concept, represented by the Greek letter zai, influences the indicator variables. As a change in the concept occurs, all indicator variables will change, this change is then captured and used to understand the concept. A popular concept that is highly studied is human intelligence, as human intelligence changes, it will be reflected in the indicator variables such as test taking abilities, reading abilities, memory abilities, etc. The delta variable refers to measurement error, this error represents the lack of correlation between the indicator variable and the latent concept. The Causal Model shows changes in indicators influencing the zai variable. This is commonly found in many measurement indices and relies solely on its specific indicators. Representation can be lost if a certain variable is removed or switched out which is not the case for the Effect Model.

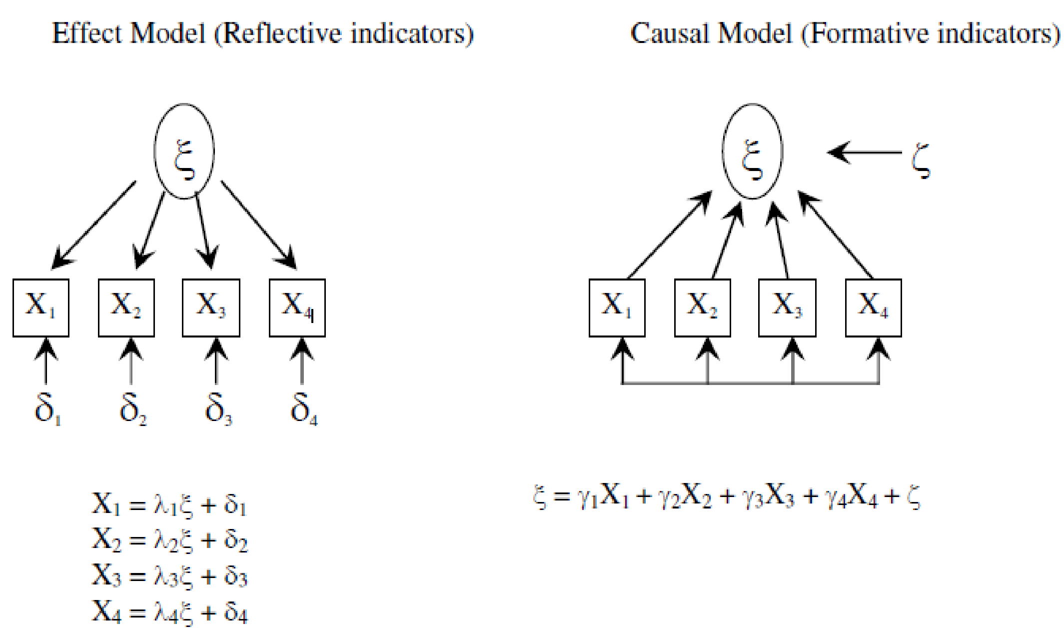


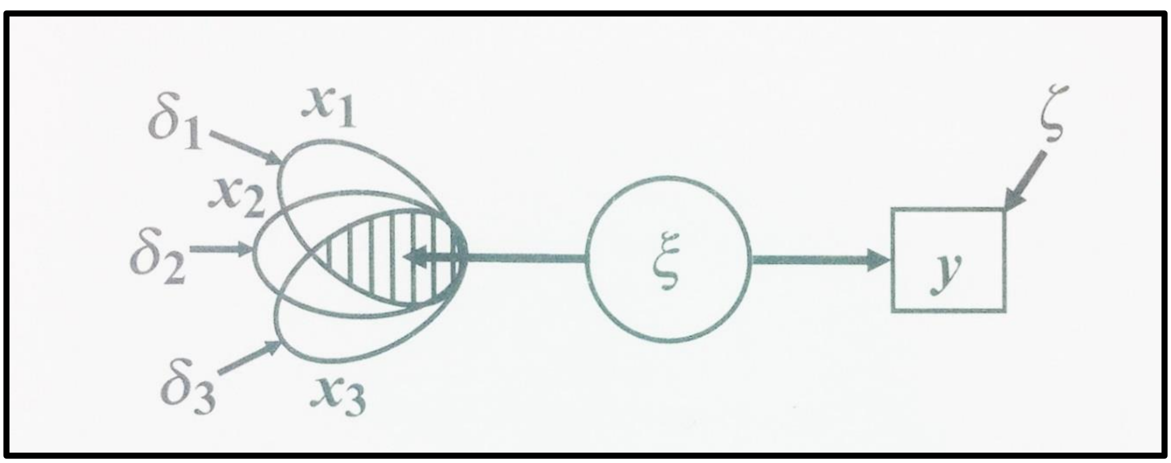
Figure 1.2. Reflective and Formative Measures (Coltman et al., 2008)

Table 1.2. A Framework For Assessing Reflective and Formative Models: Theoretical and Empirical Considerations (Coltman et al., 2008)

Considerations	Reflective model	Formative model	Relevant literature
Theoretical Considerations			
1. Nature of construct	Latent construct is existing ➢ Latent construct exists independent of the measures used	Latent construct is formed ➢ Latent constructs is determined as a combination of its indicators	Borsboom et al. (2003, 2004)
2. Direction of causality between items and latent construct	Causality from construct to items ➢ Variation in the construct causes variation in the item measures ➢ Variation in item measures does not cause variation in the construct	Causality from items to construct ➢ Variation in the construct does not cause variation in the item measures ➢ Variation in item measures causes variation in the construct	Bollen and Lennox (1991); Edwards and Bagozzi (2000); Rossiter (2002); Jarvis et al. (2003)
3. Characteristics of items used to measure the construct	Items are manifested by the construct ➢ Items share a common theme ➢ Items are interchangeable ➢ Adding or dropping an item does not change the conceptual domain of the construct	Items define the construct ➢ Items need not share a common theme ➢ Items are not interchangeable ➢ Adding or dropping an item may change the conceptual domain of the construct	Rossiter (2002) ; Jarvis et al. (2003)
Empirical Considerations			
4. Item intercorrelation	Items should have high positive intercorrelations ➢ Empirical test: internal consistency and reliability assessed via Cronbach alpha, average variance extracted, and factor loadings (e.g., from common or confirmatory factor analysis)	Items can have any pattern of intercorrelation but should possess the same directional relationship ➢ Empirical test: indicator reliability cannot be assessed empirically; various preliminary analyses are useful to check directionality between items and construct	Cronbach (1951); Nunnally and Bernstein (1994); Churchill (1979); Diamantopoulos and Siguaw (2006)
5. Item relationships with construct antecedents and consequences	Items have similar sign and significance of relationships with the antecedents/consequences as the construct ➢ Empirical test: content validity is established based on theoretical considerations, and assessed empirically via convergent and discriminant validity	Items may not have similar significance of relationships with the antecedents/consequences as the construct ➢ Empirical test: nomological validity can be assessed empirically using a MIMIC model, and/or structural linkage with another criterion variable	Bollen and Lennox (1991); Diamantopoulos and Winklhofer (2001); Diamantopoulos and Siguaw (2006)
6. Measurement error and collinearity	Error term in items can be identified ➢ Empirical test: common factor analysis can be used to identify and extract out measurement error	Error term cannot be identified if the formative measurement model is estimated in isolation ➢ Empirical test: vanishing tetrad test can be used to determine if the formative items behave as predicted ➢ Collinearity should be ruled out by standard diagnostics such as the condition index	Bollen and Ting (2000); Diamantopoulos (2006)

The three latent variables used in the SEM model are reflective. Several steps can be taken to confirm that a reflective approach is needed over a formative approach (direction of arrows or causality). First, Grace 2006 offers four questions to test the theoretical approach taken by the researcher in choosing to use the reflective concept. First, what is the direction of causality? This was addressed in the prior section and serves as a significant step. The second part to this first question is that of independence among indicators. If high independence is found, then the reflective approach is not appropriate as all indicators would not react the same to changes in the overall concept. The second question offered by Grace is, are the indicators under a single latent variable interchangeable? As mentioned in the Causal Model (Figure 1.2), the formative approach relies heavily on the representation from its indicator variables, where in the reflective approach, indicators may be interchanged with little effect. The third question is, should there be an expectation that indicators under a latent variable should covary (e.g., because of joint causality)? This reflects the second question but in relation to covariance. Finally, the last question is, do the indicators associated with a latent variable have a consistent set of causal influences? This question is related to the first question and addresses the question of common antecedents among indicators. If a common antecedent is found then this is best modeled with the reflective approach and if not then the formative approach is best.

Figure 1.3. Representation of shared variance among indicators (x_1-x_3). Hatched area is shared variance. Only this variance is linked to latent variable. Error variance can be thought of as that of what's not inside of hatched area (Grace, 2006).



A check of causality offered among other SEM scholars allows for an empirical analysis of causal direction. An investigation of the correlation matrix allows for this analysis. Within the correlation matrix, relationships between indicator variables under the same latent variable should have a high consistent correlation with each other. If this is the case, then causation should flow from the latent variable to its indicator variables. If a low or no correlation is found, then causation may flow from indicator variables to the latent variable.

1.2.3.2. Graphical model. A significant benefit in SEM is the utilization of a physical model; however, with the concept of reflective latent variables, much care is needed in the use of labels for these variables. For example, in Figure 1.1. the latent variable Household Education Level (HEL) may be interpreted as an average of school levels between the father and mother. This would in turn require the use of a formative latent variable as the indicator variables (mother and father school level) influence the latent variable (education level of the household). The intent of this variable is to represent the collective cognitive ability of the household. Much knowledge is gained outside of school and therefore school enrollment may not be effective as a sole indicator of cognition. It becomes theory driven when we start to look at locations or indicators of knowledge gain, for this reason reflective latent variables are used.

1.2.3.3. Model fit indices. A critical part of SEM is testing model fit. Testing model fit happens within the measurement model (ie latent variable relationships) and if fit is deemed adequate a second test of model fit happens within the full model. A large body of work is available for study on this subject as there are differing opinions on particular tests. However, generally, it is suggested to use Chi-Square (Bentler, P.M. 2007), root mean square error of approximation; RMSEA (Hu, L. et al. 1998), Comparative Fit Index; CFI (Hu, L. et al. 1998), and Tucker Lewis Index; TLI (Muthen & Muthen 2012). The software package MPLUS 7 was used for analysis of data (Muthen & Muthen 2012). Within this software package the authors, Muthen & Muthen recommend three of these four tests of model fit be ‘adequate’. Literature offers certain levels to each test that deem the model ‘adequately’ fit. These values are, for χ^2 $p > .05$, RMSEA $p < .10$, CFI > .9, and TLI > .9. It should be noted that models should not be assessed and adjusted by these statistical tests alone, careful consideration of concepts within the model need to be taken in combination with the results of model fit tests.

1.2.4. Step-Two Structural/Full Model. If the assessment of the measurement model produces an acceptable outcome (via researcher assessment of proper direction of regression and model fit tests) the structural model can be incorporated to create the full model. As the latent variables have been tested for robust representation the hypothesized structural model is assessed using path analysis. This assessment incorporates the original hypothesis along with the supporting hypotheses.

1.2.4.1. Path analysis. Path analysis allows for the assessment of direct and indirect effects between variables within the SEM model. The use of statistical techniques such as partial regression and maximum likelihood are used to assess the multi-dependent relationships within the SEM model. The analysis of this is reported in terms of path coefficients, either in standardized or unstandardized format. Each format offers a crucial piece of information about the model. The standardized path coefficients allow for direct and indirect effects to be totaled and reported. These totals present both magnitude and positive/negative influences of exogenous variables (observed and latent) on the endogenous variable(s) in units of standard deviation. Significance levels of each variable can be obtained through comparison of the standardized path coefficients and a rank of important variables can be established. Unstandardized path coefficients allow for the individual casual effects between two variables to be analyzed. As a one unit change in the exogenous variable occurs, the path coefficient unit depicts the amount and direction of change in the endogenous variable. Table 1.3. reports the eight rules of path coefficients that offer a guideline to using path analysis. An understanding of the data being analyzed is critical to proper analysis.

Table 1.3. Rules of path coefficients, applicable for both standardized (given) and unstandardized (Grace, 2006).

(1) The path coefficients for unanalyzed relationships between exogenous variables are the bivariate correlations.
(2) When two variables are only connected through a single directed path, the coefficient for that path corresponds to the bivariate regression coefficient.
(3) The strength of a compound path (one that involves multiple arrows) is the product of the coefficients along that path.
(4) When two variables are connected by more than one causal pathway, the calculation of partial regression coefficients is required.
(5) Coefficients associated with paths from error variables are correlations representing unexplained effects.
(6) Unanalyzed correlations between endogenous variables are represented by partial correlations.
(7) The total effect one variable has on another equals the sum of its direct and indirect effects through directed (causal) pathways.
(8) The sum of all pathways connecting two variables, including both causal and noncausal paths, adds up to the value of the bivariate or total correlation between those two variables.

1.2.4.2. Identification. To identify a hybrid model, such as the one utilized in this analysis, identification of the measurement model and the full model is needed. A model is identified if there are as many or more known values as compared to the number of parameters to be estimated. Both models within this thesis satisfy that requirement.

1.2.5. Software. There are several software packages available that allow for usage of SEM. The one utilized for this thesis was MPLUS 7. This software package has recently become more popular within fields that often work with SEM such as Econometrics and the Social Sciences. MPLUS 7 uses the same underlining equations as the more widely known LISREL package but offers more options in terms of analysis techniques.

1.2.6. Multicollinearity. An issue that arises when working with multiple robust latent variables and independent variables that are all hypothesized to regress on one or two variables is multicollinearity. This means that two or more variables that are thought to be separate concepts within the model are actually highly correlated (>0.5-0.6). While MPLUS 7 does not directly test for this, investigation of correlations between the structural model variables can identify potential issues. This phenomenon was monitored throughout the analysis and was not found to be an issue.

1.2.7. Confidence Bounds. Within the path analysis, confidence intervals were established at a 5% upper and lower limit (ie 90%). Besides the HEL variable (discussed later) no bounds were found to have a change in sign (insinuating a positive or negative flip of influence) unless the parameter was between +/-0.046 which meant that the pathway was assessed to carry negligible effects.

1.2.8. Data Source. Data from Brazil was collected using household surveys that complimented the current SEM model. Help from local practitioners and translators also enabled the collection of data. Cluster sampling was used while in the villages of Brazil. Once data was recorded, it was transferred to excel for further analysis. Data from Guatemala was also analyzed in the last manuscript. This data was collected by the Demographic and Health Survey (DHS) Program, who is funded by United States Agency for International Development (USAID). They administer household surveys throughout the world including Guatemala. Their surveys can be accessed on a per household level with a specific region of a country using cluster sampling techniques. Data access and aggregation is a big issue when analyzing developing countries and an area that needs much improvement to allow for better feedback.

1.2.9. Mahalanobis-Taguchi Strategy. Finally, MTS was utilized in the last manuscript of this thesis. MTS is a pattern recognition technique that was made popular by Taguchi in 2002 (Taguchi & Jugulum, 2002). It has been used primarily in the automotive industry, however, it shows promise in multiple disciplines. Dr. Elizabeth Cudney is a professor in the Engineering Management and Systems Engineering department at Missouri University of Science and Technology and has written multiple articles on the utilization of MTS and MTGS. We collaborated with her on the final paper in this thesis (Paper 4).

The purpose of using MTS was to verify that the correct indicators were chosen in reference to the latent variables in the SEM model from Divelbiss et al. As previously stated, the latent variables are derived from theoretical means. MTS allows for verification of the correct indicators and is unique because of the utilization of the Mahalanobis distance (MD). MD is a measure of distance that utilizes orthogonal transcriptions within the correlation matrix to eliminate issues with multicollinearity. By analyzing variances and covariances the MD approach differs from classical statistical approaches. It is also able to account for independent and dependent variables within the same set. The driven purpose of MTS is to accurately predict significant variables that show similar patterns within a multidimensional system (Taguchi & Jugulum, 2002).

Utilizing a different statistical tool to analyze different data (DHS from the same region and demographic of Guatemala) allows for a unique set of significant indicator variables to be identified. These results can then be used to either support or reject the indicators used in Divelbiss' model. The reader is referred to Paper 4 for the full report.

1.3. BACKGROUND

1.3.1. CAWST Biosand Filter. The CAWST biosand filter was originally constructed by Dr. David Manz in the 1990's at the University of Calgary, Canada. With the biosand filter becoming very popular among aid organizations the Centre for Affordable Water and Sanitation Technology was established as a professional service provider in 2001. CAWST offers support services, ranging from community demonstrations to technical papers on effectiveness.

The limited materials needed for construction offer a very promising tool to regions that may have issues with access to resources. Cement, gravel, sand, baffle, lid, piping and a bucket are all that is needed to improve water quality in a house. Once implemented in the house, water is poured in the top, over the baffle and clean water is collected out of the nose of the filter (see Figure 1.4.).

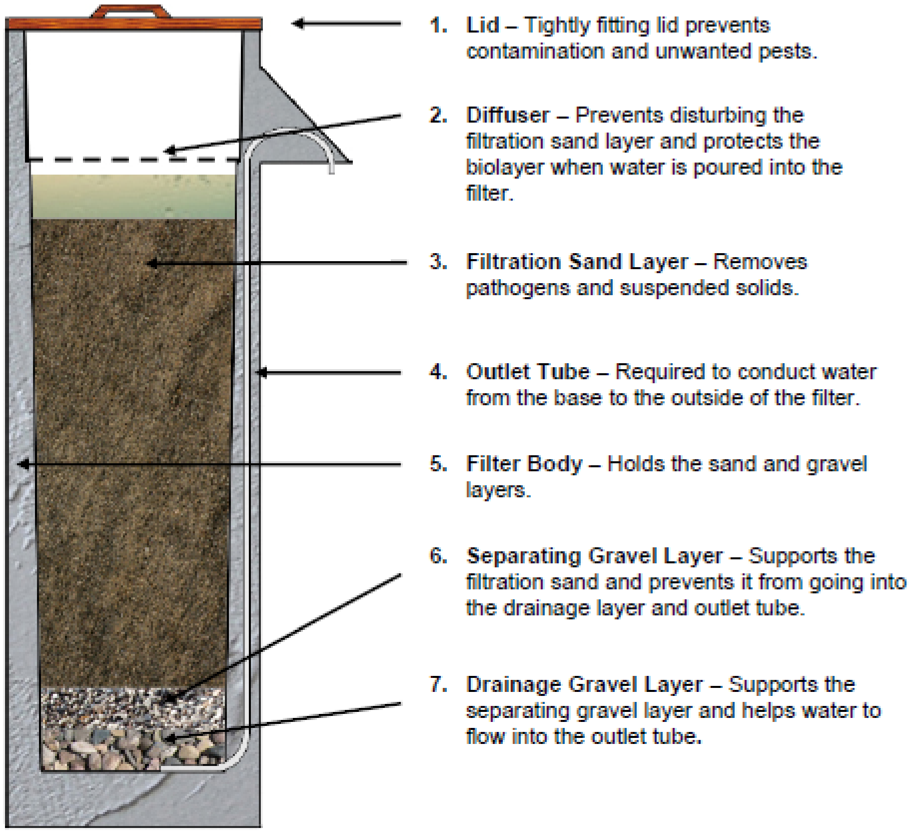


Figure 1.4. A description of what is incorporated in the CAWST biosand filter found in the CAWST manual (CAWST, 2009)

There are four specific types of pathogenic microorganism removal processes; predation, trapping, adsorption and natural death. Trapping, adsorption and natural death all occur within the body of the sand. Some organisms are too large to pass through, others are adsorbed to the surface of the sand by adhesion principles (ie mechanical, chemical, dispersive, electrostatic, and diffusive), and still others are caught in an anaerobic region for a period of time and die off. However, the most important removal process of the filter is predation. Predation happens on the surface of the sand in the biofilm or schmutzdecke layer. If the filter is utilized at least three times a week and is kept moist, the biofilm layer will continue to aid in removal of contaminants. It takes five to seven days for the biofilm layer to grow, but the filter can still be used in this time. CAWST recommends the addition of a second treatment option to ensure safety.

The CAWST biosand filter has been documented to remove 100% of helminthes and protozoa, 98.5% of bacteria, and 70-99% of viruses. It also is able to remove turbidity and iron at up to 95%. Finally, arsenic can be removed with the addition of rusty nails within the sand layer; this is referred to as the Kanchan Filer. An extensive collection of literature on the CAWST biosand filter can be found on their website (Centre for Affordable Water and Sanitation Technology, 2013).

To ensure a high level of removal efficiency, maintenance is recommended every 30 to 40 days. Maintenance includes the scrapping and mixing of the top inch of sand with a user's hand. The biofilm layer becomes clogged as it grows too thick and sediment collects on top. Once disturbed, the film layer will take another five to seven days to grow back however, flow rates should greatly improve. Once a year all of the sand should be removed from the filter casing and washed. With these basic maintenance steps potable water can be provided (Centre for Affordable Water and Sanitation Technology, 2009).

1.3.2. Locations. Once the intervention item had been selected, locations needed to be established. For the project implemented by Divelbiss et al. in Guatemala, Divelbiss had previously worked with an organization in the region of Quiche, Guatemala to build and distribute CAWST biosand filters. A return trip one year later allowed for the assessment of the implementation.

The region of Quiche, Guatemala has both the central highlands and the mountain ranges of Sierra de los Cuchumatanes and Sierra de Chuacus. The villages under study had been relocated in prior years and placed according to a grid by government officials. This created a unique dynamic in cultures and languages but still had the common problem of sickness and disease, thought, in larger part, to be attributed to contaminated water sources.

From the results of Divelbiss et al. further investigation was warranted on the original hypothesis. In the fall of 2012 Dr. Daniel Oerther was on a Fulbright Teaching Scholarship near Santarem, Para, Brazil. Through his work in the region Dr. Oerther became familiar with the work of the organization Project Amazon (PAZ). Primarily a church planting organization, PAZ had spent ten years building and distributing CAWST biosand filters for villages located along the Amazon River. With over 10,000 filters distributed, the area offered a great opportunity to study a vastly different environment.

Brazil is a middle income country with a growing economy (The World Bank, 2014). However, many pockets of low income or impoverished households can still be found. The villages studied within the Para region are located in the flood plain of the Amazon River. All houses are erected on stilts and made of wood. Depending on location of the village, the river completely floods the area for one to four months out of the year. The most common job for males within the region is fishing and for females is souvenir creation. Portuguese is the only language spoken which meant translators were heavily relied on. Most, if not all, children were found to be in school. A yellow school boat was even available for transport to and from school. Taxi boats were available for use but limited in scheduling and destination. The water source utilized by villagers was typically from the river. Personal latrines were common; they were in the form of an outhouse with a chute down into a hole in the river bed. During high water season, these would flood and downstream neighbors would unknowingly have raw waste water. There was a single power line running through most of the villages and some families utilized car batteries to run personal items. Health posts were located in larger villages but many did not have their own. It was apparent that on average these families did not live uncomfortably but still lacked access to some basic necessities.

2. NARRATIVE OF LEARNING

2.1. MY PATH TO COLLEGE

I am originally from a small town in Kansas called Hesston. I spent the majority of my life there and enjoyed much of it. I was fortunate to be brought up with both my parents and a younger brother. My father is a professor in Microbiology, my mother is a social worker at the Veterans Administration and my brother is on the pre-med track as a junior in College. I was involved in a plethora of sports in high school as was a part of the concert and jazz band throughout high school. I grew up in the Mennonite Christian Community which greatly emphasized mission work. Since a young age I have always wanted to help people and as I neared the end of my high school career I felt that a degree in Civil Engineering could help facilitate this.

2.2. UNDERGRADUATE STUDIES

I began my college career at Bethel College playing football. Bethel is a private Mennonite university as is Hesston College where I would transfer each spring semester to play baseball. For two years I transferred back and forth playing football and baseball, all the while taking classes that ranged from engineering to religion to conflict studies. In August of 2010 I transferred to MST to play baseball and continue my studies in Civil Engineering. I was able to get involved in Engineers Without Borders which allowed me to gain a new perspective on helping people. However, my path was severely changed when I met Dr. Daniel Oerther.

At the time, Dr. Oerther had also just arrived at MST and was teaching the Intro to Environmental Engineering course. Doing well in this class and knowing he was also interested in poverty alleviation I took a 390 research project with him the next two semesters. Much of my final year of undergraduate was spent in confusion of how best to help people; much of this was attributed to Dr. Oerther, lovingly of course.

I had the pleasure of working with Dr. Oerther’s graduate student, Andrew Schriner on utilizing crowd sourcing as a solution for poverty. As a civil engineer focused on building water tanks for people in developing countries I felt very misplaced in my research. In working with an inefficient platform and with a demographic that was not dying of diarrhea I struggle to visualize the bigger picture. Moving into my final semester and last undergraduate research project I still had my sights on becoming a missionary but had thoughts of graduate school. My research focus was on popular point of use water treatment methods and creating a display. This allowed me to investigate the environments in which these items were used and the positives and negatives about each. I also found myself uncertain on when each of them were appropriate to use in different situations, due to the complexity of the situations. However, this complexity was only the tip of the ice burg as I would soon realize. So, I committed to undertaking my masters with Dr. Oerther.

2.3. GRADUATE STUDIES

To begin my master's program I had the opportunity to travel to India for one month with a group of MST students and Dr. Oerther. This experience destroyed any notion of saving the world through water tanks. Working in villages, staying in five star hotels, passing families living in tents and collaborating with local college students put my little world in global perspective. However, deep down, I still held on to the idea that since my research group was testing water we were making the biggest difference. In my trip I found the hardest experience to handle was returning to the United States. My focus shifted from water tank solutions to an understanding of different solutions for different communities. But, I still believed in one 'sliver bullet' solution.

In March of 2013 I traveled to Para, Brazil to begin my thesis research. Brazil offered a very different environment than India and a very different set of problems within the rural communities. Working in these communities, asking about personal education and wealth along with health issues forced me to realize the complexity of developing regions. Analyzing these issues involved getting exposure to the world of statistics. I was fortunate to have a fantastic statistics teacher (Dr. Samaranayake) who was able to help me understand the importance of statistical analysis in effective research. With all of this knowledge in hand I returned to Brazil in September to continue my research.

My final step came through literature reviews this semester. The book, Poor Economics summarized it nicely portraying the complexity of solutions even among just two people. With this knowledge I press forward both unsure and excited about my future and how to help people.

PAPER

I. UTILIZING STUCTRUAL EQUATION MODELING TO CORRELATE BIOSAND FILTER PERFORMANCE AND OCCURANCE OF DIARRHEA IN ENSEADO DO ARITAPERA, PARA, BRAZIL

Abstract

Previously, Divelbiss and co-workers demonstrated the use of structural equation modeling to evaluate the effectiveness of point of use biosand filters to reduce diarrhea transmission in rural Guatemala. While prior research in the laboratory and in the field has documented the effectiveness of biosand filters to remove infectious agents (Palmateer *et al*. 1999, Stauber *et al.* 2006, Stauber *et al.* 2009, Aiken *et al.* 2011, Fabiszewski *et al.* 2012), experience in field sites suggests that multiple local factors greatly influence diarrhea transmission. This study employed a holistic approach to evaluate the benefit of biosand filters by including household education level, socioeconomic status, as well as maintenance and sanitation as factors to impact diarrhea transmission for households in the Amazon River basin in Brazil**.** The strongest correlation found was between the utilization of an 'improved' water source and the reduction of diarrhea within the household. Socio-economic indicators were also collected, such as household density, improved roof, and ownership of a personal boat amongst others. Of the 18 correlations amongst indicators of wealth (6) and diarrhea (3), 16 were negatively correlated, supporting the prior findings by Divelbiss et al. Also, the increased quality of the operation and maintenance of the POU was also found to have a negative correlation with the occurrence of diarrhea within the house supporting the original hypothesis that access to filters is not the primary factor influencing diarrhea transmission. While the findings within this project support the results of Divelbiss et al., more research is needed to effectively warrant change on a large scale within 'aid distribution'.

Introduction

Diarrhea is an important public health problem closely associated with hygiene conditions and how water is used (Ozkan et al. 2007). Inadequate water quality and quantity, combined with a lack of basic sanitation and poor hygiene, create risk conditions for endemic transmission of pathogens. While developed countries have created public health systems that effectively break the fecal-oral route of exposure, in developing countries the transmission of diarrhea via unclean water accounts for over 85% of the global diarrhea disease burden (Tiwari et al. 2009). These health risks, while harmful to adults, can be devastating to children. More than 6.6 million children younger than 5 years perish annually from preventable diseases (Black et al. 2003). Reports suggest that diarrheal diseases is the second leading cause of child death in the world (Tiwari et al. 2009). While a scientific understanding of clean water and promoting efficienct interventions are important, it is also important to recognize that the transmission of diarrhea is dependent upon many variables including fluids (water), fingers, flies, feces, and fields as well as the personal and collective decisions made by individuals, households, communities, and nations.

While epidemiological statistical tools such as relative risk often are used to evaluate the probably of illness among an exposures group as compared to a group protected through an intervention, the complex nature of diarrhea transmission is better studied using a more complex statistical tool. Structural equation modeling (SEM) is a useful tool for capturing the complex relationships imagined among diverse variables. SEM is a statistical procedure used to estimate causal relations based on a set of researcher specified hypotheses. SEMs share the characteristic of requiring a pre data model that theorize relationships and allow the application of both observed and latent variables. As previously described, model specification is commonly based on theories, previously published research, and experience of the researchers (Divelbiss et al. 2013). Conceptual models are used to explain complex relationships. The concepts are reduced to equations using software and the results of experimental data are used to validate all or portions of the conceptual model within a degree of statistical certainty. Previously, SEM was used to evaluate diarrhea transmission in the Ixcan region of Guatemala (Divelbiss, 2013).

In this study, a modified version of a previous SEM was created using field observation of households in Enseado Do Aritapera, Para, Brazil. Biosand filters have been distributed within the target community over the past decade in an effort to reduce diarrhea transmission. This intervention was organized by a Christian missionary team operating in Para, Brazil. The filters used in this intervention were of the design promoted by the Centre for Affordable Water and Sanitation Technology, Calgary, Alberta, Canada or known more commonly as the CAWST BSF. Previously, the BSF has proved effective in removing pathogens, parasites, turbidity and some metals (Kubare et al. 2010). The objective of this study was to evaluate if diarrhea has been reduced in the target community, and if the BSF is responsible for any improvement.

Methodology

SEM has been described as a combination of exploratory factor analysis and multiple regression (Schreiber et al. 2006). SEM allows for the analysis of relationships between independent variables and dependent variables where either type can be represented by a measured variable (directly observed) or a latent variable (unobserved, not directly observed) (Ullman 2006). SEM has been used extensively in psychological, social, and behavioral sciences (Bentler et al. 1999). However, recent work has been done to utilize the added benefits in SEM analysis such as latent variables and multiple hypothesis modeling. Traditionally, SEM has been used primarily as a confirmatory technique, but it can be used for exploratory purposes (Schreiber et al. 2006). Muthen and Muthen created the software package MPLUS to allow researchers to utilize the SEM technique quickly and easily. MPLUS 7 was used in this study.

To use SEM as an exploratory technique, a two-step approach was employed including: (step 1) the analysis of acceptable latent variables and (step 2) then the analysis of fit indices of the model as a whole (Divelbiss et al. 2013). Iacobucci reported that a minimum of 50 samples are required to populate the model correctly (2009); whereas other studies have recommended larger sample sizes (Ullman 2006, Schreiber et al 2006 and Barrett 2006). For the purposes of this study, Chi-square, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) were used to test model fit. Further information on these techniques and others can be found in Schreiber et al. 2006, Ullman 2006, Barrett 2006, Bentler et al. 1999 or Hu et al 1998.

The use of SEM as an exploratory technique arises from iterative attempts to populate the theorized model. Because repeated steps are involved, as fit is evaluated, failure occurs and iteration drives the model towards a better representation of the relationships present within the environment. Within the MPLUS7 software package, residual values are provided in tabular format allowing the user to evaluate the fitness of variables. In ideal cases, reconstruction of the conceptual model can improve fit, reduce residuals, and lead to an improved understanding of the environment.

The village of Enseado Do Aritapera is located within the municipality of Santarem in the state of Para, Brazil (2.4300 S, 54.7200 W). During the rainy season, the village is flooded for approximately three months (March, April, and May). During this time, all water is consumed from the Tapajos River directly. As the flood waters recede, wells and springs are used to provide access to shallow ground water. A native translator was recruited to provide assistance with the field study. For each group of homes, a visit was made to the local political leader, and he or she was recruited to assist in the identification of twenty 'random' homes. While the households were selected at random, it was necessary to use the assistance of the local political leader to create a list of households; hence while random the potential for bias could exist within household selection. Samples of household potable water and water supply were collected and analyzed in the field, and an interview was performed with an adult member of each household.

The field survey has been published previously (Divelbiss et al, 2013). Modifications to the survey were included based upon the details of the study site. A total of 55 questions were created, and each household was surveyed with the assistance of a translator. A complete copy of the survey is provided as a supplementary document.

To analyze the data collected from the survey, two different approaches were used. The first approach used SEM as an exploratory technique. However, with a small sample size the method of bootstrapping was used to provide a statically appropriate sample size. As expected, bootstrapping created variability within the data so multiple (60) bootstrapped samples were analyzed and the results were aggregated for reporting. This approach allowed for general analysis of the model, however, it should be noted that the use of bootstrapping with SEM should be avoided.

The second approach used an analysis of a correlation matrix. When working with latent variables, the correlations between the observable variables that collectively represent the latent variable need to correlate strongly (as compared to other correlations) and consistently (+/- .2 range) (Grace 2006). It is also possible to assess trends in correlations with other observable variables outside of the individual latent variables. While the statistical significance of these numbers can be found following most general statistical literature. But again, caution is needed correlation may not be robust, the observed trends provide corroborating support for the results observed through bootstrapping. Collectively, these results can be used in aggregate to test the validity of the conceptual model in SEM.

Results

A graphical representation of the conceptual model used previously by Divelbiss et al (2013) is provided in Figure 1. This graphical representation is based upon the following hypotheses:

- Increased household educational level has a negative effect on the severity of diarrhea burden
- Increased socio-economic status has a negative effect on the severity of diarrhea burden

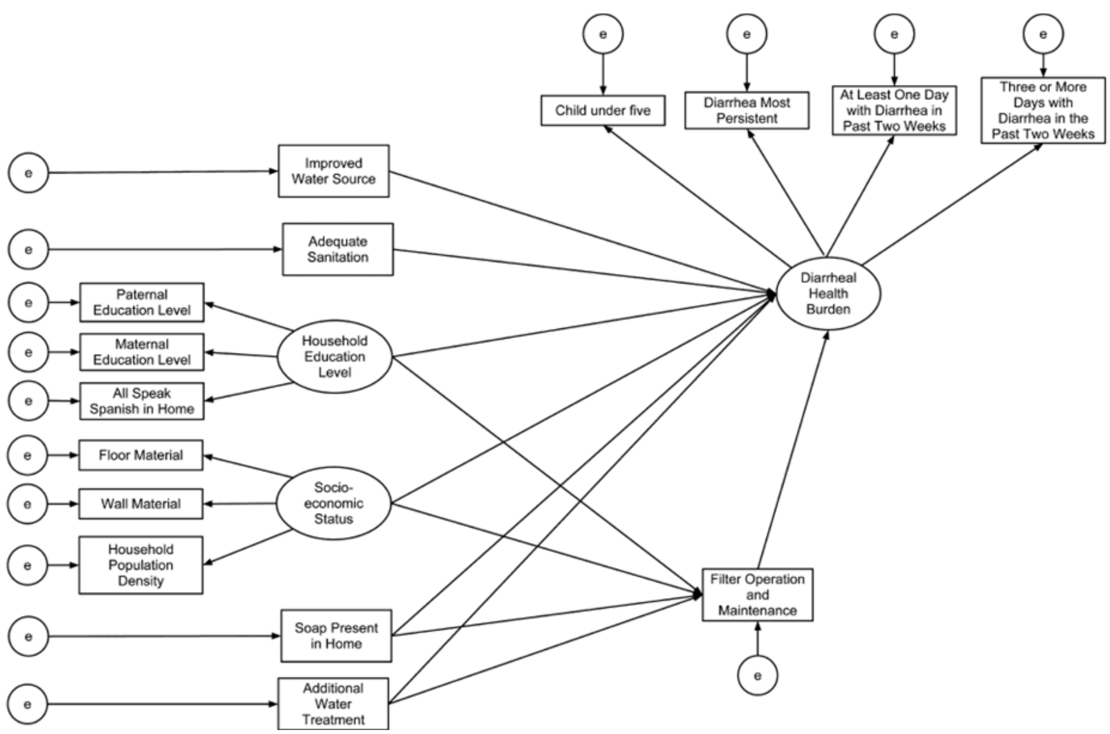


Figure 1. Divelbiss et al. 2013 SEM model used in Guatemala

- Poor hygiene practices have a positive effect on the severity of diarrhea burden
- Additional water treatment beyond the filter has a negative effect on the severity of diarrhea burden
- Access to an improved water source has negative effect on the severity of diarrhea burden
- Access to adequate sanitation has a negative effect on the severity of diarrhea burden
- Proper water storage has a negative effect on the severity of diarrhea burden

Limited research has been performed to identify filter operation and maintenance characteristics that have a specific response to various household characteristics. Therefore the following hypotheses were suggested based on expectations of how certain household characteristics would affect filter operation and maintenance, and findings in the literature associated with the impacts on diarrheal burden:

- Increased socio-economic status has a positive effect on filter operation and maintenance
- Better personal hygiene practices have a positive effect on filter operation and maintenance
- Additional water treatment has a positive effect on filter operation and maintenance
- Increased household educational Level has a positive effect on filter operation and maintenance
- Proper filter operation and maintenance is expected to have a negative effect on diarrhea burden due to the results from case-control studies.

Observations of the homes in Enseado Do Aritapera, Para, Brazil suggested that the approach used in Guatemala required modification. For example, due to the location of the study, the conceptual model was modified to incorporate the hypothesis:

- Access to an improved water source has a negative effect on the severity of diarrhea burden

The reason behind this hypothesis was the living situation of the villagers. All houses were erected on stilts, and at the time of the administration of the survey, each of the homes sites was flooded by the Tapajos River. Canoes and larger river boats were the only access villagers had to transportation. Flooding limited villagers in terms of the selection of drinking water source. Using observations from the field, a modified conceptual model was created (Figure 2.).

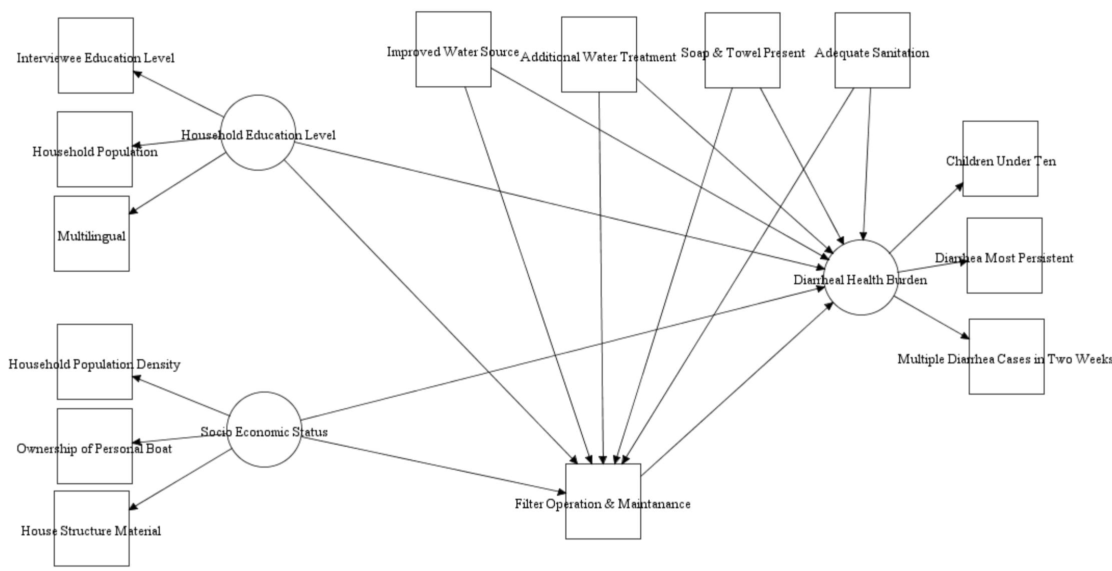


Figure 2. Altered SEM model for use in Brazil

The oral survey instrument previously developed by Divelbiss et al was modified as follows:

- only one parent was asked about education level,
- the languages spoken within the house were changed to fit Portuguese and local dialects and
- an additional question on household population was constructed.

As part of the determination of socioeconomic status, the ownership of a personal boat was added to help represent the latent variable. The observable variables for DHB (diarrheal health burden) changed slightly in hopes of allowing for a more significant correlation. Children under the age of ten were noted, and diarrhea lasting more than one day was used as a variable. In looking at the independent observable variables the only change made was to the ‘Soap Present in Home’ as the use of soap was promoted by the village leaders already within the community. An additional requirement was added in the presence of a towel. These changes were made before data collection began and then reassessed after.

The conceptual model presented in Figure 2. was modified after bootstrap analysis with the oral survey data. The final model is presented in Figure 3. From this figure the direct effects can be viewed represented by an arrow and a number. The number can be interpreted as with a one point increase in the variable in which the arrow originated the correlation factor gives the increase or decrease of the variables to which the arrow reaches to. For example, as AWT (additional water treatment) increases by one point, FOM (filter operation and maintenance) will increase by 0.3 points, therefore having a moderately positive effect on FOM. A list of the twelve hypotheses, the direct/indirect effect and correlation factor associated with the relationship, can be found

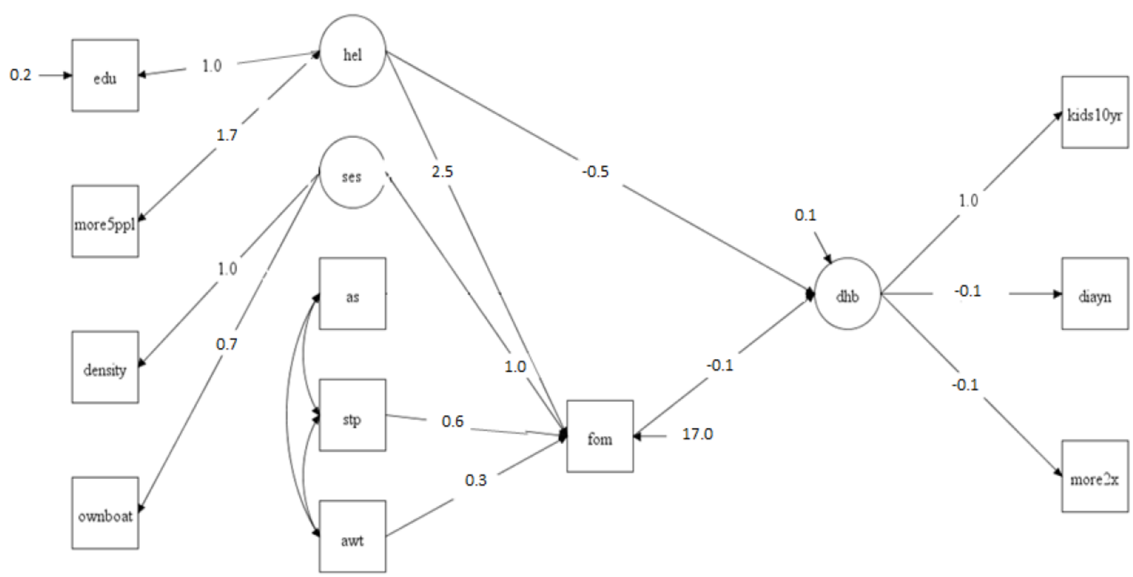


Figure 3. Fitted model that shows numeric (standardized) direct effects between variables using bootstrapped data.

in Table 1. The latent variables are represented by circles and the observable variables are represented by squares. In the model the strongest correlation between two variables is the effect the household education level (HEL) has on the filter operation and maintenance (2.5 point increase). Socio-economic status (SES) has the second largest effect amongst all relationships with a 1.0 factor increase on FOM.

For the DHB variable, the strongest correlation is that of HEL on DHB with a -0.5 factor. A negative factor in this scenario means that as the education level increases within the household the diarrheal burden decreases. Similarly with FOM, a -0.1 factor is found to directly affect the diarrheal burden in the house.

Evaluation of the latent variables was also performed. The representation within HEL and SES variables only utilize two observable variables. The recommended minimum is three observable variables with similar correlation factors (Grace 2006). The latent variable DHB has three representative observable variables, however the correlation factors are not strong indicators of good representation. Aware of these concerns, for the purposes of using SEM as an exploratory technique the trends in correlations are presented in Table 1.

Table 2. shows the correlation matrix that was computed from the data utilizing Excel 2007. The highlighted green squares show the relationships that are significant. Significant meaning the value is greater than $2/n^{1/2}$, where $n = 20$. This means that any value that is found to be greater than 0.447 can be thought of as significant (Walsh). It can be observed that the strongest correlation was between IWS (improved water source) and diarrhea. This correlation was negative supporting what is already known, ‘drinking contaminated water results in more diarrhea.’ The second highest correlation is between being able to afford water being pumped to the house and IWSt (improved water storage) within the house. This is a positive correlation. Of the six significant correlations found in the matrix, half of them were associated with the SES of the household. Of the six SES variables assessed within the survey, all had an average negative correlation with DHB. Of the 18 total correlations between the factors associated with SES (6) and factors associated with DHB (3), 16 had negative factors, four of which were of low to moderate significance. These are identified in Table 2 by an outlined box.

With the limited data, residuals were not available, however, improvement factors to model could be assessed through the correlation matrix. A pattern of consistent, moderate to high significance levels within each of the three latent variables within the model were assessed. The lack of representation within the latent variable HEL is evident in only two poorly correlated observable variables. The observable variables for DHB had better correlation but still needs improvement with the possibility that an additional variable might be needed. SES had the best representation in its observable variables, however, an increase in the correlation factors between three or four variables is needed.

Table 1. Results from physical MPLUS7 model.
[a] Original hypotheses established from previously collected data, [b] Answer to hypotheses after data analysis,
[c] If hypothesis was answered through compound relationships, [d] Path coefficient

	Hypothesis[a]	Agreed/Disagreed[b]	Direct/Indirect Effect[c]	Correlation Factor[d]
1	Increased household education level has a negative effect on the severity of diarrhea burden	Agreed	Direct	-0.5
2	Increased socio-economic status has a negative effect on the severity of diarrhea burden	Agreed	Indirect	SES to FOM to DHB
3	Poor hygiene practices have a positive effect on the severity of diarrhea burden	Agreed	Indirect	STP to FOM to DHB
4	Additional water treatment beyond the filter has a negative on the severity of diarrhea burden	Agreed	Indirect	AWT to FOM to DHB
5	Access to an improved water source has a negative effect on the severity of diarrhea burden	Not Analyzed	N/A	N/A
6	Access to adequate sanitation has a negative effect on the severity of diarrhea burden	Not Significant	N/A	N/A
7	Proper water storage has a negative effect on the severity of diarrhea burden	Agreed	Indirect	In FOM compellation score
8	Increased socio-economic status has a positive effect on filter operation and maintenance	Agreed	Direct	1.0
9	Better personal hygiene practices have a positive effect on filter operation and maintenance	Agreed	Direct	0.6
10	Additional water treatment has a positive effect on filter operation and maintenance	Agreed	Direct	0.3
11	Increased household educational level has a positive effect on filter operation and maintenance	Agreed	Direct	2.5
12	Proper filter operation and maintenance is expected to have a negative effect on diarrhea burden	Agreed	Direct	-0.1

Table 2. Correlation matrix from raw data. Green is latent variable HEL, orange is latent variable SES, and brown is latent variable DHB.

	Education	# of ppl	Density	IR	Person. Boat	Trash	WSJ	Retrieval	FO&M	Persistent	2 Weeks	Kids < 6	IMS Score	IWS Score	IWT Score	IWSt Score
Interviewee Education Level	1															
# of People in House	-0.1705	1														
Density = # of Bedrooms / person	0.0836	-0.7880	1													
Improved Roof	0.3158	-0.1595	-0.0482	1												
Onwership of Personal Boat	-0.5151	0.0685	0.1445	-0.1873	1											
Trash Disposal	-0.3126	-0.4097	0.3346	-0.1683	0.4708	1										
Type of Water Storage Jar (WSJ)	0.1325	-0.3738	0.2540	0.1325	-0.2357	0.1816	1									
Water Retrieval (Walk or Pumped)	0.1683	-0.3595	0.1968	0.0721	0.0428	0.3407	0.4237	1								
FO&M	0.1543	-0.3423	0.2818	0.0171	0.1119	0.2116	0.4747	0.0548	1							
Diarrhea Most Persistent	-0.0468	0.1664	-0.3902	-0.1873	-0.2500	-0.1712	-0.2357	0.0428	-0.4475	1						
Diarrhea Multiple Times in 2 Weeks	0.1325	0.0692	-0.1809	-0.1325	0.0000	-0.3026	-0.3333	-0.3026	-0.3308	0.4714	1					
Kids Under Six	0.1001	0.4029	-0.4324	-0.1001	-0.1336	-0.2059	-0.1260	-0.2059	0.1142	-0.1336	0.1260	1				
Improved Sanitation Score	0.2140	-0.2014	0.3090	0.1751	-0.4155	-0.3290	0.3428	-0.1512	0.0275	-0.2424	0.1469	0.0555	1			
Improved Water Source Score	0.1705	0.1799	-0.1223	0.1395	-0.0552	-0.0496	0.1951	-0.1913	0.1061	-0.6069	-0.1170	0.1917	0.2808	1		
Improved Water Treatment Score	-0.2075	0.1711	-0.1760	-0.0231	-0.0821	0.1370	0.5222	0.1370	0.4082	-0.0821	-0.4062	0.2851	0.0767	-0.1426	1	
Improved Water Storage Score	0.1792	-0.2779	0.0365	0.1195	0.0266	0.1774	0.2255	0.5868	0.0097	-0.1063	0.0752	-0.1989	-0.0110	0.3431	-0.1439	1

Discussion and Conclusion

Table 1 shows the original hypotheses and whether the data agreed or disagreed with each one. Only two of the twelve either were not analyzed or had no significance. The data shows that education plays a significant role within both filter operation and maintenance in the house but also plays a role in a decrease in diarrheal health burden within the house. As preventing sickness through consuming unclean water (ie fluids) is only one of the five major ways to contract diarrhea, it is not surprising that a general higher education level helps in reducing other areas of potential contamination. SES within the household was the second most important factor in proper use of the BSF. The more money people had in this village correlated with the ability to correctly take care of the filter. It did not, however, correlate with a significant decrease in diarrhea found within the household. The SES indicator did have issues with large variances within the bootstrapped data sets and warrants further consideration in a larger data set. Interestingly, the significance of education within the household was more important to decreasing diarrheal issues within the house then the presence of the BSF. This result is consistent with Divelbiss, and suggests that while access to a BSF may be helpful, there are benefits to other types of interventions if the ultimate objective is to reduce the occurrence of diarrhea illness.

Table 2 presents the correlation matrix as discussed in the results. The matrix provides more information regarding the SES relationships that were not significant in the model. The fact that SES observable variables were present within half of the significant correlations overall as well as having a negative correlation with 16 of 18 diarrheal correlations warrants further study of the factor SES plays in the health of a household. The UNICEP and many non-governmental organizations focus a great deal on water improvements; this is reinforced in the research project as an important topic. But, arguably equally as important is the issue of a household's socio economic status.

In comparing these results to Divelbiss' work in Guatemala we see similarities in the models. The importance of education and its effect on the overall diarrheal health burden is echoed in both projects. As well as the importance of proper household hygiene in the form of proper hand washing resources. Again, we see that SES is a non-significant factor within Divelbiss' model, but this warrants further analysis in the raw data as was apparent in this project.

From these results we see the importance of quality water within the health of the household. But, we are also made aware of the factor that education and socio economic status play within the health of the household. A greater push for education as well as job opportunities and resources to help people help themselves also plays a significant role within the larger picture of development aid.

References

Aiken, B. A.; Stauber, C. E.; Ortiz, G. M.; Sobsey, M. D. An assessment of continued use and health impact of the concrete biosand filter in Bonao, Dominican Republic. American Journal of Tropical Medicine and Hygiene 2011, 85 (2), 309−317.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. Science Direct, Personality and Individual Differences 42 (2007) 815-824

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. Multivariate Behavioral Research, 34,181–197.

Black, R. E.; Morris, S. S.; Bryce, J. Where and why are 10 million children dying every year? The Lancet. 2003, 361, 2226−2234.

Christensen, R. (1996). Analysis of Variance, Design, and Regression: Applied statistical methods. Boca Raton, Florida: Chapman & Hall/CRC. 1st Ed.

Divelbiss, D. et al, 2013, "Environmental Health and Household Demographics Impacting Biosand Filter Maintenance and Diarrhea in Guatemala: An Application of Structural Equation Modeling'," Environmental Science and Technology 47(3):1638-1645 (and supplementary electronic information).

Fabiszewski de Aceituno, A. M.; Stauber, C. E.; Walters, A. R.; Meza Sanchez, R. E.; Sobsey, M. D. A randomized controlled trial of the plastic-housing biosand filter and its impact on diarrheal disease in Copan, Honduras. American Journal of Tropical Medicine and Hygiene
2012, 86 (6), 913−921.

Grace, J.B. (2006). Structural Equation Modeling and Natural Systems. New York: Cambridge University Press.

Hu, L., & Bentler, P.M. (1998). Fit indices in covariance structural equation modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3, 424–453.

Iacobucci, D., (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. Science Direct, Journal of Consumer Psychology 20 (2010) 90-98

Kubare, M. and Haarhoff, J. (2010). Rational design of domestic biosand filters. Journal of Water Supply: Research and Technology 59(1):1–15.

Ozkan, S.; Tüzün, H.; Görer, N.; Ceyhan, M.; Aycan, S.; Albayrak, S.; Bumin, M. Water usage habits and the incidence of diarrhea in rural Ankara, Turkey. Transactions of the Royal Society of Tropical Medicine and Hygiene 2007, 101 (11), 1131−1135.

Palmateer, G. , D. Manz, A. Jurkovic, R. McInnis, S. Unger, K.K. Kwan, and B.J. Dutka. (1999). Toxicant and parasite challenge of Manz intermittent slow sand filter. Environmental Toxicology. 14.2: 217-225.

Schreiber, J.B., Stage, F. K., King, J., Nora, A., Barlow, E.A., (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of Educational Research. July/August 2006 [Vol. 99(No. 6)].

Stauber, C. E., Elliott, M. A., Koksal, F., Ortiz, G. M., DiGiano, F. A., and Sobsey, M. D. (2006). Characterisation of the biosand filter for e. coli reductions from household drinking water under controlled laboratory and field use conditions. Water Science & Technology 54(3):1+.

Stauber, C. E., Ortiz, G. M., Loomis, D. P., and Sobsey, M. D. (2009). A randomized controlled trial of the concrete biosand filter and its impact on diarrheal disease in bonao, dominican republic. The American Journal of Tropical Medicine and Hygiene 80(2):286–293.

Tiwari, S.; Schmidt, W.; Darby, J.; Kariuki, Z.; Jenkins, M. Intermittent slow sand filtration for preventing diarrhoea among children in Kenyan households using unimproved water sources:
Randomized controlled trial. Tropical Medicine & International Health 2009, 14 (11), 1374−1382.

Ullman, J. B., (2006). Structural Equation Modeling: Reviewing the Basics and Moving Forward. Journal of Personality Assessment, 87(1), 35-50

Walsh, D. A Simple Rule of Thumb for Statistically Significant Correlation. Middle Tennessee State University.

II. UTILIZING STRUCTURAL EQUATION MODELING AS AN EVALUATION TOOL FOR CRITICAL PARAMETERS OF THE BIOSAND FILTER IN A PILOT STUDY IN PARA, BRAZIL

Abstract

Biosand filters (BSF) have brought potable water to many communities throughout the world. A significant number of studies have shown the effectiveness of the filter if utilized properly. However, Divelbiss et al. 2013 investigated the BSF from a holistic view of the community. Results showed, while important, the BSF was not among the significant factors for health improvement. This pilot study, as part of a larger case study, investigated the issues behind these findings. Utilizing the SEM model that explained the complex system of interactions within the Guatemalan communities, several Brazilian communities along the Amazon River were studied. While further research is ongoing, it was found that the BSF, again while important, was not the largest factor in the improvement of health within the household. Socio-economic status and improved water source had the largest positive impact on health improvement. It was also found that additional water treatment had the highest negative effect of the usage of the BSF.

Introduction

The biosand filter (BSF) has been utilized within developing regions for over 200 years (CAWST 2012). A strong body of literature has been established on the effectiveness of the biosand filter, in particular the BSF produced by Centre for Affordable Water and Sanitation Technology; CAWST (Stauber, C. et al. 2009; Tiwari, S. et al. 2009; Aiken et al. 2011; Vanderzwaag et al. 2009). Many of these studies show that in a controlled setting, whether in a laboratory or in the field, a removal of at least 1 log is achievable for a majority of harmful contaminants within water. However, this removal is highly dependent on the users understanding of operation and maintenance of the BSF. The BSF can be found in over 60 different countries within 400,000 different households (CAWST 2012). Over this spectrum of communities that are found within these many countries, numerous cultural understandings, education levels, economic statuses, health issues, and other societal issues can be found. Divelbiss et al. 2013 studied one particular set of communities located in northern Guatemala. The studies focus was on the particular intervention, distributing BSFs, and the impact this had on the ‘multidimensional’ issue of poverty. The primary hypothesis of the BSF having the largest effect on improving community member’s health, was not found. Divelbiss reports, “the community is a complex system of interactions which directly and indirectly influence the health of its residents. Policy makers and development practitioners must recognize that single target interventions (e.g., improving water quality) have limited influence on the entire system (Divelbiss et al. 2013).” This study, being limited to northern Guatemala, warranted further investigation.

A secondary study site was selected in the region of Para, Brazil. This site allowed for significant cultural and locational differences. Within this case study, a feasibility and pilot study have been administered by Voth-Gaeddert and colleagues. Voth-Gaeddert et al. 2013 successfully demonstrated the adaptability of the Structural Equation Modeling (SEM) model that described the complex system of interactions within the Guatemalan communities to the Brazilian communities found along the Amazon River. In September of 2013 a pilot study was conducted amongst the communities in the Para region. This region was selected because a Christian organization known as Project Amazon (PAZ) distributed over 10,000 BSFs among rural villages within this area.

Divelbess used the statistical platform SEM to describe the complex system of interactions within the communities. Factors such as filter operation and maintenance (FOM) of the BSFs, household education level (HEL), socio-economic status (SES), diarrheal health burden (DHB), household sanitation, and water consumption habits could be assessed simultaneously. SEM can be either exploratory or confirmatory in its analysis approach and offers a unique combination of features. These include hidden or latent variables (ex. SES in Brazil could be represented by what material the house is made from, ownership of a boat and number of working light bulbs in the house), simultaneous regression analysis, non-summation of errors within aggregated indicators, and a robust graphical interface. Using SEM, assessments of interventions within communities are possible. The significance of understanding what factors have the largest effect within a community is crucial to a successful intervention approach.

Methodology

Two goals were specified within this pilot project, the first was to confirm what was found by Voth-Gaeddert et al. 2013 showing feasibility in the transition of a region specific survey and SEM model used in Guatemala and the application of these two tools to a region in Brazil. The second goal was to use the revised survey in the first iteration of the SEM model in an exploratory style of analysis. This survey would populate an SEM model that had been revised after the feasibility study performed by Voth-Gaeddert. However, utilizing SEM for exploratory style of analysis requires a full understanding of the techniques involved.

A devotion to a particular set of steps is crucial to producing statistically accurate SEM models. SEM has a broad range of abilities; this study focuses on a single particular ability. The reader is encouraged to study Grace 2006 or Kline 2005. In this particular usage of SEM further care has to be taken due to the use of exploratory techniques and a balance needs to be taken when working with both Exploratory and Confirmatory Factor Analysis (EFA and CFA, respectively) (Grace 2006 and Kline 2005). CFA is, as the name suggests, a confirmatory technique- it is theory driven. Therefore, the planning of the analysis is driven by the theoretical relationships among the observed (indicator) and unobserved (latent) variables. When CFA is used, one hypothesized model (outcome of Voth-Gaeddert et al. 2013) is being compared to a data driven (observed) model using the covariance matrices of both for analysis. A minimization of the difference between matrices is the goal of CFA and is tested through Chi-Square tests as well as others (Schreiber et al. 2006). CFA is used in SEM in the first step of a two-step process. A model of only the latent variables and their indicator variables is created to test for fit within the latent variables; this is called the measurement model (see Figure 1.). If the measurement model fits according to the fit indices (see Hu, L. et al. 1998; Bentler, P.M. 2007) then the structural model can be assessed.

A structural model describes the relationships between the latent variables as well as other independent observables variables that are hypothesized to have effects within the model. Again, simultaneous regression analysis is used to represent these relationships by way of linear regression (for continuous variables) or log regression (for categorical variables). In Table 1. a list of all variables within the model can be found with the type of variables (continuous or categorical) and a brief definition.

EFA is outside the scope of this paper, however it can be an effective tool when relationships between factor and indicator variables are unknown (see Grace 2006). Once a model has shown ‘adequate fit’ and a new set of data has been used to populate and confirm adequate fit, the path coefficients can be assessed. Looking at the unstandardized path coefficients, direct and indirect effects can be analyzed. The path

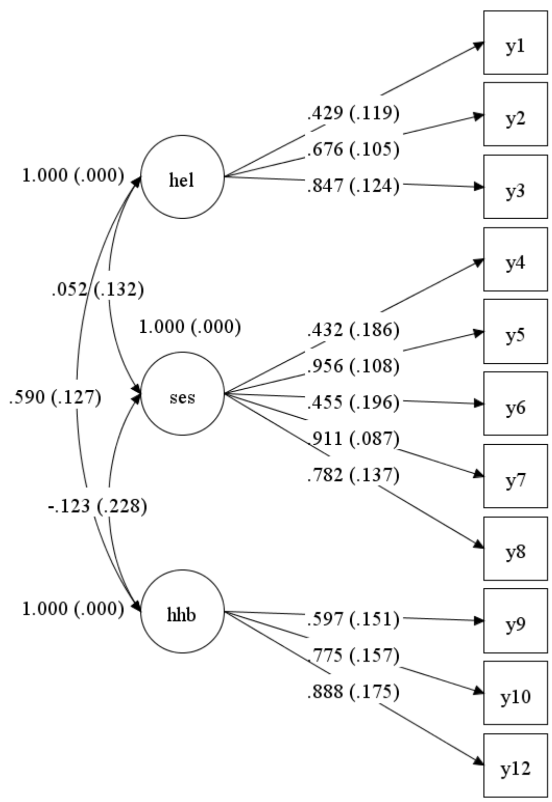


Figure 1. Measurement Model with parameter estimates, definitions in Table 1

coefficients can be interpreted as unit increases between variables. For example in Figure 2 a direct path from SES to HHB can be seen with a coefficient of -.115. This means that as SES increases by one unit (an increase in “wealth”), HHB will decrease by .115 units (improved health). However, it is stressed within the literature that causal assumptions are held for the reader to determine but can be discussed within the discussion section of a report (Kline 2005).

A significant part of this style of SEM is testing model fit. A large body of work is available for study on this subject as there are differing opinions on particular tests. However, generally, it is suggest to use Chi-Square (Bentler, P.M. 2007), root mean square error of approximation; RMSEA (Hu, L. et al. 1998), Comparative Fit Index; CFI (Hu, L. et al. 1998), and Tucker Lewis Index; TLI (Muthen & Muthen 2012). The software package MPLUS 7 was used for analysis of data (Muthen & Muthen 2012).

A continued awareness and analysis of cultural differences is crucial when working with multiple global locations. For this reason, changes to the survey and the SEM model from Voth-Gaeddert et al. 2013 were made as a deeper understanding of cultural influences were discovered. These changes occurred mainly within the observable indicator variables used to represent the latent variables in the model. All changes were made either before data collection began or at the time of transition from confirmatory analysis to exploratory analysis. Understanding impacts of BSFs in communities around the world warrants a continued devotion to cultural observations.

Table 1. List of variables and their type

Variable	Definition	Type
Y1	Reading and Writing Score	Ordinal
Y2	Paternal Education Level	Ordinal
Y3	Maternal Education Level	Ordinal
Y4	Improved Roof	Dictomous
Y5	Owns a TV	Dictomous
Y6	Owns a Shower	Dictomous
Y7	Has Multiple Working Lightbulbs	Dictomous
Y8	Improved Water Storage Material	Dictomous
Y9	Severity of Sickness	Ordinal
Y10	Occurrence of Sickness	Ordinal
Y11	Wait Period of Sickness and Doctor Visit	Ordinal
Y12	Children Under 11	Dictomous
Y13	Filter Operation and Maintenance	Ordinal
Y14	Improved Sanitation	Dictomous
Y15	Improved Water Source	Dictomous
Y16	Improved Water Treatment	Dictomous
Y17	Improved Water Storage	Dictomous

Results and Discussion

The following changes were conducted due to cultural observations made while in the field. Variables are listed and are then followed by an explanation:

- HEL observable variable 'Reading and Writing' test was administered

The previous SEM model only used two observable variables to describe HEL which in the literature is said to have a weak representation. Three or more observable variables are recommended to help explain the variance within the HEL factor. A reading and writing test does not allow the interviewee to hide illiteracy and helps accurately describe education of that household.

- SES observable variables of ownership of a TV, shower, working light bulbs, material of water storage container, and 'improved' roof, along with others were collected

A wide spectrum of SES indicators can be found in the literature. This range includes using income level, type of job, net worth of physical objects found in the household, type of objects in household or a combination of these (Jones, E.C. et al 2011; Dressler, W.W. 1998). Further study is needed within this particular subject on strength of indicators. Physical objects found in the household were the main indicators for SES within this study.

- DHB was changed to Household Health Burden (HHB) which included severity and type of illness

This transition from DHB to HHB was due to a lack of occurrence of diarrhea during the low season of the Amazon River. This could have been due to the location of latrines. During the rainy season, the Amazon River floods the latrines and carries the contents downstream past other villager's households. The top health issues were recorded for each household and these were used to rank the level of the adverse health burden found in the given household.

- Independent variables were changed; 'Soap & Towel Present' to 'Improved Storage' and definition of terms changed

While the importance of soap and towel within a household remains important, it was found that over 90% of the sample population had these items. This term was then incorporated into the 'Improved Sanitation' score partly defined by WHO (WHO 2006). After further investigation the variable 'Improved Storage' was added to test the hypothesis that if a household stores their water properly, the occurrence of diarrhea within the household will go down. Additionally definitions of 'Additional Water Treatment' and 'Adequate Sanitation' were changed to 'Improved Water Treatment' and 'Improved Sanitation' to be consistent with WHO terminology (WHO 2006).

From these changes a new theoretical model was created and data from 41 households were used to populate the altered hypothetical model. A measurement model was first run to assess the fit of observable indicator variables on their respective latent variables. Failures in model identification occurred due to poor explained variance within particular factors. After assessment of raw data and of MPLUS 7 output data (including parameter estimates, residuals, and variances) issues were highlighted and further literature reviews were done. As change in the model should not be data driven alone, alternative models were assessed based off of literature reviews (Guest, G. 2000; Undurraga, E.A. et al. 2010). The process then becomes exploratory and any further analysis should be done with caution, however it can still provide productive insight to the model.

The alternative model that had the best fit ($\chi2$= 54.5 , p>.077; RMSEA <.171; CFI = .92 ; TLI = .89) is shown in Figure 1. The model showed close to adequate fit (p>.05, <.08, .90, .90, respectively); however it needs to be noted that with a small sample size the fit indices are less accurate. The structural model could then be added to the measurement model to create the full model. Again, adjustments had to be made with relationships inside the model. The alternate model that showed the best fit still fell far short of the goodness of fit tests ($\chi2$= 54.5 , p>.0012; RMSEA <.01; CFI = .62 ; TLI = .52). However, even with a failed model, some important insights can be gained.

Even with an exploratory style of analysis observations can be gained through both the model failures as well as the final "best fit" model. These observations are;

- HEL needs the addition of indicator variables to improve representative strength, these could include reported reading habits, vocational classes taken or attendance of school for children
- SES was represented the best out of all factors, however, ownership of a shower and TV were problematic, so consideration of the usage of overall income and occupation could improve representation
- SES had a negative effect with HHB (improved health)
- HHB indicator relationship between 'severity of sickness' and 'children under 11' could have been an issue
- HHB may be changed to Activities of Daily Living (ADL) or International Classification of Functioning, Disability and Health (ICF) depending on findings in ongoing research
- FOM had negligible effect on HHB
- 'Improved Water Treatment' has highest negative effect on FOM (worsened use of filter)
- 'Improved Source' had highest negative effect with HHB (improved health the best)
- Poor definitions within the survey of 'Improved Source' may have led to some of the fit issues in the full model

It is important to provide discussion of these observations from a holistic view of the larger body of work between Guatemala and Brazil. Therefore it can be observed that HEL was not well represented within this study. This may have been the main issue in the lack of fit of the full model. The idea that an increase in education will indirectly affect the filters ability to improve overall health within the house was uncertain. Would an increase in education cause them to use alternative methods to filter water; such as using chlorine or a ceramic filter? An improved representation of HEL will allow for further analysis of this question in the next study.

SES was represented well; however, even with “good fit,” proper representation (selecting the ‘best’ indicators) is the goal. SES was shown (in Figure 2.) to have the second largest negative effect on HHB and to also have a negative effect on FOM. This could suggest that as “wealth” increases within the household alternative methods are used to acquire clean drinking water; such as buying bottled water or chlorine. This would create a negative effect on FOM but possible improve the overall health within the house.

The “Improvement” variables within the model (Y14-Y17) were all regressed on HHB but only Y15 and Y16 of these four were regressed on FOM. Improved Water Treatment had the largest negative effect on FOM. This suggests that as other treatment methods are utilized within the house, the operation of the BSF in the house decreases. A reexamination of the raw data shows that 20% of household used chlorine and of those users of chlorine, 40% used it incorrectly. The main fallacy that was found was the usage of chlorine before filtration through the BSF. This would compromise the biofilm layer that is an integral part of the contaminant removal process. Improved Sanitation and Improved Source (of water) had a positive impact on health issues within the households.

Finally, FOM had a near negligible effect on HHB. This could have occurred for several reasons. The most likely is poor representation within the latent variable HHB or the cumulative score associated with FOM. Further investigation is needed to help shed light on this issue.

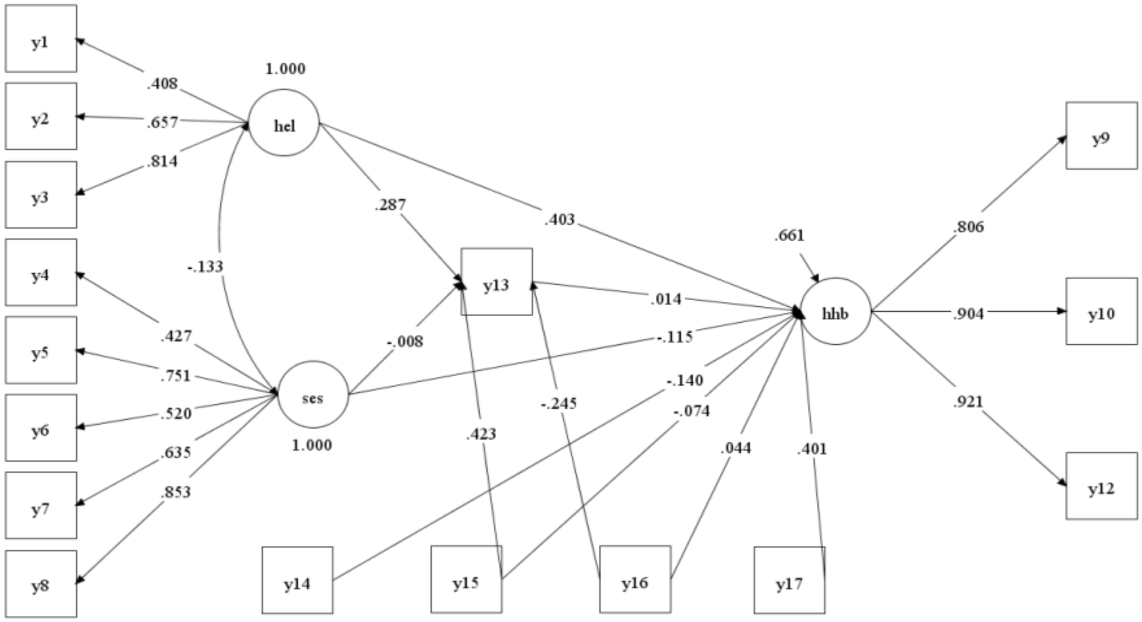


Figure 2. Full Model with standardized parameter estimates

Conclusions

The second iteration within this study allowed for a significant step to identifying the correct SEM model to represent the environment within households of this region. As outlined in the Results and Discussion section, focus on strengthening representation within latent variables will significantly improve the scores to the goodness of fit tests. The survey used in data collection will be adjusted accordingly. As with all SEM models, a deeper understanding of the scenario that is being modeled will increase the ability of the model to properly represent it. SEM has good potential to serve as a platform for a more standardized approach in analyzing situations within impoverished areas, however further analysis is needed to show what can be accomplished through SEM.

References

Aiken, B.A., C. E. Stauber , G. M. Ortiz , and M. D. Sobsey, (2011). An Assessment of Continued Use and Health Impact of the Concrete Biosand Filter in Bonao, Dominican Republic. American Journal of Tropical Medicine and Hygiene, 85(2):309–317

Bentler, P.M. (2007). On tests and indices for evaluating structural models. Personality and Individual Differences 42:825-829.

Centre for Affordable Water and Sanitation Technology (2012). Biosand Filter: History of the Biosand Filter. Cawst.org/en/resources/biosand-filter

Divelbiss, D. et al. (2013). Environmental Health and Household Demographics Impacting Biosand Filter Maintenance and Diarrhea in Guatemala: An Application of Structural Equation Modeling. Environmental Science and Technology 47(3): 1638-1645.

Dressler, W.W., Balieiro, M.C., Santos, E.D. (1998). Culture, Socioeconomic Status, and Physical and Mental Health in Brazil. Medical Anthropology Quarterly. 12(4):424-446

Grace, J.B. (2006). Structural Equation Modeling and Natural Systems

Guest, G. (2000). Using Guttman Scaling to Rank Wealth: Intergrating Quantitative and Qualitative Data. Field Methods. 12(4):346-357.

Hu, L., Bentler, P.M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. American Psychological Association. (3)4:424-453

Jones, E.C. et al. (2011). Inequality, Socioeconomic Status, and Social Support in Post-Disaster Mental Health in Mexico. Human Organization. 70(1).

Kline, R.B. (2005). Principles and Practice of Structural Equation Modeling.

Muthen, L.K., Muthen, B.O. (1998-2012). MPLUS User’s Guide. Seventh Edition. Los Angeles, CA: Muthen & Muthen

Schreiber, J.B. et al. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of Educational Research. Heldref Publications.

Stauber, C.; Ortiz, G.; Loomis, D.; Sobsey, M. A randomized controlled trial of the concrete Biosand filter and its impact on diarrheal disease in Bonao, Dominican Republic. The American Journal of Tropical Medicine and Hygiene 2009, 80 (2), 286−293.

Tiwari, S.; Schmidt, W.; Darby, J.; Kariuki, Z.; Jenkins, M. Intermittent slow sand filtration for preventing diarrhoea among
children in Kenyan households using unimproved water sources: Randomized controlled trial. Tropical Medicine & International Health 2009, 14 (11), 1374−1382.

Vanderzwwaag, J.C., Atwater, J.W., Bartlett, K.H., Baker, D. (2009). Field Evaluation of Long-Term Performance and Use of Biosand Filters in Posoltega, Nicaragua. Water Quality Research Journal of Canada. 44(2):111-121.

Voth-Gaeddert, L.E., Divelbiss, D., Oerther, D.B. (2013). Utilizing Structural Equation Modeling to Correlate Biosand Filter Performance and Occurrence of Diarrhea in Enseado Do Aritapera, Para, Brazil. International Water Association Brazil.

Undurraga, E.A. et al. (2010). Individual Wealth Rank, Community Wealth Inequality, and Self-Reported Adult Poor Health. Medical Anthropology Quarterly. 24(4):522-548.

WHO (2006). Core Questions on Drinking-Water and Sanitation for Household Surveys. United Nations Children's Fund.

III. UTILIZING STRUCTURAL EQUATION MODELING IN THE DEVELOPMENT OF A STANDARDIZED INTERVENTION ASSESSMENT TOOL

Abstract

There are numerous approaches to measuring multidimensional poverty; these include the Human Development Index and the Multidimensional Poverty Index among others [1]. However, a gap in the literature is found when intervention assessment tools are investigated. The idea of creating a standardized assessment tool would allow for a deeper understanding of poverty on a per community basis. Structural Equation Modeling (SEM) offers a robust platform in which to establish such a tool. An overview of SEM and several other general approaches to data aggregation are addressed. The notion of a standardized intervention assessment tool is discussed; this is focused on utilizing the SEM platform for this tool. Further, previous works by Divelbiss [2] and Voth-Gaeddert [3], [4] are discussed. To date SEM has shown to handle adaptability of differing environments positively. Divelbiss reported on the SEM multivariable poverty model within villages of Guatemala and Voth-Gaeddert reports on the applicability of this model used in a dissimilar environment in Brazil. These findings suggest feasibility in the utilization of a SEM platform for a standardized intervention assessment tool.

1. Introduction

The Millennium Development Goals (MDGs) offer a broad spectrum of issues that are found in impoverished areas around the world. These goals include; eradication of extreme poverty, universal education, gender equality, reduction of child mortality, improved maternal health, combating prevalent diseases, promotion of environmental sustainability, and promoting global partnerships [5]. Many unaffiliated organizations are also working on these issues throughout the world. While some of these goals are on track, unfortunately the UN reports that, “accelerated progress and bolder action are needed in many areas” [5]. To address these issues more effectively a deeper understanding of how poverty is defined, is needed.

The term poverty is a broadly used word that encompasses a broad array of definitions. Independent researchers have slight deviations when defining and measuring poverty. This has led to varying methods of poverty measurement [6]. These measurements or indices recently have started incorporating a multidimensional approach. The term multidimensional poverty was first pioneered by Bourguignon and Chakravarty and Tsui and has manifested into a larger section of literature [7]. The concept of poverty being 'multi-dimensional' is highlighted by the spectrum of goals outlined in the UN's MDG's [5]. It is now broadly understood that poverty is multidimensional in character; the Human Development Index, for example, recognizes the role of health and education in addition to socio economic levels [8].

Currently there are many non-governmental organizations (NGOs) trying to effectively implement projects as well as government agencies that spend millions of dollars a year on poverty eradication. While a representative universal method for measuring poverty will be of great value, the immediate focus should be on effectiveness of foreign aid. The UN MDG report says that, "based on a wide range of statistics, the actions of all stakeholders are coalescing in the achievement of many of the MDGs. At the same time, a majority of items on the agenda remain incomplete [5]." On January 17th Engineers Without Boarders (EWB) national staff met to, "create a common understanding of the terminology [that they] will use to define EWB-USA's impact in the world and … theory of change [9]." "EWB-USA recognizes the need for a strong and coherent planning, monitoring and evaluation framework that promotes learning and improved performance [9]." Many dedicated organizations are looking for an effective way to measure their impact on the communities they are involved in. There is a clear need for a standardized assessment plan for post implementation of projects within foreign aid.

Previous work by Divelbiss, D., Voth-Gaeddert, L.E., and Oerther, D.B. [2] [3] [4] has incorporated the utilization of structural equation modeling (SEM) as an assessment tool. After a completed analysis in northern Guatemala [2] and preliminary studies in Brazil [3], [4], early indications are good for a community specific intervention assessment tool using SEM. A consistent result through all projects has been the lack of monetary income having the strongest direct effect to poor health within these particular communities. Significant work has been done by Schriner, A. and Oerther D.B. on a market based solution to offer an increase in monetary income globally [10]. Continued analysis of health correlations among education, economic status and sanitation will provide much needed direction for foreign aid efforts.

There is a strong body of work on how to measure multidimensional poverty globally. However, this differs from a project based assessment tool and it is felt that a gap in the literature is present within this particular subject. This paper will cover a brief outline of several measures of multidimensional poverty, including SEM. However, the larger focus will be on the use of the SEM platform for a universal community focused foreign aid assessment tool and examples from previous research.

2. Measuring Multidimensional Poverty

To analyze something as complex as poverty, multiple variables are often used in measurement. These particular measures often aggregate data to simplify or score the variables so they can be analyzed. Depending on the measured scores weights can be added to variables that report a higher or lower significance. Several types of poverty measures can be found in the literature, these include Composite Welfare Indicators; CWI [11], Multidimensional Poverty Index; MPI [1], Human Development Index; HDI [12], Water Poverty Index; WPI [13] Comprehensive Poverty Index; CPI, and Bourguignon-Fields Class of Poverty Indices; BFPI [14] among others. Each of these measures uses certain approaches to adjust for issues within the data. Some of these approaches include the Fuzzy Set Approach [15], Axiomatic approach [16], Information Theory Approach [17] [18], Distance Function Approach [19] and SEM Approach [20] among others. A number of articles offered reviews of multiple approaches, those papers include; Deutsch and Silber 2005 [21], Ningaye, P. et al. 2013 [11], Walker, R. et al. 2007 [22]. These approaches are complex and many refer to textbooks to grasp a full understanding of them. A brief overview is all that is offered in this text; a more in depth look is outside the scope of this paper, however, references to further literature can be found throughout each section.

The Fuzzy Set Approach has proven to be a powerful way of dealing with the vagueness of the term poverty. Fuzzy sets attempt to address two major issues within multidimensional poverty, 1) the identification of a poverty line or threshold and 2) the choice of a unit of analysis as well as of a measure, or better put; the aggregation problem [15]. It addresses the 'grey area,' that is poverty; when is a person no longer impoverished? Also, using fuzzy aggregation methods and weights, calibrations and single scores can be accumulated. For further analysis of the Fuzzy Set Approach see Lemmi, A. & Betti, G. 2006 [23].

The Oxford English Dictionary defines axiomatic or an axiom as, “a proposition that commends itself to general acceptance; a well-established or universally conceded principle.” The axiomatic approach within multidimensional poverty is a list of rules that are followed within the literature when working in this particular subject area. For example, Tsui (2002) [24] provides a list of six axioms that unidimensional poverty measures are often assumed to satisfy: focus, symmetry, plication invariance, monotonicity, continuity and subgroup consistency [25]. This approach helps standardize analysis methods but can also lead to deviations within approaches.

Information Theory Approach is based around the concept of information expectancy. The expected result of an experience is established and a probability is then defined to have that result actually occur. The term entropy can then be introduced [26] [18], entropy is typically the expected information from the experience had. Miceli 1997 [27] was the first to apply this to multidimensional poverty. Miceli suggests that a measurement can be obtained from the distribution of the composite index. This index is an output from a function that describes the distribution of the probability of a result of an experience [21].

The Distance Function Approach is very useful in describing an outcome, such as standard of living, by a resource variable and a function variable of that resource. This is to help generalize other outcomes. In other words, using Sen’s “capability approach” [28]; two vectors can be denoted; one being a resource vector (a person’s resource) and the other being a functioning vector (how the individual uses the resource). To evaluate these two vectors a numerical representation is needed, this is typically in the form of an index [21]. A theorized distance function is then used to analyze the vectors. The issue that arises is a correlation between vectors and the composite error term. If this happens then the indicators could be biased. For more discussion on Distance Function Approach see Coelli et al. 2005 [19].

The SEM Approach is best served when analyzing a multidimensional or multivariate problem like poverty. It has been used extensively in the social science discipline but has recently become more widely used [29]. SEM enables the use of multiple regression equations simultaneously. By deriving composite indicators on the basis of the variance shared between the original (rather than by summing the variables), the attenuation of estimates caused by measurement error is avoided [22]. Confirmatory Factor Analysis (CFA), Exploratory Factor Analysis (EFA), Latent Growth Modeling (LGM), as well as other options, are available within the SEM framework. It can be used in both an exploratory or confirmatory analysis style. A further discussion of SEM can be found in the next section.

3. Structural Equation Modeling

Differences between the various approaches listed above are however much smaller as far as the determinants of multidimensional poverty are concerned [21]. This analysis may change with a new focus on using these techniques for a pre and post implementation assessment tool. To be able to effectively assess impacts of aid within a community, a large variety of factors need to be analyzed; direct and indirect effects need to be taken into account; and the assessment needs to be flexible with both time and error in measurement.

SEM allows for the incorporation and understanding of multiple relationships within a complicated reality. The use of latent variables is a concept within the area of SEM that allows the researcher to represent variables that can prove difficult to analyze through basic observations. Instead of using the idea of an index of indicators, SEM is able to avoid the errors accumulated from the summation of variables, whether weighted or not. The analysis of variance and covariance between multiple observable indicator variables allows for representation of these latent variables. For example, in a Brazilian household the latent variable, socio-economic status, can be represented in three indicator variables; building materials of the house, density of people within the house, and ownership of a personal boat. From these indicator variables a more robust representation can be obtained with the significant idea that less error will be found in the latent variable compared to other techniques. Basic statistical techniques can be used, such as maximum likelihood and others, to help estimate the path coefficients within the model.

The general SEM analysis is typically a two-step approach when working towards a confirmatory model. The definition of a ‘model’ can be vague, therefore a graphical representation of one type of SEM model is offered in Figure 1. In the figure, the latent variables are represented by circles and observable variables are represented by squares. This model is described as the full model. There are two parts to the full model, a measurement model and a structural model. The measurement model describes the relationships between the latent variables and the observable indicator variables. Using CFA (see [29]) the hypothesized model (covariance matrix) is compared to a data driven model (covariance matrix) using the Chi-Square test of model fit [30]. If the measurement model does fit the data via several tests of model fit indices [31], then the structural model can be assessed. Using the same style of model fit a model is either accepted or rejected. If the model is rejected adjustments can be made (with caution) and then retested. Once the model is accepted, direct and indirect effects can be assessed between latent and independent variables. This allows for the analysis of relationships between factors such as socio economic status, education, health, etc.

SEM can also be used in a purely exploratory style of analysis using EFA. If the researcher is uncertain as to which factor is described by which indicator, EFA allows for freedom amongst relationships within the measurement model. It is highly recommended within the literature that once a model is established through EFA, CFA is used with new data to test the model [29] [30].

It is possible using SEM to create a composite index from a set of deprivation measures gathered in one year and to fix or ‘freeze’ it and apply it to later years, thereby allowing change in the composite deprivation index to be accurately measured over time [22]. This idea is the basis for creating an assessment tool on interventions within a particular community. A robust representation of factors in a particular community can be found through the general SEM analysis. Pre-implementation surveys allow for understanding of the multidimensional issues relevant to that community. Mid and post implementation surveys would allow organizations real time feedback on the impacts of the intervention. If the intervention is long term based a Latent Growth Model (LGM) analysis could potentially be utilized to predict future effects of the intervention.

LGM allows for further analysis by predicting what effect continued intervention would have on a community [29]. Walker 2007 says LGM can be thought of as taking a repeated measure of an indicator and creating two latent variables which summarize the level and the trajectory of the indicator in question over time for each case. A greater understanding of LGM is needed as it is a recent innovation within SEM, but shows much potential [32] [30].

Being able to correlate and analyze multiple types of variables through a robust graphical model introduces a tool that has potential to be used on many different levels of experimental analysis. For users that are not trained in SEM, such as development practitioners, the graphical models produced from the analysis can provide a user friendly interface. SEM has the potential to become a simple application tool required for all interventions within impoverished communities.

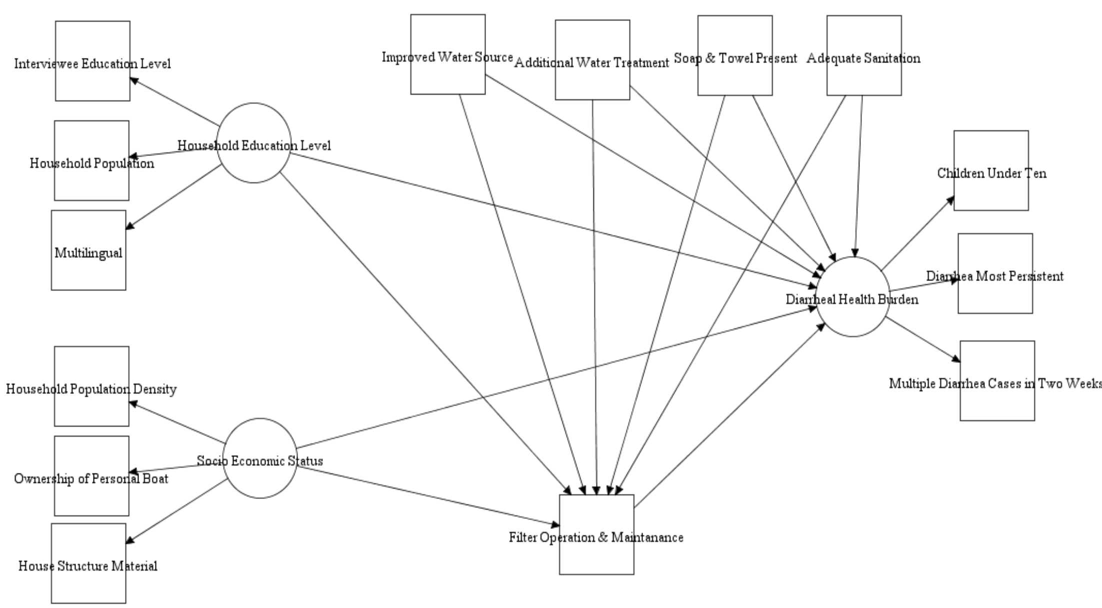


Figure 1. Graphical display of a full SEM model with latent and observable

4. Current Work

SEM represents both a different way of analyzing data, and a different way of doing science [29]. Divelbiss et al. 2013 used this approach in Guatemala to begin to observe how SEM could be used within intervention assessments. “In rural health development practice, engineers and scientists must recognize the complex interactions that influence individuals' contact with disease-causing pathogens and understand how household habits may impact the adoption and long-term sustainability of new technology [2]." The term 'new technology' refers to the CAWST Biosand Filter which was the intervention that occurred within this area. With the utilization of SEM in a dual exploratory-confirmatory analysis methodology, Divelbiss successfully described the complex relationships that are associated with daily living. He reports that, "the results illustrate how demographics, infrastructure, and practices within the home have a significant effect on proper operation and maintenance of the Biosand filter." Specifically the Biosand filter was shown to help reduce health issues within the house. However, it was not among the top three most significant effects that reduced health issues. Therefore Divelbiss also reports, "Policy makers and development practitioners must recognize that single target interventions (e.g., improving water quality) have limited influence on the entire system." The reader is encouraged to study the full report (see [2]).

Although the results are location specific, the methodology behind the study allows for a greater understanding of the application of SEM. Divelbiss also was able to highlight a potential model within multidimensional poverty. To investigate the applicability of this model to other regions a second study site was selected. In March of 2013 a complimentary feasibility study was directed by Voth-Gaeddert et al. 2013 [3] in several rural villages located west of Santarem, Para, Brazil along the Amazon River. Once the feasibility study was found to be successful a pilot study was then run [4]. The larger case study is ongoing but preliminary reports suggest the structural model developed by Divelbiss in Guatemala is suitable for a dissimilar region such as the Amazon. The underlying issue that affects all measurement tools for multidimensional poverty is the selection of indicator variables. However, SEM offers several different options to help decipher implications of different indicators.

A second finding among all three studies [2] [3] [4] was of the significance of sufficient monetary income. The factor, socio economic status, had higher direct and indirect effects on health then the intervention itself. While the intervention itself remains important, the data suggests a larger focus on monetary income and job creation. The idea of using the needs of an information-based economy to provide work through a market based strategy is being investigated by Schriner and Oerther. They are offering a platform (Pula Cloud) in which human computation work can be done benefiting both the worker (market driven to developing countries) and the requester (needs primarily from wealthier areas). Further study is encouraged to the reader in this subject (see [10]).

5. Conclusion

Continued efforts for solutions to poverty [10] and the evaluation of these solutions [3] [4] are needed. While large scale, government driven approaches are taking place, a partnered accountability towards these efforts will ensure appropriate interventions. A grass roots community based assessment tool can offer accountability on a large scale if implemented correctly. SEM has the potential to be used as a platform for this tool. The literature shows an increase in the interest and application of SEM throughout multiple disciplines [29]. While a holistic positive trend of diminishing poverty is taking place, billions are still suffering [5]. While effective, a danger of taking a holistic approach to solving poverty is the consequence of particular groups of people being left behind. Global accountability through assessments is needed to ensure a standard of living that every person deserves.

Acknowledgements

Support provided by the John A. and Susan Mathes Grant of Missouri University of Science and Technology and on ground logistical help by Gil Serique and Samuel Vasconcelos.

References

[1] S. Alkire and M. E. Santos, “Multidimensional Poverty Index,” Oxford Poverty Hum. Dev. Initiat., no. July, 2010.

[2] D. W. Divelbiss, D. L. Boccelli, P. A. Succop, and D. B. Oerther, “Environmental health and household demographics impacting biosand filter maintenance and diarrhea in Guatemala: an application of structural equation modeling.,” Environ. Sci. Technol., vol. 47, no. 3, pp. 1638–45, Feb. 2013.

[3] L. E. Voth-Gaeddert, D. W. Divelbiss, and D. B. Oerther, “Utilizing Structural Equation Modeling to Correlate Biosand Filter Performance and Occurance of Diarrhea in Enseado Do Aritapera, Para, Brazil,” in 17th International Symposium on Health-Related Water Microbiology, 2013.

[4] L. E. Voth-Gaeddert, D. W. Divelbiss, and D. B. Oerther, "Utilizing Structural Equation Modeling as an Evaluation Tool for Critical Parameters of the Biosand Filter in a Pilot Study in Para, Brazil," in 5th International Slow Sand and Alternative Biological Filtration Conference, 2014.

[5] United Nations, "The Millennium Development Goals Report 2013," New York, 2013.

[6] E. Callander, "Towards a holistic understanding of poverty: A new multidimensional measure of poverty for Australia," Heal. …, vol. 2, no. 2, pp. 141–155, 2012.

[7] F. H. G. Ferreira, "Poverty is multidimensional. But what are we going to do about it?," J. Econ. Inequal., vol. 9, no. 3, pp. 493–495, Aug. 2011.

[8] S. Smith, "The Scope of NGOs and Development Programme Design: Application to Problems of Multidimensional Poverty," Public Adm. Dev., vol. 370, no. October 2011, pp. 357–370, 2012.

[9] T. Martindale, "Planning, Monitoring, Evaluation and Learning Program," 2013.

[10] A. Schriner and D. B. Oerther, "Employment via Crowdsourced Human Computation: A Strategy for Stimulating Economic Development," 2014.

[11] P. Ningaye, T. Y. Alexi, and T. F. Virginie, "Multi-Poverty in Cameroon: A Structural Equation Modeling Approach," Soc. Indic. Res., vol. 113, no. 1, pp. 159–181, Jun. 2012.

[12] K. Malik, "Human Development Report 2013," New York, 2013.

[13] R. G. Garriga and a P. Foguet, "Application of a revised Water Poverty Index to target the water poor.," Water Sci. Technol., vol. 63, no. 6, pp. 1099–110, Jan. 2011.

[14] L. Esposito and P. Lambert, "A note on the Bourguignon–Fields class of poverty indices," J. Public Econ., vol. 93, no. 7–8, pp. 852–854, Aug. 2009.

[15] D. Neff, "Fuzzy set theoretic applications in poverty research," Policy Soc., vol. 32, no. 4, pp. 319–331, Dec. 2013.

[16] F. Bourguignon and S. R. Chakravarty, "The Measurement of Multidimensional Poverty," J. Econ. Inequal., no. 1, pp. 25–49, 2003.

[17] L. Asselin, "Composite indicator of multidimensional poverty," Multidimens. Poverty Theory, 2002.

[18] E. Maasoumi, “A Compendium of Information Theory in Economics and Econometrics,” Econom. Rev., vol. 12, no. 3, pp. 1–49, 1993.

[19] T. Coelli, D. Rao, C. O’Donnell, and G. Battese, An introduction to efficiency and productivity analysis. 2005.

[20] A. Giuffrida, R. Iunes, and W. Savedoff, “Health and poverty in Brazil: Estimation by structural equation model with latent variables,” Tech. Note Heal., no. 1, 2005.

[21] J. Deutsch and J. Silber, “Measuring multidimensional poverty: An empirical comparison of various approaches,” Rev. Income Wealth, no. 1, 2005.

[22] R. Walker, “Measuring change in multidimensional poverty using structural equation modelling,” 2007.

[23] A. Lemmi and G. Betti, Fuzzy set approach to multidimensional poverty measurement. 2006.

[24] K. Tsui, “Multidimensional poverty indices,” Soc. Choice Welfare, vol. 19, no. 1, pp. 69–93, Jan. 2002.

[25] P. Makdissi and Q. Wodon, “Defining and Measuring Extreme Poverty,” Dyn. Inequal. Poverty (Research Econ. Inequality), vol. 13, pp. 325–340, 2006.

[26] C. E. Shannon, “The Mathematical Theory of Communication,” Bell Syst. Tech J., vol. 27, pp. 379–423, 625–656, 1948.

[27] D. Miceli, “Measurement of Poverty; Theory and Application to Switzerland,” University of Geneva, 1997.

[28] A. Sen, “From income inequality to economic inequality,” South. Econ. J., 1997.

[29] J. B. Grace, Structural Equation Modeling and Natural Systems, First. United Kingdom: Cambridge University Press, 2006.

[30] R. B. Kline, Principles and Practices of Structural Equation Modeling, Second. New York: The Guilford Press, 2005.

[31] J. Schreiber and A. Nora, “Reporting structural equation modeling and confirmatory factor analysis results: A review,” J. …, 2006.

[32] T. E. Duncan, S. C. Duncan, L. A. Strycker, F. Li, and A. Alpert, An Introduction of Latent Variable Growth Curve Modeling; Concepts, Issues, and Applications. Mahwah, NJ: Lawrence Erlbaum Associates, 1999.

IV. ANALYZING INDICATORS OF MULTIDIMENSIONAL POVERTY FOR STRUCTURAL EQUATION MODELING USING MAHALANOBIS-TAGUCHI STRATEGY

Abstract

Universal access to clean water is an important and challenging goal. However, if a holistic view of the multidimensional system found in regions of need is not understood, then providing impactful aid may prove difficult. Structural equation modeling (SEM) offers a statistical platform in which to assess this system. SEM is a robust tool that utilizes latent or hidden variables that are represented by a set of indicator variables. These indicator variables are often selected through a priori knowledge. However, the Mahalanobis-Taguchi Strategy provides a pattern recognition technique for these indicators. MTS can either confirm the indicator variables in SEM models or provide seeds in indicator variables to the SEM model. A confirmatory approach was utilized for a SEM model that was successfully populated by data from private practitioners in Guatemala. Demographic and Health Survey data from this region was analyzed in the MTS method to confirm significant indicator variables. Of the six indicators used in the SEM model, four of them were in the top ten of the significant indicators found using MTS. Maternal education was found to be the most significant, followed by paternal education. MTS is able to provide a valuable tool to compliment the SEM platform being used to create an intervention assessment tool.

Introduction

By 2015, the United Nations has pledged to reduce by half extreme poverty through a set of eight broad goals referred to as the Millennium Development Goals; MDGs (United Nations, 2013). As of 2014, the only goal substantially achieved was that of universal access to clean water. While an optimistic view is taken by the UN, the MDG interim report states, “accelerated progress and bolder action are needed in many areas (United Nations, 2013).” In preparation for the post 2015 agenda, the UN and others have begun comprehensive assessments of related programs and projects all aimed at reducing poverty. For example, the Peace Corps reported, “The Comprehensive Agency Assessment refined Peace Corps’ vision, stating that the agency will be ‘a leader, in partnership with others, in the global effort to further human progress and foster understanding and respect among people (Peace Corps, 2012).'" Engineers Without Borders United States of America (EWB-USA) also reports, "[we] recognize the need for a strong and coherent planning, monitoring and evaluation framework that promotes learning and improved performance (Martindale, 2013)." These recent organizational statements suggest a shift in focus to emphasize improvements within post implementation assessments.

This shift in focus can be partially attributed to the growing attention around multidimensional poverty originally described by econometrists Bourguignon, Chakravarty and Tsui (Ferreira, 2011). While access to water is a significant factor, several other factors contribute to multidimensional poverty as well. Several measurement indices that try and address this issue are the Human Development Index; HDI (Malik, 2013), Multidimensional Poverty Index; MPI (Alkire & Santos, 2010), Composite Welfare Indicators; CWI (Ningaye, Alexi, & Virginie, 2012), and the Water Poverty Index; WPI (Garriga & Foguet, 2011) among others. Measuring and rating poverty and access to clean water is an important objective, however, an arguably as important objective is the assessment of the implementation, such as water filters, that was utilized in aiding a particular region. Currently, all international and local aid organizations assess themselves. This concept of 'self-revision' is rare for professionals where people's lives are drastically impacted. For this reason, as well as for critical feedback for the aid organization, there is a significant need for a standardized intervention assessment tool. Previous research by Divelbiss et al. 2013 and Voth-Gaeddert et al. 2014 on biosand filters offer a solution to these challenges utilizing Structural Equation Modeling (SEM).

SEM offers the ability to analyze multiple hypotheses simultaneously while excluding researcher bias in factor weights. A powerful concept that is available within the SEM framework is the use of latent variables. Latent (also called "hidden variables") cannot be directly measured (Kline, 2005). For example, if a researcher is to measure socio economic status they may query about income, career field, physically owned items, housing material or a combination of these. Using a relevant combination of these allows for a more robust measure of socio economic status. SEM utilizes Confirmatory Factor Analysis (CFA) and Path Analysis to produce a structural model that is all inclusive and easily interpreted. This graphical representation of the model is another significant attribute in SEM (Grace, 2006).

SEM is widely used within the social sciences and lately within econometrics. However, (Divelbiss, Boccelli, Succop, & Oerther, 2013) was the first one to use SEM as an intervention assessment tool for water filtration. A small set of communities in Guatemala had received biosand filters to help reduce the diarrheal health burden within these communities. Divelbiss used SEM to hypothesize relationships among important community factors such as education, socio economic status, diarrheal health burden, sanitation, water issues and biosand filter operation and maintenance (FOM). The results showed that the biosand filters did not have the largest effect on reducing the diarrheal health burden within these communities. The significant factor was found to be the socio economic status of the household. Oerther & Schriner 2014 offer a solution to this issue in their paper, “Which has greater liquidity: money or drinking water?” Due to these results a second study site was selected. This study site was located in Para, Brazil among villages that lived along the Amazon River. Voth-Gaeddert et al. (2013; 2014) conducted a feasibility study in March of 2013 and then a pilot study in September of 2013. While the larger case study is still under investigation, early indicators point towards a more complex set of relationships within these two communities. The indicators used to describe the latent variables of household education level (HEL), socio economic status (SES) and household health burden (HHB) have proven to be the topic that is most widely disputed.

Surveys are used to collect the data needed to populate the SEM models. Local help is often utilized for the collection of data. However, there are organizations that collect survey data globally. The Demographic and Health Surveys (DHS) Program is a US Agency for International Development (USAID) supported data collection organization. Utilizing cluster sampling they have collected household, region specific, surveys from Guatemala and Brazil. A brief examination of the specific survey used by DHS within the region of Guatemala that was studied by Divelbiss, shows promise in using DHS data to confirm Divelbiss’ choice of indicators for the latent variables (Instituto Nacional de Estadistica, 1999). A common and important issue within SEM models is the significance of the indicators that describe the latent variables. Collecting data for individual projects puts constraints on the sample size available for analysis. The DHS data provides robust sample sizes as well as a significant increase in the number of variables available for analysis. Utilizing the robust data set (both in sample size and variables) from DHS, significant indicators can be identified through pattern recognition techniques. These techniques can be used in one of two ways. First, the recognized indicators can be compared with the indicators used in the SEM models, in a confirmatory analysis. Alternatively, the recognized indicators could be used as a seed into the SEM model. This general step of indicator identification could prove to be a crucial factor in providing robust confirmation or seeds for latent variables.

The pattern recognition technique used to identify important indicators was the Mahalanobis-Taguchi Strategy (MTS). This technique is a pattern recognition scheme that has gained popularity in the automotive industry. MTS aids in quantitative decisions by constructing a multivariate measurement scale using a data analytic method. MTS is unique because of the utilization of the Mahalanobis distance (MD). MD is a measure of distance that utilizes orthogonal transcriptions within the correlation matrix to eliminate issues with multicollinearity. By analyzing variances and covariances the MD approach differs from classical statistical approaches. It is also able to account for independent and dependent variables within the same set. The driven purpose of MTS is to accurately predict significant variables that show similar patterns within a multidimensional system (Taguchi & Jugulum, 2002). While the combination of SEM and MTS is not found within the larger body of literature, results are promising for the incorporation of MTS into the methodology of creating a standardized intervention assessment tool for developing regions using a SEM platform.

Methodology

Mahalanobis-Taguchi Strategy

MD is a measure for the distance between two multivariate populations that accounts for means, variances, and covariances. MD takes into account the correlations between variables. MD is commonly used to measure the distance of a single observation from the center of its respective population. Observations that differ significantly from the center of its population pose the question if the observation is truly a part of that population (Manly, 1994). MTS is a discriminant analysis technique that utilizes MD values for prediction and forecasting.

MTS can be used to minimize the number of variables required for diagnosis and to predict the performance of the system. MTS is a pattern recognition technology that aids in quantitative decisions by constructing a multivariate measurement scale using a data analytic method. The main objective of MTS is to make accurate predictions in multidimensional systems by constructing a measurement scale (Taguchi & Jugulum, 2002).

In the MTS, the Mahalanobis space (reference group) is obtained using the standardized variables of healthy or normal data. The Mahalanobis space (MS) can be used to discriminate between normal and abnormal objects. Once this MS is established, the number of attributes is reduced using orthogonal array (OA) and signal-to-noise ratio (SN) by evaluating the contribution of each attribute. Each row of the OA determines a subset of the original system by the including and excluding that attribute of system. The steps of the calculation of the MD and the identification of critical factors from MD are outlined as follows.

Stage I: Construction Of A Measurement Scale.

The first step in MTS is to construct a measurement scale using the MS as a reference. To construct a measurement scale, a data set of the normal observations needs to be collected. A reference group with suitable variables and observations that are as uniform as possible is selected. This reference group is used as a base or reference point of the scale.

The collected normal observations are then standardized using Equation 1.

$$Z_i = \frac{X_i - m}{\sigma} \qquad (1)$$

where,

- m, mean of the attribute,
- σ , standard deviation of attribute,
- Zi , standardized variables, and
- Xi, normal observations.

The standardized vector is obtained from the standardized values of Xi (i=1, 2, ..., k). MD measures the distance in multidimensional spaces by accounting for the correlation among the attributes. The statistical meaning of MD is the nearness of an unknown point to the mean of the group. The following is the formula used to calculate MDs:

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \tag{2}$$

Where C-1 is the inverse of the correlation matrix which contains correlation coefficients between the variables and T is the transpose of the standard vector. It can be easily proved that the average value of the MDs is 1 for all the observations in the MS. For this reason, MS is also called the unit space (Taguchi & Jugulum, 2002).

Stage II: Assessment Of The Measurement Scale.

The second step is to assess the measurement scale. In order to evaluate the measurement scale, observations outside of MS are used, usually abnormal or test observations. The mean value, standard deviation and correlation matrix of normal observations are used to calculate the MD of the abnormal observations. For good measurement scales, the MDs of the abnormal observations are larger than the MDs of the normal observations. Dynamic S/N ratios are calculated to determine the accuracy of the scale.

Stage III: Identify The Useful Variables (Developing Stage).

In the third stage, the system is optimized. The useful variables are determined using orthogonal arrays and signal-to-noise ratios. For this purpose, OA and SN array are very useful to identify which attributes are important. In the experiment, every factor is assigned to a column in the OA, and every row represents the experiment combination of a run. A two level OA is used to represent inclusive and exclusive. In a two level OA, 1 indicates the level that corresponds to presence of a variable and 2 indicates the level that corresponds to the absence of the variable. Each attribute will be used or neglected with respect to the OA and the SN ratio is calculated.

There are many different types of SN ratios; however, MTS uses the larger the better or dynamic SN ratio. In the context of MTS, SN ratio is defined as the measure of accuracy of prediction of the scale. It reflects the severity of the abnormalities and the difference of the average SN values of each attribute when it is included and excluded. The classification ability is compared with the feed forward artificial neural network. In the aspect of data size, efficiency and time, MTS shows good performance compared to neural network. Equation (3) shows the dynamic SN ratio.

$$SN = 10\log\left(\frac{\frac{1}{r}\left(S_\beta - V_e\right)}{V_e}\right)^2 \tag{3}$$

Where,

- ST = total sum of squares

  $S_T = \sum_{i=1}^{t} y_i^2$

- r = sum of squares due to input signal

  $r = \sum_{i=1}^{t} M_i^2$

- Sβ = sum of squares due to slope

  $S_\beta = \frac{1}{r}\sum_{i=1}^{t}(M_i y_i)^2$

- Se = error sum of squares

  $S_e = S_T - S_\beta$

- Ve = error variance

  $V_e = \frac{S_e}{t-1}$

For a given attribute Xi, SN+ represents the average SN ratio of including the attribute Xi. SN- represents when Xi is excluded.

$$Gain = SN^{+} - SN^{-} \tag{4}$$

If the gain is positive, the attribute is used, if not it is neglected. After the confirmation test, the optimization results are compared with the before and after.

Stage IV: Future Diagnosis With Useful Variables.

Monitor the conditions using the scale, which is developed with the help of the useful set of variables. Based on the values of the MD, appropriate corrective actions can be taken.

Structural Equation Modeling

SEM boasts a wide variety of uses, the particular approach that is conducive to modeling multidimensional poverty is through a confirmatory model. The approach used within the proposed assessment tool has two steps to reaching a full SEM model. All of the relationships within the full model are either linear or logarithmic regressions. A graphical representation of a SEM model is found in Figure 1. In this figure observed variables are represented as rectangles and latent variables are represented as circles. This full model is constructed of a measurement model and a structural model. The measurement model only includes the latent variables and the observed variables that serve as indicators. The measurement model for Figure 1. would include the latent variables HEL, SES and HHB and the supporting indicator variables (ie for HEL the supporting indicators are a reading test, paternal education level and maternal education level). Utilizing a predetermined set of indicators for the latent variables categorizes this technique as a CFA. Using a CFA, the hypothesized measurement model (a priori) is compared to the data driven model (from the surveys) using the Chi-Square test of model fit (16). Several other measures of model fit (data driven vs hypothesized) are recommended within the literature (24, one other), these include root mean square error of approximation; RMSEA (Hu & Bentler, 1998), Comparative Fit Index; CFI (Hu & Bentler, 1998) and Tucker Lewis Index; TFI(Muthén & Muthén, 2012). If the measurement model has adequate fit then structural model can be added to it to make the full model.

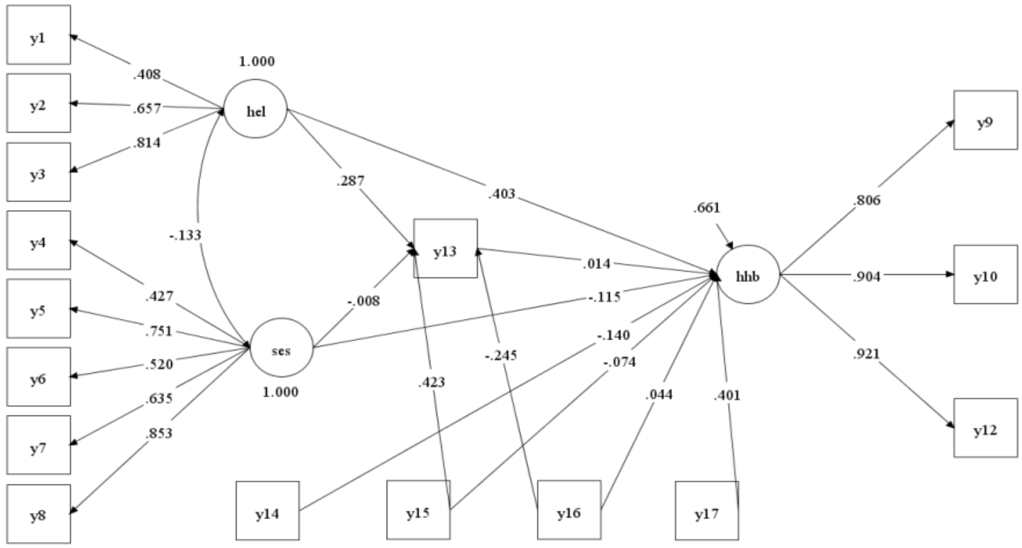


Figure 1. Full Model with parameter estimates. Variable definitions are given in Table 1.

The structural model represents the relationships between the latent variables and other independent observable variables. In the case of Figure 1. the structural model consists of the variables HEL, SES, HHB, as well as, FOM (need to turn previous into acronym), improved sanitation and improved water; source, treatment and storage. So together, the measurement model and structural model make up the full model. The same model fit tests are performed with the full model and if found satisfactory, the model can be analyzed using Path Analysis. Path Analysis allows direct and indirect effects to be assessed between variables. For example in Figure 1. SES is regressed on HHB (represented by an arrow), and a pathway value is given (along the arrow). In unstandardized terms this means that as a SES increases by one unit, HHB will decrease by -.115 units, in essence numerically representing that an increase in SES reduces the issue of poor health (HHB) on average within the sampled households.

In the case either the measurement model or full model does not fit, several options are still available. If multiple hypothesized models were made then a comparison between the fit of hypothesized models is possible. However, the popular feature of SEM is that it gives clues to what was wrong with the original hypothesized model. Adjustments can be made until an adequate fit is obtained; however, analysis of data needs to be done with caution as any changes made to the model imply an exploratory style of analysis. It is highly recommended that a new set of data is obtained and used to populate the restructured model for confirmatory purposes. Iterations of adjustment and resampling are the common statistical practice within SEM and is used in the work by Divelbiss and Voth-Gaeddert.

Table 1. List of variables and their type

Variable	Definition	Type
Y1	Reading and Writing Score	Ordinal
Y2	Paternal Education Level	Ordinal
Y3	Maternal Education Level	Ordinal
Y4	Improved Roof	Dictomous
Y5	Owns a TV	Dictomous
Y6	Owns a Shower	Dictomous
Y7	Has Multiple Working Lightbulbs	Dictomous
Y8	Improved Water Storage Material	Dictomous
Y9	Severity of Sickness	Ordinal
Y10	Occurrence of Sickness	Ordinal
Y11	Wait Period of Sickness and Doctor Visit	Ordinal
Y12	Children Under 11	Dictomous
Y13	Filter Operation and Maintenance	Ordinal
Y14	Improved Sanitation	Dictomous
Y15	Improved Water Source	Dictomous
Y16	Improved Water Treatment	Dictomous
Y17	Improved Water Storage	Dictomous

Results and Discussion

Pros And Cons Of SEM And MTS

SEM and MTS both report statistical pros and cons. To understand how these two methods can be utilized in modeling interventions in developing regions, an understanding of their respective pros and cons is needed. A detailed observation into these two statistical tools will naturally present the opportunity for collaboration.

SEM is most well-known for the use of latent variables. This has been a powerful tool in the study of social sciences and more recently econometrics. Indictor based measures have become more popular in recent years (ie. HDI, MPI, WPI), however, many of these indicators utilize a summation or weighted summation of indicators. Within the summation of indicators is also the summation of the measurement errors associated with each indicator. SEM is able to avoid this issue and allow the indicators to represent the latent variable the best. With a robust set of latent variables the structural model is that much stronger. The relationships within the structural model can be assessed using Pathway Analysis described in the methods section. Path Analysis offers the ability to not only observe significant variables in comparison with others, but to look at direct and indirect effects from these variables. This is critical when assessing the complex set of parameters that are found in impoverished regions. Finally, SEM is often represented graphically to help visualize relationships to those that are less familiar with SEM (see Figure 1.). While these traits create a strong argument for the usage of an SEM platform for an assessment tool, there are issues that underlie these benefits.

Several common issues within SEM are often cited throughout the literature. A misuse of SEM in terms of causal modeling has become more frequent. SEM does not determine the cause of one variable to another, predetermined pathways (the arrows in Figure 1.) represent the hypotheses that the researcher establishes prior to data collection. Also needed, are appropriately labeled variables in the graphical model. One issue with a latent variable, as in Figure 2. is mislabeling. For example, the term HEL is a latent variable with specific indicators. Interpretation to overall cognitive abilities was the intent

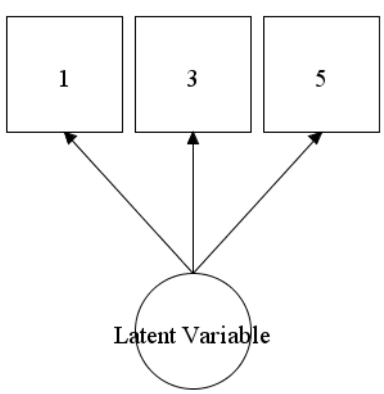


Figure 2. SEM latent variable described by three indicator variables

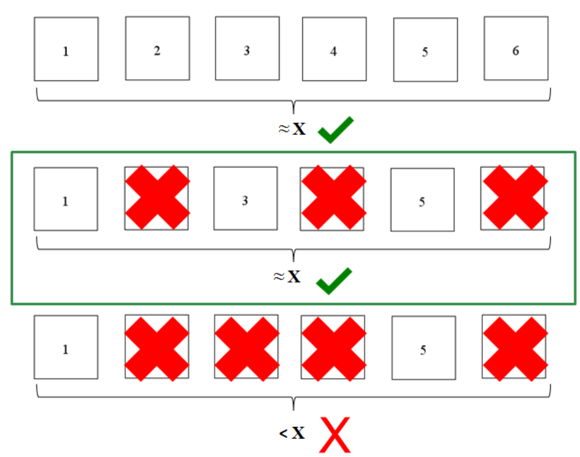


Figure 3. MTS visual simplified method of finding significant indicator variables

of this label; however, 'education level' is widely associated with school. Time must be devoted to properly attributing names to graphical variables. Fit indices to which models are measured to be adequate for interpretation have been disputed within the literature (Barrett, 2007; Bentler, 2007). For this reason, Muthen & Muthen offer four measures of fit within their software package, MPLUS 7, which has been used in the research in Guatemala and Brazil.

Finally, one of the most significant issues within SEM is in regards to the indicator variables used to describe the latent variables. Throughout much of the literature varying methods of selection are used to determine significant indicators. These methods include ranking physical assets as within the Guttman Scale (Guest, 2000), summing values of physical assets per household (Undurraga et al., 2010) or weighting income and labor type (Deaton, 1997). This is by no means an exhausted list but covers the general range of techniques whether predetermined weights (or significance) are given to the different indicators or not. Unfortunately many of these techniques bring a significant amount of either error or bias. In SEM the indicators are chosen before collection and the measurement model is then run as a confirmatory analysis. The issue becomes the high number iterations that are required to find a combination of indicator variables that satisfy the model. Even more troubling is the fact that while these indicators explain the variance within the latent variable "adequately" other indicator variables may be left out that would further explain the model. Fortunately, techniques are available that allow significant indicators to be easily identified. One such technique is MTS.

MTS is a multivariate analysis technique that takes on the basic model of a multidimensional system. It is typically utilized to minimize the number of variables required to explain the system (see Figure 3.). This is performed through an analysis of a correlation matrix. The patterns of observations in a multidimensional system highly depend on the correlation structure of the variables in the system. Where the measurement model in SEM is left to a 'guess and check' method, MTS is able to statistically determine indicator variables that are significant. This style of analysis is done through the MD methodology. MD differs from other approaches in two ways; first, variances and covariances are considered instead of averages and second, it accounts for ranges of acceptability between variables meaning it compensates for the interactions. There are very few complex systems that lack dependence between all variables which is where error can enter the analysis. Where SEM lacks statistical backing, MTS offers a robust diagnostic tool, capable of highlighting significant indicator variables within a complex system.

With all statistical analysis tools knowledge of common mistakes is critical in executing an appropriate diagnostic. For MTS, variables are normalized using the assumption of a normal distribution. However, if the distribution is known, that distribution could be applied to normalize that variable. Also, there are no set criteria currently for MDs in the abnormal observations verses the normal observations, but abnormal data can be used to verify per set criteria. Finally, the use of full and fractional factorial designs should be selected with caution, if time permits a full factorial design should be used, however, in many cases (especially within the automotive industry) data analysis is needed quickly and fractional factorial designs offer the benefit of saving time and money. With a full understanding of the methods used within MTS, the technique can offer an important guide to finding significant indicator variables for use in SEM models.

MTS Analysis On DHS Survey Data From Guatemala

The data used for analysis was collected via a household health survey that was administered per house selected through cluster sampling. There were 5587 households surveyed in Guatemala, however surveys not categorized as rural and in the Quiche region were eliminated. The sample size of this analysis was n=256. Using the method of MTS which operates with the adjoint of the correlation matrix, a ranking of significant variables in the data set was found. This method is able to handle highly correlated variables and still accurately predict patterns.

Table 2. lists the top predictions of variables that have the highest significance within the health survey used. Table 3. offers a list of variables utilized in the Divelbiss' model that were selected a priori for the SEM model in Quiche, Guatemala. The Table 2. data is presented in significance order that MTS produced. It shows that education in the house accounts for the largest amount of variability within the health survey. Socio economic status is also shown to have a significant amount of explained variability

within the health survey. This can be seen in the listing of ownership of a car, a kitchen being present in the home, ownership of a motorcycle, and the presence of a chimney in the kitchen. Of the top ten significant variables, the source of drinking water and the time it takes to get to the source of drinking water were the important water quality related indicators. Figure 4. shows a typical MTS output which allows for analysis of the indicator variables.

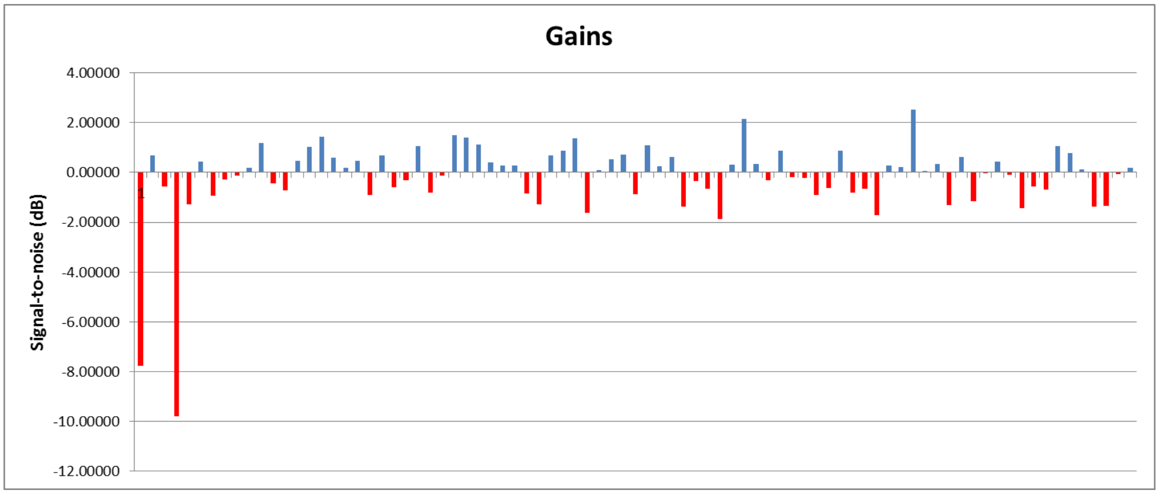


Figure 4. Results are reported in a signal-to-noise ratio (dB) where positive blue bars indicate significant variable

The MTS output can then be compared to indicator variables used within the SEM model to confirm significance. The indicators from Divelbiss et at. 2013 were the focus of this MTS analysis. Of the six indicators in Table 3. that are available for comparison, four of them are within the top ten of significant variables recognized by the MTS method. Paternal and maternal education levels, improved water source, and household population density were all confirmed as significant indicators in fully explaining what contributes to health within the household. Wall material falls farther down on the significance list but still helps to explain some of the variability within the survey. These results are promising, not only to the assistance of the SEM platform but also to highlighting the need for increased efforts in education and job creation for the health of impoverished regions. See Appendix C for extended list of significant variables.

With the limitation of magnitude and direction of the predicted variables, MTS can still be a very useful tool in the confirmation of significant indicators. It should be recognized that where MTS does not allow for analysis of direction of the correlations between variables, Mahalanobis-Taguchi Gram-Schmidt (MTGS) method can service this need. MTGS uses Gram-Schmidt orthogonalization which can handle highly correlated variables, like MTS, however, it also allows for analysis of directionality. To gain further insight to the details of significant variables, the MTGS method offers a promising option. In future work, MTGS will be utilized as a confirmation technique as well as a seed technique for SEM models.

Table 2. Indicators from MTS Analysis

Maternal education Level
Paternal education Level
Source of drinking water
Ownership of car
Number of family members
Time to water source
Sex of household
Place/room to cook (kitchen)
Ownership of motorcycle
Chimney present in kitchen

Table 3. Indicators from (Divelbiss et al. 2013)

Paternal education level
Maternal education level
All speak spanish in home*
Improved water source
Soap present in home*
Additional water treatment*
Floor material
Wall Material
Household Population Density
**Not Asked in DHS Survey*

Conclusion

SEM offers a robust statistical tool for modeling water related interventions in developing countries. Understanding the community in which the intervention is being implemented in provides the partnering organization critical knowledge of how best to offer aid. SEM may become tedious and expensive if the system being analyzed is not well understood. This is where MTS can provide crucial information to help save time and money. MTS has the potential to be used in one of two ways. It can either offer confirmation of correct a priori selected indicators or offer a seed to which decisions of indicator variables can be made. In this analysis MTS was used in a confirmatory technique in relation with SEM. MTS confirmed the previous results of Divelbiss et al. 2013 and Voth-Gaeddert et al. 2014 that education and socio economic status have a significant effect on health within rural impoverished communities.

Early indications show that MTS can provide support for analysis in SEM. The use of MTGS will allow for a more robust confirmation tool or seeding tool for SEM. While water continues to be an important variable in the system that is multidimensional poverty, a holistic focus is needed when assessing aid to people in poverty. SEM, with the collaboration of MTGS, can provide a platform in which to create a crucial intervention assessment tool that will provide international aid organizations with much needed direction.

References

Alkire, S., & Santos, M. E. (2010). Multidimensional Poverty Index. Oxford Poverty & Human Development Initiative, (July).

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. Personality and Individual Differences, 42(5), 815–824. doi:10.1016/j.paid.2006.09.018

Bentler, P. M. (2007). On tests and indices for evaluating structural models. Personality and Individual Differences, 42(5), 825–829. doi:10.1016/j.paid.2006.09.024

Deaton, A. (1997). The analysis of household survey. Research Program in Development Studies, Princeton …. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Analysis+of+Household+Surveys#2

Divelbiss, D. W., Boccelli, D. L., Succop, P. A., & Oerther, D. B. (2013). Environmental health and household demographics impacting biosand filter maintenance and diarrhea in Guatemala: an application of structural equation modeling. Environmental Science & Technology, 47(3), 1638–45. doi:10.1021/es303624a

Ferreira, F. H. G. (2011). Poverty is multidimensional. But what are we going to do about it? The Journal of Economic Inequality, 9(3), 493–495. doi:10.1007/s10888-011-9202-3

Garriga, R. G., & Foguet, a P. (2011). Application of a revised Water Poverty Index to target the water poor. Water Science and Technology : A Journal of the International Association on Water Pollution Research, 63(6), 1099–110. doi:10.2166/wst.2011.347

Grace, J. B. (2006). Structural Equation Modeling and Natural Systems (First.). United Kingdom: Cambridge University Press.

Guest, G. (2000). Using Guttman Scaling to Rank Wealth: Integrating Quantitative and Qualitative Data. Field Methods, 12(4), 346–357. doi:10.1177/1525822X0001200406

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3(4), 424–453. doi:10.1037//1082-989X.3.4.424

Instituto Nacional de Estadistica. (1999). Guatemala: Interim DHS. Rockville, MD.

Kline, R. B. (2005). Principles and Practices of Structural Equation Modeling (Second.). New York: The Guilford Press.

Malik, K. (2013). Human Development Report 2013. New York.

Manly, B. F. J. (1994). Multivariate Statistical Methods: A Primer (Third.). New York: Chapman and Hall.

Martindale, T. (2013). Planning, Monitoring, Evaluation and Learning Program (pp. 1–31).

Muthén, L. K., & Muthén, B. O. (2012). User ' s Guide (Seventh.). Los Angeles, CA: Muthen and Muthen.

Ningaye, P., Alexi, T. Y., & Virginie, T. F. (2012). Multi-Poverty in Cameroon: A Structural Equation Modeling Approach. Social Indicators Research, 113(1), 159–181. doi:10.1007/s11205-012-0087-8

Oerther, D.B., & Schriner, A. (2014). Which has greater liquidity: money or drinking water?. Internatoinal Water Association. Abu Dhabi, UAE

Peace Corps. (2012). The Peace Corps Performance Plan (pp. 1–27).

Taguchi, G., & Jugulum, R. (2002). The Mahalanobis-Taguchi Strategy. New York: John Wiley & Sons, Inc.

Undurraga, E. a., Nyberg, C., Eisenberg, D. T. a., Magvanjav, O., Reyes-García, V., Huanca, T., … Godoy, R. (2010). Individual Wealth Rank, Community Wealth Inequality, and Self-Reported Adult Poor Health: A Test of Hypotheses with Panel Data (2002-2006) from Native Amazonians, Bolivia. Medical Anthropology Quarterly, 24(4), 522–548. doi:10.1111/j.1548-1387.2010.01121.x

United Nations. (2013). The Millennium Development Goals Report 2013. New York.

Voth-Gaeddert, L.E., Divelbiss, D.W., Oerther, D.B. (2014). "Utilizing Structural Equation Modeling as an Evaluation Tool for Critical Parameters of the Biosand Filter in a Pilot Study in Para, Brazil", Progress in Slow Sand and Alternative Biofiltration Processes, International Water Association.

Voth-Gaeddert, L.E., Divelbiss, D.W., Oerther, D.B. (2013). "Utilizing Structural Equation Modeling to Correlate Biosand Filter Performance and Occurrence of Diarrhea in Enseado Do Aritapera, Para, Brazil", Water Science and Technology: Water Supply, IWA

SECTION

3. CONCLUSION

Through the prior four manuscripts the body of work for this thesis can be observed. Two separate research trips established the first two manuscripts. A literature review presented in the third manuscript outlined the general field of multivariable analysis in developing country assessments. Lastly, the fourth manuscript offered a confirmation of the researcher hypothesized indicator variables used in the original study by utilizing a pattern identification tool. These manuscripts each offered a part of the answer for the original hypothesis that the CAWST biosand filter is the most significant factor in the reduction of diarrheal occurrences in developing communities.

The analysis for the first research trip presented in Paper 1, Figure 3 utilized bootstrapping and frequency analysis to populate the SEM model. FOM did show a 0.1 reduction of DHB, meaning that as a one unit increase in FOM occurred a 0.1 reduction in the DHB occurred. However, HEL had an even greater effect on reducing the DHB with a 0.5 reduction ratio. With education being the most important the original hypothesis was not supported in the first iteration. For the supporting hypotheses (see Table 1.1.) four of the eleven were supported and the others were not found relevant enough to be displayed in the model (this will be investigated further in the next paragraph). The variables HEL, SES, STP and AWT were all found to positively affect the FOM in that order of significance. This meant that the education of the household was the most important factor in properly maintaining the filter.

A correlation matrix was also analyzed to further investigate relationships; however, as noted before, causality cannot be inferred with a correlation matrix (see Paper 1, Table 2). With this in mind, the first observation that can be made is the consistently negative relationship between variables related to SES and variables related to DHB. Of the 18 relationships, 16 of them were negative. This suggests that either people with a higher SES have fewer issues with diarrhea or that people with less diarrhea often have a higher SES. The second observation that can be made is in regards to the improved sanitation and water scores (ie IMS, IWS, IWT, and IWSt). All four variables had negative relationships with diarrhea being the most persistent health issue in the household. IWS and IWT also had a negative relationship with the question of having diarrhea in the past two weeks. These relationships helped to support the supporting hypotheses.

The model presented in Paper 2, Figure 2 depicts the second iteration completed for the region of Para, Brazil. This model populated with data from two new villages passed the measurement model fit tests but failed to satisfy the full model fit tests (the same tests are used for each stage). Investigation of this model can offer insight to both the relationships captured by the data and the reason for model failure due to misspecification. Prior to analysis, it should be noted that due to a lack of diarrheal occurrences within the village (5%) during the dry season, the DHB variable was modified to include overall health of the household (Household Health Burden; HHB). The first relationship to assess is the FOM (or y13) variable and the HHB variable. A, almost, negligible relationship can be found suggesting that the filter had very little influence on the health of the household. As compared to all other relevant variables in the model, FOM had the smallest effect on HHB. This does not support the original hypothesis. The variable that had the largest direct effect on HHB was that of sanitation. This was closely followed by the SES of the household. This infers that families with proper sanitation and a higher SES are found to be healthier. The reason for inadequate fit in this model was due to the improper selection of indicator variables for the latent variable HEL. This can be seen through investigation of residuals in the output given by MPLUS 7. This can also be confirmed through the unexplained variation found in the HHB variable. Currently the HEL variable has a positive .403 relationship with HHB, inferring that as a household's cognitive ability increases their health burden also increases. This concept has generally been rejected by most aid organizations and documented in the literature as so too (Grosse 1989). The fact that there is a large unexplained variance in HHB and that the HEL influence is significant in the wrong direction (positive not negative) is not a coincidence. This also offers the suggestion that HEL is crucial for fully understanding the HHB variable.

Finally, confirmation of the first study complete by Divelbiss' team was done through MTS. A different health data set was utilized (DHS) and a different tool, MTS, was used to identify significant variables within the data set. First, it was observed that of the six applicable variables that Divelbiss selected, four of them were found within the top ten of the 83 variables. The other two were also found to be in the top half of the data set. The second finding was that the most significant variable identified by more than twice that of the others was the maternal education level. The next was the paternal education level followed by the drinking water source. While this health survey did not analyze the usage of the CAWST biosand filter, water importance was still found to be second to education. It should also be noted that a majority of the other variables within the top ten were SES variables. These findings support the model in which Divelbiss used and was then translated for use in Brazil within this project. While these findings do not directly support or reject the original hypothesis for this thesis, it does suggest a more complex environment in which many organizations work.

From these results it can be observed that the original hypothesis that the CAWST biosand filter is the most significant factor in the reduction of diarrheal occurrences in developing communities, should be rejected. Factors that were consistently significant were household education level, socio economic status, improved sanitation and improved water source. While the CAWST biosand filter was shown to reduce diarrheal occurrences a holistic approach is needed when studying the complex environment in which it is found. Organizations that focus on implementing one project central to their expertise, need to be aware of the complex system to which they are infiltrating. While needs of people vary greatly across the developing world, it is the recommendation of this author that education or economic sustainability be part of a dual implementation if any aid is deemed necessary. Understanding the influences that pertain to the central embodiment of health is crucial to creating the largest positive impact within impoverished communities.

Finally, the traditional recommendation for further work needs to be offered. While these results are intriguing, a wider body of work is needed to create positive change within the field of foreign aid. This methodology is planned to be replicated to two other sites in vastly different regions of the world. The model in which was used has potential to provide a basic assessment tool for NGOs looking to understand the community and their implementation better. Currently many organizations assess themselves, this concept is highly frowned upon in other fields such as academics, health care, and business, yet no standardized assessment tool has been created. As the reader may now be aware, throughout the body of this article the argument is for the use of this tool and the platform of SEM for assessments of interventions. Further research will be undertaken within the a PhD.

APPENDIX A

SURVEY OF OWNERS OF CAWST BIOSAND WATER FILTERS

Survey of Owners of CAWST Biosand Water Filters

- Be accurate and write legibly
- If response is written incorrectly, strike line through it and rewrite correct response to the right
- If response is circled incorrectly, put an X through incorrect answer and circle correct response
- If response is unknown, write unknown
- Have interviewee sign document showing an understanding of research only usage of information

[MAKE SURE THEY UNDERSTAND THERE ARE NO RIGHT OR WRONG ANSWERS TO THIS SURVEY, WE ARE YOUR FRIEND, WE ARE LEARNING TOGETHER AND HOPE YOU FEEL COMFORTABLE AROUND US AND FEEL ABLE TO ASK ANY QUESTIONS YOU WANT TO US]

[INFORM INTERVIEWERS THAT IN QUESTIONS WITH CHOICE ANSWERS TO GIVE THE INTERVIEWEE A CHANCE TO ANSWER WITHOUT A PROMPT, IF A PROMPT IS NEEDED RESTATE QUESTION AGAIN AND THEN TRY AND GIVE BASIC EXAMPLES. THE LESS IDEAS THAT WE PUT INTO THEIR HEADS THE BETTER]
.

Survey Number: _________________ (Survey Set – Number in Set, Ex. 1 – 1, 1 – 2)

Interviewee Signature:

Please write answer in blank provided:

- How tall are you?
- How old where you when you had your first child?
- How many people in the household are below 20 years?
- What job is the most common among people who inhabit the particular area adjacent to the turbulent Amazon River?

Community Name: ______________________________

Name of Interviewer: ______________________________

Date of Interview: ______________________________

Household Name: ___________________________________

Water Source

1. How cups of water do you and your family drink per day?……………………………… _____________
2. What is your primary water source for drinking? **Open Well Closed Well River Bottled Water Community Distribution System Other** _____________
3. What is your primary water source for washing and cleaning? **Open Well Closed Well River Bottle Water Community Distribution System Other** ______________

4. How do you get water during the dry season?

Open Well Pump from river Walk to river Bottled
Closed Well Community Distribution System Other ______________

5. How many meters away is the water source during the low point of the dry season? ______________

6. How long does it take to walk to the water source, fill up your container, and return to your house? _______ _______

7. What water do you drink when not at home?

Unfiltered River Water Purchase bottled Open Well Closed Well Filtered River Water Bring water from house Other _______________

8. Do you or your family filter your water through your biosand filter before drinking it? **YES NO**
(If 8. is NO go to 15.)

9. What percentage of your water that your family drinks is filtered through the biosand filter? ________

10. How many days of the week do you or your family use your filter? ……………….. **7 6 5 4 3 2 1**

11. Do you treat your water before it goes through the biosand filter? ………………………………… **YES NO**

    a. If YES, how? **Chlorine Solar Disinfection Filter through Towel**
    Boil Settling/Clarification Other _______________

12. Do you treat your water after it goes through the biosand filter? …………………………………… **YES NO**

    a. If YES, how? **Chlorine Solar Disinfection Filter through Towel**
    Boil Settling/Clarification Other _______________

13. Does the individual using the filter use different containers for gathering of water from the source and collection of the filtered water from the biosand filter?

YES
NO

14. How do you store your drinking water?
Do not store water **In container with no lid or cover** **In container with lid but no spigot or tap**
In container with lid and spigot **In narrow-mouthed container** **Other** _____________ **Do not know**

Hygiene

15. What type of toilet facility is used?
    b. During rainy season **Latrine to river** **Latrine to closed pit** **Bucket/Barrel** **Bush/Field**
    Flush/Pour flush to closed pit **Piped/Septic Tank** **Other** ____________

    c. During dry season **Latrine to riverbed** **Latrine to closed pit** **Bucket/Barrel** **Bush/Field**
    Flush/Pour flush to closed pit **Piped/Septic Tank** **Other** ____________

16. Is this toilet facility shared with other families on a regular basis? ..……...…………………………. **YES** **NO**

17. How often do family members use soap while bathing in a week? …..………………………. _____________

18. What do you do with household trash? **Burn** **Garbage boat** **River** **Other** _____________

19. Where is the bathroom located? ………………………… **Inside or Connected to house** **Outside house**

House Information

20. Age of CAWST Biosand Filter: ………………………………………………………………. ____________

21. # of Bedrooms in Dwelling:………………………………………………………………………. ____________

22. What health issue is most common within your family? **Diarrhea** **Fever** **Cough** **Headache** **Other**

23. Within your family, has there been at least one case of **diarrhea** in the last week? **YES** **NO**
    a. If YES, how many different family members had diarrhea within the week? _________
    (and)

b. If YES, how many total days did the family member with the longest case of diarrhea within the household have it within the past week? …………………………………………….
____

24. How many days does a single family member wait before retrieving medicine from the health post when they or someone else within the house is sick?

____
____

25. How do you usually get to the health post or health center? **Taxi Car Taxi Boat Personal Boat**
Bicycle Walk Personal Car Other
____

Miscellaneous

26. Have you used the internet in the last month? ….…………………………………………………….. **YES NO**

27. Do you own a mobile phone? ……………………………………………………………………………. **YES NO**

Occupant Information:

	Name	Age	Highest Year Completed	English or Spanish
Father:	___________	_____	0 1 2 3 4 5 6 7 8 9 Secondary	YES NO
Mother:	___________	_____	0 1 2 3 4 5 6 7 8 9 Secondary	YES NO
Children:	___________	_____	In School Not in School	YES NO
	___________	_____	In School Not in School	YES NO
	___________	_____	In School Not in School	YES NO
	___________	_____	In School Not in School	YES NO
	___________	_____	In School Not in School	YES NO
Other:	___________	_____	0 1 2 3 4 5 6 7 8 9 Secondary	YES NO
	___________	_____	0 1 2 3 4 5 6 7 8 9 Secondary	YES NO

(Circle which person was interviewed)

Observation (to be done by interviewer)

28. Dwelling Flooring Type(s): ………………………………………...………. **Wood Concrete Dirt Other**

29. Dwelling Wall Type(s): ………………………………………..……. **Wood Concrete Dirt Metal Other**

30. Dwelling Roof Type(s): ………………………..……………… **Wood Concrete Thatched Metal Other**

31. Type of Road Leading to Dwelling: ……………………………………… **Water Concrete/Paved Dirt Other**

32. Does the family have in the house:

Television Radio Large Water Storage Tank Shower
Garden Animals/Pets MULTIPLE working light bulbs
Gas/Electric Stove Fire pit for cooking

33. How many locations are available for hand washing in the house? ………………………………… ________

34. Are these located in ATLEAST ONE of the areas where the family washes their hands? (Circle all that apply)

Water
Soap Towel

35. How far is the toilet facility from the water gathering source (meters)? …………………. ________________

36. Investigation of household hygiene:
   a. Flies present? …………………………………………………………………………………… **YES NO**
   b. Animal pen within ___***m*** of house or if elevated above water, connected to house? . **YES NO**
   c. Animals where they prepare food? ……………………………………………………………… **YES NO**

Observe Water Storage Container

37. What kind of container is used to store filtered or treated water?
Clay Plastic Metal Other _____________

38. What does the water storage container look like? (Circle all that apply)

Completely covered with lid Open, uncovered Narrow opening Spigot
Beyond reach of animals Clean Dirty Other _____________

Observe Filter:

39. Is there a diffusion/baffle plate which is undamaged and without cracks? ………………………….. **YES NO**

40. Is the filter in a suitable place? …………………………………………………………………………
YES NO

41. Is the sand surface flat and level? ………………………………………………………………………
YES NO

42. Are there leaks in the concrete filter body? …………………………………………………………..
YES NO

43. Is there a well-fitting lid? ……………………………………………………………………………….
YES NO

44. What is the depth of the water above the sand (cm)? ………………………………………..
____

45. What is the flow of the filter (secs/100mL)? ……………………………………………………
____

46. Is there a pipe or valve placed at the output of the filter? ……………………………………………..
YES NO

47. Is there a mesh fitting at the output of the filter? ……………………………………………………….
YES NO

48. Is the output of the filter clean/clear? ……………………………………………………………………
YES NO

49. Household Economic Level: Low Medium High Random

Ask if you can answer any questions they may have.

APPENDIX B

DHS HOUSEHOLD SURVEY (GUATEMALA - SPANISH)

Instituto Nacional de Estadística -INE-
Encuesta Nacional de Salud Materno Infantil 1998/99 [ENSMI-98/99]

CUESTIONARIO DEL HOGAR

IDENTIFICACIÓN CARTOGRÁFICA

Región Depto. Mplo. Sección Sector

IDENTIFICACIÓN PARA DIGITACIÓN

Paquete Hogar

JEFE DEL HOGAR: ______________________________

DIRECCIÓN DE LA VIVIENDA:	UBICACIÓN GEOGRÁFICA:
1. URBANO 2. RURAL	1. DEPARTAMENTO: ....
DIRECCIÓN:	2. MUNICIPIO: ....
....	3. ALDEA/CASERÍO/CANTÓN/FINCA: ....
....	....

RESULTADOS DE LAS VISITAS	1ra.	2da.	3ra.
FECHA Y HORA			
ENCUESTADORA			
RESULTADO*			
PRÓXIMA VISITA FECHA			
HORA			

VISITA FINAL: DIA | MES | AÑO 1 9

	CODIGO	NOMBRE
ENCUESTADORA		

RESULTADO DE LA ENTREVISTA* ____ Nº DE VISITAS ____

*CODIGOS DE RESULTADO

1- COMPLETA
2- NO ADULTOS PRESENTES
3- AUSENTE (NADIE EN CASA POR MUCHO TIEMPO)
4- POSPUESTA
5- RECHAZO
6-VIVIENDA VACANTE O LA DIRECCION NO ES VIVIENDA
7- VIVIENDA DESTRUIDA
8- VIVIENDA NO ENCONTRADA
9- OTRO ____________ (ESPECIFIQUE)

Nº DE MIEMBROS DEL HOGAR

Nº TOTAL DE MUJERES ELEGIBLES

Nº DE LINEA DE PERSONA QUE RESPONDE CUESTIONARIO DE HOGAR

	HORAS	MINUTOS
HORA INICIAL		
HORA FINAL		

IDIOMA DE LA ENTREVISTA

IDIOMA MATERNO

IDIOMA DEL CUESTIONARIO 0 1

¿SE USO TRADUCTOR? 1=SI 2=NO

01 ESPAÑOL
02 KAQCHIQUEL
03 Q'EQCHI
04 K'ICHE
05 MAM
06 POQOMCHI'
07 TZU'UTUJIL
08 KANJOBAL
09 CHORTI
10 POCOMAM
96 OTRO ____________ (ESPECIFIQUE)

CONTROL DE OFICINA	SUPERVISADO EN EL CAMPO POR:	EDITADO EN EL CAMPO POR:	EDITADO EN LA OFICINA POR:	DIGITADO POR:
NOMBRE				
FECHA				
CODIGO				

ESPAÑOL (FEBRERO 1999)

Ahora nos gustaría tener alguna información sobre las personas que generalmente viven en su hogar o que se alojan ahora con usted.

Nº DE LINEA	RESIDENTES HABITUALES Y VISITANTES	RELACIÓN CON EL JEFE DE HOGAR*	LUGAR DE RESIDENCIA		SEXO	EDAD	EDUCACIÓN PARA PERSONAS CON 6 AÑOS O MAS				SUPERVIVENCIA Y RESIDENCIA DE LOS PADRES DE (NOMBRE) PARA PERSONAS MENORES DE 15 AÑOS				LUGAR DE NACIMIENTO DE (NOMBRE)	ELEGIBI-LIDAD
								SI ASISTIÓ A LA ESCUELA	PARA TODA PERSONA MENOR DE 25 AÑOS			SI VIVE		SI VIVE		
	Dígame por favor los nombres de las personas que habitualmente viven en su hogar y de los Visitantes que pasaron la noche anterior aquí, comenzando por el jefe del hogar. *INSTRUCCIÓN:* *LLENAR ESTA COLUMNA COMPLETA ANTES DE PASAR A LA PREGUNTA (3)*	¿Cuál es la relación de (NOMBRE) con el jefe del hogar?	¿Vive (NOMBRE) perma-nente-mente aquí?	¿Durmió (NOMBRE) aquí anoche?	¿Es (NOMBRE) hombre o mujer?	¿Cuántos años cumplidos tiene? *Instrucción Menores de 1 año = 00* *No Sabe = 98*	¿Ha asistido alguna vez (NOMBRE) a la escuela?	¿Cuál es el último año de estudios que ganó (NOMBRE)?	¿Está (NOMBRE) todavía en la escuela**?	YA NO ESTA EN LA ESCUELA ¿Por qué razón dejó de estudiar***?	¿Vive la madre natural de (NOMBRE)?	Y SI VIVE EN EL HOGAR ¿Cuál es su nombre? (SI NO VIVE EN HOGAR ANOTE 00)	¿Vive el padre natural de (NOMBRE)?	Y SI VIVE EN EL HOGAR ¿Cuál es su nombre? (SI NO VIVE EN HOGAR ANOTE 00)	¿Cuál es el nombre del lugar de origen de (NOMBRE), es decir el municipio o departamento donde nació? SI EL LUGAR ES EL MISMO DONDE SE ESTA REALIZANDO LA ENCUESTA ANOTE EL CODIGO "0000"	CIRCULE EL NUMERO DE LAS MUJERES DE 15 A 49 AÑOS
(01)	(02)	(03) RELACION	(04) SI NO	(05) SI NO	(06) H M	(07) AÑOS	(08) SI NO N/S	(09) NIVEL GRADO	(10) SI NO N/S	(10 A) CODIGO	(11) SI NO N/S	(12) Nº LINEA	(13) SI NO N/S	(14) Nº LINEA	(14 A) CODIGO Y NOMBRE DEPTO. MPIO.	(15)
01			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			01
02			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			02
03			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			03
04			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			04
05			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			05
06			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			06
07			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			07

Continuación del Cuestionario del Hogar

(01)	(02)	(03) RELACION	(04) SI NO	(05) SI NO	(06) H M	(07) AÑOS	(08) SI NO N/S	(09) NIVEL GRADO	(10) SI NO N/S	(10 A) CODIGO	(11) SI NO N/S	(12) Nº LINEA	(13) SI NO N/S	(14) Nº LINEA	(14 A) CODIGO Y NOMBRE DEPTO. MPIO.	(15)
08			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			08
09			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			09
10			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			10
11			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			11
12			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			12
13			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			13
14			1 2	1 2	1 2		1 2 8		1 2 8		1 2 8		1 2 8			14

MARQUE AQUÍ SI CONTINÚA EN OTRA HOJA ☐

Sólo para estar segura que tengo una lista completa:

¿Hay otras personas como niños o recién nacidos que tal vez no hemos listado?	SI ☐	→ANOTE A CADA UNO EN EL CUADRO	NO ☐
¿Hay otras personas que no son familiares, como empleados domésticos que viven habitualmente aquí, que no hemos anotado?	SI ☐	→ANOTE A CADA UNO EN EL CUADRO	NO ☐
¿Tiene Ud. huespedes, inquilinos, habitantes temporales, o alguien más que haya dormido aquí anoche?	SI ☐	→ANOTE A CADA UNO EN EL CUADRO	NO ☐

CODIGOS PARA LA PREGUNTA 03

RELACION CON EL JEFE DE HOGAR:

01 = JEFE DEL HOGAR
02 = ESPOSA / ESPOSO
03 = HIJO/HIJA
04 = YERNO/NUERA
05 = NIETO/NIETA
06 = PADRE/MADRE
07 = SUEGRO/SUEGRA
08 = HERMANO/HERMANA
09 = OTRO FAMILIAR
10 = HIJO/A ADOPTADO/A O DE CRIANZA
11 = NO HAY PARENTESCO
98 = NO SABE

** CÓDIGOS PARA LAS PREGUNTAS 09 Y 10

NIVEL EDUCATIVO:	GRADO:	
0 = NINGUN NIVEL/PREESCOLAR	0	= MENOS DE UN AÑO DE EDUCACION APROBADO
1 = PRIMARIA	0,1, 2, 3, 4, 5, 6	= GRADOS PARA NIVELES PRIMARIA Y SECUNDARIA
2 = SECUNDARIA	0,1, 2, 3, 4, 5, 6	
3 = SUPERIOR/UNIVERSITARIA	0,1, 2, 3, 4, 5	= GRADOS PARA NIVEL SUPERIOR
4 = ALFABETIZACION	0	= ALFABETIZACION
8 = NO SABE	8	= NO SABE

*** CODIGOS PARA LA PREGUNTA 10A

RAZON DE ABANDONO DE ESCUELA

01 QUEDO EMBARAZADA
02 SE CASO
03 TENIA QUE CUIDAR NIÑOS
04 LA FAMILIA NECESITABA AYUDA
05 NO PUDO PAGAR LA CUOTA
06 NECESITABA GANAR DINERO
07 SE GRADUO/SUFICIENTE ESCOLARIDAD
08 MALAS CALIFICACIONES
09 NO LE GUSTABA LA ESCUELA
10 ESCUELA NO ACCESIBLE/MUY LEJOS
11 POR ENFERMEDAD/RAZON MEDICA
96 OTRA RAZON
98 NO SABE

PREGUNTAS RELACIONADAS CON LA CASA DE HABITACION				
No.	PREGUNTAS Y FILTROS	CATEGORÍAS Y CÓDIGOS		PASE A
16	¿Cuál es la fuente principal de abastecimiento de agua para beber que utilizan los miembros del hogar?	RED PÚBLICA ACUEDUCTO	11	18
		CHORRO PÚBLICO	12	
		OTRA FUENTE POR TUBERÍA	21	18
		PILA PÚBLICA/TANQUE PÚBLICO	22	
		POZO MECÁNICO/MANUAL (BROCAL)	23	
		RÍO/ACEQUIA/MANANTIAL	31	
		AGUA DE LLUVIA	41	18
		CAMIÓN O TANQUE AGUATERO	51	
		AGUA EMBOTELLADA	61	
		OTRO ________ (ESPECIFIQUE)	96	
17	¿Cuánto tiempo toma llegar allá, recoger agua y volver? (A PIE)	MINUTOS	[ ][ ][ ]	
		EN EL SITIO	9 9 6	
18	¿Qué tipo de servicio sanitario tiene su casa?	INODORO CONECTADO AL ALCANTARILLADO PRIVADO	11	
		COMPARTIDO	12	
		INODORO CONECTADO A FOSA SÉPTICA	21	
		LETRINA, POZO CIEGO, ESCUSADO	22	
		NO TIENE SANITARIO	31	
		OTRO ________ (ESPECIFIQUE)	96	
19	¿Tiene en su casa:		SI / NO	
	Luz eléctrica?	A) LUZ ELÉCTRICA	1 / 2	
	Radio?	B) RADIO	1 / 2	
	Televisor?	C) TELEVISOR	1 / 2	
	Teléfono?	D) TELÉFONO	1 / 2	
	Refrigeradora?	E) REFRIGERADORA	1 / 2	
20	¿Cuántos cuartos (habitaciones) usan en su casa para dormir?	NUMERO DE HABITACIONES	[ ][ ]	
21	MATERIAL PRINCIPAL DEL PISO SOLO POR OBSERVACIÓN	PISO NATURAL (TIERRA/ARENA)	11	
		LADRILLO DE BARRO O TIERRA	13	
		PISO RÚSTICO (TABLAS DE MADERA)	21	
		MADERA LUSTRADA	31	
		LADRILLO DE CEMENTO (MOSAICO)	32	
		CERAMICA	33	
		TORTA DE CEMENTO	34	
		OTRO ________ (ESPECIFIQUE)	96	
21A	MATERIAL PRINCIPAL DEL TECHO SOLO POR OBSERVACIÓN	PAJA/PAJON/PALMA	01	
		TEJA	02	
		LAMINA DE ZINC/METALICA	03	
		LAMINA DE ASBESTO (DURALITA)	04	
		LOZA/TERRAZA	05	
		OTRO ________ (ESPECIFIQUE)	96	
21B	MATERIAL PRINCIPAL DE LAS PAREDES SOLO POR OBSERVACIÓN (CONSIDERE COMO MATERIAL DE DESECHO: CARTON, PLASTICO, NYLON, ALUMINIO)	BAJAREQUE	01	
		ADOBE	02	
		BLOCK	03	
		LAMINA	04	
		MADERA	05	
		LADRILLO DE BARRO	06	
		MATERIAL DE DESECHO	07	
		OTRO ________ (ESPECIFIQUE)	96	
21C	¿La casa donde actualmente viven es propia, alquilada o prestada?	PROPIA	01	
		ALQUILADA	02	
		PRESTADA	03	
		OTRA ________ (ESPECIFIQUE)	96	

PREGUNTAS RELACIONADAS CON LA CASA DE HABITACION			
No.	PREGUNTAS Y FILTROS	CATEGORÍAS Y CÓDIGOS	PASE A
22	¿Algún miembro de los que viven en su casa tiene y utiliza para su transporte Bicicleta? Motocicleta? Carro? (Automóvil) Tractor? Bestia, caballo, mula?	SI NO A) BICICLETA ---------- 1 2 B) MOTOCICLETA ---------- 1 2 C) CARRO ---------- 1 2 D) TRACTOR ---------- 1 2 E) BESTIA, CABALLO, MULA ---------- 1 2	
23	¿Qué tipo de sal utiliza para sazonar sus alimentos?	SAL LOCAL ---------- 01 SAL EN BOLSAS CON MARCA ---------- 02 SAL EN BOLSAS SIN MARCA ---------- 03 SAL PARA GANADO/ANIMALES (GRANULADA) ---------- 04 OTRO ______ 96 (ESPECIFIQUE)	
24	¿Me puede mostrar la sal que regularmente utiliza para sazonar sus alimentos? HAGA LA PRUEBA DE CONTENIDO DE YODO A LA SAL QUE UTILIZAN PARA SAZONAR SUS ALIMENTOS SI LA BOLSA TIENE MARCA ESCRÍBALA ______ MAQUE CON UNA X EL CUADRO SEGUN EL RESULTADO DE LA PRUEBA (POSITIVO/NEGATIVO). LUEGO CIRCULE EL PORCENTAJE Y MARQUE CON UNA X EL CUADRO SEGÚN EL RESULTADO	RESULTADO DE LA PRUEBA: CON YODO ---------- 1 SIN YODO ---------- 2 OTRO ______ 6 (ESPECIFIQUE) RESULTADO DE LA PRUEBA POSITIVO (MORADO) [1] NEGATIVO (BLANCO) [2] PORCENTAJE 0 [1] 25 [2] 50 [3] 75 [4] 100 [5] +100 [6]	
25	¿Qué tipo de combustible utiliza generalmente para cocinar?	LEÑA/CARBÓN ---------- 01 GAS CORRIENTE (KEROSSENE) ---------- 02 GAS PROPANO ---------- 03 ELECTRICIDAD ---------- 04 OTRO ______ 96 (ESPECIFIQUE)	
26	¿Tiene en su hogar un lugar (cuarto, ambiente) que utilice únicamente para cocinar?	SI ---------- 1 NO ---------- 2	
27	¿Tiene chimenea en el lugar en el que regularmente cocina sus alimentos?	SI ---------- 1 NO ---------- 2	
28	¿Qué hace usted con las basura que se produce (sale) de su casa? ¿Cómo se deshace de la basura?	LA RECOJEN REGULARMENTE (SERVICIO MUNICIPAL) ---------- 01 LA RECOJEN REGULARMENTE (SERVICIO PRIVADO) ---------- 02 LA TIRAN EN EL TERRENO/LA MILPA ---------- 03 LA TIRAN EN LA CALLE ---------- 04 LA ENTIERRAN ---------- 05 NO TIENEN COMO DESHACERSE DE LA BASURA ---------- 06 OTRO ______ 96 (ESPECIFIQUE) NO SABE ---------- 98	
29	¿Tiene algún miembro de la familia alguna porción de tierra que sirva para cultivar?	SI ---------- 1 NO ---------- 2 NO SABE ---------- 8	 → 34 → 34
30	¿Cuál es la extensión de esa porción de tierra? ANOTE LA RESPUESTA TEXTUAL ______ ______	Nº DE CUERDAS ---------- [ ][ ][ ][ ][ ]	
31	¿Cultivan algún producto como granos (maíz, frijol) y/o verduras en esa porción de tierra?	SI ---------- 1 NO ---------- 2 NO SABE ---------- 8	 → 34 → 34
32	¿Cuántos productos cultiva generalmente durante el año?	NUMERO DE PRODUCTOS ---------- [ ][ ]	

PREGUNTAS RELACIONADAS CON LA CASA DE HABITACION

No.	PREGUNTAS Y FILTROS	CATEGORÍAS Y CÓDIGOS	PASE A
33	¿Qué cultivan durante el año? INDAGUE: ¿En que mes siembra o cultiva e/la (NOMBRE DEL CULTIVO)? ¿En que mes lo/la cosecha?		

CODIGO		NOMBRE DEL CULTIVO	MES DE SIEMBRA		MES DE COSECHA	

No.	PREGUNTAS Y FILTROS	CATEGORÍAS Y CÓDIGOS	PASE A
34	¿En alguna época del año por motivos de trabajo, algún miembro de la familia viaja a otro lugar (departamento, municipio o el extranjero)?	SI ---- 1 NO ---- 2 NO SABE ---- 8	 → 37 → 37
35	¿Cuántos miembros de su familia viajan a otro lugar a trabajar?	NUMERO DE MIEMBROS DEL HOGAR ---- ☐☐ TODOS VIAJAN ---- 9 5	
36	¿A que lugar o lugares viajan?		

SI TODOS LOS MIEMBROS DEL HOGAR VIAJAN ENTONCES INDAGUE Y REGISTRE UNICAMENTE LOS DATOS DEL JEFE DEL HOGAR.

SI VIAJA/N AL EXTRANJERO ANOTE 9995 EN LOS RECUADROS PARA DEPARTAMENTO Y MUNICIPIO.

ANOTE LA RESPUESTA TEXTUAL EN EL ESPACIO PARA EL NOMBRE DEL LUGAR

Nº DE LINEA		MES DEL AÑO		NOMBRE DEL LUGAR	CODIGO DEL LUGAR DEPTO.		MPIO.	

37	VEA EL NUMERO TOTAL DE MUJERES ELEGIBLES EN LA CARATULA DE ESTE CUESTIONARIO, SI EXISTE UNA O MAS ENTONCES HAGA LAS ENTREVISTAS INDIVIDUALES CORRESPONDIENTES, SI NO, ENTONCES FINALICE LA ENTREVISTA

APPENDIX C

TOP 50 SIGNIFICANT INDICATORS FOR HEALTH IN DHS SURVEY USING MTS ANALYSIS

1	Mother's Educational attainment	26	Education in single years
2	Father's Highest educational level	27	Quintiles of wealth index
3	Source of drinking water	28	Garbage disposal
4	Has car	29	Main wall material
5	Number of family members in house	30	Member still in school
6	Time to get to water source	31	Has tractor
7	Sex of head of household	32	Wealth Index from factor scores
8	Has a place/room to cook	33	Salt contains iodine
9	Has motorcycle	34	Number of household members
10	Has a chimney in the kitchen	35	Member still in school
11	Rooms for sleeping	36	Education in single years
12	Has bicycle	37	Usual resident
13	Translator used	38	Number of de facto members
14	Has telephone	39	Iodized percent
15	Main roof material	40	Usual resident
16	Sample stratum number	41	Slept last night
17	Has horse/mules	42	Age of household members
18	Type of toilet facility	43	Fuel used to cook
19	Language of interview	44	Number of children 5 and under
20	Highest year of education	45	Sex of household member
21	Age of household members	46	Has television
22	Has electricity	47	Relationship structure
23	Highest year of education	48	Relationship to head
24	Has radio	49	Dwelling ownership
25	Age of head of household	50	Has refrigerator

APPENDIX D

EXPLANATION OF VARIABLES

Improved Sanitation and Water Treatment, Storage and Source

All scoring was derived from definitions provided by the World Health Organization. For further information see (World Health Organization, 2012), (World Health Organization, 2006). A 1 was given for improved and a 0 for unimproved.

Paternal and Maternal Education Level

A number of organizations including World Bank and the United Nations utilize the school education level heavily. The specific definition may vary however, within our survey it can be seen as the last school grade or level completed (International Household Survey Network, 2009). One point was given for each grade level completed.

Reading and Writing Test

The reading and writing test consisted of four questions increasing in difficulty of reading and comprehension. The interviewee was asked to read the sentence and then write their response. A point was earned for every higher level question answered and written legibly. If the respondent asked for help on understanding or reading, the location at which the question was asked was noted (International Household Survey Network, 2009).

Improved Roof

Three different types of roves were typically found throughout the communities. Households received a 1 if they had a clay tiled roof and a 0 if they had a corrugated concrete or thatched roof.

Ownership of physical item

Several different items were investigated that were thought, from a priori knowledge and local guides, were desirable items to possess. These items included a television, radio, water tank, shower, garden, animals or pets, gas or electric stove. The possession of these items was theorized to reflect the level of socio economic status in which the household was. It should be noted that while this indicator relies heavily on the researcher, the analysis in the MTS section of this paper gives hope to the mathematical steps to identifying indicators. A 1 was given if they owned the item and 0 if they did not.

Working Light bulbs

The investigation of working light bulbs within the household was also used to indicate socio economic status. It was thought that if a house used more light bulbs this may indicate multiple things; first a need for more light at night for homework or other work, increase in room size or number of rooms, access or knowledge of the power system in the village and the money to run the system. A 1 was given if they had multiple working light bulbs and a 0 if they had one or less.

Household Density

The household density also reflected the economic status of the house. It was the number of people living in the house divided by the number of rooms in the house. A higher number reflected the potential for worse living conditions in theory.

Occurrence of sickness (or diarrhea)

The interviewee was asked if they have had sickness (or diarrhea) in the last two weeks. Two weeks was selected based on surveys from the World Bank and articles on recall periods (Arnold et al., 2013). A 1 was scored if sickness (or diarrhea) was present and a 0 if not. Sickness was measure in Paper 2 while in Paper 3 diarrhea was measured, this was dependent on the actual occurrences in the data. During the low season, diarrhea is less of an issue and therefore general health was analyzed.

Multiple cases and length

If sickness (or diarrhea) was found to be present, the interviewee was then asked how many times this occurred over a two week period. This would indicate if the contamination is consistent or if it may have been contracted outside of the home. Length of the bout of diarrhea was then asked to understand severity level. Again, a 1was given if diarrhea was an issue in either of these separate questions or a 0 if it was not.

Children under five

It can be thought that if a house has an increase in diarrheal occurrence they are more likely to have small children, as children are at a higher risk of contracting diarrheal related diseases (Bartram, 2008). If the household had at least one child under five they received a 1, if they did not, the household received a 0.

Filter operation and maintenance

A combination of ten questions and observations were based on scoring the usage of the filter by the household. Water samples were taken in the first administration of surveys to compare correlation of FOM and DHB to contamination levels (Centre for Affordable Water and Sanitation Technology, 2009).

BIBLIOGRAPHY

Ahmed, T., & Svennerholm, A. (2009). Diarrheal disease: solutions to defeat a global killer. *Path: A Catalyst for Global Health.*

Alkire, S., & Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *The Journal of Economic Inequality*, *9*(2), 289–314. doi:10.1007/s10888-011-9181-4

Arnold, B. F., Galiani, S., Ram, P. K., Hubbard, A. E., Briceño, B., Gertler, P. J., & Colford, J. M. (2013). Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. *American Journal of Epidemiology*, *177*(4), 361–70. doi:10.1093/aje/kws281

Bartram, J. (2008). *Sanitation is vital for human health*. New York.

Ben-Joseph, E. P. (2013). Kids Health. *Infections - Diarrhea*. Retrieved from http://kidshealth.org/parent/infections/common/diarrhea.html#, March 2014

Black, R. E., Morris, S. S., & Bryce, J. (2003). Where and why are 10 million children dying every year ?, *361*, 2226–2234.

Centre for Affordable Water and Sanitation Technology. (2009). Biosand filter manual design, construction, installation, operation and maintenance. Calgary, Alberta, Canada: Creative Commons.

Centre for Affordable Water and Sanitation Technology. (2013). Biosand Filter. Retrieved from http://www.cawst.org/en/resources/biosand-filter, May 2014

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262. doi:10.1016/j.jbusres.2008.01.013

Deutsch, J., & Silber, J. (2005). Measuring multidimensional poverty: An empirical comparison of various approaches. *Review of Income and Wealth*, (1).

Divelbiss, D. W., Boccelli, D. L., Succop, P. A., & Oerther, D. B. (2013). Environmental health and household demographics impacting biosand filter maintenance and diarrhea in Guatemala: an application of structural equation modeling. *Environmental Science & Technology*, *47*(3), 1638–45. doi:10.1021/es303624a

Ferreira, F. H. G. (2011). Poverty is multidimensional. But what are we going to do about it? *The Journal of Economic Inequality*, *9*(3), 493–495. doi:10.1007/s10888-011-9202-3

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford, J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *5*(1), 42–52. doi:10.1016/S1473-3099(04)01253-8

Grace, J. B. (2006). *Structural Equation Modeling and Natural Systems* (First.). United Kingdom: Cambridge University Press.

International Household Survey Network. (2009). How well is Education Measured in Household Surveys?, (002).

Martindale, T. (2013). *Planning, Monitoring, Evaluation and Learning Program* (pp. 1–31).

Ozkan, S., Tüzün, H., Görer, N., Ceyhan, M., Aycan, S., Albayrak, S., & Bumin, M. A. (2007). Water usage habits and the incidence of diarrhea in rural Ankara, Turkey. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *101*(11), 1131–5. doi:10.1016/j.trstmh.2007.05.011

Peace Corps. (2012). *The Peace Corps Performance Plan* (pp. 1–27).

Quick, R. E., Kimura, A., Thevos, A., Tembo, M., Shamputa, I., Hutwagner, L., & Mintz, E. (2002). Diarrhea prevention through household-level water disinfection and safe storage in Zambia. *The American Journal of Tropical Medicine and Hygiene*, *66*(5), 584–9.

Taguchi, G., & Jugulum, R. (2002). *The Mahalanobis-Taguchi Strategy*. New York: John Wiley & Sons, Inc.

The World Bank. (2014). Brazil Overview. Retrieved from http://www.worldbank.org/en/country/brazil/overview, April 2014

Walker, R. (2007). *Measuring change in multidimensional poverty using structural equation modelling*. doi:Project Research Report RES-000-23-1418

WHO Media Centre. (2013). Diarrhoeal Disease. *Fact Sheet*. Retrieved from http://www.who.int/mediacentre/factsheets/fs330/en/, March 2014

World Health Organization. (2006). *Core questions on drinking-water and sanitation for household surveys*. Geneva, Switzerland.

World Health Organization. (2012). *A toolkit for monitoring and evaluating household water treatment and safe storage programmes*. Geneva, Switzerland.

World Health Organization. (2013). *Levels & Trends in Child Mortality* (pp. 1–34).

VITA

Lee Voth-Gaeddert was the author of this thesis. He earned a high school diploma from Hesston High School in May of 2008 and in December of 2012 earned a Bachelor's of Science in Civil Engineering from Missouri University of Science and Technology. The degree Lee currently is working on is a Master's of Science in Environmental Engineering and will receive this in August of 2014.