Developing Location-Specific Measures of Socio-Economic Status in Partially Subsistent Economies

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Susan Gaye Dawson

July, 1997

Susan Gaye Dawson

Developing Location-Specific Measures of Socio-Economic Status in Partially Subsistent Economies

ABSTRACT

The broad aim of this study was to assess the utility of *wealth ranking* as a technique to develop valid measures of socio-economic status in partially subsistent economies. These measures were required to be location-specific, that is, reflective of geographical, climatic, cultural and economic differences across a range of locations in Tanzania. In addition, all procedures adopted to develop and validate these measures were expected to be applied by non-specialists without advanced research skills and with limited access to resources.

The broad objectives of this methodological study of measurement development were: to examine in detail the performance of the wealth ranking technique and its' feasibility as a tool for non-specialists; to assess the criterion-related and construct validity of indices of socio-economic status constructed with criteria generated during wealth ranking sessions, and to compare this with an assessment of the validity of traditional indices already in use in Tanzania. Criterion-related validity was measured using screening tests of sensitivity and specificity, and construct validity was assessed using the multitrait, multimethod approach.

The results suggested that simple wealth ranking (where households are allocated a socio-economic rank by a number of respondent groups) shows potential as a tool to identify the socio-economic status of small groups of households. Reliability between respondent groups was high. The method was feasible for use by non-specialists at the district level in that it required minimal resources and skills to conduct.

Wealth Ranking, as a tool to identify and select items for an index of socio-economic status for research purposes, requires further development. The community generated index of socio-economic status showed less construct validity than the traditional index. Neither index could be considered a suitable substitute for simple wealth ranking (criterion-relatedness). The skills required to generate indicators were more advanced than for simple wealth ranking.

Measurement theory is used to discuss the findings of the work. Gaps are identified in the theoretical and empirical base that support wealth ranking as a method to generate criteria. That is, criteria that reflect a subjective community construct of socio-economic status. An alternative model of study constructs is presented to explain the findings and to suggest that the criteria stated during wealth ranking represent a 'linguistic' construct. This construct may not fully capture the cognitive categorisation process used by respondents during the allocation of socio-economic rank to households.

ACKNOWLEDGEMENTS

I am indebted to a number of people whose support and encouragement allowed me to complete this work and develop significant, life-long friendships.

Dr. Bernard Schlecht who supervised this work and supported my interest in inequities in health.

From my supervisory panel, Dr. Dave Haran for his academic guidance and encouragement which always seemed to arrive at just the right moment, and in just the right dose. I am particularly grateful for his important contribution during the final hours of writing up; for his friendship; and his ability to provide members of his team with an excellent environment for personal development. Thanks also to Ms Korrie de Koning for her stimulating debates over delicious dinners, and her assistance during writing up. Her presence, especially in the early days, gave me the confidence to tackle such a difficult subject.

Dr. Maria-Luisa Vázquez who provided substantial assistance, intellectual input and emotional support at an important phase during my writing up period in Liverpool.

Dr. Peter Sandiford for his unshakeable confidence in my ability during my early, more insecure days.

Ms. Chris Ellis, our group secretary, who always seemed to be at the end of the fax machine or phone awaiting to respond to each and every crisis. I have valued her support and friendship over the years. Thanks also goes to Ms. Anne Gordon for her thorough and speedy proof reading and editing of the final draft.

Thank you to Mr. Male-Mukasa, Director of the African Medical and Research Foundation with whom we collaborated in Tanzania and Dr. Ahmed, our counterpart from Istituto Superiore di Sanita (Italy) who provided all the support required to get the project off the ground and completed. Dr. Upunda and Mr. Mapunda from the Tanzanian Ministry of Health, Health Sector Reform Committee for their genuine interest and never ending support. Thanks to Mr. Dave Whiting from the ODA Adult Morbidity and Mortality Project in Dar es Salaam who facilitated an extension of the project to include two new regions.

A very special thanks to Mrs. Mary Ahungu, our Research Associate, and Mr. Ramadhani Msingwa, the Kisarawe District Nursing Officer, who not only assisted in data collection throughout the entire study, but both made substantial contributions to the development and quality of this work.

Thanks to the ODA Health Sector Reform Work Programme who supported my three year Fellowship, and the EEC for the funds to carry out the research.

To all the leaders and members of the communities in which we worked, who gave up their time, and in many cases income, to contribute to this work. May their efforts not be in vain. A special thanks also to the wonderful people from the District Health Management Teams of Kisarawe, Rufiji, Mafia Island, Morogoro Rural and Hai districts in Tanzania. Their contribution and support was an important factor in this work.

To Mr. Devansh Nanda, who loved and supported me throughout the past three years, celebrated in my achievements, and provided the perfect environment in which to work.

Finally, to Dr. Drew Dawson from the University of South Australia, for showing me the importance of thinking.

TABLE OF CONTENTS

CHAPTER TWO:
Methods and Procedures

CHAPTER THREE:
Research Findings

CHAPTER FOUR:
Appropriate Applications of Wealth Ranking:
A Post-Investigative Assessment

LIST OF FIGURES AND TABLES

Figures

Tables

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ILO	International Labour Organisation
SES	Socio-economic status
LDC	Less Developed Country
ISC	Index of Status Characteristics
BRG	British Registrar General
AFES	Average Family Educational Score
NGO	Non-governmental Organisation
FUG	Forest User Group
HQ	Headquarters
SEG	Socio-economic group
SEG1	Wealthiest socio-economic group
SEG4	Poorest socio-economic group
MTMM	Multitrait, Multimethod Validation Technique
WR	Wealth Ranking

Chapter One

Socio-economic Status in Partially Subsistent Economies: Selecting an Approach for Measurement Development

1.1 Introduction and Study Overview

The broad aim of this study was to assess the utility of *wealth ranking* as a technique to develop valid measures of socio-economic status (SES) in partially subsistent economies. These measures were required to be location-specific, that is, reflective of geographical, climatic, cultural and economic differences across a range of locations. In addition, all procedures adopted to develop and validate these measures were expected to be applied by non-specialists without advanced research skills and with limited access to resources.

The study is divided into two separate, but related research activities. The first examined the feasibility and certain aspects of performance of the wealth ranking technique. Specifically, it measured the reliability of the technique; it compared the indicators of SES identified by different community respondents groups within and across a range of geographical locations; and it assessed the skills and resources required to carry out the technique successfully.

The second phase of activities involved the construction and validation of a composite index of SES using indicators identified by the community during wealth ranking exercises. A second type of index was constructed and validated during this phase for comparative purposes. This latter index is referred to as 'traditional', and was constructed using indicators identified by researchers or professionals as important indicators of SES in Tanzania. Both community generated and traditional indices were subjected to two assessments of validity, whereby the main study hypothesis was tested. Validation techniques were chosen based on their acceptability to test empirically the main study hypothesis, and the ability of typical peripheral health and social sector workers to apply them. It was considered responsible to provide users of a new technique to develop measures of SES with simple, but acceptable tools to assess the validity of their proposed measures.

The study was designed in response to the new informational requirements of the Health Sector Reform movement. In 1993, The World Development Report, an

influential document for the new directions in international health, made frequent references to the need to target public expenditure to the poor (World Bank, 1993a, e.g. p.119). This has become a popular policy in many developing countries, and as a consequence, the development of methods to identify these groups in a range of different settings, and for a range of purposes is indicated.

Accurate measures of socio-economic status are now necessary for a range of applications in health service delivery and research. For example, there has been a strong interest in measuring the equity in access to, or utilisation of, health services by different socio-economic groups. This has been prompted largely by new approaches to health sector financing such as user charges (or cost sharing) and health insurance schemes. Not only are the equity implications of interest, but many of these schemes are justified by the yet unproven assumption that exemption schemes or waivers effectively identify the poor for free health care (Dawson, 1997).

With these reforms comes a need for new information for the Health Sector. Information about who the poor are, how to reach them with health services, and their experience with the health care delivery system. We need tools for community-based surveys, as well as tools for the ongoing health management information systems. We need to be able to identify the poor who present to health facilities accurately so that waivers from health care charges are appropriately granted. These tools are needed at the national, regional and district levels. This study is focused on the informational needs at the district level in Tanzania, East Africa. It is concerned with district-specific measures rather than developing a single measure that can be applied nationally for Ministry purposes. The intention is to improve the capability of the District Health Management Teams to learn more about their own district populations. District Health Managers need practical and feasible tools to move towards these ends now, rather than waiting for sophisticated measures that are 'proven' to have uniform utility in a range of settings - if, that is, we believe such measures will ever exist.

The difficulties faced when trying to develop accurate measures of socio-economic status are well known. Most of the work in this field has been carried out in the industrialised West where measures of occupation, income and education have dominated the field. For developing countries these measures may not be appropriate where households are not earning regular salaries, where the range of occupations tend to be restricted to a limited number of agricultural-based activities, and where educational attainment may be uniformly low. Despite these obvious constraints, measures of socio-economic status used in developing countries tend to mirror those used in the West.

As the study is concerned with accurate measures for the district level, nationally applied measures may not always be appropriate. For example, livestock ownership is a commonly used measure in Tanzania and other agricultural-based economies. However, the ownership of livestock is not uniform throughout the country and in many regions the lack of ownership of livestock should not necessarily be interpreted as a sign of poverty. The same situation is faced with educational status. In many regions in Tanzania, education is not perceived to be an important nor prudent investment for families. As a result we may find most families educating their children to year seven, despite a range of socio-economic conditions. Therefore, it may be more appropriate to identify indicators of socio-economic status that are location-specific. These may vary from region to region, or even district to district according to a number of cultural, climatic, economic or geographical factors.

In view of these issues, this introductory chapter examines the literature on the conceptualisation and measurement of socio-economic status. This is done in order to select an appropriate approach for District Health Management Teams to develop relevant measures of SES for a partially subsistent economy. It begins with an examination of the theory of measurement development in general. This provides a framework for the remainder of the literature review, and consequent development of the measures in this study. The next section reviews the general approaches to the conceptualisation of socio-economic status in the industrialised

and developing worlds. After identifying the most commonly used indicators of these conceptualisations, general measurement issues relating to these indicators are examined for both industrialised and developing countries.

After a general reflection on the literature, a subjective approach to measurement development is selected to reflect a community conception of SES. Wealth ranking is a technique that has been proposed to generate indicators of such a construct. It was selected for use in this study largely because it has been promoted as an appropriate tool for non-specialists, and would therefore meet the needs and capacities of typical district level workers. The chapter closes with an examination of the literature concerned with wealth ranking; how and for what purpose it has been most commonly applied, and a review of studies that have conducted empirical assessments of its value.

Chapter two describes the theoretical framework of the study which is situated in measurement theory . It describes the research questions posed, and the main study hypothesis that indices of socio-economic status using community generated indicators are a more valid measure of the community construct of SES than traditional measures using indicators selected by researchers or social sector professionals.

Chapter two goes on to explain the study design, methods and procedures for the two main phases of the research. The remainder of this work describes, interprets and discusses the findings from this study.

1.2 Developing Measures

The body of literature concerned with the measurement of socio-economic status, class or poverty is substantial. In order to provide a framework for a focused review, it is useful to touch on some of the basic concepts and theory that underpin measurement development in general.

1.2.1 Conceptualisation of the Construct

First, it is necessary to look for a definition of measurement that can guide this work. Zeller and Carmines (1980) suggest from the social science perspective, that measurement can be viewed as the "process of linking abstract concepts to empirical indicants" (1980, p.2). In 1994, they expand on their own definition by drawing attention to the fact that measurement involves both theoretical and empirical considerations (Carmines and Zeller, 1994). Empirically we are interested in the "observable response" which can refer to a mark in a questionnaire, an answer given to an interviewer, or an observed behaviour. Theoretically, the concern is with the "underlying unobservable", and of course not directly measurable concept that is intended to be represented by the response. For the development of indicators of socio-economic status this definition is particularly relevant, as it draws attention to the relationship between the hazy concept of socio-economic status and the measures we select to observe it. Conceptualisation or definition of variables, in general, is often the most challenging step in the development of a measure. It results in a number of essential directing principles for the rest of the process of measurement development. Ghiselli (1964, p 16) explains that this definition of the variable not only provides a description of the nature of the property, it allows differentiation from other properties, and suggests the kinds of operations that should be used to categorise individuals in terms of that property. Ghiselli also points to the value of theory and knowledge as important factors in the definition of variables. As our conceptualisations and theories of certain variables change and develop, so will our definition. Variables of socio-economic status have been defined in a number

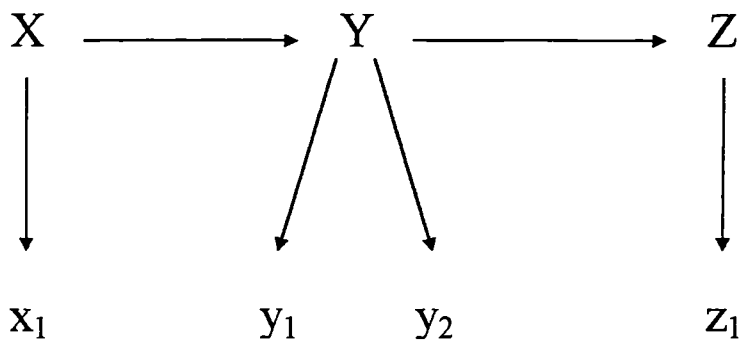
of ways according to different theories of class and social stratification, resulting in a wide variety of measures or indicators.

Blalock (1982) highlights the importance of what he has termed "auxiliary measurement theories" which stipulates this relationship between concept and indicators. He states that:

> "the process of measurement requires a set of theoretical assumptions, many of which must remain untested in any given piece of research, and that therefore the process of theory construction and measurement cannot be seen as distinctly different." (Blalock, 1982, p. 25)

One implication of this is that when a theory is tested, we find that the substantive and auxiliary measurement theories are confounded together. It is difficult to determine whether our findings are a result of the substantive theory, auxiliary measurement theory, or both. Figure 1.1, applying a causal model, demonstrates the relationships between substantive theory, constructs, auxiliary theories and measures.

Figure 1.1 *Representation of an Abstract Hypothesis and Associated Auxiliary Theories*



Source: Sullivan and Feldman (1994 p.63)

In Figure 1.1, the arrows represent the hypothesised direction of causal relationships between variables. X, Y and Z represent the abstract concepts in the

theory which is an explanation about the relationship between X and Z. To test the theory we develop x_1, y_1, y_2, and z_1. The auxiliary theories are those that link X to x_1, Y to y_1 and y_2, and Z to z_1. So the diagram represents the substantive theory (hypothesis) X is related to Z, and via the vertical arrows, epistemic correlations between concept and measure. The auxiliary theory is that there is no other factor affecting the measures or indicators other than the theoretical construct. Therefore if our hypothesis tests negative, then either there is no relationship between X and Z, or, our auxiliary theory is incorrect - our measures are not accurately measuring the construct.

Without a clear definition of what it is we want to measure, and selection of indicants based on theory and knowledge, the selection of indicants will become arbitrary. Comparisons with other studies measuring the same concept, trait, or variable may also become difficult.

Once the concept is defined, indicators that characterise this concept must be selected. This stage involves the confrontation of several measurement issues. That is, the determination of the degree to which the measure or empirical indicators characterises the concept. In terms of the indicators of socio-economic status there are a number of measurement issues involved, but at a general level two important issues are the reliability and validity of the proposed measures.

1.2.2 Reliability

Reliability, according to Nunnally (1967, p.172) "concerns the extent to which measurements are repeatable - by the same individual using different measures of the same attribute or by different persons using the same measure of an attribute". If we administer the same questionnaire measuring socio-economic status to a household several times, we would not expect to find that the results are always exactly the same. Ghiselli (1964) asserts that this can be the result of systematic or unsystematic factors. Systematic factors result in a score or rank that changes in a systematic fashion, increasing or decreasing in some regular fashion. In this case, the score may be consistently the same, although still incorrect. However, it is the observation of random unsystematic variation in scores that has led to the concern

with the issue of reliability in measurement. With poor reliability of a measure we do not have the ability to predict with accuracy the score for an individual unit on a given 'test' from any other administration of the same test, and the measures that are used have little value for describing the characteristic or trait of interest.

These unsystematic factors can be divided into two general categories: varying, and constant unsystematic factors. Description of these two types of factors is useful in a practical sense as they can guide us in the measurement of the extent of reliability, identify the causes of poor reliability and therefore suggest means to improve measurement performance. Varying factors are 'those factors whose effects are different for the same individual on different test occasions and different for different individuals on the same occasions' (Ghiselli, 1964, p. 272). These factors are ascribed to the testing situation itself, and others to the individual or unit being measured. For example, an individual may achieve different scores on the same test on different occasions for a number of reasons. For a household survey measuring socio-economic status, a respondent may be pressured by time to rush through an interview on the first occasion, and then be quite relaxed and able to complete the interview thoughtfully on another. The second scenario, where different individuals achieve different scores on the same occasion may be due to factors affecting the interview situation. For example, an interviewer may be fresh and enthusiastic at the beginning of the day's interviews, and tired and careless by the end of the day. During heavy rain, in a household with a tin roof, the respondent may experience more difficulties hearing the interviewer than a household with a thatched roof. Different individual respondents during a single occasion (survey) may also vary in unsystematic ways, such as one individual feeling particularly motivated to participate, while another unusually disinterested.

Constant unsystematic factors have a different effect on the reliability of a measure. These factors influence all individuals (or study units) in the same way at a given time. In this case, the scores for all individuals will be higher or lower than all individuals on another occasion. They are attributed to biases in raters, or biases in the instruments themselves. For example, an interviewer may have

misread a questionnaire item in the same way and recorded an incorrect score for all respondents during the first testing occasion, and have corrected this by the second. Or, on the first testing occasion all respondents were interviewed in the morning before work when they were fresh, and on the second occasion were all interviewed after a full working day when tired and less inclined to cooperate. Where reliability refers to unsystematic bias, systematic bias is discussed in the following section and is referred to as validity.

1.2.3 Validity

Validity is a second important measurement issue. The validity of a measure is how well it measures the concept that it is intended to measure. Again, Nunnally (1967, p. 75) asserts that "in a very general sense, a measuring instrument is valid if it does what it is intended to do". If a measure is less than valid, then it does not fully represent the concept, or it may represent something aside from the concept of interest. Without the definition of concept, we can go no further in the development of any measure, and the issues of validity are even more important than reliability. A measure cannot be valid if it is not reliable, but reliability alone does not imply validity.

With the issue of validity we do not need to be concerned only with the extent to which the operations measure the trait or concept of interest. It is also useful to consider what traits are being measured by operations already employed (Ghiselli, 1964). Sometimes, on a practical level, the operations that we have selected may provide additional, useful information. Therefore, the development of measures cannot be separated from the development of substantive theories. As we learn more about the constructs that our measures are capturing, we may alter theories about these constructs.

Cronbach (1971, p.447) also makes the distinction that "One validates, not a test, but an interpretation of the data arising from a specified procedure." There have been numerous categorisations of validity, but for the purposes of simplicity, I will discuss the most common as described by Carmines and Zeller (1994). This will be

followed by a brief discussion about recent developments in the conceptualisation of validity itself, and important considerations for the assessment of test validity.

Criterion-related validity (or predictive validity). As defined by Nunnally, criterion-related validity "is at issue when the purpose is to use the instrument to estimate some important form of behaviour that is external to the measuring instrument itself, that latter being referred to as the criterion" (1967, p.87). For example, criterion-validity is measured by showing how well a measure of income predicts utilisation of health services based on affordability. According to Carmines and Zeller (1994), the operational indicator of how well the measure corresponds to the criterion is usually judged by the size of their correlation. What is interesting here is that the selection of measure need not be theoretically linked to the criterion being measured. All that is necessary to establish criterion-related validity is a correlation between the measure and the criterion. This means that criterion-related validity is suitable to be used in an empirically dominated fashion where there is no demand for a theoretical relationship between measure and criterion. Messick (1980) suggests that we consider criterion-relatedness more in terms of the measure's utility rather than its validity.

There are limitations to the use of criterion-related validity, especially in the social sciences. Often attention is not given to the validity of the measure of the criterion itself. Before we can claim the validity of a measure based on assessments of criterion relatedness, there needs to be independent evidence of the validity of the criterion itself as a measure of the underlying construct of interest.

Content Validity "depends on the extent to which an empirical measurement reflects a specific domain of content" (Carmines and Zeller, 1994). For example, some would argue that a measure of well-being that focused on measures of income and assets, but neglected participation in community activities and existence of support networks would be content invalid. To develop content valid measure, there are a number of steps. First, a full specification of the domain of content that is relevant to the measurement situation. In the example of well-being one would need to specify all the different components of the construct well-being.

A review of the literature may reveal that this would involve a great number of components which could include aspects of nutritional intake, shelter, social interaction and support, status, psychological well-being, security, and so on. Then a selection of a sample of these components that would seem to capture the full domain of well-being is required. The sampling procedures must be specified and justified and ideally should be related to theory or knowledge. Finally, it is necessary to be able to construct measures or items that truly reflect the meaning behind these different dimensions of well-being.

There are many limitations to the measurement of content-validity in the social sciences. The measure of socio-economic status is a particularly good example that confronts all of the issues involved. Firstly, there is no agreed upon domain of content for socio-economic status, and what has been described usually lacks the exactness required to assess content validity. Secondly, as it is not feasible to sample content as described above, one usually selects a set of items that are thought to reflect the content of the concept of socio-economic status. At the heart of content validity assessments is the assumption that there is an "acceptance of the universe of content as defining the variable to be measured" (Cronbach and Meehl, 1955, p.282). Therefore content validity is more relevant in terms of educational achievement tests such as arithmetic tests. In this case, one would include a sample of all the operations that the student has been taught. As will be demonstrated in the following sections, for the construct of socio-economic status there is certainly no general acceptance of the universe of content. Finally, there is no agreed criterion for measuring the magnitude of content validity achieved by a particular measure (Carmines and Zeller, 1994). What actually happens is that measures are developed by selecting items that appeal on an intuitive level, rather than following rigorous procedures. Therefore, although it is recognised that a measure should have content validity, due to the lack of ways to properly assess it, content validity is an insufficient measure of validity of social science measures.

Construct Validity is defined by Carmines and Zeller (1994, p.15) as fundamentally concerned with "the extent to which a particular measure relates to

other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured". In the case of socio-economic status, it is construct validity that is the most appropriate aspect of validity where there is no agreed criterion or agreed universe of content. Carmines and Zeller describe the steps involved in construct validation.

> "First, the theoretical relationship between the concepts themselves must be specified. Second, the empirical relationship between the measures of the concepts must be examined. Finally, the empirical evidence must be interpreted in terms of how it clarifies the construct validity of the particular measure" (Carmines and Zeller, 1994, p.15).

From the above description of activities, it is clear that construct validation is theory-laden. However, Cronbach and Meehl (1955) caution against rigid adherence to the use of only well developed or elaborate theoretical frameworks. They assert that construct validation is still involved even when a construct is loosely systematised, used in ramified theory, or even simple propositions.

An important issue with construct validity is that a single study demonstrating a positive association between a measure and a concept is not sufficient on its own to demonstrate construct validity. The more elaborate the theoretical framework, the more challenging is the establishment of construct validity. In order to establish the construct validity of a measure with some confidence it is necessary to establish a pattern of consistent findings with different researchers using different theoretical structures across a range of studies.

Within construct validity, there are two aspects that are commonly described. *Convergent* and *discriminant* validity. With convergent validity there are theoretically relevant empirical consistencies (Messick, 1980). In other words, we would expect that two measures or tests that were developed to measure the same construct would correlate with each other. Discriminant validity demands that two measures of quite different constructs should only relate modestly with one another, or have a low correlation (Spector, 1992). This comes from a fundamental requirement to establish construct validity - nomological validity (Cronbach and

Meehl, 1955). This refers to whether or not our construct of interest behaves in a theoretically predictable manner with other constructs. For example, preliminary steps in establishing construct validity of SES would be to use the new measure to test theoretically accepted hypotheses. For instance, we would expect income and health status to be positively correlated. Alternatively, we can test a plausible theory between two completely unrelated traits or constructs. If the measure produces the expected relationship, then this is one piece of evidence to suggest (not prove) the measure may have construct validity.

Campbell and Fiske (1959) developed a tool in which to explore these two forms of validity simultaneously, called the Multitrait, Multimethod matrix. A detailed description of this tool can be found in Appendix 2.4, Section 2.

Angoff (1988), in a fascinating account of the evolving nature of the concept of validity over this century, describes the current status of validity as a concept. He describes the origins of validity assessment in the purely pragmatic terms of criterion-related and content-related validity. With continuing criticisms of the limitations of these methods, came the vision of construct validity as the very essence of validity itself. Messick (1980) expressed what many authors had come to propose, that "construct validity is indeed the unifying concept of validity that integrates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationship" (1980, p.1015). With this underlying principle that construct validity was an overarching term for validity, Messick goes on to assert that content validity can be seen as "content relevance" and "content coverage"; and criterion validity (concurrent and predictive) becomes "diagnostic utility" and "predictive utility". From this perspective, concurrent validity is used only to justify the substitution of one test with another test of the same construct. Messick calls this "substitutability" (1980, p.1015). In a later paper, Messick (1988) explains that the inference that can be made from a statistically significant criterion-related validity study is:

> "a dependable relationship in the particular setting between the predictor test or tests and the criterion measure. The inference the practitioner wishes

> to make, however, is that the test should be used for selection. To reach this conclusion, additional evidence is required bearing on the content relevance and construct meaning of both the predictor and the criterion and on the potential social consequences of the particular selection use" (1988. p.36).

Because there should be some hypothesis about the nature of the criterion domain that would have influenced the choice to begin with, the results of a criterion validity study can in fact contribute some evidence to the construct validity of the criterion and the predictor (Guion 1976, cited in Messick, 1988)

Therefore, we begin to see that the different strategies to measure validity are far from conceptually separate. Rather, that if we use as many strategies as is practical, we can contribute more to the developing acceptance (or otherwise) of a particular measure. And even more importantly that test or measurement validity is part of a slow and patient process involving a number of studies in a variety of settings. Messick (1988) made a simple but insightful comment when he suggested "just because a process is never-ending does not mean that it should not have a beginning" (1980, p.41).

This short overview of measurement theory has provided a framework for the development of a measure of socio-economic status for the study districts in Tanzania. To begin with, it is necessary to explore the conceptualisation of socio-economic status as it has evolved in industrialised and developing countries; and then, to examine the measures currently used. These are assessed in terms of their feasibility and relevance for partially subsistent economies. In this way, a suitable conceptualisation and measure of SES is selected for use at the district level.

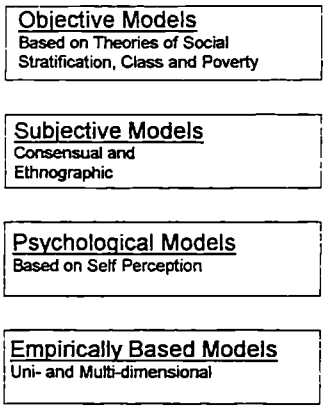
1.3 Conceptualisation of the Construct

1.3.1 Introduction

The conceptualisation of socio-economic status is a complex field. It has always been at the forefront of sociology and political science in attempts to explain the social world and the inequalities that exist within it. It is necessary to develop a

framework to explore the voluminous and diverse literature concerned with conceptualisation of socio-economic status. I have chosen to group the different approaches into four broad categories. Figure 1.2 illustrates these groupings. This section will briefly touch on some of the more important works within each general approach - it does not claim to be a comprehensive review.

Figure 1.2 Approaches to the Conceptualisation of Class



Models of social stratification are both theoretically and empirically based. It is interesting to note that although theory has played a major role in the conceptualisation of class and poverty, many of the most commonly used models are based on empirical evidence alone and not linked to theory in a coherent fashion. Indeed in many cases, there is no theoretical base at all. For this reason, I will begin with the main theoretical foundations for the conceptualisation of class and poverty, and will address other models in later sections.

The intention of this section is simply to flag the main schools of thought that have produced the most common measures of class, socio-economic status, or poverty; and identify the major conceptual issues that have commonly been ignored in current measures. Finally, a review of the most important models will suggest an approach that is most appropriate for developing a measure of socio-economic status for the rural and peri-urban context in Tanzania.

1.3.2 Objective Models Based on Theory

The theoretical bases for conceptualising socio-economic status, class or poverty are numerous and the debate has been ongoing since the beginning of sociological inquiry. There are two main disciplines in sociology that are relevant; those dealing with social stratification and the notion of 'class', and poverty research. There is a large number of relevant theories, some being full explanatory theories and others only small fragments. As an example of this complexity, Øyen (1992), in her work on comparative poverty research, identified sixteen theories relevant to a theory of poverty alone. To give a sense of the diversity and disciplines involved, these theories include: theories of inequality, resource distribution, distributive institutions, stratification, class theory, neo-Marxian theory, theory of marginalization, relative deprivation, access, gender, social change, development, modernisation, economic growth, coping and the theory of poverty culture.

Because there is still no dominant theory, it is important to focus only on those theories and conceptual issues that are most relevant to the current work. This means that the concern is with theories that deal with the general *notion* of poverty or class, rather than theories that explain the *causes* or *consequences* of poverty, or, the nature of the dynamics and relationships between socially stratified groups.

It is 'class' that is perhaps one of most widely used labels for variables of socio-economic status in medical and international health systems research. Therefore this section will begin with the work on social stratification and the notion of class, and then proceed to explore the many conceptual issues in clarifying the notion of poverty. It is well beyond the scope of this work to present a comprehensive view

of these theories. Only a selection of the most commonly accepted theorist and conceptual issues will be presented.

Modern Theories of Social Stratification and Class

In its broadest sense social stratification refers to the meaning systems which attempt to justify and explain "systematic structures of inequality" in a society. (Crompton, 1993). Inequalities have existed in all societies from ancient times to the present and only their form and extent vary. As society changes, different explanations have been offered to account for these differences. This section will begin with a brief look at the modern theorists of the 19th and 20th centuries as it is these theories that have directed the development of most of the current measures of class.

The Conservative or 'Functionalist' Tradition

The substance of the functionalist view is that stratification arises out of the needs of the society as a whole and not from the needs or desires of individuals. This means that all societies need to motivate and position individuals in the social structure so as to serve the society as a whole. Talbot Parsons, and his former student Kingsley Davis have been the major contributors to this view. The famous summation of this approach by Davis was that:

> "Social inequality is thus an unconsciously evolved device by which societies insure that the most important positions are conscientiously filled by the most qualified persons" (Davis, 1949, p.367).

If this is so, then society will find it necessary to motivate the appropriate people to fill certain positions and at the same time ensure that they are motivated to perform the duties attached to them. The positional rank is determined, and rewarded, according to both the level of importance for the society, and, by the level of training or talent required to fill the position - that is, functional importance and scarcity of personnel (Davis and Moore, 1945).

Davis and Moore (1945) in their discussion on wealth, property and labour describes how society uses unequal economic returns as the main controlling

device to ensure the entry into positions and the sustained performance of the tasks required of that position. They believe, therefore, that income is one of the main *indicants* of social status, although are quick to point out that income is not a *cause* of power and prestige. It is the result of the rewards granted to that position due to its social importance, and the scarcity of personnel for the position. They go on to say that the economic source of power and prestige is not primarily income, but the ownership of capital goods which includes patents, good will, and professional reputation. They caution that the ownership of consumer goods should not be confused with the cause of power and prestige, rather that these too are simply indices.

There are three obvious indicators of class or social status that need to be considered from the functionalist perspective. Firstly, occupation itself would appear to be the key indicator of one's position in society if, as the functionalists have stated, individuals are rewarded based on their position and 'contribution' to society. However, it is necessary to address how the functionalists are likely to categorise occupations into broader occupational groups. What are the criteria used to assess the relative ranking of a particular occupation, or occupational grouping? A position is valued according to its perceived importance or essentiality to society; the amount of training required; or the scarcity of the personnel to fill the position. For the latter, educational background becomes an obvious criterion, and it could be used either to categorise occupations, or even as an individual, single indicator.

The Radical or Conflict Theorists

In complete contrast to the functionalists, the radical, conflict theorists see social inequality from the perspective of individuals and subgroups in society seeking to meet their own needs and desires and not those of the society as a whole. This results in a struggle or conflict between the different groups (or classes).

Marx, the father of the conflict theorists, no doubt placed class in a central position in his theories. What is interesting, in the context of this discussion, is that Marx never clearly defined the concept, although throughout his writings there are

indirect definitions and references to the concept. It was in fact in his very last unfinished manuscript that Marx began, but never completed, a definition of class (Marx, 1967).

According to Marxist theory, it is the *relationship* between classes that is of most importance and it is these relationships that cause or create classes. Class relationships are found in the production relationships, or in the patterns of ownership and control of the means of production which characterise these relationships This results in two main classes, the bourgeoisie or the owners and controllers over the material means of production; and the proletariat which only own their labour which must be sold for survival (Crompton, 1994). According to Dahrendorf, a major conflict theorist of the 20th century, Marx believed that there were:

> "The owners of mere labour power, the owners of capital, and the landowners, whose respective sources of income are wage, profit and rent - thus wage labourers, capitalists and landowners constitute the three great classes of modern society based on the capitalist mode of production" (Dahrendorf, 1972).

Marx's description of the design of structured social inequality is seen in his relationships to the means of production, which in Marxist theory is how classes are identified. Dahrendorf (1972) developed Marx's theory further, perhaps to better suit 20th century economies, by drawing attention to the fact that it was *authority* or *control* over the means of production rather than simply ownership of essential property that would determine social class.

Therefore, in regard to the development of measures of class from a Marxist perspective, occupational types are usually grouped into class categories (Scase, 1992). In reality, a review of the literature reveals that Marxist class categories are seldom used to generate occupational groupings or indeed any other indicators of social class (Wright and Perrone, 1977).

The Weberian Tradition

Max Weber deserves special attention in the discussion of social class and its measurement as it has been his theories that have largely driven the use and classifications of occupation as the major single indicator of class used in industrialised countries. This is most likely, in part, due to the ease in which his notion of class can be operationalised, in contrast to a Marxist focus on more abstract human relations.

Weber explains class as a reflection of market-determined life chances (Crompton, 1994). Weber defined class as any group of persons occupying the same class status, and class status is determined by the probability that an individual or group will share similar provision of goods, external conditions of life and subjective satisfactions or frustrations (Weber, 1965). What actually determines these life chances is property, skills and education; therefore, the property and acquisition classes. Weber argued that control over different combinations of consumer goods, means of production, investments, capital funds or marketable items make up class status. These can actually be different with each variation of the combination of these possessions. This was resolved by Weber's identification of a social class which refers to the plurality of class statuses between which individuals can interchange either within their lifetime or across generations. These interchanges are not only possible, but observable. He went on to identify four main social classes: firstly at the bottom of the occupational and ownership ladder, the working class; next the petty bourgeoisie; followed by the technicians, specialists and lower-level managers; and finally the classes with property and education (Crompton, 1994).

For Weber, social position was not only determined by class, which, in his terms represents the economic base and ownership and control over resources. Weber also included status and power in this determination of social position. For Weber, social status is considered the honour or respect in which a person is held; and power, the ability to influence communal action, no matter what the content of that

action is. Many measures of socio-economic status, or occupational hierarchies are based on these dimensions of class, status and power.

Lockwood (1958) has also been influential in terms of how social class is measured. Using a neo-Weberian framework in his attempt to locate the growing numbers of lower-manual and white collar employees in the occupational hierarchy, Lockwood described class situation as being determined by three factors. These are a) market situation which is the economic situation narrowly conceived and refers to the size and source of income, degree of job security and opportunity for upward occupational mobility; b) work situation or the social relationships at work that are determined by position in the division of labour, and c) status situation or the prestige enjoyed in the wider society. Goldthorpe (1980) developed one of the most commonly applied occupational class classifications based on Lockwood's theory. In his famous study of social mobility, Goldthorpe classified occupations according to the way in which they shared market and work situations. Specifically, Goldthorpe classified occupations according to their size and sources of income, other conditions of employment, degree of economic security and chances for economic advancement, and "their location within the systems of authority and control governing the processes of production in which they are engaged" (Goldthorpe, 1980, p.40).

Types of Social Stratification in other Contemporary Societies

Western theories of social stratification have influenced the choice of measures of socio-economic status that are applied in both industrialised and developing countries. However, the following section briefly looks at alternative stratification systems in other contemporary societies, especially in societies that more closely resemble the agriculturally based economies that characterise the districts in this study. Remnants of the stratification systems described here are still seen in many parts of rural Africa.

Tuden and Plotnicov (1970) believe that it is important to avoid identifying a society with a particular type of stratification system. This is less than satisfactory when either social stratification changes within a society, or when there are one or

more types existing together. In respect to this counsel, different types, or systems of stratification are briefly summarised.

Slavery has an ancient history throughout the world. A slave is defined by the Oxford dictionary as "a person who is the legal property of another or others and is bound to absolute obedience" (Oxford, 1993). The main issue, in terms of discussions on social stratification is not the manner in which slaves are "recruited", but rather how they are employed within a society. Only where slaves are clearly defined as property, and the social binds between them and society at large are restricted that a slave group is distinct and therefore can be considered a form of stratification (Tuden and Plotnicov, 1970).

Castes, as a system of stratification, can be as varied in terminology and understanding as can class within a Western society. There are two general approaches to the definition of caste structures. One group maintains that castes are only found in the Pan-Indian civilisations (e.g. Dumont, 1972), while the other group attempts to use structural, cultural and functional similarities of caste for comparative analysis of stratification in several countries (e.g. Tuden and Plotnicov, 1970). Due to the hereditary and fixed nature of the caste system, identification of an individual is easily established (Mayer and Buckley, 1970).

Béteille (1969) summarised further developments in the understanding of caste in rural India. He drew attention to the distinction between traditional castes as defined by the Brahmin Scholars, and the operative units that the Indian people were actually living within. This resulted in an increased awareness of the ambiguities of the traditional caste system, and the prospects for mobility that were, in reality, offered by the system.

There has been an assumption demonstrated in the literature, that class systems are a necessary development with the coming of industrialisation and modern economic development. Tuden and Plotnicov (1970), assert that the application of the concept of class in Africa is not yet useful because social classes, as commonly defined in the Western context, are still in the process of formation. In addition class systems are still being mixed with other more traditional forms of

stratification or residuals from the colonial past. Other important constraints to the application of Western notions of class in Africa, is that certain arrangements of social elements are quite different in Africa. For example, in Western occupational structuring it has been considered safe to assume that the major wage earner has the same social position as his family; that residence, consumption habits and other behaviours are considered indicators of class position; and that to some extent families are hereditary groups with the same status among the close kinsmen. These are not always found in other societies. In parts of Africa families frequently hold a wide range of economic positions; a modern elite man may be married to an illiterate, uneducated wife; and the extended family residing in his house may represent the entire social spectrum. Residential areas are certainly not yet well segregated in terms of wealth, occupation or Western cultural orientations. Tuden and Plotnocov do, however, believe that the modern elite class has already emerged sufficiently to demand consideration.

Marquet, working from a more purist radical perspective believed there were four ways in which African society can be stratified. He described these as according to: a) specialised occupation (such as animal husbandry, agriculture, hunting, priesthood, war and commerce); b) possession of material goods; c) relationship to the means of production; and d) through descent (i.e. nobility or commoner). With the introduction of the colonial era, African traditional stratification systems began to break down.

Marquet (1971) asserts that with the colonisation of Africa, the most basic and general phenomena of differentiation was education. He believed that education produced a new privileged group that had a strong impact on the previous systems of stratification. Within the colonial network, Marquet also drew attention to stratification based upon the colonial networks of inequality - Europeans and the indigenous population. Despite the numerous levels within society based on education, occupation and wealth, it was always race that determined superiority when Africans and Europeans were compared. After the independence movement swept across Africa, and the white administrators left, organisations stayed

basically unchanged with the same structures and benefits. He then describes the modern African classes as: (1) the urban, educated elite; (2) the urban proletariat who were people migrating from the rural area in search of a better lifestyle; and (3) the peasant masses, now even more isolated than during the colonial period.

Lenski (1966) examined the distributive systems in a number of societies not from the perspective of stratification system type but rather from the type of society itself beginning with hunting and gathering societies, and progressing through simple and advanced horticultural societies, agrarian societies and ending with industrial societies. I will focus on his discussion of advanced horticultural societies as these will be the type of traditional systems that were found over much of East Africa in the past. Of course there has been substantial dilution with the political and economic developments since Lenksi wrote. Nonetheless, it is useful to highlight that if current stratification systems are laid upon traditional ones, and these may not share many commonalities with Western stratification systems, then a hybrid system is likely to be the result.

Lenski believed in the synthesis of the conservative and radical traditions and his work included aspects of both. Lenski differentiated between simple and advanced horticultural societies by identifying the development of significant social inequality in the latter. Their fundamental differences lay in respect to technology, especially the use of metal hoes rather than the digging sticks of simple horticultural societies. Also, differentiation is based on their huge range of edible plants such as yams and bananas, obtained from distant countries many centuries ago. This resulted in a better use of soil and allowed crop rotation and higher overall yields. Lenski links these technological developments with the demographic changes of population density, permanence and geographical extension. He believes these changes give rise, in some societies, to empire building which is made possible by the additional resources for military and political machines, and the surplus produced by potential "victims" is large enough to justify the effort.

With these demographic changes, the growth of state, and advances in trade and commerce came occupational specialisation. The larger kingdoms required a more complex governmental system being made up largely of officials and warriors with their associated hierarchies. Tools and weapons were made by craftsmen, and the courts often required other specialists such as eunuchs and a variety of court entertainers.

With trade came the development of markets and the specialisation in the merchandising of goods. It was the lack of transport through the exclusive use of man and canoe as the only means of transportation that limited these kingdoms' ability to maintain control over their more distant areas. It was the chiefs or rulers and their immediate representatives that had greatest power and wealth.

Lenksi argued that the primary determinant of status (which in his view was a combination of power, privilege and prestige) is the relationship one has to the chief and his subordinates. Another important determinant is the forces to which the king is indifferent or opposed. By this Lenski means the groups or members of the society who through either deception or fraud hold a substantial proportion of the economic surplus; or, where the chief leaves a high proportion of the surplus with individuals to encourage sustained effort. Also, those who have special relationships with religion or the supernatural, such as priests, are also accorded special rights and privileges.

Even though the 'chief' no longer holds the power in Tanzanian rural society that he once did, the relationships between the community and those with power may still be expressed in a similar fashion, along traditional lines.

Poverty Research

Introduction

Theories of poverty are largely directed at explanations of the causes, and consequences to the individual and society as a whole. This section will only concern itself with the conceptualisation and definition of poverty which began in earnest as a result of the new welfare states in the latter half of this century. There

was practical concern with the conceptualisation and subsequent measurement of poverty which was driven to a great extent by the UK and Europe. The desire to find a theoretical or scientific basis for definitions of poverty, suitable methods to identify the poor, and ultimately transfer incomes from rich to poor assisted in the rapid developments in poverty research. International measurement of poverty has also become increasingly important in the description of a nation's social conditions, and analysis of changes in poverty as a result of poverty programmes in the developing world as well as industrialised countries.

From a political perspective, poverty can be defined in a number of ways which will not only result in different methods of measurement, but ultimately different approaches to programme interventions. In addition, definitions of poverty can be influenced heavily by the different scientific disciplines employing them such as medicine, sociology, psychology or economics. McLachlan (1986) argued convincingly that it is pointless to altercate over universal definitions of poverty. He asserts that whether or not we refer to poverty in terms of income, physical needs or even socially acceptable lifestyles, clarification of the concept in the context of its application is what is of most importance. Although this argument can be justified in several circumstances, it is also important to consider the needs of comparative research where comparisons over time and location are required to advance our understanding of poverty and to compare policy performance (Knox, 1974).

This section will very briefly point to the different approaches to the measurement of poverty and the associated conceptual issues.

Subsistence and Basic Needs Approaches

Poverty has commonly been seen, perhaps due to its intuitive appeal, in terms of whether or not an individual or household is able to meet the minimal needs to sustain life. Once the minimal needs have been established, and operationalised by attaching a monetary value, then household income (wage earnings or other sources) is compared with the cost of the calculated basic needs package. Subsistence conceptualisations usually refer to a more narrow view of poverty

whereby human needs are defined as predominantly physical needs such as food, clothing and shelter, rather than social needs. The Basic Needs definitions extended subsistence to include facilities and services required within the community at large (e.g. access to health and education). As these approaches are conceptually similar, they will be discussed together.

At the dawn of this century, Rowntree (1908) conducted his study of town life in the city of York in Britain, and was perhaps one of the earliest major studies to determine the extent and conditions of poverty. Rowntree defined poverty in terms of cash earnings required to maintain 'physical efficiency'. Rowntree stated that the cash earnings required was determined by the food, clothing. and shelter that would meet this physical efficiency, and even pronounced that his interest was not in "expenditure needful for the development of the mental, moral and social sides of human nature" (Rowntree, 1908, p.87). Income of families was compared to a standard based on the cash earnings required to meet these basic needs. Rowntree himself attempted to confront several measurement issues relating to the determination of minimal needs. These included how to calculate the cost and components of the minimal diet, and other household needs; in addition to issues relating to measurement of family income.

General concerns about how we determine these basic needs have frequently been voiced, and it seems reasonable to call for agreement for this to be an acceptable approach. Rein (1977) criticised Rowntree's definition of basic needs as ruthless because there was no recognition of the needs of people to be at times frivolous, break monotony or be occasionally irresponsible.

Orshansky (1965) in a review of the poverty profile income measures constructed from the 1964 Current Population Survey by the US Bureau of the Census, identified problems with the basic needs definition. On a general level, she described the value judgements involved in determining what is "enough". When discussing the Department of Agriculture's Standards of Food Adequacy, Orshansky asked how reasonable was the standard given that a homemaker would need to be a good manager, have the time and skill to shop wisely, and prepare

nutritious and tasty foods on a budget of 95 cents per day per person, and only when all family members were eating together.

Sen (1979) argues against Orshansky's view that poverty is a value judgement by asserting that the conventions of a society at a given time are fact and not issues of morality or subjective search. He draws on Marx and Adam Smith's belief that in a given country at a given time it is practically known what constitutes the means of subsistence and goes on to suggest that what is useful is the identification of the society's dominant view.

In the context of international development and poverty alleviation programmes in developing countries, the basic-needs approach was particularly fashionable during the 1970s. The International Labour Organisation (ILO) in their Populist rhetoric suggested that the people themselves in each country should determine the scope and priority of basic needs. What we actually see in practice, however, is that the ILO economists have been very quick to offer their own definitions. For example, material basic needs include certain minimum levels of private consumption of food, clothing and shelter, and access to certain public services such as clean water, sanitation, transport, health and education services, and so on (Sanbrook, 1982). An essential aspect of the basic needs approach is the target setting, which raises many questions such as which needs are to be considered as basic needs, and, as needs must be translated into norms, what kinds of norms should be chosen. India was one of the first countries to make the satisfaction of basic needs a major focus of development strategies, however their definition of basic needs was predominantly focused on the provision of basic facilities and services at the community level (van der Hoeven, 1988).

Deprivation

Peter Townsend, advancing the relative rather than absolute thesis, has contributed a great deal to the development of this concept. He proposed that people were relatively deprived if they could not obtain the conditions of life that allowed individuals to participate in the relationships and customary behaviour which was expected from the society in which they lived (Townsend, 1994).

Townsend's (1979) concern was that poverty or relative deprivation also include what he considered the crucial social aspect of being able to participate in the customs and activities approved by the particular culture. He described the poverty line as the threshold where withdrawal from these social responsibilities escalates disproportionately to falling resources.

Deprivation is measured largely in terms of visible consumption. Deprivation measures have commonly been used to identify areas or political units of deprivation for purposes of improving the equity in resource allocation to certain areas, however these applications are not relevant to the current study. Deprivation indices can equally be applied to measure the deprivation of individual households for a range of other purposes. Deprivation indices often consider both material and social deprivation. Townsend (1987) believes that this is essential in order to be able to measure and understand the paradox frequently seen with the apparent coexistence of prosperity and deprivation. Material deprivation usually contains such types of deprivation as dietary, housing, clothing, home facilities, environment, and employment. Social deprivation measures will estimate rights to employment, family activities, integration into the community, formal participation in social institutions, recreation and education (Townsend, 1994). Townsend believes that an individual household can suffer one or multiple deprivations such as being secure on the material level, while being socially deprived.

Townsend and colleagues have been severely criticised in terms of the selection of their items for inclusion in their deprivation indices. A major challenge was related to Townsend's selection of indicators of poverty such as the consumption of a Sunday roast or taking an annual holiday. These have been considered largely functions of tastes and preferences rather than of poverty (Piachaud, 1987). Townsend defends his original index by suggesting that Piachaud's criticism had no theoretical base. Townsend asserts that the items selected for inclusion in his index were highly correlated with income (i.e. as income decreased, deprivation increased) (Townsend, 1994). But Piachaud argues that this relationship to income

tells us about people's behaviour and social or cultural differences and not necessarily reflecting true deprivation (Piachaud, 1981).

Sen and 'Capabilities' as the Standard of Living

Sen (1983) confronted the absolute-relative debate and called for a more flexible interpretation than either the rigid absolute or relativist view had offered. Sen found the relativist view could result in poor measures of poverty should the overall situation in the standard of living deteriorate. For example, during a recession, the entire population may reduce their absolute living standards significantly; however, their relative position may stay the same. Therefore a relative measure of inequality would not detect this fall in the overall situation of increasing numbers in poverty. Sen also rejected the policy definitions of poverty. By this he refers to the common approach of identifying poverty with, say, the supplementary benefits scale in the UK, which is set according to the average UK income. This can easily result in the perversity of reductions or increases in the numbers in poverty as simply as the government raising or lowering the poverty line. These adjustments in the poverty line may be determined for a variety of reasons, but commonly it is due to the feasibility of provision at certain levels as well as the influence of pressure from important groups, or policy objectives other than poverty reduction like curtailing inequalities.

Sen also rejected the utilitarian approach to valuation of the standard of living based on a person's happiness or perceived satisfaction. Sen believed that happiness or satisfaction of a poor person may be the result of social conditioning, or that a person's choices may be determined by motivations other than improving his own well-being (Frank, 1989). Because of the limitations of subjective utility as an adequate measure of well-being, and on the basis of his argument to reject relativist views, seeing poverty as simply inequality or as a policy definition, Sen proposes an alternative theoretical basis for conceptualising poverty which recognises an irreducible absolutist core.

Using a bicycle as an illustration of a commodity, Sen describes his sequence of transformation. Firstly, the bicycle, or *commodity* has the *characteristic* of

transportation, which gives the person the *capability* to move about in a certain way, which in turn gives the person *utility* such as happiness or satisfaction to move in this way. Sen believes that this capability to function should be the focus for determining the standard of living. Commodity ownership in itself does not result in the capability to function. It "may provide the basis for a contribution to the standard of living, but it is not in itself a constituent part of that standard" (Sen, 1983, p.160). In addition utility should not be seen as the focus as it reflects the use of the commodity or characteristic through mental reaction to it rather than the use itself. Sen shows that in all societies, and at different times, there is an absolute need in terms of capabilities to function; and that these require varying commodities and resources to fulfil them. For example, a carless family in a car dominated society with poor public transport opportunities is absolutely poorer than a person in an alternative poorer society with good public transport. Sen points out that basic capabilities such as meeting nutritional requirements, escaping avoidable disease, to be sheltered etc., require somewhat similar commodities across communities. However, other capabilities such as living without shame, or participating in the activities of the community may require substantially different commodities and resources in different communities, therefore recognising the notion of relativity. In summary, Sen proposes an absolute approach in the space of capabilities, but with a relative approach in the space of commodities, resources and incomes. Sen proposes some kind of efficiency-adjusted income level whereby these income "units" reflect command over capabilities rather than commodities. Unfortunately, it is quite unclear as to the way we identify the basic needs or capabilities in the first place, and then how we operationalise these. Sen's practical suggestions relate only to income and adjusted income "units", which, so far, is insufficient to provide a clear direction for operationalising the new concept of poverty as lack of capability.

1.3.3 Subjective Models

Another important conceptual issue is related to whether or not poverty or deprivation is defined in an objective or subjective fashion. Although objective and

subjective definitions are seen in a variety of ways, I will select the following orientation for the discussion. Objective definitions are those conceived by scientific observers of the social situation. Subjective definitions, on the other hand call for a consideration of the conventions and opinions of the majority of people within a society. This definition should be distinguished from what will later be referred to as psychological concepts, which is where individuals assess their own place in the social structure.

Consensual Approaches

The consensual approach or subjective determination of basic necessities is where public opinion surveys are the basis for defining the basic necessities (see e.g. Mack and Lansley, 1985). Then, empirical methods are used to identify the income at which such standards are achieved (Veit-Wilson, 1987). The term consensual comes from the fact that the definition is determined based on a consensus of the views of society as a whole. This is obviously in line with the relativist approach to definition. These were developed as a result of frustrations with the basic needs as defined by 'experts' on how one ought to be able to live, such as Rowntree, and later Beveridge in the 1940s; with the constant debates and arguments about what basic necessities are; and the role of value judgements.

There are two general forms of the consensual approach. Firstly, those which focus on the general view of the income required to avoid undefined deprivation, or income proxy approach; and secondly, those which study the general view about what necessities are required to avoid deprivation - later identifying the income level required to achieve these necessities. The income proxy approach was favoured for the ease to which it could be applied to social benefit scales as it gave a cash poverty line. However it has been criticised for not demonstrating much about social deprivations and the need for specific government interventions (Veit-Wilson, 1987).

The other approach involves the development of certain indicators of deprivation directly. Mack and Lansley (1985) developed indicators of deprivation, but rather than being interested in social exclusion and the broader definition of deprivation,

they chose to focus on personal spending only. They also did not consider any aspects of public expenditure such as those required for health, education and other social services, and define poverty as "an enforced lack of socially perceived necessities" (1985, p.39) thus avoiding arguments about whether or not exclusions from norms are pathological, or at what level inequalities become acceptable. Mack and Lansley taking a more purist consensual line, profess that although their approach may indeed obscure some possibly relevant indicators not identified through a consensual approach, this is not of serious concern. They believe that the issue is to take the choice of indicators from the hands of the experts and politicians who may attempt to manipulate definitions for their own ends. Others still argue that their approach is majoritarian and not at all consensual which implies that there are no objectors and agreement has been achieved (Viet-Wilson, 1987).

Walker (1987) drew attention to many of the deficiencies and problems with operationalising the consensual approach. A most fundamental question posed by Walker is that in order to interpret meaningfully people's responses to questions of what are the basic necessities, it is necessary to have an awareness of what concept people are trying to express and their way of expressing this concept. Secondly, Walker asks about the ability of people to determine the needs of other people and argues that these will be identified based on underlying values. For example he found in a previous study (Walker, 1981) when civil service manager were asked to determine needs for the unemployed and single parents, they placed these needs as lower than other supplementary benefit claimants. Again, this points to the need to consider and make explicit the values of respondents.

Ethnographic Approaches to Social Stratification

Related to the subjective group of models is the approach taken to consult community members to identify appropriate indicators of socio-economic status. Ethnography, the scientific description of races and cultures of mankind (Oxford, 1993), has played an important role in social stratification historically, and increasingly today. It is important to recognise the role of ethnography in the

development of a deeper understanding of social stratification in different communities. The aim of these types of studies is not simply to identify the number of ranks within the community, attach labels to them and then place each household in a category. The most important contribution ethnography can make is discovering the principles used by the culture of interest in differentiating between persons. Hymes (1964) makes the distinction between "sorting" - which can be formal and cognitively empty - and "assignment of semantic features to the dimensions of the sorting" (1964, p.117). He highlights the importance of the criteria used in the classification, without which we would simply have categorisation of people into pigeon-holes and the sorting procedure itself remains a semantic mystery.

An early example of these types of studies is to be found in Schuler (1940), who studied social and economic status in two Louisiana cotton-growing communities in the USA. He sought to describe the degree of agreement among community members concerning the social status of families in the community, and how this status was related to certain variables. Schuler first gathered data from respondent households relating to land ownership, household equipment, occupation and so on. Then, informal interviews with prominent citizen other than those covered in the original survey were conducted to learn more about opinions and attitudes of the different social groups. Later, he asked a group of nine social class respondents from households surveyed to rank individually the survey households into three social class groups. The analysis was only based on five of the nine raters and measured their level of agreement. This resulted in a social status score based on the aggregated scores of the raters. Schuler then compared these ratings with a number of variables collected on the initial survey in order to identify the characteristics of low, medium and high class families. He found, for example, that for the higher ranked households, there was land ownership (and a long history of land ownership - that is, low mobility), higher educational status of husband, wife and children, received more publications, and higher farm acreage. Number and types of possessions (e.g. livestock) was not related to social class group, whereas housing (number of rooms) and household conveniences such as radios, iceboxes,

sewing and washing machines were. Drinking water or sanitary facilities was not differentiated by social group.

Schuler, in his summary stated that social status can be objectively determined, but that the criterion of social standing in the community is a relative concept and that this can be assessed using quite subjective criterion. The rater may be using educational achievement, sexual morality or ancestor status as the main consideration. This conclusion was not clearly supported with the data. In a discussion of this study, Day (1940) also noted the lack of presentation of the schedules used, and in addition suggested that perhaps five raters was inadequate in a "pioneering study" and that at least 25 judges should have been used, and at least 10 judgements per family. He also was concerned with the elimination of judges that were not in agreement, and rightly, the choice of only higher social group judges for the activity.

A later example of ethnographic studies of social stratification can be found in Silverman's (1966) study of an Italian community. He sought specifically to identify the criteria the community members used to categorise households according to their "prestige" using techniques such as participant-observation, informal questioning, and structured interviews of key informants. The initial prestige scale constructed was a one-dimensional occupational scale, but with further work was adjusted to include other dimensions including nature of the relationship to the land, place of residence, and "well-bred" behaviour. Silverman himself lamented that his original model was too complex to be an approximation of the rules actually followed in the assignment of rank, and too simple to accommodate all the important distinctions that he observed. With this disappointment, Silverman went in to develop a new sorting method. This involved the sorting of cards with the names of community members into piles according to the level of *rispetto*, a term found to be appropriate in an earlier phase of the work. No criteria were suggested to the informants, and at the beginning the informants did not have an explicit criterion in mind. While they were sorting the cards,

criteria were elicited either through their spontaneous discussion, or through probes and questioning.

Using the results of the three key informant interviews, Silverman found a high level of agreement on the relative rank of most persons, however there were discrepancies in the number of groups, and the points at which the groups were cut off from one another. The model that Silverman developed as a result of this activity was used to generate hypotheses that were later supported with the results of the application of traditional ethnographic methods.

1.3.4 Empirically Based Models

Not all models of stratification or poverty are theoretically based, in fact many of the most popular measures are derived empirically. Theoretically based models such as Marx and Weber assume a categorical model where individuals are placed in one of a number of mutually exclusive class categories. On the other hand, empirical models have been based on a dimensional orientation and the research has usually been empirical and quantitative (Ekehammar *et al*, 1987). These can be largely divided into the uni-dimensional models and the multi-dimensional models. Due to their importance in terms of their selection and regular application in the social sciences, it is worth identifying the basis of two examples to give an sense of how empirical models are conceptualised.

Hollingshead's Index of Social Position

Hollingshead's Index of Social Position (1958) is perhaps one of the most popular measures of socio-economic status in the medical and psychiatric fields due to Hollingshead's interest in mental illness. He constructed his index in a rather innovative way. In a cross sectional survey of 552 households in New Haven, USA, the researchers interviewed each household for several hours on a number of social and economic variables of interest. At the same time, the interviewer wrote a detailed description of the family. Later, two sociologists familiar with the community's social structure, judged where the family belonged in the social stratification system of the community. Each sociologist worked separately, using

the detailed data from the structured and semi-structured interviews and interviewer's detailed statements about the family. The two sociologists compared their results and agreed on the placement of 96% of households. After this procedure, the two sociologists discussed the criteria they used to categorise each family. Although there were a number of criteria used, the two sociologists largely relied on a) place of residence, b) the way the family made their living, and c) the family's tastes, its cultural orientation, and the way it spent its leisure time.

From these criteria they developed three main indicators and sub-scales: the residential scale, the occupation scale, and the education scale. For the occupational scale they adapted the classification scheme which was then used by the United States Bureau of the Census. Education was selected based on the assumption that men and women who possess similar education will tend to have similar tastes and attitudes and will demonstrate similar behaviour patterns. Further steps were involved in the weighting of these variables and the construction of the scale, but it is sufficient for our purposes here simply to reflect on the method in which the dimensions of class or socio-economic status were identified. So, once again we see the importance of occupation and education, but with place of residence also being used to categorise households.

Benoit-Smullyan's Multi-dimensional Model

Benoit-Smullyan (1944) proposed a different conceptualisation of social position and took care to distinguish status from other forms of social differentiation which he referred to as situs and locus. Status, according to Benoit-Smullyan, is the relative position an individual takes in the three main hierarchies: the economic hierarchy, the political hierarchy, and the prestige hierarchy. This results in three primary status types which are therefore economic, power and prestige. Although he was working from an American perspective, he believed that these status types are present in most societies. Benoit-Smullyan can be contrasted with the Marxist view that tends to deny that political status can be considered separately from economic status and assumes that differences in power result from the institutional expressions of differences in wealth.

Benoit-Smullyan argues that power and economic status can be distinguished by defining power as "the capacity to make decisions which require other individuals to act in ways in which they would not act in the absence of such decisions" and wealth or income as "the possession of goods and services (or claims to goods and services) which yield satisfaction directly or facilitate the production of more goods and services" (Benoit-Smullyan, 1944, p.156). He goes on to show that there is not always a perfect correlation between power and economic status by citing examples whereby an individual may have a high political status and low economic status (for example in the early Soviet system) or individuals with a high economic position and lower political power (compared to officials or trade union leaders) through a decision not to purchase power either as a result of low political ambitions or "democratic scruples".

Benoit-Smullyan then defines five main criteria for achieving status: an object of admiration; an object of deference; an object of imitation; a source of suggestion; and a centre of attraction. He also distinguishes these criteria of prestige from sources of prestige as they are not characteristics of individuals, such as strength, beauty or skills, but rather refer to the social interactions surrounding people. Therefore it cannot be conceived in behavioural terms in the same way as economic status confers goods and services on an individual by another individual, or whereby political status is measured by his or her ability to make other individuals adapt or conform their behaviour according to command. A typical example of this may be where an academic may command a high prestige status despite the fact that there is not a commensurately high economic or political status for the individual.

This view of three primary status hierarchies can usefully influence the way in which we measure social status. Either we can chose to focus on one or more types of status, or equally, incorporate all three types in our measure. For example, when classifying occupations into "classes" we may use the three status types as the criteria for classification. For example, assess the prestige status of the occupation, the economic status resulting from income and wealth rewards, and the political

power that accompanies the position which could be viewed in terms of the authority that is granted by the occupation.

The Key Indicator Approach

In a deliberation on the multi-dimensionality of deprivation in developing countries, Chambers (1995) lists eight important dimensions that should be considered. In fact, Chambers has abandoned altogether the search for a unique simple or composite index of poverty, and instead recognises the complexity and multi-dimensionality of poverty. In 1995, Chamber's dimensions were:

Poverty as defined as a lack of physical necessities, assets and income. Social inferiority which is socially defined as genetically inferior or disadvantaged including gender, race and ethnic group, class, social group or occupation, or linked with age or intra-family relationships such as daughters-in-law.

Isolation is defined as being peripheral to society or being cut-off. This can be due to geographical location, lacking contacts and information, lack of access to social services and markets, or through a lack of social and economic supports. Physical Weakness is disability, sickness, pain and suffering. In developing countries this has added importance where the body is for many, the major resource, and where households are supporting such people without government assistance.

Vulnerability this refers to exposure and defencelessness. There is the external side of exposure to shocks, stress and risk; and the internal side referring to the lack of means to cope without damaging loss. Poor people often prefer a lower income if this means less vulnerability to loss of assets or debt. Seasonality with tropical seasonality in rural areas, many adverse factors for the poor coincide with the rainy seasons such as food shortages, scarcity of money, indebtedness, sickness and diminished access to services.

Powerlessness; the poor find difficulty in organising and bargaining, they lack influence and are subject to the power of others. *Humiliation;* a lack of self respect and freedom from being dependent on others is, according to Chambers, the most widely ignored dimension of deprivation.

The key indicator approach works under the assumption that poverty is measured over a number of different dimensions and these separate dimensions may vary at different rates and in different directions. Chambers' list of the dimensions of deprivation for the rural areas in developing countries, although not theoretically (or even empirically) based, has contributed to the way we consider deprivation for developing countries and gives us a fresh and new perspective of deprivation as discussed by Townsend and others working in a predominantly urban, industrial setting. In fact, although Chambers was concerned mainly with the rural poor in developing countries, his dimensions could be equally applied in a Western context.

The advantages of this type of approach is that it takes into account the divergent experiences of different poverty groups by monitoring the distributions of certain key indicators across various sub-groups in the population. This, in turn, allows a richer more desegregated picture of poverty. The major disadvantage is that this approach will create additional confusion and greater complexity as indicators will be both quantitative and qualitative. This can easily overload a poverty monitoring system, especially in the settings for which Chambers developed his dimensions. It would also reduce the ability to carry out comparative work (Kabeer, 1989). This approach can not yet be classified as an empirical model. It is necessary to test these dimensions in a range of settings empirically to determine their value for a more widespread usage. However their proposal is a starting point, and may prove to be an appropriate empirical model for the future.

1.3.5 Psychologically Based Models

Another distinct approach to conceptualising SES, are the psychologically based models. These emphasise the individual's perception of his or her own position in the social structure. Centers (1949) was perhaps the first to study the relationship between self-assessment and other objective measures.

Centers distinguishes between stratum, which is concerned with the economic system of a particular time (including occupations, wealth, economic and political power); and class, which he sees as a fundamentally psychological phenomena.

> "Class, as distinguished from stratum, can well be regarded as a *psychological* phenomenon in the fullest sense of the term. That is, a man's class is a part of his ego, *a feeling on his part of belongingness to something*; an *identification* with something larger than himself" (Centers, 1949, p.27).

Centres asked a respondent directly whether he considered himself upper, middle, lower or working class. Critics argued that this approach told more about Centres' class categories and less about what the respondent felt were the appropriate categories (Khal and Davis, 1955). Gross (1953) attempted to rectify this problem by asking respondents an open question regarding their class identification and comparing this with closed questions with predetermined categories. He found that class identification depended on the wording of the question asked and suggested that an open-question approach was more appropriate to allow respondents to express a lack of class identification, and to remove the problems associated with determining the appropriate number of categories for closed-questions.

Jackson and Curtis (1968) also supported this argument. They assert that although self-placement is correlated, and therefore "validated" with objective status measures, and, that by adding self-placement measures we can explain such things as political behaviour, attitudes and orientations, there is still a fundamental problem with this approach. This is also related to the interpretation of the respondent's self-placement. To be able to understand this placement, it is necessary to understand the respondent's idea of what the class structure actually is. Where respondents themselves are unclear, then analysis and interpretation of answers becomes difficult. Jackson and Curtis caution that forced self-placement questions (where respondents are forced to choose from a set number of class groupings) will not indicate how respondents place themselves in the class structure as *they* conceptualise it.

Related to the self-placement approach is the debate concerning the relevance of "feeling poor" to the categorisation of households or individuals as actually being poor. Atkinson (1989) asserts that it is neither a necessary nor sufficient basis for

categorisation and that even with information on a household's self placement, our judgements would still not be modified much having also objective information on "hard cash" or income. In addition, some people may not be aware of the options that are open to others, or may simply reduce their expectations (and therefore sense of poverty) to suit their situation (Roll, 1992).

1.3.6 Other Important Conceptual Issues

In the literature on poverty and stratification, there is considerable debate concerning a large number of conceptual issues. This section aims to provide a broad sweep over some of the more important issues that are most relevant to this study's interest in the measurement of socio-economic status at the district level.

Social Mobility and Life-long versus Life-cycle poverty

Social mobility refers to the movement of individuals up or down the social hierarchies, however they are to be conceptualised. Social mobility is of great interest to social stratification researchers, but for the purposes of health and medical research, the recognition that individuals and families can move up and down the social hierarchies has an impact on our definition of poverty. From the epidemiological perspective, it may be important to consider that the class environment earlier in an adolescent or child's life may have had a greater impact on health than, say, current occupational status of parents (MacIntyre and West, 1991).

When trying to identify the poor for the purposes of targeting health service delivery or granting waivers from fees, there is a potential issue concerning the length of time the individual or household has been in poverty. Poor people's incomes and access to resources tend to fluctuate within the year on a seasonal basis, and from year to year, especially in agricultural economies. Over the life span of people it is normal to expect periods of greater and lesser poverty. In most societies it is expected that young parents with small children and older people are often experiencing more difficulties than couples that are both working or young

working adults living at home. These types of life cycle poverty tend to be less severe because problems can be relieved through minor assistance and borrowing.

Streeten (1990) observed that this type of poverty may be less important in developing countries with extended supporting families, and where poverty is the result of achievement rather than ascribed status. On the other hand it is clear that in developing countries people are more vulnerable to the environment and sudden personal misfortune. Chambers (1989) in a discussion of vulnerability of the poor in developing countries describes the fine line between well being and sudden and persistent poverty, with which accidents and disability are commonly associated.

Life-long poverty is obviously a more severe form and the distinction has clear implications for policy in poverty alleviation programmes, or for targeting purposes. This type of poverty is often seen in households without adult males to assist in income generation or to advocate in tenancy, labour or credit problems (Streeten, 1990). Gaiha and Deolalikar (1993) defined persistent poverty as a lack of assets while temporary poverty may be the result of adverse price movements.

Drèze, Lanjouw and Stern (1992) in a study of economic mobility in India, used the notion of "apparent prosperity" to measure persistent poverty, as they argued that current income was an inadequate measure. Using this approach, households are grouped according to their living standards as determined by the two investigators independently. They used all the information that was available, incomes, and their own personal knowledge of the village that they spent 13 months investigating intensively. The two independent assessments were in fact very similar. They found that the classification of poor and non-poor was quite different depending on whether their own apparent prosperity index was used or current income.

Gaiha and Deolalikar (1993) found that persistent poverty was linked not only to lack of assets, but to certain innate disadvantages which are perhaps less obvious, such as industriousness and managerial ability. Townsend (1987) suggests that this problem can be addressed by making explicit references to duration, or implicitly by the choice of indicators.

Gender bias in the construction of the concept

Male bias in the conceptualisation of social class and poverty has been widespread over the decades. This is probably most commonly seen in the selection of indicators, and more specifically in the choice of male head of household occupation to reflect the well being of the entire household. Although occupation will be discussed in detail later, it is worth noting that often such biases creep into conceptualisation of class and poverty unwittingly. There has been a great deal written about gender bias in the concept of class (see Goldthorpe, 1983 and Stanworth, 1984 for examples of the conventional versus more radical views).

Stanworth (1984) claims that it should be the individual that is the unit of class stratification, and that it is erroneous to derive the class location of women based on the location in the hierarchy of the male family head. Erikson's (1984) study of the social class of men, women and families draws attention to the need to consider the purpose of the study as the determinant of the selection of indicators of class. Where consumption and spending is differentiated among family members, it is the family that is usually the unit of consumption.

In discussing gender bias in the developing world research, Mbughuni (1994) summarises the major criticisms of the traditional economic paradigm and the implications for the conceptualisation of poverty. One major area of concern is with the neo-classical concept of household and related data on household income. Household levels of income do not always make the situation of poverty with women apparent because intra-household dynamics are such that poor women may be even poorer than poor men. On a similar theme, Bruce (1989) reviewed the literature seeking to explore the household economy with a focus on developing countries, and specifically the societal inequalities between men and women. She demonstrated that households were not in fact a unified economy, but rather the base of several competing economies; and the gender ideologies that support the notion that men have the right to have personal spending money, where women's income is for collective purposes.

This issue has relevance when trying to identify the poor for targeting health service delivery or offering waivers for user charges. The general household may have the resources to provide adequate health care, but due to the disempowerment of women and their lack of access to and control of resources, money may not be available for expenditure on health care, a nutritional diet or other consumables required for the prevention of disease and maintenance of good health. Conceptualisation of socio-economic status can well be facilitated by a clarification of the purpose to which the measures are applied, and thought given to the implications of selection of either the conventional or individual approaches; or a household level versus a gender focus on income and consumption measures.

Cross-cultural conceptualisations

Cross cultural perspectives need to be considered when selecting a measure of socio-economic status. Does the measure make sense in the context of this community? What were the community status structures on which it was built and validated? First of all, there is the more obvious variations in cultural perspectives across nations and ethnic groups, but it is well to consider variations between communities within one nation.

In the middle decades of this century, there were several studies carried out that aimed to establish differences between rural and urban centers (e.g. Form and Stone, 1957). In these studies the empirical differences were not considerable, however it could be said that as population increases the average income, education and occupational ranks and rank differences between city centers and the suburban fringe also increases (Schnore and Varley, 1955). Later, Haller *et al* (1972) noted that rankings in remote areas of Brazil differed substantially from the urban setting, while Haller and Lewis (1966) found the same differentiation in Costa Rica and Japan. It is also possible for stratification systems to vary across regions due to differences in growth rate, climate, economy, culture and historical circumstances (Artz *et al*, 1971).

In an earlier study on social stratification and class in communities of various sizes, Laswell (1959) showed that the different criteria used for ranking households

class situation (such as occupation, education, housing etc.) differed in importance according to the size of the community. For example, in the smaller communities, community activities and choice of associates were more important than type of housing and residential address, as in the larger towns. Occupation was considered important in all communities. Laswell also showed that conceptions of class in general were influenced by the size of the community, and therefore a standardised number of classes and classification system may not be appropriate. Despite the age of this work, it still appeals to our common sense, and especially in the context of this study where we are particularly interested in the identification of criteria of socio-economic status across large geographical locations, and within communities with differing demographic compositions.

Moving from intra country cross-cultural differences, the following discussion will relate to the application of concepts of socio-economic status from one type of society to another. This chapter will deal with specific measures and their relevance to partially subsistence economies in more detail later. At this point I will focus on the broader issue of the application of general social stratification schemes using a Western-industrialised conception of class to developing nations with predominantly rural, agrarian households.

Faunce and Smucker (1966) investigated and demonstrated cross cultural variations in community status structure across nations. They carried out their comparative study of status structures in a peasant village in Guatemala, a factory town in Costa Rica and a rural village in Michigan, USA. Of interest to this study was their findings that work-related values (or occupational position) for status assignment was important where there is social structural support for such systems, which is mostly seen in the urbanised, industrial setting. In such communities there is a highly differentiated occupational structure and a system of social relationships that forces people to be constantly in contact with people at different occupational levels. This occurs through the employee-employer relationship as well as outside work like the roles of doctor-patient, voter-politician, banker-customer and so on. This reinforces a status assignment system based on work-

related values. Faunce and Smucker argue that smaller communities in industrial societies, although exposed to the same set of values, are less supportive of a status assignment system based on work-related values. The range of occupational differences is generally much narrower, there are fewer people involved in relationships of unequal occupational status, and of particular importance and interest is the functional diffusion of relationships. This simply means that say, where a doctor treats a patient, he may also be getting good advice on where to go fishing (which may in fact be the introduction of another status-assigning criteria) and so the inequality is diluted somewhat.

In a thoughtful paper on the conceptual problems in studying class in rural India, Béteille (1969) stated that success in understanding the nature of social classes in India would depend on the categories selected to describe them. He asserted that neither the experiences of Western capitalist societies nor the demands of socialist ideology would prove to be adequate. Rather, that Indian society itself would need to be examined to identify the appropriate categories. Béteille argues that the anthropological approach of describing native categories (resulting in a caste classification) is still rather narrow. It does not exhaust the categories which the Indian villager uses in terms of how he thinks and acts. Béteille believes that a class concept is not to be excluded in this context.

Much of the stratification research in agrarian societies has concerned itself with land tenure, but again, this may not be useful across a range of settings. The adoption of land tenure is most probably due to the fact that occupation, the choice of measure in urban prestige hierarchies, is of little practical use in a society where the majority of people are farmers. Some authors have challenged the role of land in determining peasant prestige or socio-economic status, and argue that security, and minimisation of risk, not land is the main issue for people living in poverty (Chambers 1989). Seligson (1977) points out that in some Latin American countries, although land ownership may provide some degree of security, there are other avenues for obtaining security such a unionised casual labour. In the Tanzanian setting, many believe that landlessness is not at issue, rather the means

to work larger acreage through improved technology or the capacity to employ casual labourers (Sarris and van Den Brink, 1993).

Chambers (1989), on the other hand cautions against emphasising any particular dimension determined by outsiders, be it income, consumption, security or vulnerability. He strongly argues the case for using the criteria of the poor themselves to determine well-being or deprivation. This results in conceptualisations and measures of deprivation being determined by the villagers or community members themselves allowing a range of unexpected criteria to emerge. Chambers (1995) continued this argument is relation to the concept of poverty, and called for development professionals to question conventional concepts and realities of poverty, and to adopt a new professionalism where the poor can analyse and express their own reality, and then to place that reality first. In a similar vein, Beck (1989) showed that loss of self-respect was of equal importance, or worse, as hunger to the poor in a West Bengal village. Although Beck discussed this as if it were a new finding of great interest, if we look at a small piece written by Taine (1874) about the workhouses in Manchester in the 1860s we see that self-respect was considered one of the major reasons for people not entering the workhouses. People would prefer their home at any price, and they could not bear to be shut up and subjected to discipline. They preferred to be free and to starve.

There is a large body of literature that concerns itself with the measurement of socio-economic status of rural rather than urban households using other measures such as level of living and standard of living, but this literature has mostly confined itself to rural USA. Some examples of these measures will be detailed in the following section on measurement and validation issues. However, from a broader perspective it can be said that the cross-cultural application of these models is often hampered by the lack of common materials and cultural items that would be considered equal in status and value across communities. It is the very fact that such items are temporally and culturally specific, that construction of such indices requires selection of items that fit the context in which they are

applied (Karim, 1990). In culturally diverse countries such as India or Tanzania, then this becomes an even more complex exercise.

Writing from the perspective of economic anthropology, Firth (1968) was able to capture the very essence of the problem of applying concepts and measures from Western industrialised settings to other cultures and economies. He argues that there are very few principles in economics that are truly general or universal, and although many claim to be general, they have in reality been constructed within an industrial, capitalist system. He describes this system as including machine technology, a monetary medium of exchange, elaborate credit systems, developed private enterprise and an individualistic Western social structure. On the other hand, in many peasant societies we have a limited range of transactions that use money, simple technologies, and enterprise, credit and income generation practices which are in sharp contrast to Western economies. This all results in real dangers in applying common measures available to Western economists. Recognition of these issues is at the heart of this work.

1.4 Measurement or Operational Issues

1.4.1 General Issues

The theoretical and conceptual issues discussed in the previous section have provided a foundation for understanding the derivation of currently accepted indicators of socio-economic status, as well as flagging issues that need to be considered when defining the construct. This section will examine the most commonly used indicators of socio-economic status in more detail, and explore general measurement or operational issues for each one. There will be an emphasis on issues that effect the selection of indicators for partially subsistent economies.

There are fundamental measurement issues related to socio-economic status aside from those related to conceptualisation of the notion of socio-economic status or class. Liberatos *et al* (1988) identify some of these in their important contribution to the measurement of social class in epidemiology. Firstly, it is still not agreed whether or not the three most commonly used indicators (occupation, education

and income) are measuring the same concept, or whether or not these are distinct components measuring multiple aspects of the concept of class. Most agree however that social class does have multiple aspects, and that the three most commonly accepted indicators are interdependent. Secondly, whether measures of class should be seen as discrete or quantitative; how many categories should be used; and where the cut-off points should be for classifying units into a construct category. Finally, what should be the best measure against which any social class scale should be validated. This issue will be discussed in more detail below. Other practical issues are also important such as ease of measurement, ability to measure for every study unit, and low variability over time (Hauser, 1994)

On a practical level, Kroeger (1983) in his landmark article on health surveys in less developed countries, points to the major difficulties encountered operationalising the concept where there is almost a complete absence of existing data. Durkin *et al* (1994) also identified an important issue relating to the measurement of socio-economic status in developing countries for epidemiological studies. Although several studies reviewed were able to show associations between health outcomes and education, land ownership, household income and/or expenditure; others were not able to establish a relationship between health and income and household possessions. Durkin *et al* correctly asks whether this is a true finding, or related to fundamental measurement problems of reliability and/or validity.

The utility of social class or status measures for use in developing countries for health service management and medical research can be assessed using three criteria (Zurayk *et al*, 1987). Firstly, our interest in the feasibility and practicality of the measure. In developing countries there is very little data already in existence and data collection and questioning of respondents poses more challenges than in many Western settings. Secondly, we are interested in how sensitive the measure is to life conditions that are known to affect health status or access to health care. Thirdly, there is now a more intense interest in the ability of the measure to reflect

life conditions for the "family as a co-residential unit of living" (p.174) as it is now accepted that individuals are affected by the family, residential unit.

1.4.2 Single Indicators

Occupation and Occupation-based Measures

Despite the lack of decisive empirical evidence, there has been considerable agreement over the years that occupation-based measures are the single most reliable and valid measure of an individuals economic, power and prestige dimensions. Blau and Duncan (1967) argued for occupation as the most important indicator in industrialised countries because it is the basis of salaries and wages, it grants occupational control and authority, and also differential prestige that is attached to numerous occupations. Morgan (1983) describes the advantages of occupational measures in community medicine as: its ability to distinguish groups of reasonable size, occupational classifications have shown clear differences in rates of morbidity and mortality, and the close association between occupation and many other dimensions of inequality.

Developed Countries

Due to the vast number of occupation-based measures, only a small selection of the most commonly used and accepted measures will be included in this review. Occupation can be measured based on public opinion of their level of esteem (prestige perspective), or by their educational requirements and monetary rewards (socio-economic perspective) (Liberatos *et al*, 1988). These two approaches can be linked to Weber's Status and Class domains respectively.

The first accepted occupation-based measure of social class was the *British Registrar General's Scale*, developed in 1911. It was developed to measure the relationship between class and infant mortality and overall mortality. It classified occupations into five classes based on the degree of skill required to perform the tasks involved, as well as the social position the occupation suggested. These classes were professional, intermediate, skilled (manual and non-manual), partly skilled, and unskilled (Liberatos *et al*, 1988). This measure was reviewed every ten

years due to the changes in the occupational structure in British Society over the decades, and has resulted in difficulties in comparison over time as the lower classes have become smaller, and occupations changed. For example, complete comparability between the 1970 and 1980 classifications has not been achieved due to the fact that some of the 1980 occupational groups contain occupations that were assigned to different classes in 1970; separation of occupational groups has occurred in these 10 years and in some cases, one occupation of the group is assigned higher than another from the same group; and new occupations emerge as a result of the developing technology of the day (Boston, 1980).

Still unresolved questions have been posed in response to the development of this measure. Firstly, Jones and Cameron (1983) questioned how the judgements of the social positions of each occupation were made, asserting that rankings of occupations were made in accordance with prejudices of "narrow-minded" professionals. Even today, the methods used to categorise occupations hierarchically are still far from agreed upon. Bland (1979) identified the properties desirable in such a measure (and thus requiring validation) as: a) the need to ensure homogeneity of categories, and b) that occupations in higher categories actually do experience "higher general standing" than lesser occupations. He raised the issue of validity of the measure when he asked how the Office of Population Census and Surveys ensured that each category is in fact homogenous in terms of its standing within the society. At the time of his writing he found that the measure had never been validated throughout its entire history!

Another popular occupation-based measure is the Edward's Social Economic Grouping of Occupations, developed in the US in 1917 as the US Bureau of the Census classification scheme, and is still the basis of the current scheme commonly used by epidemiologists in the US. Edwards based this classification on his own ideas using the distinction between intellectual and manual workers. In a review of measures of socio-economic status for epidemiology, Mueller and Parcel (1981) cautioned against its use due to the heterogeneity within the occupational categories.

Siegel's Prestige Scale was developed in 1971 using a different approach to classification of occupations. He developed a scale using prestige ratings of census occupational titles by randomly selected respondents in three surveys of the US population. Although in Muller and Parcel's (1981) review the scale was found to be reliable and stable over several decades, the main problem identified was a conceptual issue related to the fact that the criteria for classifying occupations were largely income and education and therefore reflected a socio-economic dimension rather than status or prestige dimensions, as the name implied.

A related measure, Treinman's Standard International Occupational Prestige Scales, was developed in order to facilitate cross national comparative research requirements. In addition to the surveys carried out in the US, Treinman also conducted similar surveys in 55 developed and developing countries. According to Liberatos *et al* (1988) the scale had not been updated since the early 1970s, and, although the scale was developed for international use, there may be shortcomings in socialist and predominantly agricultural societies. If this is the case, then one may well question its application internationally, where many countries are predominantly agricultural. The main issues for prestige scales using these methods is that respondents are classifying occupations using idiosyncratic perceptions and have varying degrees of familiarity with the specific occupations. Catell (1942) also showed that graduate students and labourers, with different relative familiarity with certain aspects of the social world, increased the dispersion of occupations within their particular region. Bowden *et al*, (1934) found that women rated educators more highly than men, who gave higher prestige to businessmen. In addition, it has been shown that respondents are greatly influenced by the sex of the majority of the workers for a particular occupation (Liberatos *et al*, 1988).

In a general discussion of occupation-based measures, Jackson and Curtis (1968) also identify fundamental issues still unresolved in this area. For example, how successfully are occupation-based measures capturing theoretical concepts such as prestige and power? Given that there may be subdimensions such as earning

power, training and skill, functional importance, authority and working conditions, how can these be projected into a single theoretical rank system? They also question the criteria that is used when occupational scales are revised. Other rather crucial issues were identified by Morgan (1983), such as how are we to categorise retired people? We have to decide whether or not we are interested in longest life-time occupation, versus the most recent job. Even between these, current social class and especially socio-economic status can be quite different depending on the situations an individual has faced in the last few years of his or her working life, and socio-economic status can change dramatically within a working life. Morgan also addresses the issue of categorising married women. Previously, women were predominantly situated in the home, but by 1980 56% of married women aged 25-64 in Britain were found to be employed outside the home. This may or may not be important depending on the research question, and certainly for child health, mother's occupation and educational class may be more important that the father's.

Osborne and Morris (1979) draw attention to the problems of households with no economically active head. While some propose that the occupation of the father at the time of birth of the household head may be a suitable proxy, Osborne and Morris argue that this is too distant and may be largely different from the socio-economic status of the household at the time of investigation. In addition, they question the assumption that occupational classifications for men and women have the same socio-economic implications. They caution on studies that remove retired people and women from analyses by asking if women or retired persons are classless or have no relevant contribution to make to topics relating to social class. One final issue was that if occupation alone is used as an index of class, then how is it to explain individual social differences such as culture, norms, mores, economic advantage, power, influence and education. Although this is a strong weakness of occupation, it also a strength in that it is indicative of many social and economic differences.

On a purely practical level, the measurement of occupation can require some level of expertise. For example, researchers must be able to apply a recognised coding scheme to maximise confidence in the measure. There are at least seven questions required to code occupation appropriately which is clearly impractical where occupation or class is not a focal construct of the study. In defence of occupation as a variable of choice refusal rates for occupation questions can be low, and test-retest reliability is often high (Liberatos *et al*, 1988).

Less Developed Countries

For Less Developed Countries (LDCs), occupation is used less frequently than other measures, and there are even fewer studies that have developed measures of occupation empirically. The measures described above are not applied frequently, in LDCs. However, occupation as a variable of class or socio-economic status is used, probably more in transition or newly industrialised countries. In some instances occupation is applied more simply still as manual versus non-manual (Durkin *et al*, 1994). In most instances it is a product of the individual research team's ideas and experiences rather than through scientific development. This review was not able to identify many studies that developed occupational measures empirically, or otherwise, specifically for developing countries.

One example of occupation scales developed specifically for Brazil can be found in Bills *et al*, (1985). They developed an index of socio-economic status based on data the 1973 Brazilian National Household Survey. Discriminate analysis of the education (occupational requirements) and income (occupational rewards) of Brazilian workers was used to develop status scores for 94 occupational categories covering each of the 264 specific occupations identified by the Brazilian National Household Surveys. To provide evidence of concurrent validity, the authors correlated their scale with a number of other variables such as education, income, farm status and class position (self-employed or employing others), income, occupational status of respondent's first job, and respondent's father. They carried this out separately for men and women, and found that the scale behaved as expected given any "reasonable assumptions about Brazil" (p.247).

Coleman (1966) argued effectively against the influential hypothesis that inter-societal similarities exist in occupational prestige hierarchies. He identified problems with research technique in studies that showed high correlations between different countries occupational structure, thus providing us with more reason for concern with occupational indices developed in urban, industrialised settings applied in developing or even transitional economies. These studies commonly ask respondents from each country to rank a limited number of *translatable* occupational titles. Coleman questions whether or not it is possible to yield similarity of occupational prestige structures except for those occupational titles that can be directly expressed in each language. If activities for occupational roles are different between societies then they can only be identified in occupational titles by complex and difficult phrases. This has led research only to compare occupational titles that have the same meaning in each country, thus falsely inflating the similarities. Even with a high correlation between the two, one country may have a more complex structure than the other; or high correlations may be due to the fact that unique occupations in each country have been ignored.

Commonly we see that the usual Western professional categories have been applied world-wide, and concepts such as employment, unemployment, work place and work force are concepts and categories taken from the urban, industrial societies of the West. Since the exploration and examination of the informal sector some writers have proposed livelihood as a better word as this is a larger and more universal word that captures the complex and diverse realities of the poor (Chambers, 1995). Thus, occupation or employment should be seen as one component of livelihood. The reality for the rural poor in many developing countries is that they do not undertake a single income-generating activity or 'job'. Because of the complex range of factors that create vulnerability and risk, rural poor households often diversify and complicate their livelihood strategies. Chambers (1988), drew attention to the problems with formal questionnaire surveys systematically understating the extent to which households have multiple activity strategies and a wide range of income generating activities, and overstating the reliance on agricultural income. The review of the literature suggests the

determination of occupation is of questionable use in partially subsistent economies in rural and peri-urban centers in developing countries. For the purposes of this study, the complexities in identifying and coding occupational categories is also a major disadvantage for typical district level workers.

Education

Developed Countries

Education is a relatively popular measure of socio-economic status, largely because of its ease of measure, but also because it is associated with many lifestyle characteristics, behaviours and attitudes (Morgan, 1983), health related variables and disease, and is more stable over one's lifetime than occupation or income (Liberatos *et al*, 1988). Disadvantages with education as a measure of socio-economic status is that educational status varies according to which age group you are studying. For example, older people can be more successful with less education than younger people today. Also of interest, is that there is increasing homogeneity in the level of educational achievement due to mandatory school leaving ages and easier access to higher education (Morgan, 1983).

There is no need to review all the different educational measures as the most common measure of education (almost exclusively) is number of years of schooling - either grouped into categories or as a continuous variable. However, Faia (1981) objected to the use of number of years schooling. Faia found, using national sample survey data, that actual certificates and qualifications was a better predictor of occupational prestige than years of schooling, especially among Negroes and women in the USA. James *et al* (1987) found that educational status (completed high school, or not) was not associated with job prestige for black men in the USA. This group were less able to translate high school diplomas into skilled or white collar positions. They opted to use a measure of socio-economic status that combined education with occupation to capture race differences.

Less Developed Countries

Education has long been accepted as having a strong relationship with child health in developing countries, particularly maternal education (e.g. Victoria *et al*, 1992) Although probably one of the most feasible indicators from a practical view, education has some limitations in severely underdeveloped rural areas or depressed urban areas in LDCs. This is mostly due to the fact that in such cases, a large proportion of the population have had little or no education at all, and thus the item has poor discriminate ability. For less depressed urban areas and countries in transition this may not be as relevant. Another important point in terms of developing countries and countries in transition, is that education can have high heterogeneity within a single household. This can be seen where males are usually given preference to females for education, and children receive higher educational opportunities than their parents. With this in mind, the educational status of the head of household alone will most probably be a poor measure, and average family educational attainments may be more appropriate (Zurayk *et al*, 1987).

Although it has long been held that education is a prerequisite to occupational success (and presumably economic reward), Clignet (1977) demonstrated that this may not be the case for developing countries, specifically, for Cameroun. Data was used from the 1964 government survey of 35,000 workers and several hundred firms in the modern sector of the economy. The survey was seeking to determine a better fit between enlargement of educational institutions and demand for skilled workers. Clignet found that there were significant differences in the determinants of occupational success among manual and non-manual workers. Although for non-manual workers, there was a stronger relationship between education and occupational achievement, the importance of educational qualifications is not the same for the urban and rural areas; nor, for certain sub-sections of the population showing that education is rewarded more for persons with maximal exposure to social change. In addition, employer expectations concerning training and education of local wage earners varied with the type of enterprise, size, activity, salary policies and characteristics of the employer (Camerounian versus European). Clignet's contribution highlights the need to continually remind

ourselves of the special nature of these economies and societies, and the dangers of applying assumptions based on Western experiences.

It has been widely held, anecdotally, that education is considered important as an investment for rural peasants, and therefore education of children is also seen as a reasonable indicator of status or economic well-being. However, in a study of landless agricultural labourers' asset strategies and investments, Heyer (1989) found that education was not a priority investment for most of the population in her study area in Tamil Nadu, India. Most children did not get beyond Standard Five. Commonly children of agricultural labourers did not succeed when sent to school, or were taken away at an early age to help generate income for the family. In addition, costs were too great and the period over which they had to be borne were long. Most relevant to this high cost was the fact that parents perceived the risk of failure high. Finally, as Clignet (1977) found, employment opportunities to match the qualifications (or investment) were not that high and agricultural labouring was considered a good option given the good agricultural markets in the area. Although education appears to provide a more feasible measure of socio-economic status in partially subsistent economies, Clignet's assertion seems particularly relevant to Tanzania where there is a somewhat uniformly low level of education, and opportunities to reap the benefits of high educational investments are relatively few.

Income and Expenditure or Consumption

Developed Countries

Income has been the single most popular measure in poverty research given the nature of the concept of poverty as discussed earlier. It has also been used extensively as an indicator in composite indices which will be addressed later in this section. Income as a single measure of class or socio-economic status has utility if and when the particular economic dimension of class or status is made explicit, although at the same time income has been shown to be a powerful predictor of prestige (Liberatos *et al*, 1988).

There are a number of disadvantages with income measures. Income can be relatively unstable over time especially where individuals take lower paying jobs with future potential, go back to school, get divorced, or are temporarily unable to work. Equally, sudden peaks can result from occupational promotion or taking on a new, more highly paid occupation (Liberatos *et al*, 1988). If we are interested in the ability of people to access and utilise health services, for example, then we are interested in their economic capacity to achieve this aim. Income has been considered a poor proxy for expenditure, which reflects actual material style of life which would presumably include utilisation of health services. Friedman's Permanent Income Hypothesis (cited in Slesnick, 1993) was the original work that stimulated the income versus consumption debate. This hypothesis was based on years of work demonstrating that:

> "...households in the lower tail of the income distribution are disproportionately represented by those with temporary reductions in income and typically exhibit high ratios of consumption to income in an effort to maintain their standard of living" (Slesnick, 1993 p.2)

Townsend (1977) discusses the phenomenon of expenditure surplus and refers to the consistent excess of consumption over income for all income groups. Explanations for this are that income tends to be underestimated and consumption overestimated. Income underestimation tends to be related to the difficulties in tracing sources of income. Sources are so various it is difficult to question respondents on all possibilities (such as money given to children by relatives and income from goods sold). Definitions of income tend to be narrow and do not include all sources, and income tends to be underestimated by the higher income groups. Excess in consumption is often caused by respondents trying to impress interviewers or research teams with as many positive answers as possible to consumption questions. There tends to be over-reporting of expenditure on necessities and under reporting on luxuries, particularly alcohol and cigarettes. There is also a tendency to inflate expenditures over the year where the last expenditure incurred was higher than for the rest of the year, such as with utility bills. Another important issue identified by Townsend was that the poor are less

able to keep records of expenditure through lower educational levels and pressures in the home and are therefore under-represented in surveys.

Non-response for questions relating to personal or family income are common, and indeed questioning becomes extremely complex when trying to achieve meaningful results. For example, in the Current Population Survey of the US Bureau of the Census, 30 components of annual income are collected, each requiring a separate question (Hauser, 1994). Even where fewer questions are asked, decisions need to be made about what types of income are included. This once again becomes complex as there are a great variety of sources of income aside from wages and salaries.

Less Developed Countries

Income measures of poverty are still extremely important in the development arena in less developed countries. Chambers (1995) summarises the possible reasons for this despite the obvious difficulties in application in many settings. Firstly, economists still dominate the development discourse and hold a major proportion of positions in multi- or bi-lateral aid agencies. Economic measures and concepts of poverty (income and consumption being the two major measures) are still considered the norm, and become accepted as the norm in other disciplines and professions.

Secondly, these measures were developed in the industrial north where poor people are mainly urban, and have tended to rely on cash income. Much of their economic status can be captured with this cash income, or cash-based consumption. However, the use of these concepts and measures in developing countries assumes that the same conditions prevail.

Thirdly, poverty defined as income or consumption can be measured, and allows the use of the well accepted poverty lines. The measures have a high utility for development professionals as they allow comparisons over time and across nations. However, the use of income measures standardises the diverse and excludes the divergent and different.

Finally, some argue that the worse off the poor are, the more they are concerned with income and consumption and with the need to meet basic needs. Once these have been met, then we should develop a broader definition of human need.

Many of the previous issues discussed in the context of industrialised countries are relevant for LDCs, however in most cases the issues are simply more exaggerated. Of particular note is the difficulty in gathering reliable income and expenditure data from rural populations. The World Bank has been particularly noteworthy of its recent attempt to measure poverty in Tanzania (World Bank, 1993b). This survey asked of it's respondents a staggering 41 income source questions and 76 questions relating to annual, monthly and weekly household expenditure. To exacerbate the problem, these questions were only part of an 89 page questionnaire. There is little need to discuss the reliability and validity implications of this approach to questioning for interview respondents with pressures to generate food or income on a day-to-day basis.

Another issue more relevant to LDCs than industrialised countries is the nature and size of the second economy. This may not only affect people's willingness to divulge such information, but in many cases especially in the rural areas people are not able to give income and expenditure information. For example, in Tanzania, the second economy is extremely active in both the legal and illegal sectors. For the legal activities, many sources of income may be difficult to translate into Shilling values such as household informal activities with barter exchange, and taking other jobs outside of normal working hours where payment is concealed. In many circumstances tax avoidance, receiving rent from illegal price rises, and production and distribution of goods forbidden by the government is practised (Maliyamkono and Bagachwa, 1990). How likely are we to get reliable and valid data considering these activities which account for a large proportion of the economic activity in the country?

The issues of consumption versus income measures have also been addressed for less developed countries. Deaton (1982) found for Sri Lanka, like many previous

studies in developed countries, that household rank by income and then by consumption differed considerably.

Despite the obvious problems facing income measurement in LDCs, studies have been conducted using income measures to determine inequalities. As an example, House and Phillips-Howard (1990) used monthly cash income per adult equivalent as a measure of welfare in a rural economy in Sudan. It is interesting to note that they found similar ordering of households using food security, ownership of material possessions and livestock ownership. For calculating the income measure they did not appear to include the cash values of home production, however the precise questioning was not reported. The authors did not report any difficulties in collecting income data.

Fields (1994) made some suggestions for improving income-consumption measures of poverty for developing countries with a particular emphasis on measures that will be useful for comparison over time. Among other issues, he recommends consumption measures to avoid problems with season variations in income and discusses the importance of including other inputs aside from cash income such as food and other goods that are home produced, subsidised services such as free rations of staple foods, or health and housing benefits from employers. He also points to the need to consider family size adjustments, and adjustments for large regional differences in the cost of living which is common place in developing countries. None of these suggestions are particularly new, and what Fields fails to do is address the difficulties in translating these other inputs into cash values.

In summary, income and consumption measures for use by health professionals at the district level is clearly not an acceptable option due to the severe difficulties in collecting *accurate* information, and the skill level required to consider all sources of income and translate these into a manageable list of questions that can be asked of a respondent in a sensible period of time.

1.4.3. Composite Measures

Developed Countries

The literature on the development and use of composite indices is truly vast. In this section, I will touch on a few important indices to give a sense of some of the actual measures and the measurement issues associated with them.

One of the early composite measures of social class or status was the Chapin Living Room Scale (Chapin, 1935) which was a development from Chapin's earlier work in the 1920s. He asserted that:

> "socio-economic status is the position that an individual or a family occupies with reference to the prevailing average standards of cultural possessions, effective income, material possessions, and participation in the group activities of the community" (Chapin, 1935, p.374).

With this definition, Chapin measured socio-economic status using four scales: a) cultural equipment, measured by a weighted list of articles of household equipment; b) effective income using a previously developed scale called the Sydenstricker-King Scale; c) material possessions measured by a weighted list of articles of household equipment; and, d) participation in the community was measured by scores allocated to memberships, groups attended, groups financially supported, committee membership, and officership. Chapin then found that the totals of the weights given to living room equipment correlated highly with the combined weights of the four indices, and so was then taken to be a "fair" index of socio-economic status.

The actual scale developed to measure living room equipment was then developed based on the assumptions that the living room is the most central room for family interaction, it reflects cultural acquisitions, material possessions, and the attitude of visitors can be influenced by the proper display of objects in the living room. Therefore it not only assumed that these articles reflect the family attitudes, but they also influence the attitudes of others towards the family and therefore their social status within the community. The 1931 scale had three groupings of items

which were distinguished in the following ways: a) the background of fixed features, b) the functional items of standard furniture, c) the dynamic factors of furnishings and cultural resources (such as pillows, lamps, books and newspapers), and d) the condition or distinctive personality of the room and its aesthetic atmosphere (such as cleanliness, orderliness and condition of furniture, and the visitor's impression of good taste). A shorter, more concise 1933 version was finally developed.

This scale provides a wonderful example of efforts to measure socio-economic status that are heavily value-laden and culturally specific. There are obvious problems with families who choose to express themselves differently from others in the community. The reason to include this apparently obscure example is that the lively debate the measure stimulated, in respect to issues of its value-ladenness and ethnocentricity, seems not to have been recognised even today in typical measures of socio-economic status developed by outsiders in developing countries.

An example of an early attempt to develop a scale for rural settings, is Sewell's Farm Family Socio-economic Scale. Later, Sewell (1943) adapted and shortened the scale due to the demand for a briefer instrument. The final scale contained 14 of the original 36 items relating to the husband and wife's education; house construction; room per person ratio; various housing amenities such as piped water, lighting facilities, telephone; household equipment such as power washer, refrigerator, radio; reading of daily newspaper; car ownership; and attendance of husband and wife at church. This scale must also be considered location specific and was devised using 1937 data. Belcher (1951) carried out a revision of the scale due to its popularity and lack of revisions since the original scale was developed. After carrying out his evaluation and revision to form a new scale, he found the original scale was still a valid and reliable measure of Oklahoma families in the US.

There are numerous examples of such scales as this, and mostly they are simply variations on the same theme of arbitrary composite definitions of socio-economic status with a focus on material levels of living rather then attached to any

theoretical relationship between power, prestige and life chances (Kaufman *et al*, 1953). Kaufman *et al* identified some major methodological problems with the measurement of social stratification in rural societies using socio-economic scales and prestige ratings. They pointed to the culturally-specific nature of these material levels of living scales and that economic and technological trends render many items in these scales as invalid. In geographical locations where there is a relatively high or low average levels of living, the scales may well lose their discriminatory power at one end of the range. They go on to recommend that scales are developed based on cross-community studies, and include a range of regional and local conditions. Forty years later, we still see the same types of problems in the development of such scales, and in many circumstances the justification for item selection is based only on the researchers' knowledge and perception of the community.

Less Developed Countries

One of the few examples of the application of Western empirical models to rural households in the developing world was carried out by Karim (1990). Using three popular instruments to measure socio-economic status developed in the USA, Karim adapted these and applied them in four Bangladeshi villages. His objective was to determine the utility of the adapted measures for developing countries using the basic principles behind the Western models.

Karim adapted the three instruments developed by Warner, Chapin and Sewell (referred to in the previous section). This resulted in a list of weighted items that covered: type of dwelling; level of living (type of lighting, furniture, cooking and eating utensils, possession of watch or radio, presence of books, contribution to religious obligations); family educational attainment; landholding size; membership in co-operative; self perception of status; occupational rating; educational attainment of head of household; food self-sufficiency. Each item was allocated a score and a total of the scores added to give the final status score. Each item was then correlated with the total score of the remainder of the index, to identify the items most related to the overall score. In order to determine the

contribution each item made in predicting household socio-economic status, step-wise regression analyses were conducted. The final index resulted in eight items. Overall, Karim found that the adapted index, using the principles developed by the earlier models in the USA, was useful in the Bangladeshi context. Karim asserts that using similar indices where items have been selected for the cultural setting can be successfully applied. What needs to be asked, however, is by whom and by what method are these items selected, especially if we want to avoid issues of value-ladeness and ethnocentricity.

1.4.4 Other Measures Used in Less Developed Countries

There are other measures that have realised popularity in less developed countries that will be briefly referred to here. Once again, these have not been well studied and their empirical value as a measure of socio-economic status is not well established.

Land tenure and size of landholdings is frequently seen as a measure of socio-economic status, usually as part of a composite index. However, through the variations in land availability there may be variable applications for the use of land as a suitable indicator. In land-rich countries, people are in poverty through lack of labour power or perhaps injury. In land-scarce countries, there are the same individuals in poverty, but more importantly, there are those who are in poverty through limited access to land (Illife, 1987, cited in Amis and Rakodi, 1994). This may indeed be true in many circumstances, but even within countries there are marked differences. Tanzania, a country where there are regular assertions that land is not in short supply, is also subject to intra-regional variations in landlessness. Although Collier *et al* (1986) showed that only one percent of households surveyed in Tanzania reported no land holdings, Cooksey (1994) points out the geographical variations in land availability. He directs our attention to the regions within the country, such as Kilimanjaro and Morogoro, where there is higher population densities and more highly commercialised areas.

Another important consideration when using land holdings as an indicator, is the productivity of the land. Assuming that size of land holding is intended to reflect

yield or income from that plot, we need to consider the great variations that exist intra-regionally. Again, in Tanzania, this variation is considerable. The highland farming systems of Tanzania, which produce bananas and coffee as well as livestock, are intrinsically more productive than the semi-arid, pastoral, agro-pastoralist or cashew/cassava systems of the coastal regions (FSG/Sokoine University, 1992).

Livestock ownership is another popular measure in less developed countries as this can be a major source of income for large segments of the population. Collier *et al* (1986) argued that lack of livestock was the best single indicator of an absence of marketable assets in Tanzania, and showed that the population could be divided into two sizeable groups based on this ownership. However, as with land tenure and landholding, livestock ownership has marked variation throughout the country based on geographical variation and culturally determined income generating activities.

1.4.5 Validation of Measures

This section briefly reviews the typical approaches to validation of measures of class or socio-economic status. It does not intend to be a comprehensive review, as such would be well beyond the scope of this chapter. Rather it provides examples of the most common methods of validation in order to give a backdrop to the validation techniques adopted in this study, and discussed in the following chapter. Only a small number of validation studies of measures of socio-economic status were identified for less developed countries.

Developed Countries

There are surprisingly few studies that have tested the validity of proposed measures of socio-economic status, that is, if one considers the huge body of literature concerned with measuring SES. One of the main challenges lies in the fact that there is no "gold standard" or widely accepted criterion with which to compare new measures. This is insufficient justification for not attempting to increase the theoretical and empirical base for construct validity. Different general

approaches have been taken to the measurement of validity and will be reviewed in this section.

An example of some early reliability and validation work of a scale of socio-economic status can be found again in Chapin (1935) when he attempted to validate his Living Room Scale, a scale already discussed in the section on composite scales. To measure reliability he conducted a simple test- retest with two independent interviewers on two separate occasions. When the two judges were compared the correlation for the two samples was .90, and when the first and second interviewer's test- retest scores were correlated the result was .98 and .99 respectively. To measure the scale's validity, Chapin compared the average scores obtained from relief cases with those from professional men's homes which were supposed to represent widely different classes, and found a large difference. He carried out the same comparison between the same classes in different locations and found small differences. Using the extreme group approach does not tell us much about the measure's ability to discriminate between other, less extreme socio-economic groups (Striener and Norman, 1986) Chapin compared his scale with other indicators accepted as measures of social status at the time, these being income, occupation and participation in community affairs. He compared the level of agreement between these measures and his scale using simple correlations with data from different households in different locations, and found a high level of association. This is an interesting early example of a simple approach to establishing concurrent construct validity, although it may be criticised on the grounds that there were no tests of significance performed.

Kahl and Davis (1955) used the more sophisticated factor and cluster analysis to compare nineteen indicators or indices of socio-economic status. Their concern was with the number of research-specific indices that were being generated with no indication of how they related to other measures. Their aim was to provide sociological field researchers with the an indication of the relative efficacy of the measures. According to Khal and Davis, factor analysis "partitions a set of inter-correlations into a few basic components and shows the relationship of each

variable to those components" (p.322). The factor analysis revealed that there were two common factors; those which were related to occupation (education, self assessment, and interviewer's impressionistic assessment); and the second composed of ecological measures (house and residential area) plus those of the status of the respondents parents and wife. It is worth mentioning here the limitations of factor analysis where several basic assumptions such as the *postulate of causation* (that observed variables are linear combinations of some underlying causal variable) must be considered. The researcher is obliged to support this with some other substantive knowledge of the data (Kim and Mueller, 1978). Factor analyses must, therefore, be very carefully interpreted, and requires a sophisticated level of skills to apply appropriately.

Haer (1957) sought to compare five conventional indices or indicators of socio-economic status by examining their ability to predict 22 variables known to be related to socio-economic status. Haer used the "coefficient of relative predictability" to determine how much more effectively a given variable could be predicted with the use of a stratification index, than on the basis of that variable alone. For example, if we take the variable "has travelled outside the US" in Haer's study, 138 respondents had travelled, and 182 had not. Therefore the modal category prediction is that respondents had not travelled, resulting in 138 errors (total error of 43%). When travel was predicted on the basis of 23 educational groups, error in prediction was reduced to 96. The percentage improvement using education is then computed as:

$$\frac{\text{Total error - error remaining}}{\text{Total error}} = \frac{138-96}{138} = .30$$

The coefficient only suggests the relative worth of the variable or indices. Because of the mixed nature of the indices, other statistical measures would not conform to the assumptions (normal distribution and an underlying quantitative continuum). Haer found that the composite index, the Index of Status Characteristics (ISC) (weighted score for occupation, dwelling area, house type and source of income) provided the best predictive utility, followed by education. Occupation, Centre's

class identification question (self assessment into upper, middle, lower, working class) and an open-ended class self assessment question (subjective) showed low predictive utility. Haer explained the predictive utility of the Index of Social Characteristics was due to the variety of measures employed capturing the multi-dimensionality of construct, and, because the ISC Score through weighting and scale scores provided a continuous series of ranks thus highlighting variations within categories.

Abramson *et al*, (1982) compared nine indicators of social class in Jerusalem, of which four were based on occupation. They determined their value from correlations among the indicators, and between them and 12 selected health-relevant variables. They found little difference between the occupational measures (British Registrar General's (BRG); two Israeli prestige scales; and an adapted Duncan Socio-economic Index) as they were all strongly correlated with one another. When correlated with health related variables, a similar pattern emerged with the exception that the BRG classification correlated poorly with serum cholesterol, possibly due to the reduced number of categories of the BRG. The non-occupational measures (education, authority rating, family income, room-person ratio, and housing amenities score) were moderately correlated with the occupational measures and with one another. When correlated with health variables correlations were also not strong, though some discrepancies were seen such as between education and smoking, and, housing amenity score and consumption of meats, poultry and fish. With some exceptions they found, broadly speaking, that there were similar patterns of association between the health variables and occupation, education and income. Abramson *et al* (1982) suggest that although conclusions about health and class are not insensitive to the measure used, similar results are achieved with these three class variables. They also recommend the use of composite indices to capture the relationship between health variables and class rather than a single measure, due to the lower correlations between non-occupational measures.

An unusual example of the validation of a very well accepted and used scale is the Hollingshead Index of Social Position (Hollingshead and Redlich, 1958). This scale was construct validated by comparing respondents' use of the mass media (such as television, radio, magazines and newspapers) with their allocated social class position. The authors of the research judged that the associations corresponded with their own pre-stated expectations of the association. That is, for example, that higher social classes are more likely to read the *New York Times* and listen to "serious" music than lower social classes. They did not provide evidence to support this theory. In addition, the scale was developed for a specific New England community where Hollingshead only identified 300 occupations from the 450 on the US Census. The measurement and coding reliability, as well as the quality of weighting schemes for larger, well-defined populations is unknown. This type of validation is rather weak, but the reason for looking at it is to see that measures can become very well accepted, while claims of validity are based on very questionable foundations.

Less Developed Countries

In a recent study attempting to validate various measures of socio-economic status in Bangladesh and Pakistan, Durkin *et al* (1994) employed a number of approaches. Firstly, reliability was tested using classic test-retest data (2 surveys at a 2 week interval) and factor analyses; and, again factor analyses to confirm that their questionnaire measured multiple dimensions of socio-economic status. For each set of items appearing to measure a single dimension, internal consistency reliability was measured using Cronbach's *Alpha*. Logistic regression was used for the bivariate and multivariate analyses of the indicators of low socio-economic status and child death. Potential indicators of socio-economic status were taken largely from the World Fertility Survey and were:

- mother's education (years of schooling)
- mother's ability to read a newspaper
- mother's occupation and employment status
- head of household's occupation and employment status
- household possessions (radio, TV, bicycle, motor cycle, car, cow)
- land ownership

- housing conditions (electricity, ratio of rooms/person, floor material, toilet and water supply, urban versus rural, tenure
- father's education

One interesting finding from this study was that no single indicator was associated with child death in all four of the sub-populations. In addition there were strong interactions between several variables and urban/rural residence suggesting that variables have different meanings in different locations. Electricity had opposite associations in rural and urban areas, as did tenure. Tenure was explained by the ownership of shanty town housing in urban areas. Maternal education was associated with child death in urban and rural Bangladesh, and urban but not rural Pakistan. The authors proposed that the lack of association in rural Pakistan may have been due to the small sample and lack of variability as most mothers were not educated at all. Durkin *et al*'s findings suggest that the construct of socio-economic status can vary from community to community and they recommend the use of multiple indicators for successful control of socio-economic status. From their study they suggest useful indicators for the study populations to be occupation (manual/non-manual), education (number of years schooling of mother and/or father), wealth (land ownership and household possessions), and floor material (earth versus other).

In a study focusing on the discriminatory power of two education-based measures of class with health-related variables in Jordan and Lebanon, Zurayk *et al* (1987) found that average educational attainment was a better measure of social status than educational status of the head of household. They applied regression analysis (in the form of multiple classification analysis) to measure the association between the two measures and family health-related variables. The models using the average family educational score (AFES) explained a larger percentage of the variation in 7 of 12 analyses, however the authors noted that the differences were small. The AFES reached significance in seven of the analyses, but the educational score for head of household reached significance for eight. For the four of the seven analyses where both measures reached significance, the AFES showed better

discrimination in respect to "the distance between the adjustments to the average rate of the family health variable induced by the categories of the measure" (p.179). This led the authors to conclude that the AFES was a more sensitive and feasible measure, although not consistently superior to the educational score of the head of household.

What is important to note in this section are the approaches taken to measure validity, and the skills and experience required to conduct such studies. Not only is this study concerned with the construction and validation of an index of socio-economic status, but it is also interested in demonstrating validation techniques that are appropriate for health or social sector workers at the district or regional level. There are very few health professionals that are interested in developing measures of socio-economic status for their locations, that have the resources, training, experience and exposure to carry out meaningful validation on their proposed measures. This study aims to apply validation techniques to its own proposed measures that can be replicated by non-specialists in test validity - quickly and with a satisfactory degree of confidence. It was considered insufficient (or perhaps unethical) to propose a new method to develop a measure of socio-economic status without providing the necessary tools to conduct preliminary validation checks on these new measures. Two alternative validation methods are described and applied in this study which may meet the needs of the peripheral levels in consideration of their financial and human resource limitations.

1.5 Selecting the Approach to Developing a Measure of Socio-economic Status in Tanzania

Thus far, I have examined the major approaches to conceptualising socio-economic status, reviewed the most commonly used measures emerging from these conceptual models, and discussed practical issues related to these measures for both industrialised and developing countries. This section will briefly discuss the approaches (based on the four models presented) and select a model that would seem most appropriate for typical non-specialist researchers and/or social sector

workers at the peripheral level in developing countries. This will then direct the focus for the remainder of this chapter.

Theoretical models to conceptualise and measure socio-economic status have been extremely popular in the industrialised world. Indeed, as discussed earlier, linking the selection of measures to theory and knowledge is sound practice. However, the literature indicates that caution is required when applying these measures in developing countries. Few of the main theories of social stratification have been tested in partially subsistent economies, and their performance is unclear. African societies, Tanzania providing a good example, are in a process of change and multiple stratification systems may be operating where traditional, colonial and modern systems are mixed together. Furthermore the social elements are known to be different in developing countries, with residential areas and even households containing a wide spectrum of social groups within them. What do we actually know about how to deal with these complexities, and is it reasonable to expect those interested in simple, operational research to contemplate them? Basic Needs is extremely difficult to define. It is beyond the capacity of district management teams and non-governmental organisations (NGOs) to spent time determining the scope and priorities of basic needs in their communities. Deprivation approaches are largely concerned with visible consumption, which is again too difficult to determine when socio-economic status is not the main variable of interest to health workers.

On a more practical level, the measures emerging from these models were examined in some detail. It was noted that there were a number of difficulties in identifying categories and coding systems for occupation in alternative settings; that education may have limited value in subsistent economies; and accurately measuring income is a highly complex task. Solving these problems and meeting the challenges posed is clearly beyond the capacity, and indeed interest, of district level health workers.

Empirical models, although quite common, result in similar measures to those described above. In addition, the composite measures have not been validated well

in agricultural communities and their value cross-culturally is not known. What we do know is that great caution is required to adapt these to distant settings, and the skills required to develop such tools are not likely to be found at the district level.

Psychological models, where individuals are asked to place themselves within the social hierarchy, are obviously of little value for either health service monitoring, or granting of exemptions from user chargers. In developing countries where research is often perceived to be linked to some benefit, false reporting is even more likely. In addition there are many practical problems relating to individuals' conceptualisation of the social hierarchy, and their place within it. It also brings up the complications involved in deciding whether 'feeling' poor or 'being' poor is of interest.

For the Subjective models, consensual approaches to determining basic needs are also beyond the reach of the typical district health manager. This requires large inputs to ensure that suitable numbers of people are interviewed to reach a decision by the 'majority' on what basic needs are important across the entire community. It also assumes that there is an awareness of what concept the people are trying to express when interviewed. Although this is appealing, in that it addresses some of the problems associated with imposed models based on Western stratification systems, it is less feasible.

The ethnographic approaches may offer a suitable model for district workers to follow. It addresses the issues of culturally specific criteria being used to determine socio-economic status. It also takes up the challenge posed by many in the development community who call for measures that reflect the experience and lives of the people themselves. It avoids the application of ethnocentric conceptions and measures imposed by outsiders.

Other of the conceptual issues are addressed through enthographic approaches. It is the community conceptualisation of socio-economic status that is of interest, and it may capture the differences in conceptualisation between different cultural groups. This simplifies the task for the district health managers. It is no longer his or her role to conceptualise socio-economic status, nor develop operational measures that

are linked to that conception based on theory or other empirical knowledge that is not readily available.

With these factors in mind, a specific method was required to generate an index or measure of socio-economic status from the perspective of the community members. In order to identify a suitable tool, I turned to the field of participatory research where a specific technique has already been developed for this purpose.

1.6 The Wealth Ranking Technique

1.6.1 Participatory Research and the Perspective of the Poor

Robert Chambers (1995) discusses in detail the importance of incorporating the perceptions and criteria of well-being from the poor themselves. He asserted that professionals and the poor experience and construct very different realities. He summarised some of the more common contrasting realities which will clearly affect the concept of poverty for the two groups. For the professionals' realities, these were driven by tendencies towards universalism, simplicity, reductionism, standardisation, the physical (in contrast to the experientiality of the poor), the quantifiable, income, employment or other single dimensionality. For the poor, their realities are constructed from having a local and specific perspective, a sense of complexity, holism, and diversity; they are not concerned with quantification, and see deprivation as a multi-dimensional phenomena. Chambers follows with his argument that concepts that are universal, standardised, measurable, generated by and designed for conditions in the urban industrial north cannot be universally applied in the more rural and agricultural south, and questions whether they fit or distort the diverse and complex realities of most of the poor.

However, 'the poor' are not an amorphous and undifferentiated mass. The outsiders concept of poverty may require modification to incorporate local concepts and priorities. Poor people have many criteria of well-being, and it has been outsiders that have tried to simplify them down to one or two dimensions or criteria. There is a common stereotype that the lives of the poor are 'simple and

uniform' but they may actually vary by region, community, social group, household, gender, age, season and time in history (Chambers, 1989).

Participatory research, involving community members in the research process, was developed partly in response to these types of issues. Frustrations faced with conventional research failing to incorporate local priorities, processes and perspectives, resulted in a shift in the research paradigm for many researchers. In fact, in the philosophy of science a more general level, in recent decades there has been a dramatic shift in emphasis away from positivist towards more interpretive paradigms, not just within the participatory research movement (for a more detailed discussion in this point, see section 4.7.)

The recognition of the value of 'local voices' or indigenous knowledge began to gain momentum in the 1970s. One of the earlier works that drew together research that examined the capacities, skills and rationales of peasant farmers and pastoralists was Howes and Chambers (1979). In a discussion on the elicitation of indigenous technical knowledge, Howes and Chambers point to the problems of questionnaires imposing outsider categories. For example, when asking for yield per acre, we might be ignoring the fact that a farmer may consider yield in terms of units of labour. Later, other works were published that called for the involvement of local people and the inclusion of their knowledge systems into the process of technical development. Cernea (1985), writing for the World Bank, highlighted the need to 'put people first' in the design and implementation of development projects in order to give them more opportunities to participate effectively in development opportunities and control the activities that effected their lives. In a strong critique against development studies in the 1980s, Edwards (1987) identified the complete divorce between research output and the subject matter of that research, the poor themselves. These calls are embedded in a desire for political action and social change and are the foundation of participatory research. Well into the 1990s we find those (Chambers, 1997) who are still frustrated with the lack of movement towards ensuring the realities of the poor or oppressed groups are incorporated into development research and intervention projects.

Participatory research is seen as one of the major mechanisms by which local voices are heard, but it is not the characteristic by which research is judged participatory.

According to Guijit and Cornwell (1995, p.4) participatory research (e.g. Participatory Rural Appraisal – (PRA)) is intended to be a "flexible and adaptive approach to learning and action". Participatory research has also been described as integrating three important elements: research, education and action (Green *et al.*, 1995). Participatory research should be considered an attitude rather than a method, or group of methods. According to Green *et al.* (1995) participatory research must involve those who are affected by the issue being studied *for purposes of education and taking action or effecting social change* (my italics). Chambers (1992, p.1) writes that the participatory mode is underpinned by an attitude or behavioural principle rather than an epistemic principle. It is based upon 'a family of approaches and methods to enable rural people to share, enhance, and analyse their knowledge of life and conditions, to plan, and to act'. The purpose of participation is to stimulate awareness, and the role of the researcher or facilitator is one of catalyst. In this sense, participation should not be considered co-operation with data collection. The behavioural principles at the basis of participatory research are: facilitating – investigation, analysis, presentation and learning by the people themselves; self-critical awareness by facilitators; and sharing of information between all the actors or stakeholders in the investigation (Chambers, 1992).

The task that participatory research charges itself with is addressing power and powerlessness, and it aims to confront the established and power-holding elements of society. It appears that issues of epistemology and method are secondary to this concern (Reason, 1994). Reason (1994) summarises the Participatory Action Research (PAR) literature and identifies the imbalance between discourse on ideology with discussions on what actually takes place. He found that PAR uses diverse methods, both qualitative and quantitative. It is the purpose, aim or

behavioural principles that underpin the approach that makes it participatory, rather than the method or technique employed.

Participation in Poverty Assessments has become a popular approach in some of the major organisations such as the World Bank. Poverty Assessments are now an essential component in the Bank's country economic and sector work, contributing to the wider process of poverty-related analysis and the formulation of all aspects of country strategy. The rationale for this move to participatory poverty assessments is that involving a range of stakeholders, including the poor themselves, can help the following:

- Improve understanding of the cultural, social, economic, and political dynamics that perpetuate poverty in a given country
- Ensure that strategies identified for poverty reduction reflect the real concerns voiced by the poor
- Promote ownership of the proposed solutions by a variety of stakeholders
- Build in-country institutional ability for ongoing analysis of poverty and the design of measures to reduce it (World Bank, 1996).

Participatory Poverty Assessments (PPAs) (as distinct from participation in poverty assessments) have also become popular with The Bank. These refer to the use of specific qualitative research techniques to discern the perceptions and attitudes of the poor themselves. To date, PPAs have been designed specifically to enrich the poverty profile by illustrating local experience and understanding of poverty and vulnerability. Like other institutions, the World Bank has recognised that the quality and comparability of income/consumption data is of concern, and that more needs to be done to develop systematic and reliable poverty monitoring systems in all developing countries (World Bank, 1995).

One of the tools used by the participatory research movement is wealth ranking. This tool focuses on the wealth and poverty (or well-being) of community members, usually based on a community conceptualisation reflecting their own realities and experiences.

The purpose of this research project was to identify indigenous classification systems of wealth, in order to develop indicators of socio-economic status for use in surveys. I have taken a *method* from the PRA tool kit, rather than adopting the *attitude* which is the hallmark of participatory research. However, wealth ranking has been used in a variety of settings with a range of purposes and under differing research paradigms (see the following section). I have chosen to use this technique - without adopting the participatory attitude - for several reasons. Firstly, there was no intention to apply the wealth ranking technique in order to develop awareness of the causes and dynamics of poverty within the community, nor develop an agenda for action. This would have been an unethical approach given that the project was not involved with implementation of any further activities in these villages in the future. Secondly, it is possible for typical district level health or social workers interested in investigating their communities to manage the technique. The technique seems well suited to allowing the community members to express and describe a construct of socio-economic status that would be based on their own experiences and reality. In this way, we are listening to local voices, without an agenda for social change or action. Finally, discovering the value or performance of techniques applied within different research paradigms is beneficial.

Another interpretation of Participatory Research could refer to the Tanzanian researchers participating in the conduct of a study with an external institution. This is an interpretation often used by the World Bank (no date). In this case, the local institution would participate in the identification of research problem, design, data collection and analysis. The project design was also inconsistent with this form of participation, as all researchers were from outside the communities in which we worked, both in terms of place of origin, and exposure to higher education and other styles of living over a long period. All sessions were facilitated by a social scientist who was not from any of the districts in which we worked. In addition she had been educated in European and North American universities and had been living in the commercial capital for many years. The research assistant, had a limited connection with the communities in which we worked and their way of life

in real terms, although, of course, there was an understanding, in a broad sense, from occasional visits to the larger villages in an official capacity. None of the local district health management team members that accompanied us to each village were involved with research design, or data analysis. Their role was largely one of introduction and organisation of sessions. Wealth Ranking as applied in this time-limited study was intended to be extractive rather than participatory.

1.6.2 Applications of Wealth Ranking

Grandin (1988) was the first to adapt the early ranking methods and publish a practical manual for specific application in small holder communities in developing countries. She has called this wealth ranking, which is the term generally applied to the ranking of individual households into economic, or wealth groups by community respondents. Wealth ranking has been used for a number of purposes: to identify poor households in order to target development project interventions in small populations (e.g. Pretty, 1992); evaluate the effectiveness of programmes reaching the poor (e.g. Sarch, 1992); stratify households for other research purposes (e.g. Mearns *et al.*, 1992; Chadwick and Seeley, 1996; Scoones, 1991); studying community perceptions of rural poverty (e.g. Mukherjee, 1992); understand the dynamics of rural differentiation to suggest poverty alleviation programme interventions (e.g. Turk, 1995); or identify criteria suitable to measure socio-economic status through surveys (e.g. Seeley *et al.* 1995). Despite the popularity of the wealth ranking technique, research into, or formal evaluation of the utility of the method is still in its infancy. The remainder of this section highlights a number of wealth ranking studies selected to reflect a range of applications and differing methodological issues.

Jodha (1988) used an adaption of the wealth ranking technique to study the changes in economic status of households over a 20 year period. He asked villagers in Rajasthan to cite their own categories and criteria for changing economic status. Table 1.1 lists Jodha's 12 important indicators as:

Table 1.1 Indicators of well-being/ill-being in two Rajasthan villages.

1. One or more members of the household working as attached or semi-attached labourers
2. Residing on patron's land or yard
3. Taking seed loans from patrons
4. Taking loans from others besides patron
5. Marketing farm produce only through patrons
6. Members of the household seasonally out-migrating for job
7. Selling over 80% of their marketing produce during the post harvest period
8. Making cash purchases during slack-season festivals etc.
9. Household adult members skipping third meal in the day during scarcity period
10. Women and children wear shoes regularly
11. House with only impermanent traditional structure
12. House has separate provision of stay for humans and animals
Source: Jodha, 1988.

Groverman (1990) applied the wealth ranking technique in Swaziland to identify the poorest households (potential beneficiaries) in order to approach them about project activities and encourage involvement in a participatory development project. They tried the wealth ranking method because of its rapid approach, it did not require a high skill level, and was not threatening in consideration of the sensitive nature of the topic of poverty. Groverman carried out wealth ranking in four communities, using three informants or rankers in each community. Informants were selected based on age. They were asked to place households in piles according to wealth, the number of piles being determined by the informant. During this procedure they were asked to give criteria used to rank the household, which were recorded by the research team. This reflects a typical application of the wealth ranking activity following Grandin's (1988) recommendations.

Using Grandin's (1988) analysis techniques, Groverman calculated an average score of the three informants. This is achieved by ranking each household from one to, say, one hundred and then allocating the "score" based on the rank position. This score is then averaged over the three informants used, and a final list constructed with the average rank position. Then, the lists were divided into a number of wealth strata, not exceeding the original number of piles as determined by the informants. According to Grandin, this is achieved through a search for

natural breaks in the scores. It is interesting to note that Groverman found no natural breaks in the scores demonstrating clear strata, rather a continuum. Groverman was able to construct a list of criteria the informants gave during the ranking, but the author did not report on the specific method used to collect the data, the importance given to different indicators, or how the data was collated and analysed. Groverman did state that all the informants had mentioned the same type of criteria. Finally, Groverman recommended that a thorough training in methods was required to carry out wealth ranking successfully, in addition to allowing more time to ensure adequate preparation of the community.

In The Gambia, Sarch (1992) used wealth ranking as part of a larger rapid rural appraisal activity to evaluate an agricultural programme. The wealth ranking activity was specifically used to identify which farmers had participated in the programme, and also, to explore what differentiated farmer households. The team worked in two different villages, one with 101 compounds, and the other with only seventeen. In the first village, Sarch used two very large respondent groups in each site selected on the basis of sex. The respondent groups were extremely large with between 20 and 35 participants. After a number of preliminary sorting procedures, Sarch ensured that each household was left with a ranking score from each of the two respondent groups, then an average taken for the final score. In the second village, five individual informants were used. Households were ranked in ascending order and an average score taken for the five informants.

Sarch discussed a few determinants of well-being that resulted from the study such as food availability, and animal labour in one village, and high numbers of workers per household, and assistance from others outside the village in the other. Unfortunately the author did not explain how this information was collected, processed and analysed. Sarch closes with a few simple comments on the technique related to groups versus individual respondents, and attempted to account for the diversity in results such as the different climatic conditions and the differences in the size of the study units in each case.

In response to the difficulties of using official poverty measures to identify the poor, Mukherjee (1992) applied an adapted wealth ranking technique to stratify a community in India. Using simple maps drawn by respondents, she asked respondents to identify the poorest of the poor, then the next level of poverty and so on until all households had been ranked. Respondents were selected by "drawing the villagers together near a school or tea stall" (p.23). Again, the technique is unclear, but it seems that villagers were allocating households to a strata, before having decided on how many strata there would be. This could result in confusion and much re-ranking as criteria may change in the minds of respondents as each next level is identified. This may have been done, but it is not described. Criteria were described by the respondents during the ranking process, but there was no description of how this data was handled. One of the aims of the study was stated as: "to document discrepancies, if any, between villagers' perceptions of poverty and the official ones used for targeting beneficiaries of poverty alleviation programmes" (p.22). These official indicators were income-based measures. The author did compare the criteria generated and those perceptions of the respondents in terms of the criteria generated in a descriptive fashion, but a comparison between rankings given by the respondents and a household survey using the income-based measure would have been very useful.

Another alternative to card sorting was employed by Schaefer (1992) using beans. Ten to twenty community respondents were asked to divide a pile of beans into groups with the same level of well-being. Questions were then posed about the characteristics of each group of beans, but these appear to be structured questions rather than open questions, such as sources of income. This will result in an imposition of an outsider's construct of SES or the dimensions the research team considered important in determining SES. Some interesting questions were posed relating to social mobility, credit sources and credit seekers, intermarriage between groups, casual labour employers and workers and means of payment. In addition, questions were included relating to community changes in status during recent years. Local criteria for differentiation included capacity of families to educate their children, occupation, and frequency of employment of agricultural workers,

or employment by others. This data may have provided a wealth of useful information about social and economic relationships, but unfortunately this was not reported or discussed. Finally, only one exercise was carried out in each community, which is certainly inadequate on which to base any meaningful analysis or interpretation.

Tung and Baliña (1993) tried two methods of the wealth ranking technique in the Philippines, the specific purpose of the exercise was not clearly stated. They compared the performance of individual ranking of 52 households, with group ranking by a number of informants at one time. Firstly, in the group approach, the researchers asked informants to sort the cards into three wealth groupings. Note that each informant took some of the cards each and put them into three piles, this was done with no reported cross-discussion between group members. There was a cross-check activity carried out, but the construct of SES and associated criteria that each respondent was using may have been different. Informants may have been less inclined to change piles extensively, although the authors noted that it was necessary. After the review they were asked to give features of livelihood in each wealth group, and then a discussion about the differences between the categories.

The authors "confirmed" the important wealth indicators generated during the wealth ranking in a household survey by wealth category, as house structures, land ownership, livestock ownership and size of land. There was no explanation how this was confirmed or any validation techniques utilised. With the single informant approach, which was not explained in the research report published, results were compared with the group approach. The authors found that 75% of the households had "similarity" in ranks. This was not explained, although a table suggested that the similarity actually meant the same rank. Extreme differences were not noted between the two methods, such as rich by one method and poor by another. The middle and poorer categories resulted in most disagreement, and richest in most agreement.

In a study of soil fertility management in Zimbabwe, Carter *et al* (1993) used wealth ranking to stratify 100 households according to wealth status. The objective was to identify differences in access to and management of resources for soil fertility management by wealth group. They published a report on the usefulness of the method. They used three different groups of informants (2-5 participants) to rank the same households simultaneously without any discussion allowed between groups. After sorting, the groups were asked to discuss piles according to their wealth and soil management techniques. The three different groups identified different numbers of strata, four, seven and eight. The authors used Grandin's average score, and natural break technique described above, and divided the households into six strata. Again this study found high agreement for the richest households, but also the poorest. They showed some large disagreements (richest by one group, poorest by another) in only a few cases, but they stated that may have been due to confusion about the type of pile (rich or poor) during part of the exercise. They reported considerable variation within the middle wealth groupings and suggested that this may have been due to economic mobility, although this was not elaborated upon. Unfortunately the claimed accuracy of the resulting stratification was only justified based on the "ensuing interviews with 22 households selected randomly" (p.48). These were not described. In terms of the criteria of wealth generated, cattle holding was identified as the most important criteria. The authors noted that visible signs of wealth such as housing and livestock were given, but atypical sources of income ignored (such as prophesying at apostolic events). The reported assistance from wealthier kin was overestimated resulting in higher ranking of younger households, conversely, widows and widowers were ranked lower than their status in terms of food and income security. Recent and drastic changes in status were also ignored. None of these interpretations were discussed further to assist the reader in understanding the basis of the assertions. Finally, the authors gave a detailed description of the problems encountered in setting up the exercise where adequate community preparation was not carried out.

Sharrock *et al* (1993) conducted wealth ranking in the eastern hills of Nepal for an agricultural research centre. The aim of their study was to understand the relationship between farm household type and their acceptance of the centre technologies. They used the Wealth Ranking technique to categorise farmers into farmer-defined categories to enable the centre to develop research agendas based on these categories. Households were ranked by two informant groups in each of 18 sites. Each respondent group was asked to sort household cards into groups based on their understanding of the differences between categories of villagers within the site. The groups generated from between three and nine household groupings. Discussions were then held to determine the basis of the categorisation. The two groups per site were then bought together to discuss discrepancies in the preliminary rankings, and developed a final rank for the site. The following day, group interviews with respondents from the household categories were carried out to confirm the rankings and to identify issues of concern. The results of this confirmation were not explained further in any depth, or the tool applied described. Although the authors stated that food availability was the most appropriate basis for categorisation, the report does not clearly state what information was given to the farmers in terms of developing the household groups. There was no direct reference to socio-economic or wealth status, although this is assumed to be the basis of the activity.

In a summary of methodological issues, Sharrock *et al* (1993) pointed out the difficulties in analysing wealth ranking data, and the possibly superficial nature of the ranking exercise. They cautioned on the adoption of results without further exploration of community perceptions, and called for real expertise in data collection, interpretation and application.

More recently, in a study of household coping strategies in Uganda, Seeley *et al* (1995) applied the wealth ranking technique in an innovative way to assist in developing confidence in the appropriateness of survey items in an index of socio-economic status. After the usual ranking procedure discussed in the previous studies, the respondents were asked to identify the criteria they used to categorise

households. These community generated criteria were used to select the most appropriate variables identified for a socio-economic baseline survey. The items identified were: housing construction materials, total land acreage available for cultivation; a household item index (ownership of seven items); and ownership of cows.

1.6.3 Reliability and Validity Assessments of Wealth Ranking

The literature was searched thoroughly for examples of studies that were designed to rigorously evaluate the performance of wealth ranking technique for whatever purpose it was being applied. Aside from the studies by Scoones (1995), Rajaratnam *et al* (1992), and Grosvenor-Alsop (1991), other studies reported here are examples of research projects that made efforts to use their results to provide supportive information on wealth ranking performance, rather than conduct research specifically for the purpose of measuring reliability or validity of wealth ranking. There may well be studies conducted by different groups such as private voluntary organisations, other non-governmental organisation, or government departments; but a sustained effort to identify such studies was not fruitful.

Reliability

In a larger study that explored the ways Nepalese farmers describe and classify soils, Chadwick and Seeley (1996) used the wealth ranking technique to stratify a sample of farmers according to their socio-economic status. Wealth ranking in this study was conducted in the following manner. In three different Forest User Groups (FUG), six individual respondents were chosen according to their sex, from a range of ethnic groups, and those who would be expected to know the study households well. Each respondent carried out the wealth ranking exercise independently. At the beginning of the session, the respondent was given rough criteria by which to categorise wealth. Respondents were asked to rank all the study households (94, 118 and 123 in the three FUGs) and were allowed to decide on the most appropriate number of socio-economic groups. Scores were normalised during the analysis into four groups. Scores of the six respondents for each household were summed and a mean rank calculated. During the ranking

exercise respondents were asked to give criteria (or rationale) for placing each household into a socio-economic group. This resulted in a list of criteria as stated by the different respondents.

In later analyses the authors examined the level of agreement of rank (household socio-economic status into four groups) between the respondents. This was approached by comparing the number of households that were placed in each of the four socio-economic groups. In two of the three FUGs, the authors found a marked difference in the proportion of households in each socio-economic grouping. For example, on one of the three FUGs, the proportion of households in the wealthiest category ranged from 1 to 33%. In order to describe the absolute deviations in agreement between the six respondents in one FUG, the authors found that only 24% of households were consistent to within one point on the four point ranking scale.

Chadwick and Seeley also examined the number of households not recognised (either unwilling or unable) by the respondents. They found that in the first FUG, between 59 and 96% of households were ranked (known); 17 and 85% in the second; and 42 and 83% in the final FUG. The authors pointed out that as the study had only a small number of respondents, a meaningful comparison of criteria generated by different respondents was not feasible, and comparison of criteria across FUGs was therefore descriptive.

The authors went on to provide an evaluation of the wealth ranking method, and found it only able to provide a crude estimation of socio-economic status of households. They pointed to problems with the normalising of data (where respondents are allowed to determine the number of socio-economic groups), and allowing respondents to determine their own criteria for assessing wealth. They suggest that one or two criteria are given to respondents to facilitate wealth ranking for purposes of stratifying samples for further investigation.

Scoones (1995) conducted a study to compare simple wealth ranking (allocating a rank only) and the criteria generated from respondents during focus group discussions with conventional survey indicators. In this study there were seventy-

one households ranked in one site, using three respondent groups of between nine and twelve participants. Two groups were men and women taken from the households studied and a third group which was comprised of local development workers familiar with the households. Although Scoones described certain results in terms of reliability and validity, the term reliability was not being used in the scientific sense. However, in an earlier section of the paper, Scoones did discuss the 'degree of agreement' between the respondent groups. He found that there was a 'reasonable' degree of agreement, however he did not elaborate further on what was considered reasonable. Scoones also pointed out that there was 'greatest degree of agreement between groups...at the top and bottom ends of the wealth scale' (p. 76). That is, in the four forced categories of wealth, the wealthiest and poorest were easiest for respondents to agree upon. No data was presented to determine the degree of agreement across all categories of wealth. Scoones was able to compare the criteria used by the different respondent groups. Finding that these differed between the groups, he recommended multiple ranking sessions capturing the diversity of local perceptions.

Grosvenor-Alsop (1991) examined the applicability of the wealth ranking technique in India. This was rightly justified by the fact that India was spatially and culturally very distant from the areas where wealth ranking had previously been applied. The study was particularly concerned with whether or not the position of the respondent within the complex hierarchy would influence their ranking of other households. A comparison between the resulting wealth rank and other conventional indicators of wealth was also made.

The study took place in a single village with 87 households. Five individual respondents were selected to independently rank all households in the village, and were selected from high caste, low caste, very low caste and Muslim sections of the community. Grosvenor-Alsop describes high correlations between the ranks allocated by each respondent. These pair-wise correlations ranged from 0.86 and 0.94. It is not clear which coefficient was used. The ranks were also examined to detect patterns in differences in ranks allocated. For example, the author noted that

there were lower correlations between the higher caste and Muslim informants. None of the differences described were statistically significant. Grosvenor-Alsop interpreted these high correlations as an indication that wealth ranking is "an appropriate tool to use for social analysis in a stratified society" such as the village studied (p.7). This interpretation is perhaps a little premature as the study was only conducted in a single village of 87 households, and with five individual respondents. The term social analysis suggests a more complex activity, and, decisions on the appropriateness of the tool would depend on more than an agreement on rank.

Validity

Looking at construct validity, Grosvenor-Alsop (1991) in the same study as described above, compared the wealth rank allocated to households with a number of conventional measures of wealth collected during a previous survey. These were: a) income, and b) assets (land, livestock and number of people in paid employment). A Spearmans Rho test was used to determine the correlation between the wealth rank and the conventional measures. It is not clear how many categories the wealth ranking scores were divided into, or whether they remained as a list of raw scores. Nor is it clear how all the conventional variables were measured, although income and land were divided into five categories. The author reported strong associations between allocated wealth rank and per capita household income, but weak associations with total household income, land holding, livestock ownership and persons employed in the household. Although the study provided some interesting information, and did attempt to compare wealth ranks with survey data, interpretation of the data generated from this study should be cautious. However, the author states that "The results were conclusive in that Wealth Ranking is a useful technique of stratification for the area of India in which it was tested.....Wealth Ranking is a technique that can be used with confidence in India". A more rigorous approach, and a number of additional studies carried out is clearly indicated before developing such levels of confidence.

Scoones (1995) also attempted to address the issue of validity. He compared the allocated ranks with data previously collected on cattle ownership and access, goat ownership, field size, cash and remittance incomes, crop yields and sales, age of the household head, household size, and other asset ownership levels for the same 71 households ranked during the wealth ranking sessions. These were compared separately for the respondent groups and a summary rank for men and women (although the summary statistic was not clarified). Comparison was demonstrated by correlations, but the type of correlation coefficient used was not identified to enable an assessment of correct management of ordinal data. Setting aside our inability to assess the interpretation of the findings, Scoones was able to demonstrate a high degree of correlation (0.50 - 0.61) between the allocated rank, and several of the survey indicators. Significance levels were calculated.

Although there were weakness in the reporting of this study, Scoones takes a first important step towards building up confidence in the construct validity of wealth ranking. Scoones was also able to draw some very insightful explanations about his findings using the information from focus groups discussions. Particularly, Scoones asserts that wealth and well-being are 'complex and dynamic, with multiple local meanings and interpretations' (p. 85).

Rajaratnam *et al* (1992) carried out a larger study to compare the criterion-related validity of simple wealth ranking and four traditional indices of SES constructed using commonly accepted measures of SES in Tamil Nadu, India. The research team selected a *panchayat* of 412 households residing in three different clusters. The main village consisted of 205 households, and two neighbouring settlements, one with 90 households and the other with 117 households.

The study included the following steps. Firstly, four composite indices of socio-economic status were developed based on 'traditional' measures used, presumably, by the research team in the past. The basic items that were selected and included in the four indices were: a) caste, b) type of roof, c) type of house, d) highest educational status in the household, e) highest occupation in the household, f) occupation of all household members economically active, g) size of land holding,

h) per capita income from a number of sources, i) average set of dresses per person in the household. Details of how these items were operationalised, and scoring techniques were given. The different items were arranged in different combinations during analysis into four different indices. Each index resulted in a sum score. Survey enumerators preparation was described and met usual practice standards.

Next, households in the three clusters were allocated a SES category according to allocated rank during the wealth ranking technique. In each site two respondent groups (one male and one female) each ranked the households, using cards, into three pre-determined groups. During the ranking activity respondents gave criteria for categorisation, and were free to debate the placement of individual households. The authors reported high agreement between the groups which was determined by simple inspection of the allocated ranks. No reliability tests were performed. It is important to note that where the second group disagreed with the rank of a household from the first group, the researchers intervened to determine the discrepancy. If this was an overt question, indicating that there had been a discrepancy, then this would most certainly affect statements on reliability between the two groups. The authors justified the technique as "scientific" because of "triangulation at every stage". I assume this refers to the fact that the rankings were cross-checked by the respondent group after completion, and the second group was able to discuss discrepancies from the first group. This is not triangulation, which should involve the collection of data using different data collection methods to verify information obtained.

It was unclear as to whether exactly the same households were ranked as were surveyed. While discussing the wealth ranking activity, different numbers of households were given than in the introductory description of the research population. That is, more households were ranked than were previously stated as being present in some clusters.

Household scores from the four professional indices were matched with the ranking obtained from the wealth ranking exercise. It was found that an average

agreement between the four professional indices and the community rank was 62%. Unfortunately the statistics applied in this study were not able to determine the amount of error by calculating the percentage of agreement that would be expected by chance. Those households that were given different ranks by the different methods were further investigated by the research team and the true status was "verified" by the following method.

The verification activity took place only in the main village which accounted for approximately 50% of the study households. For each household the investigators interviewed various influential persons listed as village health volunteer, leaders, and the Rural Community Officer and visited the household to conduct interviews and observe the household. Although unclear, it seems the interviews and observations were focused on clarifying the reasons for the discrepancies elicited from the ranking and survey activities. The investigators did specifically ask about criteria given by the two methods to clarify which was correct. This appeared to be a very useful and informative exercise. Adequate details of interview or observation schedules were not included in the report, but the authors referred to information obtained on gold jewellery, type of utensils in the kitchen, type of food cooked, quality of housing etc. How reliability of judgement was maintained was not apparent. For example, observers were to comment on the quality of the dresses worn. These criteria were justified as "in India these tend to reflect the economic status of a family" (p.9). No references or other justification was given to support this assertion, or to explain how 'quality' would be determined. In-depth interviews and observation is not an unreasonable approach to validation, but the study would have had more impact if it had clearly stated the methods employed for this important component of the work.

Of the 38% of households where a discrepancy existed between the two methods in the main village, 50% were sampled for the verification. The four professional indices achieved an average accuracy of 57.0%, while the wealth ranking approach achieved 96.7% accuracy. The study report concluded with some interesting reasons for misclassifications by the two methods. For example, in the survey

approach, the authors asserted that respondents were not giving accurate information on ownership of land or assets, or income. However, it is not clear how the authors developed confidence that on household verification, the 'truth' was given. They found through observation that some rich people were living under thatched roofing, while some poor people were living in fully tiled, inherited houses. They also noted that caste was a poor predictor of status. In terms of the wealth ranking, the authors noted a problem with respondents classifying relatives as poorer than reality, giving emphasis to roofing material which was shown to be a poor indicator, and judging illness as an important determinant of poverty, even when the family is quite well off.

This study was a good start at looking at the comparative performance of different the approaches. Particularly, the in-depth 'verification' procedure shows promise, and could be developed further. However, the report was unclear about the actual methods and tools employed, calculations of statistical significance would have greatly enhanced the ability to interpret the data, and it would have been difficult to replicate the study for the same population.

1.6.4. Wealth Ranking as a Valuable Tool for Identifying the Poor

As can be seen by the studies and reports discussed above, wealth ranking is becoming a popular tool to identify poor households for a number of practical applications, to understand more of the determinants of poverty and wealth, and explore social and economic relationships within small communities. The potential benefits are obvious. However, there is a dearth of rigorous studies carried out that have attempted, specifically, to understand the performance of these techniques. Most studies that have contributed to the discussion of certain aspects of performance of the wealth ranking technique have included this as additional information generated from larger studies. These studies were, most likely, unable to dedicate the resources required to assess performance more comprehensively.

Identifying and applying community perceptions of wealth and poverty certainly fits well with current development philosophy and rhetoric. Community involvement is high on the agenda of the current Health Sector Reforms that have

swept the developing world in the last decade. Governments and donors are rushing to support community-based projects and wherever techniques such as those applied in this study are used we feel happy that the community is being involved, and their perceptions and realities considered. But what do we know about techniques such as wealth ranking? There have been numerous reports on practical issues of conducting the exercises, such as gathering and selecting respondents, encouraging good participation in discussions, and descriptions of the types of criteria that are generated. Although these have enabled wealth ranking to gain popularity, little is yet known. Some research groups have made preliminary assessments of reliability and validity. Although these have been important contributions, validity assessment is an ongoing and lengthy process of research. Aside from lamenting the inadequacies of traditional measures, are we ready to assert that the criteria generated from such techniques can be used to replace traditional measures in some circumstances, or are they useful only for a limited range of small-scale projects working with very small populations? Do we have sufficient evidence about the mechanisms that are operating during the wealth ranking process? Is there a unified community concept, or is Scoones (1995) correct in asserting that there may be several?

Conversely, should these techniques prove useful in generating appropriate indicators of socio-economic status for surveys, targeting the poor, or other health system delivery purposes, then they should be brought to the attention of a wider population of health professionals and researchers beyond the community development arena. They require fewer human and financial resources than more cumbersome measurement development techniques. Importantly, they may be able to provide appropriate indicators for specific locations, especially in less developed countries where the need is greatest. We also need to encourage a commitment of adequate resources and research expertise to answer some of these questions. As validity testing is an ongoing process, we need to test wealth ranking in a variety of settings for a variety of purposes. We need to bring wealth ranking and similar techniques into the mainstream scientific literature. More importantly, we need to ensure that we are recommending reliable, valid and well tested tools

for the less experienced, more peripheral social sector professionals who are not in a position to conduct sophisticated validity or reliability assessments.

The aim of this study is to contribute to the theoretical and empirical knowledge base concerning the utility of one participatory research method - wealth ranking. Despite our limited knowledge about the performance of wealth ranking, this study only addresses a small number of issues, with a focus on the utility of wealth ranking to identify location-specific criteria that can be used for survey purposes.

The broader aim was to identify a method for non-specialists to develop a valid measure of socio-economic status for their specific location. The method must be appropriate for the skills and resources of typical district level health or other social sector workers. The 'package' must also include the means to conduct a preliminary validity assessment of the proposed measures, by researchers without advanced research skills. This is especially relevant where there are potential negative social consequences from applying poorly developed measures.

Chapter Two

Methods and Procedures

2.1 Introduction

This chapter will begin with a description of the study aim and the theoretical framework within which the research was located. The two broad objectives of the study are stated, and linked to the two main phases of research activities. The research questions for phase one, and main study hypothesis for phase two are stated and discussed. The remainder of the chapter describes the design, methods and procedures for each research phase separately.

2.2 Study Aim

The broad aim of this study was to assess the utility of the wealth ranking technique as a tool to develop valid measures of socio-economic status (SES). These measures were required to be location-specific, that is, reflective of geographical, climatic, cultural and economic differences across a range of locations. In addition, all procedures adopted to develop and validate these measures were expected to be applied by non-specialists without advanced research skills, and with limited access to resources.

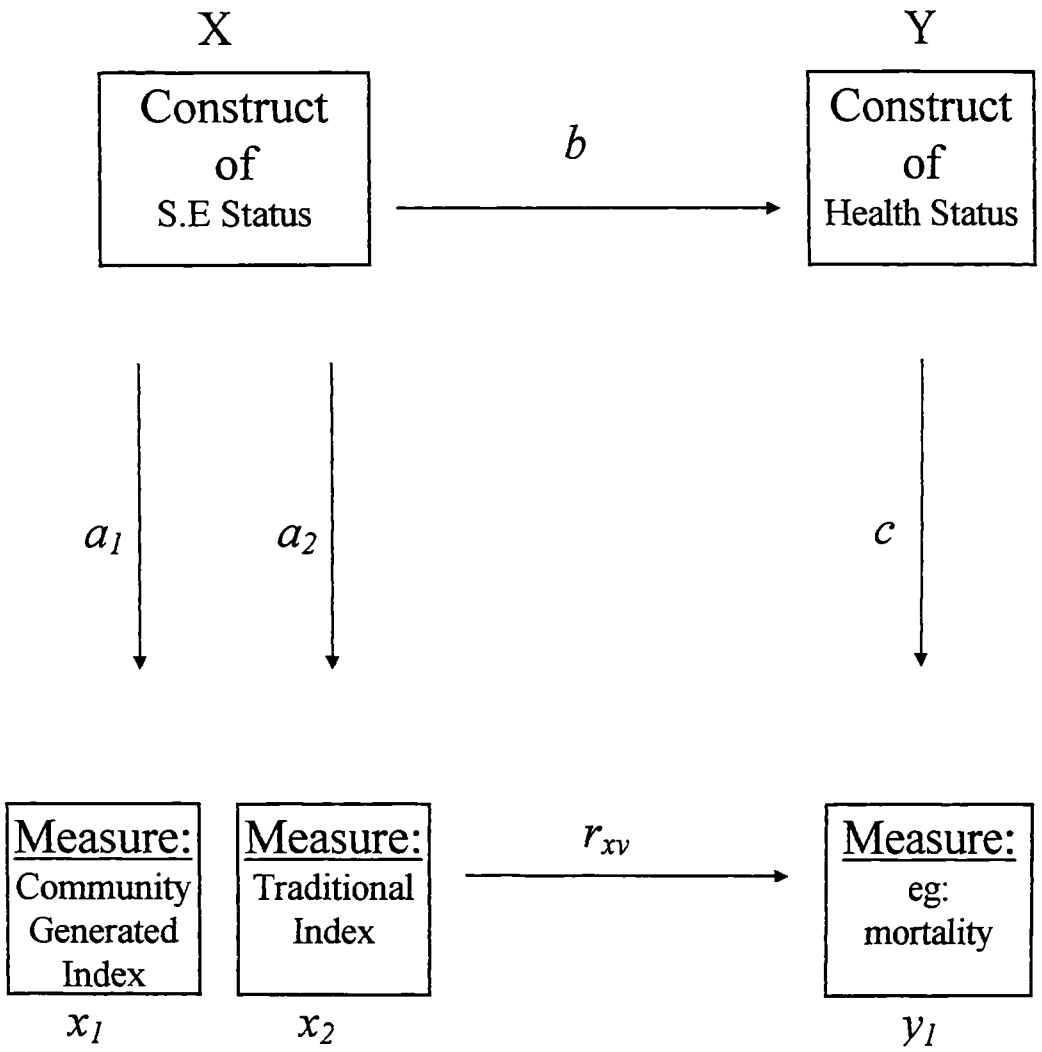
2.3 Theoretical Framework of the Study

General measurement theory has been used to guide this work. A brief discussion of this framework will provide a basis for the choice of study design, research questions and the main study hypothesis. A causal model, based on Sullivan and Feldman's (1994) model of substantive and auxiliary theory in measurement presented in Section 1.2.1 is used as the framework of this study. Figure 2.1, on the following page, is a model of a hypothetical, but typical study designed to measure inequities in health. Investigators concerned with health inequities may conceive of such a model to describe the substantive theory or hypothesis, shown as (b), that SES is related to health status (X is related to Y). Firstly, the constructs of SES and Health Status are defined. Then a measure of the construct SES is selected (x), and another for Health Status (y). Using this hypothetical model, the researchers would

measure r_{xy} in order to make inferences about the true relationship between SES and Health Status.

These measures, x and y, should be chosen according to some auxiliary theory or epistemic correlations previously determined (Blalock, 1982). However, there is a dilemma for most researchers in developing countries to decide on their measure (x) as little auxiliary theory has been developed. This model shows two possible options; x_1, a community generated index of SES, and, x_2, a traditional measure using indicators selected by professional researchers or individuals from outside the community under study. This study is concerned with a preliminary exploration of a_1 and a_2.

Figure 2.1. Hypothetical Theoretical Framework of an Equity Study



As Chambers (1995) has argued, the community itself is best placed to describe it's own realities and experiences. For poverty research this implies that community members are able to describe the causes, nature and dynamics of poverty in accordance with their own experiences and in the context of their political, cultural, geographical and economic environment. This leads us to consider whether or not a traditional index can adequately capture a conceptualisation of SES that reflects this community reality; or, whether an index constructed with indicators identified by community representatives would be a more appropriate measure. Wealth ranking is a technique that has been widely applied to capture this community conceptualisation of SES (or wealth), and in this study the technique was used to construct an index using the indicators identified by community respondents during wealth ranking exercises.

When choosing a measure for whatever purpose, its value is assessed in terms of its ability to measure what it is intended to measure. Therefore, in the exploration of a_1 and a_2 in the model, validity of the proposed measures is the central issue.

In accordance with Messick's (1988) recommendation of selecting more than one strategy to investigate validity, I have chosen to measure concurrent criterion-related (or diagnostic utility) but more importantly to focus on construct validity. The literature review highlighted that criterion-related validity does not fully address whether or not the measure is measuring the same construct. In this study, criterion-related validity is examined to assess whether or not the community generated index could be considered a suitable *substitute* for wealth ranking individual households. It is obviously not feasible to rank each and every household within a district's population. Wealth ranking limits the investigators to small groups of households, and certainly not more than 100 households at a time. If a community generated index can be constructed and applied over larger populations, then this would be a practical substitute for conducting many series of wealth ranking exercises. Therefore, in the context of criterion-related validity, the rank allocated by respondents during wealth ranking exercises was identified as the criterion.

This criterion was also selected based on an hypothesised relationship between the criterion and the underlying construct of interest, in this case a community construct of SES. It is recognised that the auxiliary theory linking an allocated rank using the wealth ranking technique to the community concept of SES is in its earliest stages of development. However, two studies were identified that provided preliminary empirical evidence of this relationship with "expert" constructs.

As construct validity is of a more fundamental concern, this study measured the convergent and discriminate construct validity of the two study indices, community generated and traditional. This was achieved by applying the Multitrait-Multimethod technique used extensively in psychometrics, and to a lesser extent in health research. Although no technique can *prove* the validity of a measure, it is a powerful tool to develop confidence in a measure - depending on certain characteristics of the measures and how they relate to one another.

2.4. Broad Objectives

The study had two broad objectives:

- to examine in detail the performance of the wealth ranking technique and its' feasibility as a tool for non-specialists;

- to assess the validity of indices of SES constructed with criteria generated during wealth ranking sessions; and to compare this with an assessment of the validity of traditional indices.

The study was carried out in two phases according to these broad objectives. More specifically, the first phase examined the reliability of wealth ranking; compared the criteria generated within and across geographical locations, and assessed the skills and resources required to conduct wealth ranking exercises. Phase one was conducted from August 1994 to July 1995.

In the second phase, indices of SES were constructed using the criteria that were identified during wealth ranking exercises in the first phase. The study then carried out assessments of the criterion-relatedness and construct validity of an index constructed with these community generated indicators. These findings were compared with identical validation assessments of a second index constructed with indicators commonly used by professional researchers. Phase two was carried out between August 1995 and September 1995.

2.5 Research Questions and Main Study Hypothesis

To meet the first broad objective, the study posed a number of research questions about the performance and feasibility of the wealth ranking technique. For the second objective, the main study hypothesis was posed.

2.5.1 Phase One: Performance and Feasibility of the WR Technique

The following research questions are divided into three main groups. Firstly, questions relating to the ranking procedure itself, where a number of households are allocated a socio-economic rank by five different community respondent groups (selected according to sex, socio-economic status, and community leadership). These questions are categorised under the heading - *Simple Ranking.*

The second group of questions relate to the indicators or criteria used by respondent groups during the ranking session to determine a household's socio-economic status. Although these criteria were used in the second phase to develop the community generated indices of socio-economic status, the first phase was concerned with the use of these criteria by respondent groups within and across geographical locations. These questions are found under the heading *Criteria Used by Respondent Groups to Determine Household SES.*

The final group of questions are concerned with the *Skills and Resources Required* to carry out the wealth ranking exercises as applied in this study.

Simple Ranking

1. Do different community respondent groups rank their neighbours wealth status reliably?

Claims are made about the reliability of the wealth ranking technique to allocate a socio-economic category to clusters of 50 to 100 households. However, many of these studies were carried out in a small number of locations within single regions with similar cultural, climatic, agricultural or topographical characteristics. It is valuable to measure reliability using statistical techniques that allow us to determine the significance of our reliability estimates, and across a diverse range of locations.

Secondly, if respondent groups do reliably rank their neighbours' SES, then we may assert more confidently that respondents may be using similar criteria to allocate these ranks. If this is so, then we could consider the potential of capturing these criteria and constructing an index of SES for survey purposes. If reliability is poor, then it would seem reasonable to expect that respondents are either conceptualising SES differently, or are using different criteria to make judgements about SES rank - thus putting into question the value of constructing an index of SES using these criteria that are generated by the community.

It was beyond the scope of this study to determine the validity of the simple wealth rank as a measure of a community construct of SES. However, if simple wealth ranking is highly reliable, then we have taken the *first* of many steps in the assessment of the validity of the measure in Tanzania. The measure could not be valid if it is unreliable.

2. Are different community respondent groups who participate in the wealth ranking sessions (wealthier men, poorer men, wealthier women, poorer women) actually composed of respondents of the correct socio-economic status?

When studying the results of different respondent group ranks, or examining the type of criteria generated by different respondent groups, comparisons are made between the groups based on sex or SES. Before we can make meaningful comparisons, it is necessary to determine whether or not the respondent groups are in fact composed of individuals of the correct SES. This is an important question for other studies using wealth ranking to explore different perspectives within the community - that is, where analyses are based on the assumption that respondents have been correctly classified. This will also provide us with additional information on whether the selection process used in this study was effective.

3. Do different respondent groups (based on socio-economic status, sex and leadership) who participate in the wealth ranking have a particular bias in the way they categorise households' socio-economic status?

4. Do particular respondent groups (based on socio-economic status, sex and leadership) have difficulty in recognising households for ranking?

It is frequently heard anecdotally that leaders and women are well informed about the characteristics of members of their community. Especially, that women's knowledge has been under-utilised. An examination of possible biases in ranking will add more detail to our knowledge of the wealth ranking technique in Tanzania, and of community groups' knowledge of their community. One of the rationales for using wealth ranking is that community members are more likely to know about their community than outsiders. This implies that they are likely to know, in some depth, about their neighbours wealth status and general conditions of life. If this is the case, then we would expect a high level of recognition of households to be ranked.

Criteria Used by Respondent Groups to Determine Household SES.

5. Do different community respondent groups within a single village location use *similar* criteria in ranking their neighbours' wealth status?

6. Do community respondent groups across a range of geographical locations use *different* criteria in ranking their neighbours' wealth status?

There have been frequent references to the variety of social stratification systems, cultural norms and practices across geographical locations (e.g. Artz *et al*, 1971); and therefore a need for location specific measures of SES. If this is so, then we would expect to see that the criteria cited by respondents would vary a great deal across a large geographical area, (encompassing a range of ethnic groups, languages, economic activities etc.). If criteria do vary substantially, then this would support the argument for location-specific indices of SES.

Skills and Resources Required

7. What skills are required to carry out wealth ranking as a method to develop location-specific indicators of SES?

The proponents of wealth ranking assert that it is a simple and rapid technique ideal for use by non-specialist researchers. However, there are few studies that have reported on this in detail. If wealth ranking can be managed by personnel with minimal training or preparation, then the technique should be widely and enthusiastically promoted. However, if the technique requires special skills for certain applications, then these should be made clear, and the information made accessible to potential users. Poorly applied research tools, with faulty analyses can destroy the reputation of methods that may have been useful at certain levels of expertise.

8. What resources are required to use wealth ranking as a method to develop location-specific indicators of SES?

Only one study on wealth ranking (Rajaratnam *et al.*, 1992) calculated and reported the actual cost of carrying out the wealth ranking exercise. This information is an important factor in the decision to adopt a method for severely constrained budgets. What has been reported as cheap for some institutions funded by or working in transitional or industrialised settings may be well beyond the reach of those locally financed institutions in Sub-Saharan Africa. Details of what the budget was spent on in dollar estimates eliminates confusion from subjective reporting of cost.

2.5.2 Phase Two:

Construction and Validity Assessments of Community Generated and Traditional Indices of SES

The final phase involved the construction and validation of the two study indices. Preliminary assessments on the criterion-related and construct validity of the two study indices were carried out, and the main study hypothesis was tested:

> indices of socio-economic status using community generated indicators are a more valid measure of the community construct of socio-economic status than traditional measures using indicators selected by researchers or social sector professionals.

The next section will describe the research sites and their selection for both phases of the study. As the study was conducted in two phases, each addressing one broad objective, the remainder of this chapter will be divided into two main sections. The first will describe the design, methods and procedures for the first phase of the study concerned with the performance and feasibility of the wealth ranking technique. The second section will describe these same aspects of the research for the second phase which was concerned with the construction and validation of the study indices.

2.6 Design

This was a methodological study of measurement development in two parts. Part one assessed the reliability of the wealth ranking technique using kappa statistics; it also applied kappa statistics to compare levels of agreement on the importance of indicators of SES, generated by the wealth ranking technique, within and across geographical locations; and finally, conducted a descriptive, subjective assessment of the skills required to apply the technique appropriately, and estimated the cost of wealth ranking sessions per site in US dollars.

The second part constructed and assessed the validity of two indices of SES. Indicators for the first index of SES were generated from community respondents during part one. A second index was constructed using indicators commonly used by research groups in Tanzania. Criterion-relatedness was assessed using household surveys to calculate the sensitivity and specificity of the alternative measures. Construct validity was assessed using household surveys and applying the Multitrait, multimethod validation technique.

2.7 Research Sites

The study was carried out in three of the five districts within the Coast Region of Tanzania, and in one district each for Morogoro Region and Kilimanjaro Region. Figure 2.2 shows a map of Tanzania with the research sites marked 1 to 5. This imbalance in the number of districts per region, a limitation of the study design, occurred because these were the only sites where permission was granted to conduct research within the larger projects with which we collaborated. The first phase of the research, assessing the performance and feasibility of the wealth ranking technique, was carried out in all five study districts. The final, validation phase was conducted only in the final two districts.

The three districts in Coast Region were selected from the five districts in the region based on their general levels of socio-economic status, and their cultural

and geographical variation. This was determined in consultation with Regional Ministry of Health personnel and general demographic data available.

Kisarawe District is located about 40 kms or 50 minutes drive from Dar es Salaam, the national commercial capital. Road access and transport services are good. This district was categorised as a middle level of socio-economic status. The main food crops are cassava and maize, while cashews and oranges are the main cash crops. The main ethnic group are the Wazeramo. Islam and Christianity are the main religions in the area. Provision of essential social services are about average for Tanzania.

Figure 2.2 Research Sites in Coast, Morogoro and Kilimanjaro Regions

Rufiji District is located some 180 kms from Dar es Salaam, and was accessible only by a very poorly maintained partly-sealed road. Despite the short distance from the capital, it takes about 6 hours to reach the District Headquarters (HQ) by four wheel drive vehicle. This journey includes a ferry trip across the large Rufiji River, the successful crossing being dependent on the ability of the traveller to provide the required diesel at certain times of the year. Flooding also creates a complete barrier to accessing the District HQ by car during the rainy season. Transport and road infrastructure in the region is a major problem inhibiting economic development, and the district can be thought of as representing a lower socio-economic environment. Rufiji is an agricultural area with rice, cassava and beans being major food crops, and cashews representing a large cash crop earner. The main ethnic group in the area is the Wandengereho. Islam and Christianity are the major religions in the area. Essential social services are weak and access to health care difficult.

Mafia Island was chosen to represent the better off populations in the Region. Mafia Island is located 25 kms off the coast of the mainland and 150 kms from Dar es Salaam. It is accessed either by ferry, light aircraft, or privately operated boats. The main cash crop on the island is coconuts, and to a smaller extent, cashews. The main food crops grown are cassava and rice. Fishing is the most important income generating activity both for small and large scale enterprises. There is no specific ethnic group on the island, and the population there do not identify with any particular tribe. They see themselves as largely a mixed group of people that came to the island over the centuries, and, through free intermarriage have consequently lost their original tribal identities. The large majority of the population are Muslim. Essential social services are comparatively good despite their isolation from the mainland.

In Morogoro Region, we worked in *Morogoro Rural District*, the district surrounding the Municipality of Morogoro, a large regional commercial centre. Morogoro Rural is approximately 200 kms from Dar es Salaam, four hours drive on a reasonably well maintained, sealed road. Transport and access to services is

better than for two of the three districts in Coast Region. Although there is still a high reported incidence of poverty in the rural areas, Morogoro Rural District itself can be considered a district of middle socio-economic status. Main food crops are maize and cassava, while cash crops are maize, and a number of vegetables such as potatoes and tomatoes. The district is located at the foot of the Uluguru Mountains, enjoying a cooler, less humid climate than Coast Region. The main ethnic group in the area is the Wauluguru, and Islam and Christianity are the dominant religions.

Kilimanjaro Region is one of the wealthiest regions in the country. Here, we worked in *Hai District*. The research sites were located in the foothills of Mount Kilimanjaro, about 600 kms from Dar es Salaam, eight to ten hours drive on well sealed roads. The main food crops are maize and banana, while the main cash crop is coffee. Land availability is restricted in this area. The main ethnic group is the Wachagga, who have adopted a progressive lifestyle with more substantial investments in education and a desire to accumulate material possessions which is less notable in the Coast Region. The majority of the population are Christians. Transport and access to services in Hai district is reasonable, with some difficulties encountered in the higher mountain locations. Hai District was categorised as a higher level of socio-economic status.

2.8 Methods and Procedures

2.8.1 Phase One: Performance and Feasibility of Wealth Ranking

A. Overview

Wealth Ranking exercises were carried out in ten villages in five districts in Tanzania. In each district, one peri-urban and one rural village were selected. Fifty households were purposefully selected based on their close proximity to one another. These households were ranked into one of four socio-economic categories ranging from the wealthiest (SEG1) to the poorest (SEG4) by five respondent groups in each village. Four respondent groups were selected based on their relative socio-economic status and sex, and a final group was composed of local leaders in the village.

During these sessions, respondents were asked to state the criteria for allocating a household a particular socio-economic rank (SEG1-4). This resulted in a list of criteria for each respondent group, representing all criteria stated during the ranking of the fifty households.

Using kappa reliability statistics, the reliability with which respondent groups allocated a socio-economic category to each household was determined. There were three measures of reliability calculated. Firstly, the overall reliability by which the five respondent groups ranked the fifty households. Secondly, the reliability with which all respondent groups ranked the four different socio-economic groups (rich, comfortable, poor and very poor), in order to identify the socio-economic groups or categories where there was most agreement. Thirdly, the performance of each respondent group in terms of their level of agreement with other respondent groups to enable identification of the most reliable respondent groups.

The criteria that were generated during the exercises were compared within each village to assess the level of similarity or differences between criteria used by different respondent groups. Each criterion was allocated a category of importance

according to the relative frequency of use by the respective respondent groups. Using kappa statistics, the level of agreement about each criterion's degree of importance across the respondent groups was calculated.

Comparison of criteria generated across geographical locations was also made. This was achieved by aggregating the criteria used within a single village by all respondent groups, and similarly allocating a category of importance based on relative frequency of use. Using kappa statistics, the level of agreement across locations was calculated.

Feasibility of the wealth ranking technique to simply allocate a SES rank, and to generate criteria was also assessed. This assessment was made based on the technical skills required to conduct the activities, and the resources required to do so.

Focus group interviews were conducted after each wealth ranking exercise to provide additional information about the most important criteria cited by respondents during the ranking activity. Information from both the wealth ranking exercises and the focus group interviews were used to identify items for inclusion in an index of SES for *potential* use by each district in the future. The wealth ranking and focus group interviews were also conducted in this phase to perfect techniques designed to select, weight and construct response categories for index items; and to assess the feasibility of collecting and managing such information by typical district level health workers.

B. Methods: Wealth Ranking Technique and Focus Group Interviews

Wealth Ranking was used to determine the relative SES of fifty households according to five community respondent groups in separate sessions. A session facilitator, using cards displaying the name of the household head, asked respondents to place each of the fifty households, or cards, into one of four wealth groupings: rich, comfortable, poor or very poor. This was done one household at a

time. Respondent groups were not given a definition of wealth by the facilitator of the exercise. They were given freedom to determine their own concept of SES, and only through discussion amongst themselves, decide upon a suitable socio-economic category for each household.

A final, sixth wealth ranking session was conducted in order to determine a socio-economic rank for each household that would best represent the knowledge and perspectives of the five respondent groups. This session was called the 'verification' session and focused on discussing households where there were discrepancies between the respondent groups. One respondent from each of the previous five wealth ranking sessions was selected to participate in this session. Households where discrepancies existed were discussed in more detail and respondents would debate the socio-economic status of the household based on the information presented in each of their respective respondent groups. A final SEG would be allocated to households that reflected the final group decision. This final community allocated rank was used in phase one to judge respondent group bias, and to assess whether respondents participated in the correct socio-economic respondent group.

The Wealth Ranking session was also used to generate a list of criteria that respondents stated they used to make their categorisation. After each household had been allocated a socio-economic group or category, the facilitator asked respondents to state the criteria they used to allocate the SEG. In other words, what were the characteristics of the household that respondents used to decide on SES, relative to the other households. In phase one, these criteria would be compared and similarities and differences described for criteria used within villages by different respondent groups, and also across geographical locations. In phase two these criteria would be used to construct a community index of SES.

After each wealth ranking session with each respondent group, focus group discussions or interviews were conducted in accordance with Dawson *et al* (1993). Focus group discussions were not conducted with the verification sessions. I have used the term Focus Group Interviews to indicate that in some situations a full

intra-group discussion was not achieved, rather the group answered specific questions posed by the facilitator.

Focus group interviews were conducted to extract more detailed information about the criteria respondents used to categorise households. This information was used to cross-check the importance of criteria generated during the wealth ranking exercise, and to provide additional information that would be used to construct an index of SES using these community generated criteria, during the second phase. The data generated by the focus group interviews was not used during the first phase. Rather, the technique was refined, and then assessed in terms of its feasibility as a complementary technique to wealth ranking for construction of indices of SES.

Across all ten sites in phase one, a total of 60 wealth ranking sessions and 50 focus group discussions were conducted.

C. Data Collection Procedures

Definition of Terms

Household

The study definition of household was a family residing together, eating from the same cooking pot, and *likely* to be sharing resources (in acknowledgement that intra-household sharing of resources is not always the case). Clarification was simple as in all research sites there was little confusion with what constituted a household. There were very few compounds with sub-households within. In these cases, households were determined as the sub-household fulfilling the above criteria. There were a few first and second-wife households for an individual household head in Islamic areas. In these cases the two households were identified as separate units.

The Construct: Socio-economic Status

In order to capture a community construct of socio-economic status, it was important to select a word that allowed a broad interpretation. One that did not force respondents into narrow conceptions based on a uni-dimensional construct possibly inferred by the Kiswahili term chosen. The notion of socio-economic status was best captured by the Kiswahili term *uwezo*. This choice was based on extensive informal interviews within the regions with people whose mother tongue was Kiswahili. It is fascinating to note that *uwezo* translates directly into English as 'ability'. This hints at the same notion of 'capability' that Sen (1983) described (see Section 1.3.2).

Community Entry

After introductory meetings and securing permission from Regional officials, the research team entered each district and held discussions with District officials to introduce the project, gain support and to identify the villages in which the team would work. One rural community was chosen, and one peri-urban. Once the two villages were identified, the team travelled to the sites and spent two to three days in preparatory and introductory meetings with local ward and village leaders, in addition to public meetings with the community. The objectives of these meetings were to gain approval for entry, explain the purposes and use of the data, to gain support for the project, and reduce suspicion that often is associated with the sensitive nature of economic information. The meetings with leaders were also used to identify households for study, and the respondent group participants.

Selection of Study Households

Fifty households were required in each of the ten study villages to conduct the wealth ranking. These were selected purposefully in each village based on their close proximity to one another. For wealth ranking to be successful, households must be neighbours and respondents chosen from within the fifty households, likely to be reasonably familiar with one another.

In Tanzania, the smallest political units are referred to as '10-cell' units, comprising ten households. This unit was functioning well during Nyerere's Government, although this unit has been disintegrating since the change of government in 1985. Where the units were still present, five neighbouring 10 cells were chosen in each village. A 10-cell unit was chosen if households were clustered together and was adjacent to the other 10-cells selected. Where these units had disintegrated, the more recently developed *kitongoji* unit was used which can be as large as 150-200 households. We identified sufficient neighbouring households to reach the required number. A list of the household heads was prepared by the leaders of each 10-cell unit from which households were selected, or from the *kitongoji* leader. As this study was not aiming to measure the incidence of poverty, or make generalisations from the study population in any way, probability sampling was not appropriate.

Selection and Recruitment of Respondents

Five respondent groups were required to reflect a range of perspectives from within the community. Respondents were selected, where possible, from the fifty households identified for the study, so that they were more likely to have an in-depth knowledge of the study households. In a few situations, sufficient respondents could not be identified from the fifty households, and were chosen from other households immediately adjacent to study households.

As with other studies employing participatory methods (e.g.Welbourn, 1992) respondents were selected and grouped according to their socio-economic status and by sex. This separation of the community respondent groups was designed to aid in the open and frank discussions. It reduces inhibitions to participate based on perceived power or better knowledge of individual group members which is usually associated with socio-economic status and gender in these communities. According to Status Generalisation Theory (Berger *et al*, 1972), authority and control in interactions is automatically assigned to individuals of higher status. This occurs even when the interactions or group activities have no relationship

with status position. This separation also allowed perceptions of wealth and poverty from quite distinct groups within the community, and an examination of respondent group performance based on sex and socio-economic status. This classification of respondents resulted in four respondent groups labelled 'Wealthier Men', 'Poorer Men', 'Wealthier Women', and 'Poorer Women'. A final fifth respondent group was selected comprised of the leaders from all ten cell units selected for the study, or immediate neighbouring political units.

With the list of the fifty households, leaders were asked to identify between five and eight respondents to participate in each of the four groups based on sex and socio-economic status. They were not asked to rank households into socio-economic categories, rather select a number of respondents that could be considered to represent the wealthier or poorer households on the list. The accuracy of this procedure was assessed with one of the study research questions.

For the final, verification session as described above, one respondent was selected from each of the five previous respondent groups after their respective sessions were completed. These representatives were chosen according to their ability to participate well in a group discussion, and their apparent knowledge of the fifty households.

Respondents were contacted by a member of the research team one or two days before the session and invited to attend, and contacted again on the morning of the session. Local leaders accompanied the team during these household visits. During the initial contact they were given an introduction to the project objectives, use of the information, and planned activities. Respondents were told from this early contact, that no direct benefits would accrue to the village as a result of the research, and that their co-operation was voluntary.

Session Preparation

A site was selected to conduct the wealth ranking and focus group interviews by local leaders. The location was chosen to maximise privacy and to reduce the

number of observers who may inhibit open discussion. Sites were either school rooms, health clinics after working hours, or the government administrative office for the location. Session schedules were organised over as few a days as possible to minimise discussion in the evenings and comparison of ranks allocated by respondent groups. This could not be eliminated.

Each site was prepared for the following session. Furniture was assembled to ensure comfort during the ranking and discussion activities. A table was used to place the cards for ranking, and chairs or benches for respondents. Cards were pre-prepared with the name of the head of household on a single card for each of the fifty households. These cards were reused for all sessions in the site. A tape recorder and microphone was tested and positioned for use during the focus group interviews.

Respondents who had been recruited assembled at the site, or were collected from their homes by community volunteers or the co-ordinating leaders. Upon arrival respondents were greeted and the research team talked informally with participants to improve rapport.

The Session Activities and Tools

The research team for the wealth ranking and focus group interviews included a sociologist who facilitated all sessions, a research assistant who observed and recorded sessions, and myself, who provided technical assistance to the facilitator and co-ordinated all activities.

Once the participants had been shown to their seats, the facilitator introduced the team, project aims and planned activities within the village. This introduction was carefully planned and was given in the same way for all sessions. Appendix 2.1 summarises the major points and the rationale for each. It is important to state these explicitly so that the reader can determine how, and if, respondents were influenced to respond in a certain way.

Participants were asked to categorise households into one of four groups according to their *uwezo*. Each SEG was given a Kiswahili label reflecting the English terms 'rich', 'comfortable', 'poor' and 'very poor'. The discussion facilitator allowed the respondent group to discuss these labels and definitions among themselves. They were not instructed how to categorise households, or what would distinguish one socio-economic category from another. This part of the session would usually take between ten and fifteen minutes.

The facilitator then held the pile of cards and presented one household card at a time to be ranked. She showed the name of the household to the group, but also called out the name to enable illiterate respondents to participate. Sometimes nicknames were given (and written on the card) to enable other group members to identify the household. As the household was allocated a rank into one of the four socio-economic categories, participants were asked to identify the criteria they used to make their judgements. The criteria stated for each household was recorded by the observer. During the ranking sessions the observer also recorded on a separate sheet the criteria being used most commonly. These would be explored further during the focus group interviews. As a second observer, I provided support to the facilitator, to maximise facilitation performance during criteria generation. Although I was not fluent in Kiswahili, my comprehension level was sufficient to follow most of the discussion, and identify areas where the facilitator required support. I also recorded the final rank allocated to each household on a separate sheet to cross check with the observer's records.

Usually, after about five or six households had been ranked, or when cards appeared in all four piles, the facilitator reviewed the rankings allocated thus far. This was done to ensure that the allocation criteria had not shifted as the participants became more familiar with the exercise and the way in which they were allocating ranks. At the end of the exercise, all households were again reviewed to allow re-allocation of ranks if necessary. The majority of households were not allocated a different socio-economic rank.

Respondent groups were not asked to make socio-economic groups or piles of equal size as is often seen in wealth ranking exercises. I decided to allow respondents to place as many, or as few households per pile as seemed appropriate for the households selected for study. If respondents are forced into making strata of the same size, then this must result in forced misclassification for many households.

Immediately after the wealth ranking activity, focus group interviews were conducted. The sociologist facilitated these session and the research assistant acted as the discussion observer. For this activity, he made notes on group participation, assisted in managing people gathering outside, prepared and checked recording equipment and made notes for improving questioning and other practical issues with the session.

The objective of this part of the session was to explore further the criteria already generated during the ranking procedure with two sets of distinct questions. The first set of questions sought to identify common characteristics of the socio-economic groups (rich, comfortable, poor and very poor) in the community at large. These questions were designed to cross-check information collected during the wealth ranking on individual households to assess whether the criteria generated reflected characteristics of SES in the wider community.

The second group of questions were designed to explore the main criteria discussed by the group so that response categories for a final survey questionnaire could be constructed in phase two. For example, if the size of land was an important criterion for the group during ranking, then questions were designed to identify the actual size of land held by the different socio-economic groups in the community. See Appendix 2.2 for an example of a focus group discussion question guide. As this is an iterative process, focus group question guides changed to some extent with each session, although not substantially. Changes were made in order to improve the questioning technique rather than content.

D. Data management and preparation for analysis

During the Wealth Ranking in the Coast Region districts the research assistant (observer) recorded all the criteria given for each household during the session. For the final two regions we improved the quality of the data by tape recording this part of the session and entering lists of criteria per household directly onto the data entry charts from the recorded sessions. Lists of criteria per household were put on a matrix using MS Excel Version 4 for all five sessions in the site to facilitate easy checks on primary data. See Appendix 2.3 for an example from Rufiji District for the first 10 households in that site. For each household the final rank allocated by the group is also entered in the top right hand corner of each household box.

Focus group discussions were all recorded with high quality recorder and 360° microphone. At the end of each day the facilitator or research assistant would translate the full 30-40 minute session, while I entered a full *verbatim* transcript in English into WordPerfect for Windows Version 6.

After each session, the household SES rank allocated by each respondent group was entered into Crunch 4 statistics package in order to determine reliability of the technique using the kappa module of the programme, specially designed to calculate kappas for multiple judges.

Once primary data was entered into the computer software packages described, two backups were made in the field, and a third backup in the office in Dar es Salaam. Hard copies were also printed. Names of households and participants were kept secure and confidential, household ranking cards were destroyed after site completion.

E. Quality Assurance of Data

Research Team Training and Pilot Testing

The research team was given a seven day training that covered the aims and objectives of the study, a general overview of methods and procedures, and a review of each members' roles and responsibilities. The research assistant, a District Nursing Officer, had no previous exposure to research. A third project officer also attended this general introduction, although his role was to manage logistical aspects of the work, rather than data collection activities.

Although the sociologist was experienced in social science research methods, she had no previous experience of the wealth ranking technique and only limited experience with focus group discussions. A fourteen day, full-time training course was conducted separately to the general seven day training for the team. This training focused on qualitative questioning and group discussions.

Before commencing field work, and as an additional component of the training, the team conducted the wealth ranking and focus group discussions in a pilot village in Kisarawe District, a village not included in the study. Two respondent groups were used for the pre-testing of activities and the discussion question guide, and to develop basic experience in the team. This involved all procedures from site selection, village preparation, household and respondent selection, conduct of sessions, translation and transcription of the recorded sessions. A short de-brief amongst team members was carried out immediately after the session, and more comprehensive office debriefs were held after full transcriptions of the sessions were available for examination.

Ongoing training was provided, and special attention was given to the session facilitator. This training was quite extensive and was facilitated by the translation and transcribing of sessions. As the facilitator translated the tape recordings, specific examples were identified to demonstrate improved questioning or facilitation techniques. At least four hours per day was spent on this activity as

transcripts had to be prepared and preliminary data analysis conducted before the following sessions.

De-Briefing Sessions

A ten to thirty minute de-brief or discussion was conducted at the research site after all sessions. This allowed the team to review quickly and document any major issues of concern encountered during the session. At the end of each day (or before the next session) a full de-brief was conducted by the entire team where these issues were recorded and solutions proposed to improve the quality of the data. Transcripts were also examined and used to provide specific examples where questioning or discussion facilitation could be improved.

F. Data Analysis Procedures

Simple Ranking

Research Question 1: Do different community respondent groups reliably rank their neighbours wealth status?

The major aim of this analysis was to measure *overall* reliability of ranking household SES by the five respondent groups. Additional analyses were carried out concerning the reliability of particular respondent groups, the reliability of categorising specific socio-economic groups (rich, comfortable, poor and very poor), and the level of reliability achieved in different sites based on urban or rural location, or the socio-economic status of the district.

To meet these objectives, I applied kappa statistics which are now well accepted as an appropriate test for multiple judge reliability (Kraemer, 1980 and Posner *et al*, 1990). See Appendix 2.4 for a more detailed discussion of the different kappa statistics applied in this study.

Research Question 2: Are different community respondent groups who participate in the wealth ranking sessions (wealthier men, poorer men,

wealthier women and poorer women) actually composed of respondents of the correct socio-economic status? That is, is the process using leaders to identify and group respondents effective?

Respondents not residing in study households that we ranked according to SES were eliminated from this analysis. Using the community allocated rank from the final verification session in each site, respondents' 'true' SES was determined. This was then compared to the actual respondent group in which they participated. Those ranked as either rich (SEG1) or comfortable (SEG2) were classified as 'Better-Off' and those ranked poor (SEG3) or very poor (SEG4) were classified as 'Poor'. Simple proportions were calculated showing what percentage of respondents were correctly placed for each of the four respondent groups - wealthier men, poorer men, wealthier women, and poorer women.

Research Question 3: Do different respondent groups (based on sex and socio-economic status) who participate in the wealth ranking sessions have a particular bias in the way they categorise household socio-economic status?

For each respondent group session, the rank allocated to each household was compared with the community allocated rank from the final 'verification' ranking session in each site. The number of households where the particular respondent group overestimated, underestimated and gave the same rank was calculated across all sites by respondent group.

Research Question 4: Do particular respondent groups (based on socio-economic status and sex) have difficulty in recognising households for ranking?

For each respondent group session, the number of households that were 'not known' were calculated and aggregated across all sites for each respondent group.

Criteria Used by Respondent Groups to Determine Household SES.

Research Question 5: Do different community groups in both rural and peri-urban locations within a single district use *similar* criteria in ranking their neighbour's wealth status?

A list of criteria used to allocate a household SES was generated for each respondent group in each site. The number of times it was used over the fifty households was recorded. A criterion was only recorded once if it was used by several respondents during the ranking of a single household. See Appendix 2.5 as an example of how the data was aggregated. To enable comparisons of the importance of different criteria across the respondent groups according to their frequency of use, scores were transformed into proportions, then finally into categories of importance. Categories were determined based on the proportion that a particular criterion accounted for all criteria cited during the respondent group session. Categories were allocated in the following manner:

Proportion of all Criteria	*Frequency of Use*	*Category of Importance*
0.100 +	Very Frequently	1
0.050 - 0.099	Frequently	2
0.015 - 0.049	Occasionally	3
0.001 - 0.014	Rarely	4
0.000	Not Used	5

After each criterion had been allocated a category reflecting its relative frequency of use, it was possible to measure the level of agreement across respondent groups using the kappa statistic.

The choice of kappa as a statistic to summarise similarities and differences (agreement or disagreement) between respondent groups has been made cautiously. More usual measures of dispersion or variation such as standard deviations, coefficients of variation or ANOVA require a larger number of units for analysis. The five units representing the five respondent groups would be insufficient to determine the extent of random error. In the case of kappa, the

criteria themselves become the unit of study and the respondent groups the judges allocating a category of importance to each criterion. Kappa statistics do assume that the units of study are independent. In this case, the criteria are categorised according to a proportion, which by it's nature, is dependent. However, no statistic is currently available to describe this type of data, and the size of the data sets means that a qualitative description is not feasible. The kappa has been chosen as most suitable from what is currently available. Its limitations are recognised, but any negative implications of its use are considered minimal, compared to other measures of variation.

Research Question 6: Do community respondents identify *different* criteria in ranking their neighbour's wealth status in different geographical locations?

The data for cross location comparisons was analysed at two levels. Firstly, data was aggregated for all respondent groups for the peri-urban and rural sites into a single district set of criteria. Proportions were calculated and categories of importance were allocated in the same manner as the intra-site comparison. The district, rather than the respondent group is now the 'judge' for the kappa statistics.

Secondly, comparisons were made at the site level. That is, comparisons across the ten villages were made without aggregating data to the district level. It should be noted that the more judges that are included in a kappa statistic, the higher the reliability. Therefore when comparing a five judge kappa with a ten judge kappa, the behaviour of the kappa needs to be incorporated into the interpretation.

Other comparisons were made that examined the levels of agreement of specific respondent groups over the ten sites; and comparing the criteria used according to site type (i.e. rural versus peri-urban).

Skills and Resources Required

The project attempted to use the absolute minimal amount of resources to apply the wealth ranking technique as a method to construct an index of SES. A description of the resources actually used in this study is given, in addition to my own subjective assessment of the skills required to replicate this work.

2.8.2 Phase Two: Construction and Validity Assessment of Community Generated and Traditional Indices

A. Overview

Indices of SES were constructed and questionnaires developed using the techniques developed and refined in phase one. This was carried out for the final two of the five districts studied in phase one. District specific community generated indices were constructed for each of the two districts. A second index was constructed using traditional indicators of SES applied commonly in Tanzania by respected research institutions. The value of these two indices as measures of a community construct of SES was compared using two different validation strategies.

Four new research villages were selected for the validity assessments, but from the same two districts where the criteria for the indices had been generated. One peri-urban and one rural village were selected. In each site eighty households were purposefully selected based on their close proximity to one another. These households were surveyed using the community generated and traditional indices, and a socio-economic category was allocated to each household, according to each index or measure. Households were categorised into one of three categories: rich, poor, or very poor.

Firstly, criterion-relatedness, or 'substitutability' was assessed. The criterion or measure that the index was to be substituted for, was the wealth rank allocated by community respondents. In this case, wealth ranking was carried out on the eighty surveyed households in the same way as in the previous phase. This time, however,

criteria were not recorded or intentionally generated, and only three socio-economic groups were allowed. This resulted in five SES rankings being allocated to each of the eighty study households from each of the five respondent groups. A final, sixth session was carried out to discuss households where discrepancies occurred and to decide upon a final SES rank that would reflect the views of all the community respondent groups. This final rank, the community allocated rank, is the criterion by which the substitutability of the index was judged. Sensitivity and specificity tests were carried out for this purpose.

Secondly, convergent and discriminate construct validity for the community generated and traditional indices were assessed using the Multitrait-Multimethod approach (MTMM). For this approach two traits are measured using two methods. The two methods were community allocated rank, and household survey; while the two traits were SES, and skin tone. Convergent validity was assessed correlating two different measures (or methods) of the same construct or trait. Discriminate validity was assessed by correlating two theoretically unrelated constructs or traits; that is SES, and skin tone, using the same methods. Skin tone was required to be measured by the same two methods, community rank and survey. For the ranking method, each household was ranked into one of three categories according to the skin tone of the head of the household (black, brown and white). This was carried out in exactly the same manner as ranking according to SES, again by the five respondent groups, with a sixth verification session. Skin tone was also measured during the household survey as a single item question. Method variance was assessed using the different-trait, different-method correlations on the MTMM matrix. Correlations were calculated using Kendall's Tau *ß*

The following section will first explain the method of constructing the community generated and traditional indices, and will then present the methods and procedures followed to conduct the validity assessment of these indices. Indices were constructed using data collected from the wealth ranking and focus group discussions in the first phase. Although some of the data collected in phase one are

presented here incidentally as index items, this section aims to describe the method of index construction.

B. Constructing the Indices

The Community Generated Indices

Selecting index items from the criteria generated during the wealth ranking exercises

Taking all the criteria generated during the wealth ranking exercises in the final two districts of phase one, a list of criteria was constructed which reflected their relative importance. Importance was determined by the frequency of use during these sessions as described in the previous section comparing intra and cross site use of criteria by respondent groups. As an example, Appendix 2.5 shows the criteria generated for the peri-urban (H2) and rural sites (H1) for Hai District.

Next, focus group transcripts were examined. Data from focus group interviews were used to cross-check the findings from the wealth ranking sessions. The questions relating to the common characteristics of each socio-economic group were analysed for this step. Using the indexing facility of WordPerfect for Windows Version 6, transcripts were marked and coded. All statements made, mentioning a socio-economic characteristic of a socio-economic group, were indexed. See Appendix 2.6 for an example of the indexed data.

The number of different focus groups where criteria were mentioned was used, rather than the frequency with which they were used within each focus group. A criterion was selected for inclusion in the focus group data summary if it was mentioned in at least four of the ten respondent groups for the district (five groups in each rural and peri-urban site).

The criteria from the focus group summary and the wealth ranking exercises were then put together into summary tables, reflecting their importance during each

method of data collection. Tables 2.1 and 2.2 show this summary for both Morogoro Rural and Hai Districts. Only criteria reaching a minimum proportion of 0.015 (1.5% of all criteria generated) during the ranking sessions, and only those criteria mention in four of the ten focus groups were included.

Table 2.1 Summary of Main Criteria selected via WR and FGDs for Morogoro Rural District

Criteria Generated During Wealth Ranking Activity	Proportion of Total Criteria	Criteria Generated During Focus Group Discussions	No. Focus Groups
Occupation	0.122	Crop Yield	9
Crop Yield	0.115	Ability to Assist Others	8
Dependence on Casual Labour	0.060	Food Security	5
Age	0.050	Ability to Employ Casual Labour	5
Ability to Employ Casual Labour	0.044	Dependence on Casual Labour	5
Dependence on Others	0.043	Income	4
Roofing Material of House	0.042	Physical Status	4
Type of Crops	0.040	Assistance Received from Others	4
Food Security	0.040		
Income	0.035		
Physical Status	0.035		
Condition of House	0.027		
Variety of Sources of Income	0.025		
Assistance Received from Others	0.023		
Tenure of House	0.022		
Size of Farm	0.022		
Health Status	0.019		
Employee (on regular salary)	0.016		
Marital Status	0.016		
Ability to Assist Others	0.015		

Table 2.2 Summary of Main Criteria selected via WR and FGDs for Hai District

Criteria Generated During Wealth Ranking Activity	Proportion of Total Criteria	Criteria Generated During Focus Group Discussions	No. Focus Groups
Type of Crops	0.126	Tenure of House	8
Occupation	0.093	Size of Farm	7
Size of farm	0.082	Quality of House	7
Tenure of House	0.067	Income	7
Age	0.052	Type of Crops	6
Tenure of Farm Land	0.052	Tenure of Farm Land	6
Dependence on Casual Labour	0.052	Livestock Ownership	6
Livestock Ownership	0.036	Education	6
Wall Material of House	0.029	Food Security	5
Receive Assistance from Others	0.027	Condition of Farm	5
Education	0.023	Occupation	5
Condition of the House	0.019	Work Ethic	5
Work Ethic	0.019	Physical Status	5
Variety of Sources of Income	0.019	Dependence on Casual Labour	4
Dependence on Others	0.018	Variety of Sources of Income	4
Physical Status	0.018	Dependence on Others	4
Roofing Material of House	0.016		
Employee (on regular salary)	0.016		

On two occasions, criteria were aggregated together to form a single item. Ownership of livestock, number of livestock and type of livestock were all grouped together into one item called 'Livestock Ownership' and carried the proportion that all three criteria accounted for. The other was 'Dependency on Others' combined with 'Dependency on Children'.

The next step was to select the final criteria or index items. As with any measure or scale, items are selected based on, among other things, the feasibility of transforming them into successful items for the questionnaire. That is, that a question can be constructed with appropriate response categories, and with consideration to the likelihood of accurate responses from people interviewed (Streiner and Norman, 1986). Certain criteria were rejected, and for these districts it is those criteria that are underlined in tables 2.1 and 2.2. This decision was based on several considerations.

Firstly, the literature on measurement issues discussed in Section 1.4 pointed to severe difficulties in using occupation and income in partially-subsistent

economies. Secondly, 'Size of Farm' was found to vary dramatically over Hai district with highlands of Kilimanjaro producing higher yields per acre than lower lands, and more pressure for land in the highlands. Thirdly, 'Age' was mentioned frequently in both districts, but the researchers perceived this to be a descriptor of the household head in many circumstances and not always a criteria for allocation of socio-economic category or group. As it was impossible to determine in each case, it was removed from the list of eligible criteria. Finally, some criteria could not be operationalised satisfactorily. 'Work Ethic', although important in Hai district, was not considered an item that could practically be measured on survey. And, 'Health status' is a complex variable, and not feasible where it is necessary to develop a simple question line to elicit accurate information.

Allocating Weights to the Index Items

The proportion (based on frequency of use during ranking sessions) of each criterion was used as the basis for weighting. This decision was based on a proposition that criteria that are used more frequently than others are more important indicators of SES to the respondent groups. This study is only concerned with techniques that can be carried out by non-specialists with limited statistical or even mathematical skills. More orthodox techniques employed in index item weighting such as factor analysis cannot be employed by district level workers without post graduate qualifications. Although the proportions during ranking were the basis for weighting, the focus group data was used to confirm its importance and in some cases to increase the allocated weighting. The objective was to develop an index that was likely to perform well, so judgement is required for the successful marriage of the data from both sources. Tables 2.3 and 2.4 show the final indices selected for Morogoro Rural and Hai Districts. It is similar to tables 2.1 and 2.2, however these tables show the final items selected, and how the weighting was adjusted using the focus group data. The first column lists the criteria which reached the desired frequency of use (a proportion of 0.015 or more). The second column shows the weight allocated. Where the focus group data

was used to adjust the proportions, the items have been underlined in the tables and the adjustment are discussed below each table.

Table 2.3 Final District-Wide Index with Allocated Weightings, Morogoro

Item	Weight
Crop yield	0.115
Dependency on casual labour	0.060
Ability to employ casual labour	0.044
Dependence on others	0.043
Roofing material	0.042
Type of crops	0.040
Food security	0.040
Ability to assist others	0.040
Physical status (ability to work)	0.035
Condition of house	0.027
Variety of sources of income	0.025
Assistance received from others	0.023
Tenure of house	0.022
Employee (salaried income)	0.016
Marital status	0.016

No additional items have been included from the focus group discussions, however 'Ability to assist others' has had its weighting increased from 0.015 to 0.040 to reflect the relative importance during the discussions where it was mentioned in eight of the ten respondent groups in the two sites.

Table 2.4 Final District-Wide Index with Allocated Weightings, Hai

Item	Weight
Type of crops	0.126
Tenure of house	0.067
Tenure of farm land	0.052
Dependency on casual labour	0.052
Food security	0.050
Condition of farm	0.050
Number and type of livestock	0.036
Wall material of house	0.029
Assistance received from others	0.027
Educational status (of children)	0.023
Condition of the house	0.019
Variety of sources of income	0.019
Dependence on others	0.018
Physical status (ability to work)	0.018
Roofing material	0.016
Employee (salaried work)	0.016

'Food security' and 'Condition of farm' were added to the criteria list from the ranking session as a result of their importance during five of the focus groups. They have been allocated a weighting of 0.050.

Designing the Questions and Response Categories for the Questionnaire

The next step was to develop questions in Kiswahili that would successfully capture the item's intention as given during the wealth ranking and focus group discussions. In some circumstances, such as 'Roofing Material', this was relatively straightforward, but for other items such as 'Dependence on Others' and 'Assistance from Others' the actual intention of these (as perceived by the respondents during the wealth ranking) proved far more challenging to preserve. Appendix 2.7 shows the final English translation of the questionnaire for Hai District.

Question design was carried out in consultation with all team members. Questions were first written in English, then translated into in Kiswahili. A 'back translation' into English was carried out by a translator unfamiliar with the study to ensure the intent was preserved.

Next, response categories for each question item were required. For example, for the item 'Food Security' it was necessary to establish what the level of food security would be for the different socio-economic groups or categories within the district. This can vary greatly from district to district, and these response categories are an important component of the location-specific nature of the index. The focus groups were used to generate this data, specifically, the second series of questions asking respondent groups to describe in detail certain important criteria they cited most frequently. Appendix 2.8 shows an example of the data generated in one respondent group session to the question asking how many months food security per year was achieved for the different socio-economic groups.

Questionnaire items are labelled in the Appendix 2.7 as either 'Community Generated' and 'Traditional' to assist the reader distinguish between items for each

index. The questionnaire was developed to ensure logic of flow to the respondent, and therefore items were not presented according to their respective indices.

Scoring for Response Categories

The final index sought to categorise people into three socio-economic groups: rich and comfortable; poor; and very poor (SEG 1-3). A score of 1, 2 or 3 was allocated to each response category to reflect these socio-economic groupings.

Not all question items should necessarily carry a possible score of 1, 2 and 3. Some items such as 'Dependency on Casual Farm Labour' are most useful in identifying the poorer members of the community. Just because a household is *not* dependent on casual labour does not mean it should be scored as 1, or wealthy. However, a dependence on casual labour for most of the year does indicate economic hardship. For this item, only scores of 2 and 3 were appropriate. This reduces the likelihood of overestimation of SES. Conversely, some items, such as 'Employee in the household' are useful in identifying the better-off households as it implies a regular income which is very important in some locations. But, not having an employee in the household cannot necessarily be interpreted as being poor. In this case only scores of 1 and 2 are possible. This reduces underestimation of SES.

Table 2.5 shows the possible scores for each item. An 'X' indicates that a score is possible for that particular SEG score, 1 through 3. A blank area indicates that no score is possible for that SEG score for that item. It is also important to note that the indices for the two districts have items with a different emphasis on identifying the richer and poorer households. For example, in Morogoro, 11 of the 15 items allow a possible score of 3. In Hai, 14 of the 16 items allow a possible score of 3. An equal number of items that can identify the well-off, and very poor households would be ideal, but as the purpose was to develop indices reflecting a community construct using only community generated items, it was not appropriate to re-select items on this basis.

For both districts there were more items that could identify SEG3 than SEG1, however this was considered acceptable as the usual intention is to identify the very poor households with greater accuracy, and with less interest in the discriminatory power between rich and comfortable households and the middle category.

Table 2.5 Possible Scores SEG1 (1); SEG2 (2); SEG3 (3) for Index Items in Morogoro Rural and Hai Districts

Morogoro Rural				**Hai**			
Index Item	**1**	**2**	**3**	**Index Item**	**1**	**2**	**3**
Employee in the Household	X	X		Employee in Household	X	X	
Variety of Sources of Income	X	X		Variety of Sources of Income	X	X	
Dependence on Casual Labour		X	X	Dependence on Casual Labour		X	X
Ability to Employ Cas.Lab.	X	X		Tenure of House		X	X
Tenure of House		X	X	Roofing Material	X	X	X
Roofing Material	X	X	X	Wall Material	X	X	X
Condition of House	X	X	X	Condition of House	X	X	X
Type of Crops	X	X	X	Type of Crops	X	X	X
Crop Yield	X	X	X	Condition of Farm		X	X
Food Security		X	X	Land Tenure		X	X
Marital Status		X	X	Food Security		X	X
Physical Status		X	X	Physical Status		X	X
Dependence on Others		X	X	Dependence on Others		X	X
Assistance Received fr Others	X	X	X	Assistance Received fr Others	X	X	X
Ability to Assist Others	X	X		Educational Status (children)	X	X	X
				Livestock Ownership	X	X	X

In Hai District Livestock Ownership was identified by the community respondent groups as an important indicator. As this is a sub-scale (requiring several variables to be measured), and is also found on the traditional index, details of how this was constructed are discussed in the section on constructing the traditional index. Appendix 2.7 (Hai questionnaire) shows the scores allocated to each response category on the right hand side of the questionnaire. These were added to the questionnaire for explanatory purposes only to assist the reader and were not found on the actual questionnaire used. They were allocated after data collection, and interviewers were only required to mark in the box the appropriate response category.

Calculating the Final, Summated Score

Where respondents answered 'don't know' for a particular item, an average of all other items was calculated and allocated to that item (Streiner and Norman, 1986). There were no non-responses in any of the survey sites. As each question item is allocated a score of 1, 2 or 3, these are summed to calculate a final score for the index. Where weighting is applied in the analysis, these items' scores are multiplied by the allocated weights, and then summed for all items.

In order to use the final score to allocate a socio-economic group to the household, cut-off points in the list of scores are required. Deciding where to place the cut-off points for each socio-economic group would be dependent on the purposes of the index. I have proposed and applied two alternative cut-off points in this study which will be described in detail following the next section on construction of the traditional index.

The Traditional Index

Item Selection, Weighting and Response Categories

Under usual circumstances, researchers select a number of indicators of SES that seem appropriate for their locations. In addition, the literature shows that health personnel also decide somewhat arbitrarily what their indicators of SES are when developing exemption schemes, targeting vulnerable households or carrying out health systems research. I aimed to duplicate this approach using the items developed by other research groups. I selected 12 indicators from two well respected research groups currently undertaking research in Tanzania, The Economic Research Bureau of the University of Dar es Salaam, the Adult Morbidity and Mortality Project supported by the University of Newcastle (UK) and the British Overseas Development Administration. I also included a small number from the 1988 National Census, and the World Bank Tanzania Poverty Profile (1993b).

As weighted and non-weighted community generated indices were constructed, a weighting scheme was also applied to traditional indices. Items were weighted according to the research team's best knowledge of the research locations, which is in line with usual practice. For example, educational status is known to be uniformly low in Tanzania, and less important than, say livestock ownership in many areas. The final index and allocated weights are shown on table 2.6

Table 2.6 Traditional Index with Allocated Weightings

Index Item	**Weighting (when applied)**
1. Livestock Ownership	0.060M/0.100H*
2. Density of Living	0.015
3. Material Possessions	0.100
4. Tenure of House	0.040
5. Roofing Material	0.060
6. Wall Material	0.080
7. Access to Drinking Water	0.015
8. Toilet Facilities	0.020
9. Cooking Fuel	0.020
10. Size of Farm (in acres)	0.150
11. Educational Status of HH Head	0.040
12. Literacy Status of HH Head	0.020

* 0.060 for Morogoro District, 0.100 for Hai District

Response categories were in accordance with those used by the 'source' research groups and can be seen on the questionnaire in Appendix 2.7. Traditional items are marked and clearly distinguished from Community Generated items. In most cases, response category scores carried a possible allocation of 1, 2 and 3, which is usual practice with items currently in use. Exceptions to this were 'Tenure of house' as ownership of a house in these communities cannot imply the highest socio-economic status. In this case, the response category of inherited house was allocated a score of 2 rather than 3 as in the community generated index because traditional indices did not distinguish between outright ownership and inheritance of homes. During the wealth ranking and focus group sessions, respondents considered owning an inherited home as a sign of dependence or lack of ability to construct a home during a parent's lifetime. With 'Literacy Status' it is generally accepted that being literate does not imply a higher socio-economic status. During

the Nyerere administration most people received sufficient education to read and write. However, illiteracy is likely to imply quite a low socio-economic status.

Sub-Scale Construction

Three sub-scales were required to complete the items for the traditional index. Livestock ownership, material possessions index, and a density of living index. Sub-scales are uni-dimensional index items that are made up of a number of variables. These will be described below.

> *Livestock Ownership*

Scores are applied to different types of livestock and a total livestock score is obtained. According to Collier's *et al* (1986) Livestock Index for Tanzania, scores are allocated to the varieties of livestock as shown on table 2.7

Table 2.7 Collier's Livestock Index

Type of Livestock	**Score**
Bulls (local)	100
Bulls (improved)	159
Oxen	117
Cows (local)	90
Cows (improved)	251
Sheep	14
Goats	16
Pigs	37
Chickens	3
Ducks and Geese	3
Donkeys	36

Each household is asked how many of the listed livestock they hold. This number is multiplied by the weight or score shown on table 2.7. A summated score was then calculated. This resulted in a range of scores that required categorisation into three socio-economic groups. Collier does not suggest how this can be achieved, nor is it explained in the World Bank Poverty Profile Report how they determined cut-off points, so I chose to set cut-off points using the World Bank mean livestock

ownership scores for socio-economic groups as a guideline. These means, however, were quite low compared to our sample of households. I also chose to apply a different weighting scheme in the two districts for livestock ownership as there is very little livestock keeping in the Morogoro Rural District (see table 2.6). Again this was done to maximise performance where items were known to have less utility in some areas. This was, however, the only weighting adjustment made.

> *Ownership of Material Possessions*

A list of material possessions was constructed that would reflect the reality of the areas in which the research was carried out. These reflected the typical choices of important material possessions in Tanzania. After the list was constructed, an estimated value in Tanzanian shillings was put on each item. This score was then converted into number of months income per household for each item. The World Bank Poverty Profile (1993) estimated that the 1991 per capita income was approximately Tsh 55,000. This roughly translates into a monthly household income of Tsh 28,875 with an average household size of 6.3. Scores for material possessions can be found on table 2.8.

Table 2.8 Material Possession Index

Material Possessions	**Weight or Score**
Car	167.0
Motorcycle	13.3
Refrigerator	11.6
Television	8.3
Sewing Machine	3.3
Hardwood Wardrobe	2.6
Covered Sofa Set	1.6
Modern bed with Mattress	1.3
Bicycle	1.0
Softwood Wardrobe	1.0
Uncovered Sofa Set	0.8
Small Radio	0.2
String Bed	0.1

This also resulted in a score which needs to have cut-off points applied. Cut-off points were chosen where there were large breaks in the scores, and where

material asset ownership is likely to be for the rich and very poor. For example, we would not expect the very poor to have accumulated many months of material assets.

Often material possession indices are simply a single point score for each item possessed by the household. I have used the approximate cost of these items to capture the differences between the ability to purchase disparate items such as a car and a small radio, and weight them accordingly to improve performance.

> *Density of Living*

This was calculated by dividing the number of people living in the household by the number of rooms, excluding the kitchen and bathroom, if any. Cut-off points for these scores were chosen as:

Rich or Comfortable	=	< 1 person per room
Poor	=	1-2.99 persons per room
Very Poor	=	3 or more persons per room

Transformations of All Study Indices

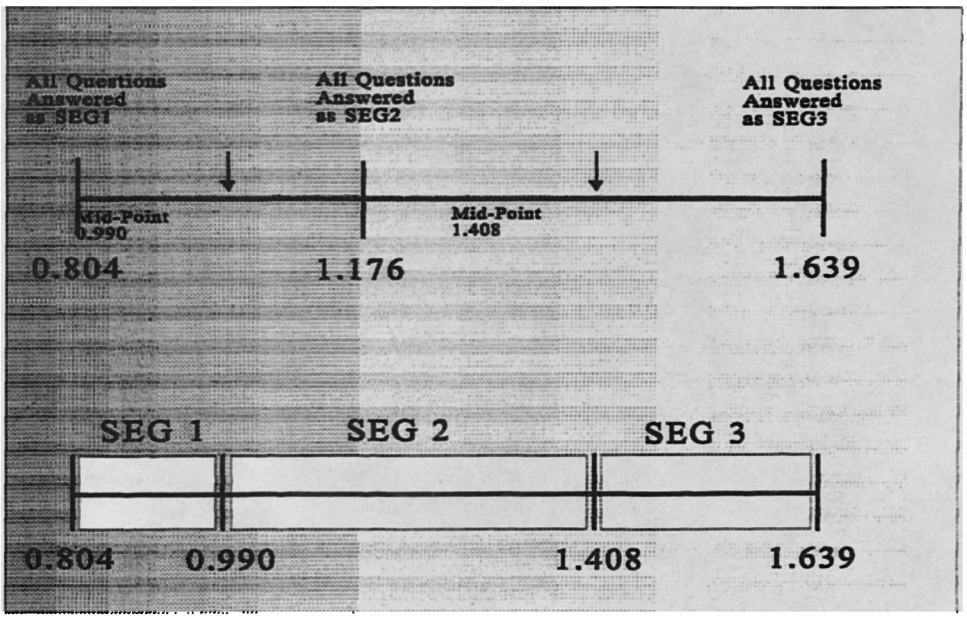
The community generated and traditional indices were subjected to two transformations, resulting in eight study indices. Firstly, indices were weighted and unweighted; and each index was subjected to two different approaches to determining cut-off points for the final summated scores. The indices were required to classify households into one of three socio-economic categories or groups (SEG 1-3) with the wealthiest household being allocated to SEG1.

The classic approach to determining cut-off points is to take the full range of scores and divide by the number of categories required, in this case three. As the survey carried out in this study was not a population based survey, it is not possible to take this approach. Two different cut-off point approaches were taken. The first approach is referred to as the 'mid-point' approach, which could be applied when the whole data set is not known. The possible scores for each item were calculated as if the respondent were answering all questions as SEG1, then all questions as

SEG2 and finally, SEG3. This resulted in a range of *possible* scores with a 'central' score for SEG2.

For example, Figure 2.3 shows the Weighted Community Index for Morogoro Rural District. The *possible* range of scores was 0.0804 to 1.639. A score of 0.0804 would be where a household received the lowest possible score for all questions. The lower the score, the wealthier the household. A score of 1.639 would be the results of receiving the highest possible score for all questions. The final score was calculated for SEG2 (to all questions) as 1.176. Next, the *midpoint* between these figures was identified. The following figure should clarify this more clearly.

Figure 2.3 *Identifying the Mid-Point Cut-Offs for Weighted Community Generated Index, Morogoro Rural District*

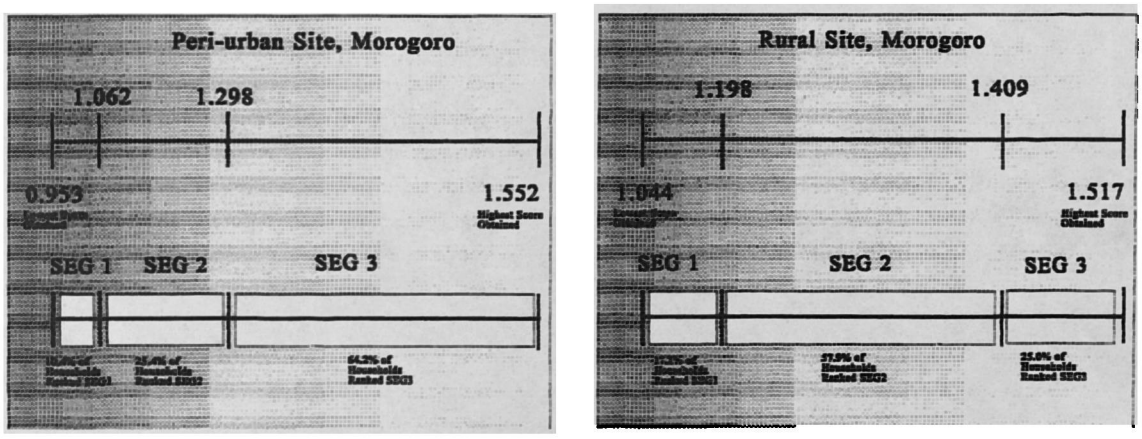


The second, 'proportional' approach used the data from verification wealth ranking sessions from the same survey households. The verification session resulted in classification of the study households into three socio-economic groups. The proportion of households in each of these groups was determined. These

proportions were used to categorise the list of household scores from the index. As households were not randomly selected and we know that, say, 30% of the households were not rich, 30% poor and 30% very poor, this approach allowed the index scores to be categorised in a way that was more likely to reflect the actual groupings for the households selected. The approach was used to allow a more realistic assessment of the study indices' performance. It is not recommended as a method of determining cut-off points for general index construction.

Figure 2.4 Proportional Cut-Off Points for Rural Site in Morogoro

Figure (a) *Figure (b)*



Figures 2.4 (a) and (b) shows an example of how the proportional cut-off points were determined for the same index discussed above for Morogoro Rural District. In the peri-urban site, the final community rank allocated to households during the verification wealth ranking session resulted in 10% of households being categorised as rich or comfortable. The poor households accounted for 25% of households ranked, and 64% were very poor. Taking the total list of index scores attained by each of the households on survey, the top 10% were categorised as rich or comfortable, 25% were categorised as poor and the remaining 64% were categorised as very poor. The same procedure was followed in the rural site where

17% of households were rich or comfortable, 57% were poor, and 25% of households were ranked as very poor.

The Final Indices

The study was concerned with validating community generated and traditional indices, however, because of the two main transformations (weighting and cut-off points), there were eight indices validated. The eight indices are shown on table 2.9.

Table 2.9 Description of the Eight Study Indices Validated

		Transformation	
Label	**Type of Index**	**Cut-Off Point**	**Weight**
COM 1	Community Generated	Proportional	Yes
COM 2	Community Generated	Mid-Point	Yes
COM 3	Community Generated	Proportional	No
COM 4	Community Generated	Mid-Point	No
TRAD 1	Traditional	Proportional	Yes
TRAD 2	Traditional	Mid-Point	Yes
TRAD 3	Traditional	Proportional	No
TRAD 4	Traditional	Mid-Point	No

Control Measure of SES

A final measure, referred to as SELF RANK, was included in the study for validation. This was included to act as a 'control' measure. During the household survey, a respondent was asked to categorise the household into one of three socio-economic groups. Self assessments are generally thought to be poor measures of SES (e.g. Atkinson, 1988), and in the Tanzanian context, people often perceive benefits to accrue from false reporting. This measure was expected to perform the least well of the study indices, and was intended to act as a measure with poor criterion-related or construct validity for comparative purposes.

C. Methods and Procedures for Validation Assessments

Methods: ***Household Survey and Wealth Ranking***

The two approaches used for validation assessments were the sensitivity and specificity tests assessing criterion-relatedness, and the multitrait, multimethod approach assessing construct validity. For the sensitivity and specificity tests, the results of a household survey using the study indices were compared with the criterion - community allocated rank using the wealth ranking technique. For the multitrait, multimethod approach, two traits were measured (SES and skin tone) by two methods - the household survey and the wealth ranking technique. Constructing a correlation matrix as described in Appendix 2.4, convergent and discriminant construct validity, and method variance is assessed. Using these methods, eighty households were surveyed and subsequently wealth ranked in a peri-urban and rural site in each of the final two districts, Morogoro Rural and Hai Districts.

Data Collection Procedures

Household Survey

> *Selection and Training of Enumerators*

Six enumerators were chosen from each district to work in both the rural and peri-urban sites. A second, well experienced research associate was recruited from the faculty of the Medical School at the University of Dar es Salaam to supervise the survey activities and enumerators. The same research assistant (District Nursing Officer) employed in phase one was also contracted to work as a survey supervisor and again as the wealth ranking observer. Six survey enumerators were identified by the District Medical Officer in each district, and were all either from the District Health Management Team, or were local school teachers.

In each district, a four day training was conducted in Kiswahili and English by the second research associate. Project objectives, application of findings, survey activities and procedures were discussed. There was a special focus on

interviewing techniques and adequate opportunity for practical experience with role plays designed to deal with the most common problems during interviewing. This training also included a review of the questionnaire as already developed to double check a number of design issues as outlined by Abramson (1990). These included appropriateness of the sequence of questioning, suitability and perceived success of the introduction, quality of questionnaire formatting for enumerator ease of use, clarity, sensitivity, and likelihood that the respondent would be able to answer the question. In addition, issues related to question wording (in Kiswahili) as reviewed in Moser and Kalton (1971) were addressed. These included questions being sufficiently specific and in accordance with the stated intention of the question, simplicity of language, lack of ambiguity, clarity of wording, level of potential embarrassment or sensitivity to the question.

The final training day involved a pre-test of the questionnaire in a village near to the district headquarters and not included in the survey proper. Enumerators were expected to conduct at least six interviews, and were observed by the supervisors conducting an entire interview from introduction to exit. At the end of this session, the group met again to discuss problems with logistics, technique, questionnaires or community response. All questionnaires were reviewed by myself and problems were clarified, and the questionnaire adjusted.

> *Supervision of Survey*

For the six enumerators, I allocated two supervisors for observing interviews throughout the day to maintain quality, while I collected and reviewed all questionnaires for completeness, correctness and inconsistencies in responses to certain questions. Daily de-brief sessions were held after data collection at the site to discuss any problems. More urgent errors were dealt with immediately.

> *The Questionnaire*

The questionnaire contained all items that were developed for the community generated and tradition indices. In addition questions relating to self-perception of

socio-economic status (SELF RANK) and skin tone were found at the end of the questionnaire. The Hai District questionnaire can be found in Appendix 2.7 (note: response category scores did not appear on the questionnaire when administered).

Wealth Ranking Activity

For this phase wealth ranking was carried in the same manner as in phase one, however this time, criteria were not generated and respondents were asked to allocate a rank to each of eighty study households. In this phase, only three socio-economic categories were given: rich or comfortable (SEG1), poor (SEG2) and very poor (SEG3). Respondents were encouraged to discuss these ranks amongst themselves, but the criteria they used during the discussion was not recorded or analysed. This ranking without criteria generation is referred to as simple wealth ranking. The same respondent groups were used as in phase one, wealthier men, wealthier women, poorer men, poorer women and community leaders. Households for ranking and survey were selected in the same manner as phase one, as were the respondent groups for the ranking activities. A final, sixth 'verification' session was also carried as in the previous phase. The final rank allocated to each household during this session became the variable 'Community Rank' used in the validation assessments.

As the multitrait, multimethod approach required that two traits were measured using both methods, respondents were also asked to rank the same eighty households according to the skin tone of the head of the household. Skin tone categories were defined using the Kiswahili words that are used in normal conversation when describing an individual's appearance. The ranking procedure for skin tone was carried out simultaneously with the wealth ranking. That is, after each household was allocated to a socio-economic group, the same household was then ranked according to skin tone. The second rank of skin tone was recorded on a separate sheet, while the household cards were placed on the table in front of respondents still in piles according to their socio-economic rank. The same activity was carried out during the verification session, and this final rank was used as the

community rank for skin tone in the multitrait, multimethod validation assessments.

Data Management and Preparation for Analysis

Household Survey Data

At the end of each day of data collection, I collected and re-checked all the questionnaires and personally entered all data into SPSS for Windows Version 6. Four separate files were kept for each of the survey sites. Enumerators were asked to revisit households where discrepancies or contradictions in responses were noted during data entry that were missed by supervisors during the day. Questionnaires were then transported back to Dar es Salaam where they are stored safely to ensure confidentiality. Data was entered twice and discrepancies checked manually and corrected against original survey forms. Full lists of all variables were printed and all data sets were manually cleaned.

Wealth Ranking Data

At the end of each day of data collection, the final community rank for both socio-economic status and skin colour were entered into SPSS for Windows Version 6, along with household identification. These were entered into the same files as the survey data.

Quality Assurance

Quality of the survey data was enhanced by the supervision measures described above. In addition, data was double entered into statistics programmes and manually checked for discrepancies. I was directly involved in all aspects of the survey, from training and pilot testing, to field supervision and all data entry. I am able to report confidently on quality issues from full participation or observation of all aspects of data collection and management.

Chapter Three

Research Findings

3.1 Performance and Feasibility of Wealth Ranking

This section presents the results from the first study concerning the performance and feasibility of the wealth ranking technique. Results are presented in accordance with the research questions posed in Chapter Two. These are related to the reliability of the simple ranking procedure (section 3.1.1), the criteria generated within and across the ten villages (section 3.1.2), and the skills and resources required to use the wealth ranking technique (section 3.1.3).

3.1.1 Simple Wealth Ranking

Reliability of Simple Ranking

1. Do different community respondent groups rank their neighbours' wealth status reliably?

Overall Reliability

Overall reliability refers to the reliability of the five respondent groups or judges, ranking all households into four socio-economic groups. The alternative measures of reliability will be outlined with each analysis in this section.

Simple wealth ranking was highly reliable. The following data will support the hypothesis that community groups do reliably rank their neighbours' socio-economic status.

The kappa statistic in table 3.1 represents the agreement between the five respondent groups (or judges) and is calculated over all the four socio-economic categories or groups together. The appropriate interpretation of the strength of agreement as suggested by Landis and Koch (1977) is also shown. Sites are arranged by region, and the relative socio-economic status of the district is indicated in the first column. Households are removed from the kappa analysis if they were not recognised or ranked by any one of the five respondent groups. The number of households included in the analysis is shown on table 3.1 with each kappa statistic.

Table 3.1 Reliability: Overall Strength of Agreement in all Sites

District	Peri-urban Sites			Rural Sites		
	Kappa	**p-value**	**Strength of Agreement (S.O.A)**	**Kappa**	**p-value**	**Strength of Agreement (S.O.A)**
Kisarawe (Middle SES)	(N = 40) 0.594	0.0000	Moderate	(N = 44) 0.604	0.0000	Substantial
Rufiji (Lower SES)	(N = 47) 0.432	0.0000	Moderate	(N = 49) 0.424	0.0000	Moderate
Mafia (Higher SES)	(N = 50) 0.812	0.0000	Almost Perfect	(N = 46) 0.721	0.0000	Substantial
Morogoro (Middle SES)	(N = 34) 0.618	0.0000	Substantial	(N = 44) 0.635	0.0000	Substantial
Hai (Higher SES)	(N = 20) 0.924	0.0000	Almost Perfect	(N = 50) 0.699	0.0000	Substantial

Overall rankings achieved a kappa over 0.600 in most sites. Exceptions were Rufiji District (both sites) with kappas of 0.432 and 0.424, and Kisarawe peri-urban with a kappa of 0.594. This should be interpreted as a high level of agreement for the following reasons. Firstly, the construct of SES could be considered an ill-defined notion when comparing kappa statistics with those from medical diagnostic tests. Secondly, respondents were not given a definition of socio-economic status; and finally, respondents groups were not instructed how to determine SES before the exercise.

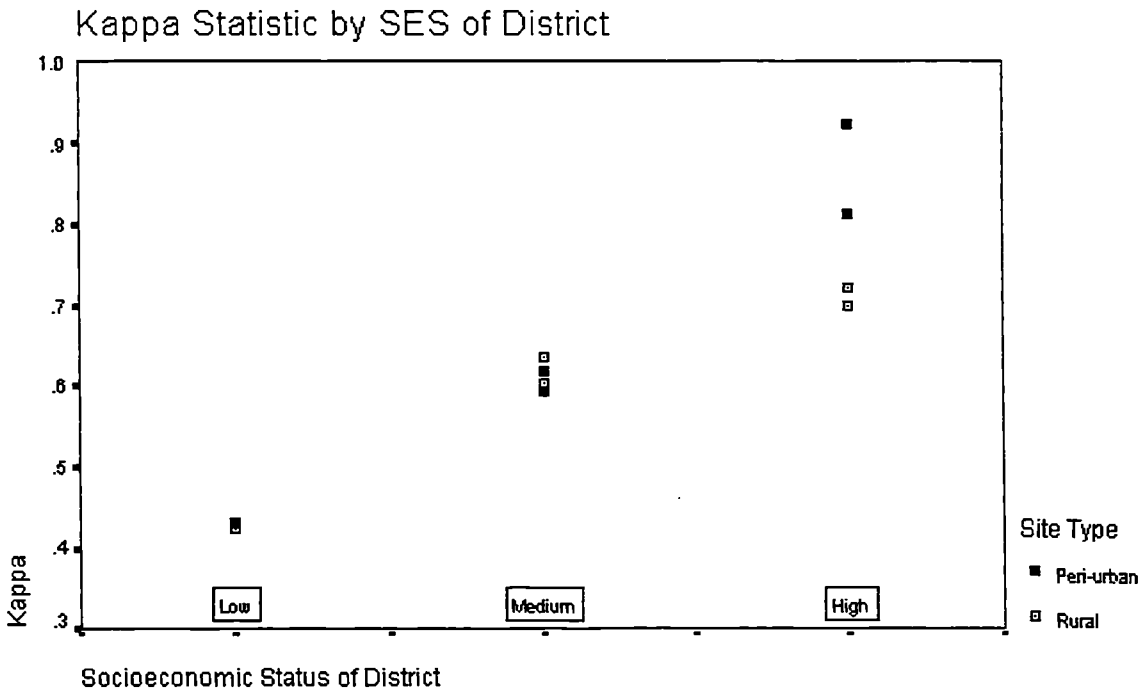
Although it is not acceptable to compare kappa statistics across different sites by calculating a mean kappa, it can be seen that the peri-urban sites achieved slightly higher scores in three of the five districts. This is in contrast to the common assertion that wealth ranking is less reliable in peri-urban areas than rural areas.

Reliability of Ranking in Different Locations

It is useful to examine wealth ranking reliability in locations with different levels of socio-economic status. Due to the small number of sites, a simple least squares linear regression would not be an appropriate test. A scatter plot was constructed, however, using the socio-economic status of the district as the independent

variable, and the Kappa as the dependent variable. A visual examination of the evidence suggests that there may be a relationship between reliability and the socio-economic status of the location.

Figure 3.1 Scatter Plot for Kappa Statistics by District Socio-economic Status



Reliability of Ranking by Respondent Groups

The next question is to determine if particular respondent groups are more reliable at ranking their neighbours than other groups. The following table summarises the data shown in Appendix 3.1 and ranks respondent groups from best to worse for each site. A rank of '1' means that this respondent group in the corresponding site on the table was the most reliable respondent group at ranking their neighbours' SES. A rank of 5 indicates the least reliable respondent group.

Table 3.2 Respondent Group Reliability Ranked from Best to Worst by Site*

Site	Leaders	Wealthier Men	Poorer Men	Wealthier Women	Poorer Women
Kisarawe, peri-urban	3	2	5	1	4
Kisarawe, rural	5	2	3	1	4
Rufiji, peri-urban	2	1	3	5	4
Rufiji, rural	1	3	2	4	5
Mafia Is., peri-urban	4	1	5	3	2
Mafia Is., rural	1	5	2	4	3
Morogoro, peri-urban	4	5	2	1	3
Morogoro, rural	4	3	5	2	1
Hai, peri-urban	4	3	2	5	1
Hai, rural	4	1	5	2	3
Total	**3.2**	**2.6**	**3.4**	**2.8**	**3.0**

* 1 = most reliable; 5 = least reliable

Table 3.2 shows, with mean rankings by respondent groups, Wealthier Men were most reliable, followed by Wealthier Women, Poorer women, Leaders and worst, Poorer Men. However, an examination of the table shows little consistency in performance. This suggests that specific groups were not consistently superior in their ability to reliably rank households. Note the comparatively poorer performance of community leaders, commonly assumed to be well placed to identify the socio-economic status of their own community members. They ranked worst or second worst in six of the ten sites.

Reliability of Different Socio-economic Groups

The measurement of the reliability of different categories, or socio-economic groupings, was conducted to assess the level of agreement about the poorest households. Table 3.3 shows the kappa statistic by socio-economic category for each of the ten sites. As described in the methods chapter it is not meaningful to calculate mean kappa statistics across sites. Shading has been used instead to highlight the top two scores for each site. The darker shading represents the highest score, the lighter represents the second highest score.

Table 3.3 Reliability of Ranking Socio-economic Groups by Site

Site	SEG 1 Rich	SEG2 Comfortable	SEG 3 Poor	SEG 4 Very Poor
Kisarawe Peri-Urban	0.571	0.127	0.126	0.427
Kisarawe Rural	1.000	0.530	0.166	0.333
Rufiji Peri-Urban	0.267	0.531	0.069	0.184
Rufiji Rural	None	0.395	0.235	0.335
Mafia Peri-Urban	0.581	0.488	0.526	0.662
Mafia Rural	0.492	0.423	0.092	0.563
Morogoro Peri-Urban	0.629	0.304	0.062	0.280
Morogoro Rural	0.560	0.279	0.203	0.453
Hai Peri-Urban	0.875	0.778	0.577	0.709
Hai Rural	0.537	0.326	0.194	0.481

In line with expectations, the wealthiest socio-economic group achieved highest kappa statistics in six of the nine sites where all four categories were identified. In Rufiji peri-urban site, no SEG1 (Rich) households were identified, but in this instance, SEG 2 (Comfortable) achieved the highest kappa. Overall, the richest households were well identified. The very poor households (SEG4) were most reliably identified in two sites only, but scored the second highest kappa statistic in four other sites.

Comfortable households were also well identified, but SEG3 (poor) did not yield high scores in any of the ten sites. Although the poorest households were not as well identified as the wealthiest ones, the poorest category (SEG4) performed next best overall. We would expect that the wealthiest households would be the easiest to identify due to their usual prominence in a community.

Effectiveness of the Process of Selecting Respondent Group Participants

2. *Are different community respondent groups who participate in the wealth ranking sessions (wealthier men, poorer men, wealthier women, poorer women) actually composed of respondents of the correct socio-economic status?*

Of the 151 respondents that were from the study households, 78.8% were correctly identified as poor or better-off and participated in the appropriate respondent groups.

Table 3.4 summarises the socio-economic status of respondent group participants across all sites, according to the rank allocated to their household during the final verification wealth ranking session.

Table 3.4 Respondent Groups Participants' Socio-economic Status: All Sites

Respondent Group	Total All Sites	Number Correct	Proportion Correct
Wealthier Men	37	27	73%
Poorer Men	43	42	98%
Wealthier Women	33	12	36%
Poorer Women	38	38	100%
Total	**151**	**119**	**78.8%**

Overall, poorer respondent groups had more correctly placed participants, while wealthier groups had participants with mixed SES. This was less marked in the wealthier male groups. Wealthier female respondent groups were in fact largely composed of poorer women.

Respondent Group Bias in Categorising Household SES

3. *Do different respondent groups (based on socio-economic status, sex and leadership) who participate in the wealth ranking have a particular bias in the way they categorise households socio-economic status?*

There were mixed biases in over- and under-estimation of SES for households. However, there is no clear pattern in the direction of over- or

under-estimation of SES based on respondent group sex or socio-economic status.

Table 3.5 shows the total over, under and correct estimates of SES for each respondent group for all sites. The final row totals are not equal as some respondent groups did not recognise certain households, in which case households were not ranked. This will be addressed in the following section. Appendix 3.2. shows for each site the number of households where socio-economic status was either over-estimated or under-estimated for all respondent groups. Once again, the final community rank from the sixth verification session was used as the standard by which biases were judged.

Table 3.5 Ranking Bias for all Respondent Groups: All Sites

Direction of Bias	**Leaders**		**Wealthier Men**		**Poorer Men**		**Wealthier Women**		**Poorer Women**	
	No.	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Over Estimated SES	56	11.5	76	15.6	101	22.0	73	16.0	64	14.0
Correct	331	68.0	343	70.6	295	64.3	308	67.5	294	64.5
Under Estimated SES	100	20.5	67	13.8	63	13.7	75	16.5	98	21.5
Total	**487**	**100**	**486**	**100**	**459**	**100**	**456**	**100**	**456**	**100**

Overall, there was a narrow range of ranks that agreed with the verification session, from 64.3% for Poorer Men to 70.6% for Wealthier men. This also shows that misclassifications (according to the standard) accounted for almost a third of households ranked by all respondent groups. Appendix 3.2 shows that most of these misclassifications were within one socio-economic category from the standard.

Both the Leaders and the Poorer Women tended to under-estimate the SES of households, over all sites. An examination of Appendix 3.2 shows by site that Leaders underestimated SES in seven of the ten sites. For Poorer Women,

households were also underestimated in seven of the ten sites. Conversely, Poorer Men tended to overestimate household SES, but only in half of the sites. The remaining groups did not show a general bias in either direction. Both wealthier respondent groups tended to show a balance between over and under estimation, although for the wealthier men this was slight. The poorer respondent groups showed opposing direction of bias. Both male respondent groups tended to overestimate household SES, while only the poorer women showed a bias (toward underestimation). Appendix 3.2 shows that there is no clear pattern established between the type of group (based on socio-economic status or sex) and the direction of the bias.

Difficulties in Recognising Households for Ranking

4. *Do particular respondent groups (based on socio-economic status, sex and leadership) have difficulty in recognising households for ranking?*

   Female respondent groups had more difficulty in recognising all households for ranking than male groups. Leaders and Wealthier Men recognised most households.

The Leaders and Wealthier Men had far fewer households that they did not recognise (and rank) compared to the rest of the respondent groups. Across all sites, both female respondent groups did not recognise 32 households each of the 488 households ranked. The poorer male groups could not identify 29 of the total households. An examination of table 3.6 shows that for the poorer men, the high number of households not recognised was concentrated in a single site - the peri-urban site in Hai District. This was also the case for the two female groups; however, for the female groups there were six other sites each where households were not known. Table 3.6 provides a breakdown for all respondent groups, for all sites.

Table 3.6 Households not Recognised by Respondent Groups

Site	Leaders	Wealthier Men	Poorer Men	Wealthier Women	Poorer Women
Peri-urban					
Kisarawe	0	1	5	3	2
Rufiji	0	0	1	1	1
Mafia Island	0	0	0	0	0
Morogoro	0	0	4	7	8
Hai	1	1	18	11	13
Rural					
Kisarawe	0	0	1	3	5
Rufiji	0	0	0	0	0
Mafia Island	0	0	0	4	1
Morogoro	0	0	0	3	2
Hai	0	0	0	0	0
Total	**1**	**2**	**29**	**32**	**32**

Table 3.6 also shows that, according to common assumptions, peri-urban sites had more households that could not be recognised. However, Hai peri-urban accounted for the majority of unknown households. Looking at the remainder of the sites separately, for both the female groups, unknown households were evenly spread over rural and peri-urban sites, but for poorer men, there was only one rural household not known, and none for leaders and wealthier men. In these research sites, men in rural areas were better able to recognise households.

3.1.2 Criteria Used to Allocate Socio-economic Rank

Intra-Site Comparison of Criteria Used

5. *Do different community respondent groups within a single village location use similar criteria in ranking their neighbours' wealth status?*

   Different community respondent groups within a single location used similar criteria to rank their neighbours' wealth status.

The criteria used by each respondent group within a single site were aggregated and totalled as described in the methods section. Over the ten sites, a total of 29 criteria were generated, with a small number of criteria being allocated to the criterion labelled "other" on the list. Appendix 3.3 (a) - (j) shows for each site the

criteria generated, and what proportion of the total criteria generated it accounted for. The data is presented by respondent group. Table 3.7 shows the kappa statistics calculated on the agreement between respondent groups on the relative importance of criteria used to allocate a SES rank to households. Importance was determined by relative frequency of use.

Table 3.7 *Agreement between respondent groups on the importance (i.e. frequency of use) of criteria used to allocate a SES rank*

Region	District	Peri-urban Sites		Rural Sites	
		Kappa	S.O.A	Kappa	S.O.A
	Kisarawe	0.576* (p= 0.0000)	Moderate	0.636 (p= .0000)	Substantial
Coast	**Rufiji**	0.753 (p= 0.0000)	Substantial	0.717 (p=0.0000)	Substantial
	Mafia	0.744 (p= 0.0000)	Substantial	0.660 (p=0.0000)	Substantial
Morogoro	**Morogoro Rural**	0.601 (p= 0.0000)	Substantial	0.574 (p=0.0000)	Moderate
Hai	**Hai**	0.724 (p= 0.0000)	Substantial	0.680 (p=0.0000)	Substantial

* four respondent groups only

Table 3.7 shows substantial strength of agreement in eight of the ten study sites. Respondents were not given any definitions of wealth or poverty; they were free to cite any criteria they considered important to allocate a SES rank to the study households; and they could use these criteria as often or as infrequently as they chose. Although the open nature of this data collection tool should reduce our expectations for high kappa scores, in most sites a substantial level of agreement was achieved.

Peri-urban sites, overall, had higher Kappa scores; however, these differences were minimal and could not be interpreted as an indication of higher levels of agreement in peri-urban areas.

Cross Site and Cross District Comparison of Criteria Generated

6. *Do community respondent groups across a range of geographical locations use different criteria in ranking their neighbours wealth status?*

   Community respondent groups used different criteria across geographical locations, however these differences are not substantially greater than differences in criteria used within a single site.

The same technique is used to compare the criteria used across locations as in the previous question where criteria used by different respondent groups within a single site were compared. For this section I carried out two different types of cross site comparisons. Firstly, the aggregated district comparison where peri-urban and rural site criteria were aggregated together into a single district list of criteria. This results in a comparison of five districts. Secondly, I conducted a cross-site comparison where sites were compared without aggregating to the district level. This resulted in a comparison of the ten individual study sites. Secondary comparisons were made across peri-urban and rural sites separately; and the strength of agreement between single respondent groups across all sites was compared.

Cross District Comparison of Criteria Generated

As the final indices were constructed using both the rural and peri-urban criteria together for each district, a comparison of the district-wide criteria was carried out first. The data for aggregated criteria (rural and peri-urban sites) can be found in Appendix 3.4. It is presented as a proportion that each criterion accounted for the total criteria generated for each district.

The strength of agreement across the five districts, using kappa statistics was 0.550 ($p<0.00001$), or moderate. This finding does not support the assumption that community respondent groups use different criteria in different geographical locations. In contrast, the kappa for the level of agreement on the importance of each criteria (frequency of use) fell within the Moderate strength of agreement

which was only slightly below the range of kappa scores for the intra-site comparisons.

In order to explore the data and search for plausible explanations for these findings, several further analyses were carried out. These are found in the following two sections looking at cross-site and cross-respondent group comparisons.

Cross Site Comparison of Criteria Generated

Firstly, the data was analysed in its desegregated form. That is, a cross-site comparison was made with all ten sites. This analysis was performed to determine whether the aggregation of the criteria to the district level increased the similarity by merging the types of indicators identified in rural and peri-urban settings. The raw data of proportions cross-site can be found in Appendix 3.5. Across all ten sites this results in a kappa of 0.435 which should be interpreted as still within the range of a Moderate strength of agreement. It is important to note again here that when kappas were calculated for the cross-district comparison, there were only five judge groups, or districts. For the cross-site comparison we have ten judges, or individual sites. The nature of the kappa statistic is that the more judges that you compare, the higher the level of agreement across all judge groups. Therefore, despite this doubling of judges, we still maintain a drop in kappa score.

As the cross-site comparison did drop to a level lower than the intra-site comparison, the next analyses sought to explain whether the high aggregated district result was related to the type of site. That is, do rural (or peri-urban) community respondents tend to use similar criteria across geographical locations, and therefore it is the type of location rather than the geographical location in the country that will influence the criteria used to allocate socio-economic ranking to a household.

When only peri-urban sites were compared across the five districts, a kappa score of 0.448 ($p<0.00001$) was achieved, which is within the Moderate strength of agreement. The rural site comparison achieved a kappa of 0.481 ($p<0.00001$) or

Moderate strength of agreement. A magnitude of agreement similar to the cross-site comparisons. Again, as peri-urban and rural site comparisons were for five sites each, compared to the ten cross-site comparisons, there is some improvement in the strength of agreement within rural and peri-urban sites, but still this is not substantial and is still not as high as the intra-site comparisons.

Respondent Group Comparisons of Criteria Generated

This raises a further question as to whether it may be respondent group characteristics that determine the type of criteria used to rank the socio-economic status of households, rather than the location. For example, do community leaders, wealthier men or poorer women tend to identify the same types of criteria, no matter where they live?

The results of these kappa statistics are presented in table 3.8. Column two shows the results across all sites, then columns three and four for peri-urban and rural sites respectively. The cross-site kappas are shown for comparative purposes.

Table 3.8 Respondent Group Comparisons of Criteria Generated

Respondent Group	**All Sites (10) Column 2**		**Peri-urban Sites (5) Column 3**		**Rural (5) Column 4**	
	Kappa	**S.O.A**	**Kappa**	**S.O.A**	**Kappa**	**S.O.A**
Community Leaders	0.322 (p<0.0001)	Fair	0.381 (p<0.0001)	Fair	0.361 (p = 0.0000)	Fair
Wealthier Men	0.386 (p< 0.0001)	Fair	0.451 (p<0.0001)	Moderate	0.356 (p = 0.0000)	Fair
Poorer Men	0.366 (p< 0.0001)	Fair	0.345 (p<0.0001)	Fair	0.410 (p = 0.0000)	Moderate
Wealthier Women	0.298 (p< 0.0001)	Fair	0.231 (p=0.0001)	Fair	0.397 (p = 0.0000)	Fair
Poorer Women	0.321* (p< 0.0001)	Fair	0.332** (p<0.0001)	Fair	0.347 (p = 0.0000)	Fair
Cross Site Comparison	0.435 (p< 0.0001)	Moderate	0.448 (p<0.0001)	Moderate	0.481 (p = 0.0001)	Moderate

* 9 Sites only ** 4 Sites only

As can be seen from table 3.8 all except one cross-respondent group kappas were lower than the cross site Kappas. This suggests that, for these regions in Tanzania, it is more likely to be the location than the respondent groups that determines what criteria are most important.

3.1.3 Skills and Resources Required for Wealth Ranking

Skills Required for Wealth Ranking as Applied in this Study

7. *What skills are required to carry out wealth ranking as a method to develop location-specific indicators of SES?*

   Typical district or social sector workers would not be capable of handling all aspects of the sessions without substantial support during data collection activities for their first practical experience. Simple wealth ranking required fewer skills than criteria generation or the focus group discussions for index construction.

I would recommend the following areas to be considered if attempting to replicate this work at the district level or with smaller non-government organisations without experienced social scientists familiar with qualitative methods.

General Comments

This project employed an experienced social scientist as the facilitator of all sessions (simple ranking, criteria generation, and subsequent focus group discussions), however she was not experienced in qualitative methods. This facilitator underwent a well planned fourteen day training programme with complex role plays incorporated into the training design. During the course of the field work, a substantial amount of 'on-the-job' training was required which was achieved largely through instant feedback during the *verbatim* transcription of all sessions, and to a lesser extent, through the daily extended de-briefing sessions. I estimate that she was able to perform the sessions expertly and unassisted after we finished the third district (Mafia Island). This was after completion of six sites and 30 sessions. Therefore, I would expect that for the first experience, the target users of this method as defined by the study, would require a principle investigator who is able to provide on-going and daily support with structured on-the-job training. This is not to say that the training of peripheral workers is not feasible, but simply the recognition that substantial support is required for the exercise to be successful and workers adequately trained. I consider that the on-the-job training is where the

trainees truly gain skills, and that the classroom training is simply an exposure and orientation to the method. Conducting short workshops is inadequate if we are serious about developing people's skills to work successfully, independently, and train others in the future.

Planning Skills, and Knowledge of Research Sites

In terms of community entry and logistics, certain skills and local knowledge are required and not always recognised by planners. These are quite well covered in the literature covering PRA methods, however, there are specific skills and knowledge required for this work. The ability to plan the project and be ready to consider a broad range of issues related to logistics and community entry is essential. Some issues not anticipated by this research team are described below to indicate the types of problems even experienced researchers and community development workers can face.

Firstly, acquiring sufficient knowledge of local factors that may inhibit community openness in discussing sensitive issues, such as socio-economic status of their neighbours. This is obviously important, but can easily be missed if not looked for carefully. In this study we encountered problems such as local elections, and village loan scandals which were not carefully investigated or anticipated beforehand. These events had a deleterious effect on the success of a small number of sessions.

Local conditions for travel and accommodation need to be planned for well ahead. For example, we were not aware of local flooding of the Rufiji River until days before planned departure. This resulted in a severe delay of project activities, despite the fact that the project team included Tanzanian nationals with substantial knowledge of seasonal conditions. The point here is that we could have sought information on local conditions more systematically, rather than accidentally.

A final area not well anticipated in this study was the scheduling of national and local holidays. For example, in countries that celebrate Islamic holidays and festivals, these are usually determined by the moon. We had several public

holidays announced at midnight, the night before planned departure. These type of delays cannot be avoided, but could have been built into the general timetable even though the specific date was not yet known. At village level we were often not informed of scheduled weddings and funerals until the time of the planned sessions with respondent groups. Better planning with specific pre-prepared questions with local leaders could have avoided these lost days. By this I refer to the need to ask specifically about certain types of events, rather than asking about events in a general way. Village informants do not necessarily recognise what information is important to researchers. During such events other necessary project activities such as preliminary analyses could have been attended to, rather than waiting for respondents to arrive, or travelling to the villages a day or two before meetings or sessions could be held.

Simple Wealth Ranking and Criteria Generation

Simple wealth ranking, asking respondents to allocate a socio-economic group or category to each household, was clearly the most straightforward, and required few advanced skills. I would consider typical health and social sector workers capable of conducting the simple ranking sessions with few difficulties if well prepared. However, the application of simple ranking of households alone is only useful when trying to identify household socio-economic status in projects working in very small communities.

For criteria generation during these sessions, a few issues were noted. It was noted in most sessions with all respondent groups a tendency to repeat each other's criteria simply using different words. Also, once a respondent group had identified a few criteria, then all households within the site would be ranked according to those limited number of criteria. For example, once the housing materials and size of farm was used for the first few households, the group would tend to state the same criteria for every house. The session facilitator would need to bring the group back constantly to consider other criteria that they were actually using, without, at the same time, putting undue pressure on the group to invent criteria just to please the facilitator. Without a high level of expertise in session facilitation, this could

easily result in a limited construct of SES being reflected by the allocated rank, a very small number of criteria generated, or an index with items that do not truly represent the range or type of indicators that people actually use. This problem cannot be over-emphasised as it was present in almost all of the 60 sessions. If not anticipated and managed appropriately, this would limit the value of criteria generation. In the female respondent groups this was more apparent as more difficulty was encountered articulating criteria, than in other respondent groups.

The interest in this study was to identify indicators of socio-economic status that may be used at the district level for survey purposes. This required the conduct of focus group discussions to explore criteria generated during the wealth ranking in more depth.

Focus Group Discussions for Index Construction

Let me begin by stating that I do not consider that one needs to be an experienced social scientist to conduct effective focus group discussions. However, I think it is prudent to recognise the complexity of facilitating these sessions, and the need for high quality training, substantial support in the form of on-the-job training, and plenty of experience. Many people without tertiary education are highly skilled at group facilitation, and specifically people who work at community level, but these skills alone do not necessarily result in successful focus group discussions conducted for research purposes. Specific problems are described here, but I would anticipate a great deal more difficulties with a facilitator who had not been exposed to general principles of research, a number of complex role plays, and continued support during the exercises. I will not discuss general facilitation issues with focus group discussion as these are well documented in the literature (e.g. Krueger, 1994).

The most common problem was the ability of the facilitator to recognise important information given by respondents that was necessary information to pursue to ensure that the questions were well answered during the session. Equally difficult is the recognition that some information is less important and distracting the group from discussing the questions on the prepared discussion guide. I consider this

skill only comes from extensive discussion of the research objectives, giving trainees guidance on how better to deal with respondent answers using specific examples in the session transcripts, and by a great deal of supervised practice. Although the recording and transcript of sessions is not essential for data collection purposes, an important value of this practice rests with the ability to provide specific examples on better discussion techniques for trainee facilitators.

Some specific problems were noted with the questions relating to the development of response categories. These series of questions were extremely valuable for the construction of response categories for the ultimate survey questionnaire for a community generated index. Because of the usefulness of these questions, it is worth anticipating problems that could severely limit the quality of the data produced.

After the ranking session participants were asked, among other questions, to describe the criteria they had given in more detail. For example, where size of farm was used frequently, respondents were asked to state the actual size in acres that each of the wealthier to poorest socio-economic groups would be likely to hold. Although this sounds straightforward, we encountered major problems with the facilitator being able to remember all of the major criteria generated during the wealth ranking session. When data was analysed each evening, there were several gaps noted for a number of criteria. For example, in several sites food security was given as an important criteria by all groups, but the facilitator had omitted the question relating to food security from the focus group discussion for some groups, even though this had been a frequently cited criterion. After noting this problem, we developed check sheets in advance and asked the facilitator to write down frequently cited criteria during the wealth ranking session, and then fill in boxes under each socio-economic group. Although this decreased the problem, it was far from eliminated. I would suggest that an observer be given the task of noting the important criteria, as the moderator is likely to have difficulty facilitating the session and remembering to note important criteria correctly at the same time. The

observer can then hand over the prepared checklist to the facilitator at the appropriate time during the focus group discussion.

Inexperienced facilitators and study planners may not recognise the very real difficulties faced in facilitating focus group discussion, especially collecting the correct information to meet the study objectives. The burden associated with trying to carry out any activity in addition to the complex tasks of formulating questions 'on the run' and facilitating group discussion may result in poor quality data in all areas.

Preparation and analyses of data

Good planning skills are all that are required to ensure that data generated from sessions are accurately recorded in an easy to use form. Simple record sheets need to be prepared beforehand to list the criteria generated during the simple ranking session. Alternatively these sessions can be recorded and criteria entered directly onto simple matrices such as that found in Appendix 2.3. Both approaches are quite straightforward.

Translation and/or transcription of recorded focus group discussions (if required) is a skill that demands a higher skill level and develops from extensive practice. In terms of translation, not only does the translator require a high level of fluency in both languages, but the tendency to summarise long dialogues and over-interpret respondent meaning is very hard to eliminate. Transcriber interpretation also occurs frequently even where no translation is carried out. This should be expected, no matter how experienced the transcriber, and needs to be checked as part of the routine quality assurance activities.

In the methods section I have detailed a simple method to identify and weight criteria generated. This component only requires that the researchers have very basic mathematical skills, patience, and good common sense in working out how they will aggregate the criteria generated. I strongly suggest that only one person is selected to decide on the aggregation of criteria, or that two people work together at the same time. During this study, where people were working independently on

different sessions or sites, major problems arose in how specific criteria were aggregated. This resulted in the need to conduct a complete reworking of the data, the initial aggregation of criteria. To clarify this point further, in a single site there may be 90 or more individual criteria generated and recorded almost *verbatim.* Some of these criteria are obviously very closely related, such as 'iron roof sheets' and 'rusted iron roof sheets'. One individual may aggregate these into a single criteria called 'roofing material' while another may separate these into two criteria: 'roofing material' and 'condition of roofing material'. This obviously would result in chaos when criteria are aggregated or compared across sites.

A simple method for the construction of response categories is also outlined in the methods section. In this area we did not encounter any difficulties that typical district workers could not handle, assuming of course, that they are well supervised in the first few sites, and that data collection and record keeping during the sessions is well planned, as discussed in the previous section.

Resources Required for Wealth Ranking as Applied in this Study

7. *What resources are required to carry out the WRT as a method to develop location-specific indicators of SES?*

How many people are required to conduct this type of research?

It is possible to conduct this research with a minimum of three people, assuming that they are experienced in the method, local protocols and conditions. The minimal requirement is one facilitator, and one observer for the sessions, and another person to supervise village assistants gathering respondents for subsequent sessions. Of course it is encouraged to use local leaders to ensure the smooth running of sessions and the arrival of participants. Another staff member would be required to contact and prepare subsequent research sites if the project is to be carried out over several sites as quickly as possible.

How long does it take to prepare the community for data collection?

Site preparation was described in the methods section. It usually required from two to three days. This is crucial to the success of data collection through community acceptance and participation. If multiple sites are being studied (which is highly recommended), while one site is being completed, the second site can be contacted and prepared simultaneously.

How long does the data collection take per site?

The combined wealth ranking and focus group discussion take from one hour to 90 minutes for each respondent group. However, it is important to schedule time for research workers with no experience. I found, to begin with, the team could only conduct one session per day. This allows for the team to: a) iron out problems in the selection of session location, b) gather the participants, c) conduct the field de-briefs, and d) conduct preliminary data analysis where problems become more apparent.

Session Site Selection: We often encountered problems with session site selection in the early phases of the project. Although adequate preparation of the research team will reduce this problem, sometimes it is not until one arrives at the site (usually identified by local leaders) that problems emerge. The major problems we encountered were lack of privacy for the participants and lack of suitable furniture. 'Suitable' furniture is dictated by local expectation, and in our case people preferred to sit on chairs with a central table to place the ranking cards on. An additional problem we encountered in some sites was that leaders had identified the village government office as the site for sessions. As this was during election time, many village people interpreted our presence and discussions held in the party offices as political in nature. The neutrality of sites is important in politically sensitive places, especially during political campaigns. This of course seems quite obvious, but during the data collection for this study, Tanzania was not considered an environment where political sensitivities demanded special planning.

Gathering participants: This was a lengthy process in many sites throughout the study districts. Although we spent substantial time with community leaders, conducted public orientations to the sessions, and selected times of the day deemed suitable for the respective respondent groups, many times participants did not arrive on time and had to be collected from their homes directly by leaders or volunteer assistants recruited from the village. Individuals identified and recruited for participation in sessions are, of course, not necessarily so acutely interested in the project's activities that when the time comes for the session, they make special arrangements to attend. Often, individuals are highly pressured by day-to-day demands that make attendance at a particular time extremely difficult, despite previous agreements to participate. In other circumstances, individuals may be apprehensive and unsure and decide at the last minute not to attend. These problems are not usually serious, but it is helpful to make realistic schedules that allow community respondents some flexibility.

Field de-briefs: Our practice was to conduct a short field de-brief immediately after each session to discuss urgent points. For the early sessions, this could last from 30 to 45 minutes. With practice, this was reduced to five minutes, and in some cases not at all.

Preliminary Data Analyses: As this type of research should be considered iterative, it is essential to carry out the session transcripts and preliminary data analyses after each session. If an attempt is made to conduct more than one session per day in the first few sessions, the opportunity to correct facilitation problems is lost before they are repeated. Necessary redesign or improvement of the focus group discussion question guides cannot be carried out in time. Session transcripts usually take from three to five hours to process. This depends on availability of computers, ability of staff to type or write quickly, and the speed at which translators can work accurately with a *verbatim* transcript.

Number of session per day: As sessions only take about 90 minutes, it would appear that it is possible to conduct several session in one day. This is not the case. The demands on the session facilitator are great, with a high level of concentration

required to ensure the desired information is obtained. This results in facilitator fatigue. In addition, it necessary to have sessions transcribed and preliminary analyses conducted to detect weaknesses in facilitation, especially during the earlier sessions. Planning for more than two sessions per day would be unrealistic.

In the initial site for data collection and preliminary analyses, allow six working days. To this, add two to three days for community preparation, and consider days where people do not traditionally like to gather. In our case this was Sundays and Friday afternoons. Also prepare for local community events, national holidays, and allow for loss of at least one day per site for 'no-shows' or other unforeseen problems. Research staff must also have a routine weekly break which may fall on different day(s) as community holidays and events. Although wealth ranking is promoted as a rapid research tool, in actual fact, it can take a lot longer than initially anticipated.

How much does it cost?

This will depend largely on the level of transport required and available, and whether or not it is considered appropriate to pay respondent group participants. In our study community leaders gave up several days of work on their routine income-generating activities, and were paid honorariums to compensate for this loss of income. This project attempted to limit general expenditure and payment to respondents to avoid problems associated with externally funded projects that cannot be replicated by district budgets. Payments for community 'participation' markedly reduces willingness for communities to cooperate in future projects initiated by severely resource constrained, locally funded projects. We encountered substantial resistance in one site where UNICEF had previously paid very high incentives for community 'participation' in research activities. We lost two working days meeting with the community to negotiate voluntary collaboration by respondents. I would recommend as a minimum serving refreshments as a courtesy, and ensuring detailed feedback to the village on research outcomes. Our policy was also to spend time with community members in informal gatherings in the evenings to share stories of our experiences outside the village, and

demonstrate a genuine interest in their lives and general circumstances. In most countries, people from small often isolated villages enjoy an opportunity to talk informally with people from other places. This can increase community interest in project activities and acceptance of the research team, without draining limited research budgets. Residing in research villages during data collection and preliminary analyses not only facilitates a deeper understanding of the community and the quality of the data collected, but also means that more can be achieved with limited resources.

For this project, table 3.9 shows that the average cost per site was $282.71. Transport costs were calculated from district headquarters to the sites, rather than from the commercial capital where our project was based. In this project, per diems for food and accommodation were in line with Ministry of Health recommendations. This only allowed us to accommodate ourselves in the most basic village guest houses, often with very poor facilities, and poor quality and availability of food. For some organisations, this standard may not be acceptable to project staff.

It is also necessary to add to this figure training costs including payment of attendance fees for project staff during training sessions, as this is accepted practice in many countries. Finally, capital equipment such as the tape recorder, if purchased for the activity, need to be costed in to the budget. I have assumed that computers would not be purchased for this activity and only used if already available. Computers are not essential for data collection and analysis for this phase.

Table 3.9 Average Project Costs Per Site (seven days)

Item	T. Shillings	US$ ($1=Tsh580)
Co-ordinating Community Leader (x1)	3,500	6.03
Staff Accommodation and Food (x3)	80,000	137.90
Fuel	45,000	77.58
Vehicle Maintenance and Repairs	12,000	20.69
Stationary	5,000	8.62
Audio Cassettes	9,000	15.52
Respondent Refreshments	6,000	10.34
Batteries for Recorder	3,500	6.03
Total	**164,000**	**282.71**

How cheap is cheap?

As this figure is the calculated cost per site after salaries and training costs, the decision on whether or not it is affordable rests with the number of sites included in the research project, and of course the available budget. For a single district, it would be desirable to include a number of sites. Assuming that at least four sites are studied over the district, this translates into a *minimum* cost, before training and salaries, of about US$1,130 - in Tanzania. For many externally funded non-government organisations this is relatively cheap, but for a typical district health budget in Tanzania this would be considered a large sum of money. The benefits would need to be clearly demonstrated where districts are not usually allocated funds for research activities designed and conducted by the management teams or health workers themselves, and where numerous competing priorities exist.

3.2 Validity Assessments: Community Generated and Traditional Indices

3.2.1 Criterion-Relatedness or Substitutablity of Measures: The Screening Tests

The findings show that both the community generated and traditional indices are poor substitutes for the wealth ranking technique as a measure of a community construct of SES.

Levels of sensitivity and specificity were not encouraging for either the community generated or traditional indices. Importantly, the probability of very poor households being correctly identified by the indices as very poor, was small.

In the Sensitivity tests, the number of test positives (i.e. an index successfully identifying the very poor) is divided by the 'true' positives (i.e. the criterion: final community rank allocated during the verification the wealth ranking session). High sensitivity means that there are few false negatives, or, households incorrectly classified as better-off. In the Specificity tests, the number test negatives (i.e. index successfully identifying the better-off) is divided by the 'true' negatives. High specificity means that there are few false positives, or, households incorrectly classified as very poor. These tests provide us with information that is helpful at a policy level where measures are chosen for their performance in two important aspects, and results should be studied in that context. First, with Sensitivity, we get a clearer picture of the number of very poor that are (and are not) protected using the particular measure. Secondly, with Specificity, the amount of leakage where better-off households are identified as very poor. This is very important in terms of the cost of using the proposed measure. Table 3.10 shows the sensitivity and specificity of all indices across all study sites.

Table 3.10 Sensitivity and specificity of Study Indices (with 95 % confidence intervals)

	Morogoro Rural District		Hai District	
Survey Index	**Peri-urban**	**Rural**	**Peri-urban**	**Rural**
COM1 (weighted, proportion)				
Sensitivity	77.8% (58.5-89.7)	16.7% (02.9-49.1)	40.0% (17.5-67.1)	50.0% (09.2-90.8)
Specificity	43.8% (20.8-69.4)	81.6% (65.1-91.7)	83.9% (65.5-93.9)	91.4% (80.3-96.8)
COM 2 (weighted, midpoints)				
Sensitivity	54.8% (36.3-72.2)	16.7% (02.9-49.1)	13.3% (02.3-41.6)	0.00% (0.00-60.4)
Specificity	87.5% (60.4-97.8)	81.6% (65.1-91.7)	100% (86.3-100)	98.3% (89.5-99.9)
COM 3 (unweight, proportion)				
Sensitivity	83.9% (65.5-93.9)	33.3% (11.3-64.6)	60.0% (32.9-82.5)	75.0% (21.9-98.7)
Specificity	37.5% (16.3-64.1)	81.6% (65.1-91.7)	83.9% (65.5-93.9)	89.7% (78.2-95.7)
COM 4 (unweight, midpoint)				
Sensitivity	41.9% (25.1-60.7)	16.7% (02.9-49.1)	0.00%	0.00%
Specificity	93.8% (67.7-99.7)	92.1% (77.5-97.9)	100%	100%
TRAD 1 (weighted, proportion)				
Sensitivity	93.5% (77.2-98.9)	41.7% (16.5-71.4)	40.0% (17.5-67.1)	50.0% (09.2-90.8)
Specificity	68.8% (41.5-87.9)	94.7% (80.9-99.1)	80.6% (61.0-91.9)	91.4% (80.3-96.8)
TRAD 2 (weighted, midpoint)				
Sensitivity	06.5% (01.1-22.8)	25.0% (06.7-57.2)	06.7% (00.3-34.0)	50.0% (09.2-90.8)
Specificity	93.8% (67.7-99.7)	94.7% (80.9-99.1)	96.8% (81.5-99.8)	100% (92.3-100)
TRAD 3 (unweight, proportion)				
Sensitivity	77.4% (58.5-89.7)	91.7% (59.8-99.6)	46.7% (22.3-72.6)	50.0% (09.2-90.8)
Specificity	62.5% (35.9-83.7)	39.5% (24.5-56.6)	83.9% (65.5-93.9)	89.7% (79.2-95.7)
TRAD 4 (unweight, midpoint)				
Sensitivity	0.00%	08.3% (0.4-40.2)	06.7% (00.3-34.0)	50.0% (09.2-90.8)
Specificity	100%	100%	96.8% (81.5-99.8)	100% (92.3-100)
SELF RANK				
Sensitivity	29.0% (14.9-60.4)	27.3% (07.3-60.7)	40.0% (17.5-67.1)	50.0% (09.2-90.8)
Specificity	87.5% (60.4-97.8)	92.1% (77.5-97.9)	100% (86.3-100)	98.3% (98.5-99.9)

Medical diagnostic tests aim for sensitivities in the order of 85-95%, although for the measurement of a relatively ill-defined construct such as socio-economic status we would tolerate lower sensitivities. In this case I have chosen 60% to represent the lowest acceptable level of sensitivity. Before an index is considered acceptable, it must have an associated specificity of no less than 60%. This

decision was guided by the fact that it would be unreasonable to expect the same high levels of sensitivity of medical diagnostic tests. This was balanced with a consideration of the negative consequences to individual households, and health service efficiency, of setting standards of sensitivity and specificity too low. Confidence intervals are also presented on the table. These are not narrow due to the small number of households (46 - 62) in each study site. A much larger study requiring substantial resources would need to be carried out to achieve narrower confidence intervals. However, these results provide sufficient information to determine whether further investigation is justified. The results discussed in the following paragraph have been highlighted on table 3.10.

For the community generated indices, table 3.10 shows that only the unweighted, community generated index with proportional cut-off points (COM3) achieved an acceptable level and balance between sensitivity and specificity. This was only achieved in the peri-urban and rural sites of Hai District.

For the traditional indices, the weighted index with proportional cut-off points (TRAD1), and the unweighted index with proportional cut-off points (TRAD3) achieved an acceptable level and balance between sensitivity and specificity. For both indices this was only seen in one of the four sites - Morogoro, peri-urban site.

As expected, the SELF RANK measure performed uniformly badly in all sites.

For both the community and traditional indices there is substantial variation in sensitivity and specificity across sites for all indices. Neither type of index could be considered as a satisfactory substitute for wealth ranking.

3.2.2 Construct Validity: The Multitrait, Multimethod Matrices

The main study hypothesis that indices of socio-economic status (SES) using community generated indicators are a more valid measure of the community construct of SES than traditional measures using indicators selected by researchers or social sector professionals, was not supported by the evidence.

Multitrait, multimethod matrices were constructed for all indices in all four survey sites. Appendix 3.6 shows all indices constructed (four community generated, four traditional and a single measure SELF RANK) in all four sites (= 32 matrices). All correlations shown are Kendall's Tau β, as discussed in Appendix 2.4.

Appendix 3.7 shows four tables which indicate whether or not for each site, each index met the three individual criteria required to be considered for further validity assessments. Appendix 2.4 explains the rationale for the criteria, but to refresh the reader these are restated below and allow an interpretation of the columns in the tables in Appendix 3.7.

1. The values in the validity diagonals should be sufficiently different from zero and large enough to encourage further examination of validity (Column 4).

2. Each validity coefficient should be larger than *all* of the heterotrait, heteromethod correlations which are in the same rows or columns as the validity coefficient itself (Column 5).

3. Each validity coefficient should be larger than the heterotrait-monomethod correlations which involve the same variable as the validity coefficient (Column 6).

A summary of Appendix 3.7 can be found on table 3.11. It shows whether or not each index met all three criteria, in addition to the validity coefficient for the measure of SES via the two methods.

Table 3.11 Summary of Multitrait, multimethod matrices

INDEX	**Morogoro Rural District**		**Hai District**	
	Peri-urban (N = 47)	Rural (N = 50)	Peri-urban (N = 46)	Rural (N = 62)
COM 1 (weighted, proportional cut-off points)				
Met all three criteria	No	No	No	No
Validity coefficient	0.28	0.29	0.35	0.32
COM 2 (weighted, midpoint cut-off points)				
Met all three criteria	Yes	No	No	No
Validity coefficient	0.45	0.07	0.25	0.01
COM 3 (unweighted, proportional cut-off points)				
Met all three criteria	No	Yes	Yes	Yes
Validity coefficient	0.27	0.42	0.47	0.50
COM 4 (unweighted, midpoint cut-off points)				
Met all three criteria	Yes	No	No	No
Validity coefficient	0.46	0.16	0.33	0.29
TRAD 1 (weighted, proportional cut-off points)				
Met all three criteria	Yes	Yes	Yes	No
Validity coefficient	0.67	0.62	0.40	0.39
TRAD 2 (weighted, midpoint cut-off points)				
Met all three criteria	No	No	Yes	Yes
Validity coefficient	0.25	0.27	0.41	0.74
TRAD 3 (unweighted, proportional cut-off points)				
Met all three criteria	Yes	Yes	Yes	Yes
Validity coefficient	0.43	0.49	0.43	0.44
TRAD 4 (unweighted, midpoint cut-off points)				
Met all three criteria	No	No	No	Yes
Validity coefficient	0.39	0.20	0.27	0.66
SELF PERCEPTION				
Met all three criteria	No	Yes	No	No
Validity coefficient	0.18	0.43	0.38	0.34

From table 3.11, it can be seen that the community generated indices did not prove to be more valid measures of a community construct of SES than traditional ones. In fact, the traditional indices successfully met all three criteria on ten of a possible 16 occasions. Only one traditional index was validated in all sites (TRAD3), and one was validated in three of the four sites (TRAD 1). These were both indices with proportional cut-off points. A review of the validity coefficients shows that the traditional indices achieved strong coefficients (0.60 or higher) only on four of the possible 16 occasions, and TRAD 3 on no occasion.

The traditional indices with proportional cut-off points may possess some degree of construct validity, but the results should be interpreted with some caution as overall, validity coefficients were not convincing in magnitude. Traditional indices were able to discriminate between traits on all occasions.

The community generated indices successfully met the three criteria in only five of a possible 16 occasions. Only one (COM3), was successfully validated in three of the four sites, and no other index in more than one site. This index also used proportional cut-off points. None of the community indices had very strong validity coefficients of 0.60 or higher. An examination of Appendix 3.2 reveals that with the community indices, discriminant validity was not achieved on five occasions. This suggests that method variance was more important than trait variance. As expected, the self rank item performed quite poorly, meeting all criteria in only one site.

Looking at the performance of all indices (including self rank) by site, both peri-urban and rural sites validated indices on eight occasions each. In addition, both Morogoro Rural and Hai Districts (peri-urban and rural together) validated indices of eight occasions each. This shows that there is unlikely to be any relationship with the type of site (peri-urban v. rural; Morogoro v. Hai Districts), and the successful validation of indices.

The community index that was successfully validated in three of the four sites (COM3) was an unweighted index, and for the traditional indices, TRAD1 was

weighted and TRAD 3, unweighted. All successful indices had proportional cut-off points, suggesting that the midpoint cut-off approach has little value.

These results do not support the hypothesis that the community generated indices are a more valid measure than traditional measures. The results suggest that these community generated indices developed in this study, irrespective of transformations, do not show promise for future use in the districts for which they were developed. The results showed that the traditional indices did capture some aspect of a community construct of SES.

Chapter Four

Appropriate Applications of Wealth Ranking: A Post-investigative Assessment

4.1 Overview

This chapter discusses the suitability of wealth ranking as a tool to identify the poor in a number of contexts for which such a tool might be used. Relevant issues are identified for each of these contexts. Four propositions are considered that derive from the results presented in chapter three:

1. Simple wealth ranking (allocating a rank to specific households) shows merit for use in small populations.
2. Wealth ranking shows potential as a method to complement validation studies of other proposed measures.
3. It is possibly unnecessary to generate location-specific indicators of socio-economic status through wealth ranking for the range of populations in this study.
4. Wealth ranking as a tool to generate indicators for use in survey indices, may not be an appropriate choice given the limited theoretical and empirical base underpinning the technique.

Measurement theory is used to discuss three possible interpretations of the results of the index validation assessments: a) that the indices lacked construct validity, b) that there was poor construct validity for other variables in the analysis, and c) that the methods used to assess validity were inappropriate. The discussion argues that a lack of construct validity of the indices, and problems in test construction were the most probable explanations for the findings.

The chapter closes with a discussion of the tension encountered when using both qualitative and quantitative methods, and an assessment of the value implications and social consequences of measures of SES used in health service delivery, and discusses the ethical burden on users of measures of SES in general.

4.2 Simple Wealth Ranking as a Tool to Determine the SES of Specific Households

Simple wealth ranking as a tool to identify the SES of specific households shows promise. High levels of reliability, the ease at which respondents could rank the poorest households, and the few resources and skills required to carry out the technique demonstrate the value of simple wealth ranking. In this context, the evidence would suggest that an investment in further examination of construct validity is justified. At the same time, the results indicate that there is a need to select a range of respondent groups for ranking exercises, due to the variability of respondent group performance. That is, in terms of reliability, bias in over- and under- estimation of SES rank, and the number of households recognised. All these issues will be discussed in turn.

Overall Reliability

Of the studies that have reported levels of agreement between respondents (or respondent groups) during wealth ranking few have been designed specifically to measure reliability empirically. They have mostly described data collected to answer other research questions, but have contributed useful preliminary information about the wealth ranking technique applied during the research activities.

The findings reported from these studies describing levels of agreement were inconsistent. For example, Chadwick and Seeley (1996) found low agreement between respondents during wealth ranking exercises, while Grosvenor-Alsop (1991) found high levels of agreement. These studies applied different measures of agreement so comparison of their findings should be made cautiously. Chadwick and Seeley (1996) conducted their investigation in several areas, while the Grosvenor-Alsop (1991) study was conducted in a single village with 87 households. Although this latter study claimed to be specifically measuring agreement, impact on our advancing knowledge of the reliability of respondent ranks has been limited due to the size of the study, the number of respondent groups that participated, and the nature of the statistics applied to test reliability.

The strategy and design of the current study provided a more robust assessment of the reliability of respondent group ranking. The study was conducted in 10 villages over a large range of geographical locations affected by a number of cultural, climatic and economic factors. Four hundred and eighty eight households were ranked by five respondent groups each and again by a final 'verification' session. This resulted in 2,928 individual ranks performed. In addition, the study applied kappa statistics which are accepted as an appropriate measure of judge agreement, and especially suitable for multiple judge analyses. Statistical significance was calculated to allow further confidence in the findings of high reliability.

Although the findings show high levels of reliability, it is insufficient to simply measure and report on reliability statistics. Part of the routine process of examining reliability is to identify areas of weakness and suggest way of improving reliability in the future. Chadwick and Seeley (1996) suggest reasons for low agreement between respondents in their study. They point to the possible varying knowledge of respondents of the households being ranked, mistaking household names, and guessing in an effort to conceal knowledge gaps. They recommend the use of respondent groups rather than individuals to reduce these possibilities. These are factors which may have affected reliability in other studies, and were also noted during wealth ranking sessions in the current work.

Although the results in the current study demonstrated high reliability, a number of factors operated that reduced the chances of a perfect score, and are likely to be operating in other wealth ranking activities in other projects. Using Ghiselli's (1964) theoretical framework of reliability discussed in chapter one, I have identified possible causes for reliability levels that are less than 'Almost Perfect'. In table 4.1, I have listed these factors, along with some indication of the level of control we had (and exercised) to reduce the effect of these factors on reliability. Table 4.1 also shows factors where no control could be exerted. These factors are likely to contribute to less than perfect agreement. Considering all these areas for potential breakdown, it is quite encouraging that such high levels of reliability were demonstrated.

*Table 4.1 Factors Affecting the Reliability of Wealth Ranking and Levels of Control**

Factor	Testing Situation	Control	Individual Households	Control
VARYING	Participants		Recognised by Respondents	
	> interest/motivation/fatigue	++	> profile in the community	Nil
	> personal agenda of individuals	+	> isolated or near other households	+
	> influenced by previous respond. groups	+	> participation in community	Nil
	> knowledge of the households ranked	+	events	
	> confusing names of household heads	++		
	> concept of SES of individual participant	Nil	Demonstration of wealth	Nil
	> criteria used by individuals to rank	Nil		
	Facilitator		Relationship to participants	
	> dominant participants only rank	+++	> relative or friend	Nil
	> speed of session	+++	> popularity in the community	Nil
	> demonstrated interest in respondents	+++	> envied/pitied	Nil
	Observer			
	> recording of ranks accurately	+++	Ranked Early or Late in Session	Nil
	Environment			
	> privacy/quiet for session	++		
	> pressure to attend to other demands	++		
CONSTANT	Facilitator			
	> instructions given for each session	+++		
	Participants		Not Applicable	
	> characteristics of respondent group (Sex or SES bias')	Nil		
	> group concept of SES	Nil		

* Nil = No control: + little control: ++ some control: +++ good control

Respondent motivation to participate, or fatigue levels obviously determine the amount of time and serious thought respondents dedicate to determine the rank allocated to each household. It is obvious that good facilitation will reduce this problem. Sometimes, respondents have personal agendas in terms of what they perceive to be the use of the information given to the researchers. If they perceive benefits will accrue to households ranked as poor, or the village at large, then this could influence the rank they allocate.

One would expect where respondent group sessions are carried out over several days, there may be animated discussions after sessions in the community about what was being said during the wealth ranking sessions. Respondents may discuss the ranks they gave to certain households, the criteria they used, or other information

discussed during their session. This could influence the ranks allocated to households by some respondents and not others. This can be reduced by holding sessions as closely together as is practical, and by selecting respondents from different households.

It is also possible that individual respondents have a different concept of SES to each other, and are using different criteria to rank. There is little the facilitator can do about this if a community concept is the one of interest, and prompting is not permitted.

In this study confusing the name of the head of household was rather common, especially in the Islamic areas where many people shared the same names. We used nicknames to distinguish households from one another, but often we were not able to identify the correct nickname with sufficient confidence. As a result, like the Chadwick and Seeley (1996) study, this may have accounted for some unsystematic variation.

For the constant factors from the testing situation, there may be some bias in ranking based on the characteristics of the whole group. Women or men may rank in a different way, or alternatively the richer and poorer respondent groups. I have examined this issue separately, but was unable to demonstrate a consistent pattern in the bias.

For those varying factors that related to the individual household, some households may be more or less recognisable than others to the whole group. For example, households with high profile members, households located nearer to the particular respondents in the group, or community members that are often seen participating in community events may be more recognisable to the whole group of respondents. This is conceptually different to individual respondents not knowing a household. In this case it is something about the particular household that causes all respondents to respond in a similar fashion.

Another of the varying factors related to the individual households are the cultural norms that determine demonstration of wealth. For example, on Mafia Island, some households expressed a fear of "bad luck" if they were to show off their wealth. This can result in quite wealthy households concealing wealth and living in very similar

conditions to poorer members of the community. This may make it more difficult for respondents to judge each other's SES if they do not possess a more intimate knowledge of the household.

Finally, when households are ranked earlier in the session, before the respondent group has discussed their criteria at length, it is possible that the group holds a different conception of SES. As the group discussion proceeds, the act of listening and discussing the criteria may easily result in some shift in the conceptualisation of the group. Therefore, households ranked early in the session may be ranked according to one set of criteria, and households ranked later in the session by a slightly different set of criteria. This can be reduced by checking the rank allocated during, and at the end of the session. In my experience, however, people are so tired by the end of the activity that they rarely make many changes to their original ranks.

Overall reliability was slightly higher in peri-urban areas contrary to the general belief that wealth ranking may only be useful in rural areas. This finding could be an artefact of the data due to the small number of comparisons. Alternatively, it could be because households in peri-urban areas are closer together than in rural areas with more opportunity to observe the lifestyle and experiences of each other, and are more community minded than in fully-urban settings. Peri-urban areas are usually market towns; people have many opportunities to speak to each other while passing through the market, or selling produce. This would provide more opportunities for interaction and sharing details of personal life than for people working on more isolated farms. Finally, it may be related to the SES of the area as peri-urban areas are usually better off than rural areas in Tanzania.

The results suggest that reliability in ranking appears to improve with the general socio-economic status of the location. Although the comparisons were few and certainly not enough to make any generalisations, the results suggest that this positive relationship exists. This may be driven by a number of factors. People living in areas of higher socio-economic status may be more familiar with outsiders and more comfortable articulating about abstract concepts with limited preparation. Also, in more isolated locations (which are usually poorer) people may lack confidence in performing the exercise or task, and rush to give answers to avoid looking foolish. Alternatively, different stratification systems may be operating that

are easier to classify, for example, a more detailed occupational structure, and more visible consumption practices.

Reliability in categorising Highest and Lowest Socio-economic Groups was Best

In most applications of wealth ranking, an important requirement is the identification of the 'very poor' households. Commonly, we are concerned with identifying the poorest households to target project activities, or ensure that the project interventions have reached, and benefited the poorest households.

Chadwick and Seeley (1996) found relatively low levels of agreement between the number of households placed in the highest and lowest socio-economic groups or categories in all four study sites. I have applied different measures of reliability based on levels of observed and estimated chance agreement across all households while the Chadwick and Seeley study compared proportions of households in each of the four socio-economic groups. This may account for the difference in findings. The findings in the current study were relatively stable over the ten sites, and were consistent with expectations. The wealthiest households were most reliably ranked, followed by the poorest households, with middle categories of SES least reliable.

Wealthier households in our study sites were often obvious through higher levels of visible consumption, general status, providing loans for profit, and philanthropy (providing assistance through non-interest loans, donations and gifts). Poorest people were less visible, but may have been identified more easily than the middle classes through their requests for assistance, non-participation in events, or serious crises that demand attention of the community. Chambers (1995) refers to several of these dimensions of deprivation when he discusses the humiliation experienced when dependent on others, and when isolated from the community.

The Method for Identifying Respondent Group Participants

Leaders were asked to identify respondent group participants based on their perception of participants' SES. They were asked to identify two groups for each sex, one representing the 'better off' households and another the 'poorer' households. Most of the participants were correctly placed. However, the misclassifications that did occur were largely in the wealthier respondent groups for both sexes.

This suggests that care is required in making statements about how wealthier respondent groups perceive wealth, especially where methods used to classify respondents are not well tested. There were a number of possible explanations for this misclassification of wealthier respondents. Firstly, wealthier men were hesitant to leave their business activities to participate in the session, claiming that they were too busy. Of course the poorer respondents were also pressured with the more urgent demands of generating income on a daily basis, but were generally more willing to attend the sessions. Secondly, we required five male and five female respondents for each of the wealthier groups. We were often not able to identify enough wealthier participants from the fifty households. Leaders were most likely identifying respondents from middle socio-economic groups in order to identify sufficient respondents. In addition, in the Muslim areas, women were not normally found outside the household compound and their attendance at the sessions may have been a sensitive issue. In the Christian Kilimanjaro Region there were fewer misclassifications of wealthier women.

Respondent Group Performance (Reliability, Bias and Recognition of households)

Although the simple wealth ranking appears to be highly reliable, some of the findings suggest that a number of community groups should be selected to participate in the wealth ranking sessions. Using respondent groups (versus individuals) where discussion and debate are possible, may improve reliability and ultimately validity. This recommendation is supported, in part, from the examination of the factors previously discussed that may have reduced potential reliability scores. A range of respondent groups is also recommended as respondent groups showed variations in reliability, bias in over- or under-estimation, and recognition of households. Because no regular pattern was identified, selecting a range of groups would seem prudent.

Respondent Group Reliability

No studies were identified that examined performance in terms of which respondent group was most reliable over a number of sites. If some groups (based on sex and

SES) are in fact more reliable than others, it may be more effective to use only those groups for specific purposes where resources are limited.

This study applied statistics specially designed to measure multiple judge performance based on the ratio between the proportion of times a judge agrees with all other judges and the proportion of times all the other judges agree with each other. Judges in this context are the respondent groups. This is calculated for all judges and provides an index for each judge which is used to assess performance.

It would seem reasonable to interpret the findings as reflecting little difference in ranking behaviour between groups of respondents based on sex or SES. However caution is required in interpretation. It is possible that there may be differences based on sex or SES, but these may have been obscured by the way in which respondent groups were selected in this study. For example, although there are wealthier and poorer respondent groups, we have no way of knowing *how* wealthy or how poor the individuals were. We may have had participants that were located more centrally in the stratification hierarchy. The wealthier men may not have been substantially wealthier than those who were placed in the respondent group for poorer men.

It may also have been that within a single district you may identify patterns, and that certain factors are operating at a district level that would result in patterns of reliability. By comparing respondent group patterns across districts and regions such differences would not be detected. To test this hypothesis, it would be necessary to carry out wealth ranking in numerous sites within a single district. We may then find that certain respondent groups based on sex and socio-economic status may be more reliable judges of SES. Conducting wealth ranking with a range of respondent groups is desirable as it is not yet appropriate to make generalisations about whether or not wealthier or poorer respondents; males and females; or leaders are "better" in these, or indeed other locations.

Bias

The Grosvenor-Alsop (1991) study attempted to determine if the position of the respondents in the social hierarchy influenced their allocation of rank. If it does, then this is a strong argument for multiple respondent groups from a range of SES

situations. In addition it would provide additional justification for investigating differences within the community from a range of perspectives. Do certain groups display regular, predictable ranking behaviour? Grosvenor-Alsop asserts that there were no differences in the way different respondents (based on caste) allocated a SES rank. However, the methods used to determine this were unclear and an assessment of the appropriateness of this interpretation is not possible.

Although the number of sites in this study is not sufficient to make broad generalisations, and tests of statistical significance were not performed, it may be sufficient to suggest trends. Leaders and Poorer Women underestimated SES in 70% of the sites. It could be that leaders report lower socio-economic status of their households as they are more highly motivated to attract benefits to their villages, which would certainly reflect well on their general performance. Leaders were often younger men placed in these positions by the government as administrators and often did not perform true leadership roles within the community. They came from a variety of socio-economic backgrounds, and their ranking behaviour should not be related to their position in the socio-economic hierarchy.

As Poorer Men overestimated and Poorer Women under estimated SES of households, it cannot be said that over- or under-estimation was likely to be a function of SES. An examination of the ranking behaviour of poorer people only revealed that they are not uniformly ranking people higher or lower than themselves. It is also necessary to consider the previous discussion concerning the actual location of participants in the socio-economic hierarchy. If indeed our respondents were located more centrally in the hierarchy, then these comparisons are not meaningful.

A further, alternative explanation of these findings is that the verification session that was used as the 'standard' may have been a poor measure of SES, thus obscuring patterns in respondent group ranking. This possibility is discussed in detail later.

Households Not Recognised

Giving attention to the number of households recognised is important. If there are a large number of households not recognised, then reliability and validity must be in question - and the ultimate value of the activity becomes dubious.

Several studies looked at the number of households not recognised by respondents, but as the type and composition of respondents across studies were so different it is not possible to compare results. In some studies respondents were individuals, in others groups of individuals were selected according to sex, socio-economic status or age. What was uniformly found, however, was that the number of households not recognised was quite low. An exception was Chadwick and Seeley (1996) who demonstrated a wide variation of households not recognised across sites (4 to 83% unrecognised).

The current study found that, in all but one site, almost all households were recognised by most respondent groups. However, the female respondent groups accounted for more of the unknown households than the other three respondent groups. Explanations for these findings may concern some cultural norms in the study sites. For example, in most of the Coast Region and Morogoro, the majority of the population is Muslim. In these areas, unless households are particularly poor, women are not usually found far from their own compound.

As proposed by Chadwick and Seeley (1996) it may be that men pretend knowledge of households. This is a less plausible explanation in these research sites as reliability is unlikely to be so high where respondent groups are guessing the socio-economic status of households. Another possibility to explain low levels of unknown households is the proximity that households are from one another. In other wealth ranking studies, households may have been located at greater distances. A selection criteria for households to be ranked in the current study was the distances between the houses. Only houses that were immediate neighbours were selected, and in most circumstances households were within a few metres of one another.

In a single site in Kilimanjaro there were larger numbers of households not recognised. This was most likely due to the selection process of households to rank than true lack of respondent knowledge. During preparations for the wealth ranking exercises, several of the leaders appeared to have limited knowledge of the households in their areas of responsibility. They were not able to easily identify the name of household heads and the construction of the list of households was confusing.

This study did not evaluate the *validity* of simple wealth ranking to determine the SES of individual households. It should be considered as simply contributing additional knowledge about the technique. However, if reliability had been poor, then the validity of the allocated rank could be ruled out for these respondents groups.

Feasibility of Simple Wealth Ranking

Simple wealth ranking appears to be quite feasible as a tool for non-specialists to identify specific households' SES. It is relatively easy to train people to perform the technique, organise the activities, recruit willing community respondents, conduct the sessions and analyse the data. It is also relatively cheap, especially if transport is not an issue, and no additional personnel are employed. *Simple* wealth ranking without criteria generation or additional focus groups could be carried out in one hour per respondent group, and several group sessions could be carried out within a single day. Simple wealth ranking, although demanding basic facilitation skills, does not require the concentration and skill level of focus group discussions, nor any advanced skills in recording or analysis. Simple wealth ranking is far more accessible to non-specialists than other techniques to identify the poor. From a practical perspective, its limitation is the small number of households that can be ranked within a given population.

4.3 Wealth Ranking As a Tool to Generate Location-Specific Indicators of SES

A comparison of the criteria generated through wealth ranking across large geographical areas was not identified in the literature. Most commonly, studies using wealth ranking (e.g. Jodha, 1988) describe the criteria that are generated by respondents. In other studies (e.g. Scoones, 1995) the criteria used by different respondents within research sites is compared. These comparisons are useful if examining the dynamics of poverty in particular locations from a range of

perspectives, but an important question remains about the need for location-specific indicators. If criteria used across locations are substantially different from those used within a single site, then this would provide a convincing argument for developing location-specific indices for targeting purposes.

This study results revealed that there was a moderate level of agreement about which indicators were important within the sites by the five respondent groups. This supports the assertion that similar criteria of SES are likely to be used within a single site where respondent groups experience the same causes and effects of poverty, and share a similar interpretation of their reality. If different criteria are being used by different respondent groups, then this would suggest that either their construct of SES is different; or, that the different perspectives result from different dimensions of SES being given different levels of importance, even though the underlying construct is similar.

With the cross district comparison a similar level of agreement about which criteria are important (based on frequency of use) was found. This is the result of merging the peri-urban and rural criteria to form a single district set of criteria. This result suggests that although the cross-district kappa statistic of 0.550 is lower than the intra-site kappas (0.574 - 0.753), the difference is not substantial and would not justify the added resources being applied to develop location-specific criteria. There are three possible causes for these findings, and I will argue that elements of all three may be operating.

Firstly, there is, in fact, less difference than expected in the criteria that respondents state they use across geographical locations in rural and peri-urban Tanzania. An examination of the criteria generated in each site provides some additional insight. For example, during data collection, education *seemed* to the researchers as far more important in Kilimanjaro region than in Coast Region. This is an impression driven, perhaps, by the higher educational status of people in the region. The Wachaga ethnic group from Kilimanjaro Region have a reputation for enjoying a higher status and often securing higher positions in the public and private sector in Tanzania. They are more likely to have a higher educational status, are more entrepreneurial, and commonly enjoy a relatively higher SES than other ethnic groups in the country. Investigators come to the area with these prejudices and expectations.

Education was mentioned as a criterion more often in Kilimanjaro than in Coast Region. However, if you look at its importance compared to other indicators such as agriculture-related indicators, it is still not used as frequently as other criteria, and is of relatively minor importance. This was a good example of the researchers' impressions can be misleading. Care is required in making generalisations about the need for location-specific indicators from impressionistic information only. This issue is not whether education is *correlated* with a measure of SES. Education may prove to be a good predictor of SES, without necessarily being amongst the criteria people say they use to decide on their neighbours' SES.

A second factor explaining the higher than expected levels of agreement cross-districts may be a function of merging the rural and peri-urban data. Data was combined from the two sites into a single list of criteria with the frequency of use being a total of that used in both the rural and peri-urban site within each district. If this is the case, then perhaps the specific nature of these types of sites (peri-urban versus rural) would need to be reflected in indices developed on a sub-district basis, a rather impractical proposal. Alternatively, it may be that only rural and peri-urban sites would require different indices reflecting the differing economic influences in each type of site. By merging sites with possibly different indicators of SES together there may be a flattening out of differences. Therefore the cross-site levels of agreement, before aggregation, were examined.

The level of agreement on the importance of criteria across all ten sites was slightly lower (kappa = 0.435) than the cross district comparison (kappa = 0.550) where the rural and peri-urban criteria were aggregated. It should be noted that kappa statistics do improve with the number of judges in the calculation. In this discussion we have a comparison of a kappa calculated with five judges (the five districts with aggregated rural and peri-urban criteria) with a kappa calculated with ten judges (the ten sites before aggregation to the district level). Therefore, we should interpret the cross-site comparison as a still lower level of agreement than the cross-district comparison. Lower, that is, than what a simple comparison of the kappa statistic implies. I consider that this is still an insufficient drop to justify a number of location-specific exercises which wealth ranking appears to lend itself towards.

As mentioned, merging the rural and peri-urban criteria together may have accounted for the higher levels of agreement for the cross-district comparisons. This leads us to consider whether it is the type of site (rural/peri-urban), rather than the geographical location that is more important for developing site-specific indices of SES. To assess this proposition, rural and peri-urban comparisons were made across regions. Although the level of agreement improved slightly, it was still not substantial. The type of site had some effect, but not enough to suggest that it would be more appropriate to develop separate indices for each type of site.

Respondent group characteristics (i.e. sex and SES) may have been the determinant of the criteria used rather than the type of site or the geographical location. Perhaps women use certain criteria, while men use other criteria during the categorisation process. An examination of the levels of agreement between respondent group types across sites revealed a more substantial drop. This may provide some preliminary evidence that sex and/or SES may have less effect on criteria used to determine SES than geographical location or the level of urbanity. Before this could be asserted with confidence, it would be necessary to conduct a more thorough, in-depth analysis of perspectives, using more appropriate ethnographic methods.

A final explanation of these findings is that the respondents did not *articulate* the criteria they actually used when allocating a SES rank to the households. This possible interpretation of the data will be discussed in detail in the following section.

In summary, the criteria generated through wealth ranking across sites was not substantially different from those generated within sites in these three regions of Tanzania. Therefore, applying the wealth ranking technique across large numbers of locations in Tanzania would not be a prudent use of scarce resources. It would be more useful to identify criteria that are appropriate across wide areas, either at the national or regional level, and invest in research by specialists with advanced research skills that can determine their validity in a range of locations. Specialists would be better equipped to confront issues concerning the conceptualisation of the community construct of SES; what construct the proposed criteria or indicators are measuring; or generate criteria to measure a specific construct that has been shown,

theoretically, to be related the behaviour of interest, such as utilisation of health services.

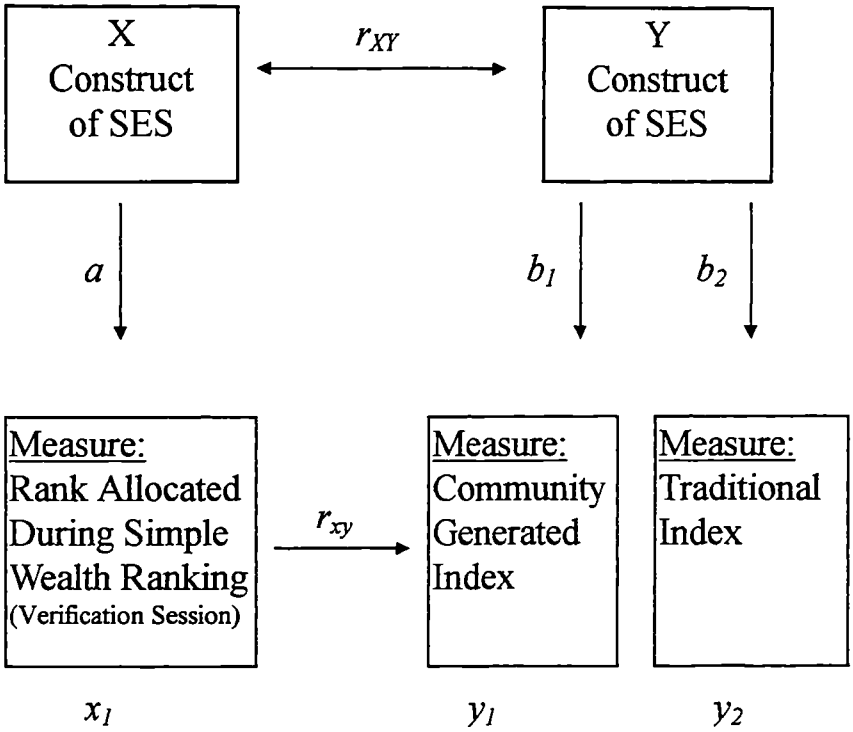
4.4 Wealth Ranking as a Tool to Develop Indices of SES by Non-Specialists

The findings of this study suggest that wealth ranking, as applied in this study, is not an appropriate tool for non-specialists to develop indices of SES. Both criterion-related and construct validity of the indices developed using the wealth ranking method, were low. The main hypothesis, that indices of socio-economic status using community generated indicators are a more valid measure of the community construct of SES than traditional measures, was not supported.

Only the Rajaratnam *et al* (1992) and Scoones (1993) studies confronted validity issues of wealth ranking directly. The study carried out by Rajaratnam *et al* in India was dedicated to this purpose, and measured criterion validity of the allocated wealth rank; while Scoones (1993) measured convergent construct validity, where the rank allocated to households during the wealth ranking session were correlated to a number of traditional survey items independently. The researcher could find no studies to demonstrate the validity of an index constructed with indicators generated by wealth ranking.

The results of this research did not provide supportive evidence of construct validity of the community generated indices. Although the evidence for traditional indices showed some promise, it was not convincing. There are three possible interpretations of the findings from the validity assessments of both indices: 1) the indices lack construct validity; 2) there was lack of construct validity or unreliability of some other variable in the analysis; or, 3) the method or procedures used to test the hypothesis was faulty or inappropriate. In figure 4.1 these interpretations are clarified.

Figure 4.1 Study Constructs and Measures



1. If the two indices lacked construct validity, then we would expect that r_{xy} is small for each of the two measures of Y; but, that *a*, is supported by auxiliary theory. That is, that the rank allocated during simple wealth ranking is a valid measure of the construct of SES.

2. If there was a lack of construct validity or reliability of some other variable in the analysis, this could infer, for example, that the rank allocated during simple wealth ranking (x) was a poor measure of the construct of SES, or that *a* was not supported by the auxiliary theory.

3. Finally, it may have been that the methods and procedures used to assess validity were inappropriate or poorly applied. That refers to the screening test, the multitrait, multimethod matrices, and the statistics used during the analyses.

I will argue that 1 and 2 are likely to be the most appropriate explanations of the findings. This discussion will focus largely on the community generated index, as one would have expected that this index would have been more likely to be measuring a community construct of SES than a traditional index reflecting an outsider's construct.

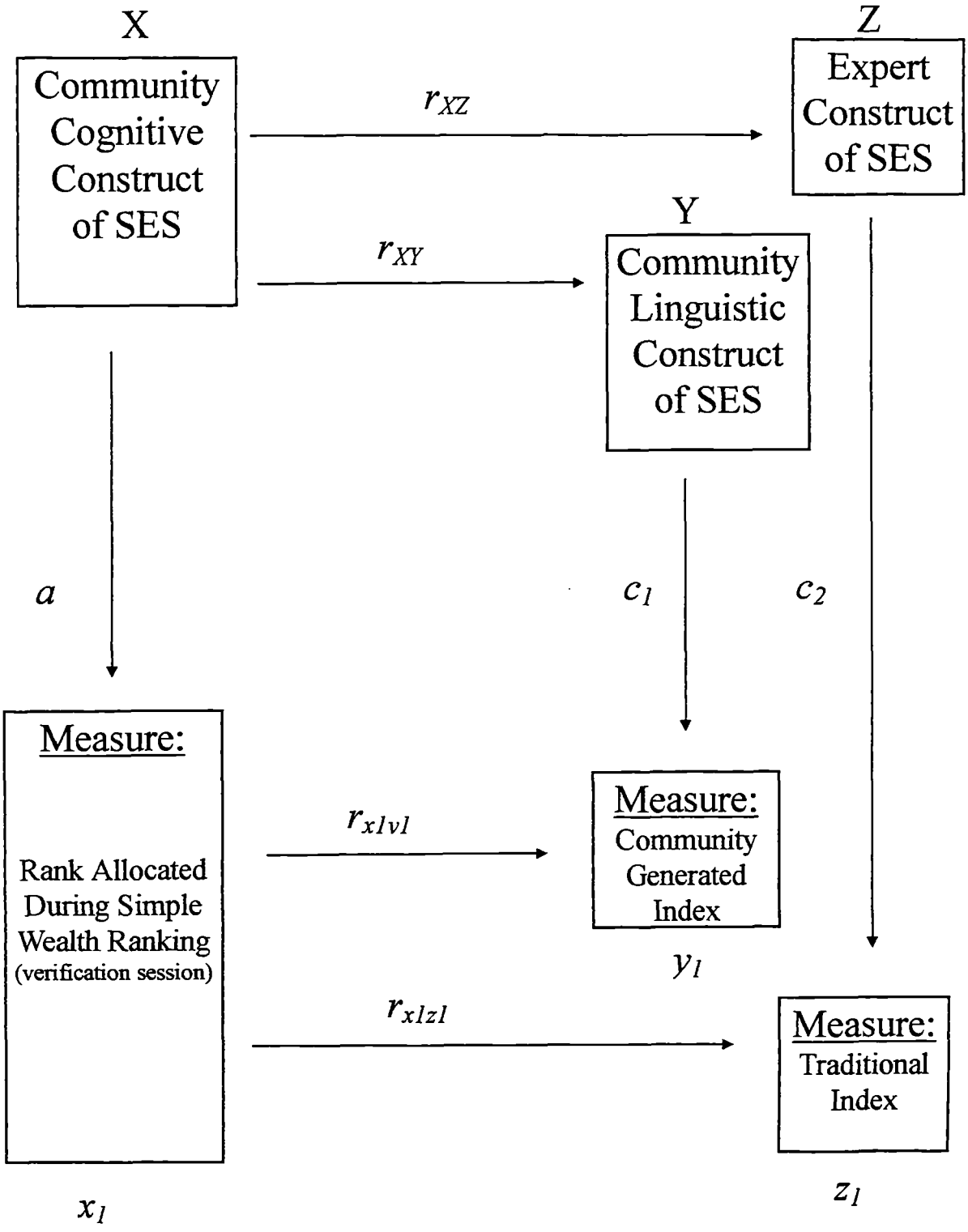
4.4.1 Interpretation 1: The Indices Lack Construct Validity

In this study, the community generated indices lacked construct validity. In other words, they were unlikely to be measuring the same underlying trait or construct that was being measured by the rank allocated during simple wealth ranking. The results suggest that the indices were measuring *some* construct rather than randomly allocating categories to households since the ability to discriminate between two different, unrelated traits was demonstrated by the multitrait, multimethod matrices. If this is the case, then what constructs could the indices be measuring? Figure 4.2 is a proposed explanation of the constructs being measured by the indices, and their possible relationships.

I propose that there were three constructs being measured in this study. At the outset, it was hypothesised that the traditional index, which reflects an outsider's construct of SES would not be as closely related to a community construct of SES as measured by the community generated index. Hence the hypothesis that indices of socio-economic status using community generated indicators are a more valid measure of the community construct of SES than traditional measures using indicators selected by researchers or social sector professionals. Using Figure, 4.2, 'Z' symbolises this expert construct.

The findings suggest that the community generated index and the allocated wealth rank from the verification session are not measuring the same trait or construct. They may in fact be measuring two *distinct*, but perhaps *related* constructs. The validity coefficients on the MTMMs were above .30 on a number of occasions, and it would therefore be reasonable to suppose that these two constructs are somewhat related, possibly even measuring different dimensions of the same underlying community construct.

Figure 4.2 *Alternative Model of Theoretical Relationships between Constructs and Measures*



In the model, as shown in figure 4.2, 'X' shows what could now be termed the cognitive construct. This is the construct that is being measured by the rank allocated during the final wealth ranking session. It is the construct that is in

people's minds during the wealth ranking activity; that is, during the *cognitive process* used to categorise a household into a socio-economic group. It cannot be clearly defined as it reflects a community conception, one we are still exploring. It can be thought of as the construct of most interest, the concept that most reflects the realities and experiences of the people, and is not influenced by outsiders.

Next, in the model, 'Y' represents the linguistic construct. This refers to the construct that is determined by the criteria that community respondents *articulate*. For example, when a respondent is asked to rank a neighbour in terms of their SES, the thought processes are likely to be highly complex. They may consider a tremendous variety of factors, based on varying levels of personal knowledge. They may consider the household's family relationships and commitments, abilities to plan for the future, work ethic, private and public responsibilities outside the household, last year's bad luck, next year's potential gains, in addition to more familiar criteria such as income and living conditions. They may also consider things of a private or sensitive nature based on their more intimate knowledge of the household, and probably many other dimensions we cannot even anticipate. During the cognitive process all these factors are given weights, or, considered in terms of their overall importance.

When we ask respondents to tell us what criteria they used, it is perhaps unreasonable to expect that they will explain all these thought processes comprehensively, logically and coherently. The respondents, most probably, would never have had to consider the dimensions of wealth in such detail before, let alone articulate them systematically giving each dimension a clear, and accurate weighting of importance. There may be a gap between thought and language. In the model r_{XZ} shows this hypothesised relationship, while r_{xlzl} refers to the study's measure of this relationship.

I propose that there are a number of factors that prevent people from explaining all the criteria involved in their decision making process. These could be factors that affect their *choice* to articulate criteria, factors affecting their *ability* to articulate criteria, or other factors could be those related to the *exercise* itself, or the *individual respondents*. Table 4.2 lists potential factors that seem plausible for the communities in which this study was carried out. It is necessary to generate as many

factors as possible that might explain the findings of this study in order to provide more substance to our discussion of wealth ranking, and to suggest future directions for testing the relationship between how individuals categorise households according to wealth status, and what they tell us were the indicators or criteria they used to do so.

Table 4.2 Proposed Factors Affecting the Criteria Generated by Respondents

Factor Group	Factor
Choice	> certain dimensions are *perceived* to be of more interest to the investigators > adherence to Grice's (1975) 'Co-operative Principle' of Quantity > the use of culturally determined inference techniques > willingness to deconstruct the cognitive process quickly > sensitive information withheld
Ability	> ability to deconstruct the cognitive process quickly > ability to recognise the most relevant aspects of the cognitive process > ability to remember the most relevant aspects of the cognitive process > sufficient knowledge of the specific household under discussion
Exercise Situation	> time given to respondents to reflect on cognitive process and articulate the most important aspects > general facilitation skills > often respondents use what appears to be criteria when actually only helping other respondents to identify a household (e.g. age, occupation)
Individual Respondents	> motivation to participate in a demanding exercise > poor task equivalence (Hughes, Seidman and Williams, 1993) > citing criteria already used by other respondents to 'get it right' > making something up if the household is not known in an effort to appear knowledgeable

Choice

Respondents may make conscious, or unconscious choices about the criteria they articulate to the investigators. It may be that respondents *perceive* that some criteria may be the types of criteria that outsiders are more interested in. For example, it may be assumed by the respondent that the size of the farm land would be of more interest to the researchers, than a personal problem an individual has developed with an influential person in the village; one that may influence the ability to secure a loan, or transport produce to market.

It is also possible that informants choose only limited information that they consider best represents the complex cognitive process. According to Grice (1975), people

usually adhere to the "Co-operative Principle" in conversation which refers to the quantity, quality, relevance and manner in which people communicate with each other. For the quantity component, he shows how we: a) only say as much as is required, and b) do not make our contribution to the discussion more informative than required. This is partially linked to the previous point, only the respondent's perception of what may be most relevant to the investigator may not be the only factor affecting the choice of criteria mentioned.

According to this "Co-operative Principle", a speaker does not spell out every conceivable item of information in their speech, but rather leaves some information unsaid, or implied. The listener infers the implied information using the context and their own knowledge of the world. What are the implications of this when the listener may be the session moderator coming to the conversation with a different knowledge of the world? If implied information is determined by the local context and world-view, perhaps important criteria are missed by the investigator. Certain statements made during the session may imply something different to what the investigator has understood.

Jodha (1988) discusses what he calls the "communication gap" between what respondents tell a researcher and what is *understood* by that researcher. In his discussion of the factors responsible for misinterpretation of information by researchers, he highlights an important factor that may be relevant in this study; that is, differences in the connotation of the same concepts as they are understood by the respondent and researcher. Jodha gives an example of the connotation of 'man-hours of labour input' as understood, estimated and reported by farmers who are less time-conscious than researchers. In our case, these differences in connotation may result when information is implied rather than clearly stated.

Another factor that could be affecting the choice of criteria given by respondents is the motivation to deconstruct the cognitive process. Given the complex nature of the task respondents are being asked to perform in a very short space of time, they are sure to have varying levels of motivation to recall the cognitive process, and then carefully select those criteria which best capture that process.

Finally, it is clear from observing several thousand individual rankings, that often respondents withhold what they consider to be sensitive information. For example, in our study sites, women were often willing to discuss male expenditure on alcohol and mistresses and the effect this had on household cash flow. Men, however, rarely identified such consumption practices. No doubt there would be a number of criteria that all respondent groups chose not to discuss with the investigators.

Ability

That community members are the most knowledgeable about their own realities and experiences (Chambers, 1995) is a tenable proposition. However, the ability to articulate clearly on such matters *quickly*, assumes some experience in doing so. For example, studies in medical anthropology are often engaged in testing hypotheses about health belief (e.g. McNee *et al*, 1995). Using ethnographic techniques we can ask respondents to explain their beliefs about the causes, and progress in the body, of certain diseases. Faced with such questions, respondents give a range of explanations and descriptions. As researchers we record these statements and then make inferences about the community perceptions and develop models of health belief.

It may be that most of the respondents may never have considered this question before, and have no clear belief at all. In order to satisfy the investigator they give some explanation that seems plausible, and may or may not be composed of fragments of ideas previously considered. But this 'deficiency' in having a well developed health belief is not restricted to isolated villages in developing countries. Try asking colleagues or friends (without medical training) to discuss the cause and progress of respiratory infection that comes from going outside with wet hair - a health belief commonly shared in many societies. You may well find that people are not able to clearly describe the related physiology or progress from wet hair to pneumonia.

I would propose that people have not ever considered their perceptions of SES in sufficient detail to articulate clearly *in a limited amount of time*. This is not to say that community members do not have a rich understanding of their own

community's reality. Rather, that articulating the cognitive process used to describe the important dimensions of a highly complex construct such as SES may be somewhat difficult when using a rapid technique that does not allow for sufficient depth in exploration and analysis that more sophisticated ethnographic techniques allow.

Next, it may be useful to consider that a respondent is being asked not only to *remember* the cognitive process used, but to assess, select and identify the *most relevant* criteria. This assumes the ability to assess what is 'most relevant'. Assuming relevance is measured in terms of the community construct of SES, it may be extremely difficult for respondents to select that which best captures this construct. It is quite possible that the criteria that are more visible and easy to articulate with the available linguistic tools are used; for example those criteria related to material consumption, ownership of property and so on. This would not mean that they are necessarily the most relevant. Vulnerability, relationships and obligation within the community may be equally or even more important, just more difficult to explain quickly in a few words, and in a short space of time.

The proposition that respondent groups may not have the ability to articulate the classification system applied during the cognitive process can be supported by work carried out in the field of psycho-linguistics. For example, Cole and Scribner (1974) described Scribner's earlier, unpublished work that examined the relationship between how respondents categorised a group of objects, and their verbalisation of the categories they used for the activity. Classification of the objects could be based on either colour, shape or function. The adult respondent populations studied were high-school students, non-literate adults from transitional villages holding cash jobs, and non-literate rice farmers who lived five miles from the nearest road. Scribner also included matched groups of school children and non-school children from two age groups (6 to 8, and 10 to 14). She found that both the non-literate adult villages (transitional and isolated villages) and the 10 to 14 year old school children were similar in their practical classifying activities, that is they put objects together in groups in a similar manner. However, they were very dissimilar in the verbal explanations they gave for these activities. The younger people with schooling reflected the category nature of their groupings in the way they described them,

while the villagers without schooling did not. The unschooled villagers often used 'inadequate' explanations such as "I like them this way" or "my sense told me to do it this way" (p. 120). These findings were also reflected in this study. On several occasions in the rural villages respondents answered that their criteria for categorisation was "It is God's will that he is in this pile". Often, no additional criteria was elicited from the respondent when probed for further explanation. Cole and Scriber suggest that it may be useful to make a clear distinction between the way individuals *perform* sorting operations, and the way they *describe* their own operations.

The wealth ranking technique also pre-supposes a high level of familiarity with the households under study. The earlier phase of this work was able to show that in these study sites, recognition appeared quite high. However, it is possible that a respondent may know which household is being ranked, but have insufficient knowledge to cite criteria confidently. It may be that respondents often simply cite criteria that have already been used for previous households when they are clearly not very familiar with the household under discussion.

Exercise Situation

The wealth ranking situation itself can obviously affect the ability of the respondents to generate meaningful criteria that are appropriate indicators of the community construct of SES. For practical reasons it is necessary to find a balance between allowing respondents sufficient time to reflect on the cognitive process and articulate the 'most relevant' criteria - and spending too long with each household, boring respondents and reducing the quality of the data. General facilitation skills are very important.

Task equivalence is another factor which may affect the criteria stated. This refers to the "extent to which respondents' familiarity with, or interpretation of, the assessment situation and task demands are similar" (Hughes *et al*, 1993, p. 693). Hughes *et al* argue that respondents in different cultural groups have different notions about what is expected of them in the research situation. This would seem plausible for wealth ranking activities where different respondent groups (based on

sex and SES) may interpret the exercise differently resulting in the articulation of different aspects of the cognitive process. Or, that the cognitive process itself is different, based on their interpretation of the instructions given by the facilitator.

A very important quirk during the generation of criteria was noted during the wealth ranking sessions. When respondents were asked to rank the household into one of four possible socio-economic groups, they were citing the criteria for categorisation at the same time. In many instances when a household was not immediately recognised by the group of respondents, some participants would attempt to identify the household to other participants using what could be construed as criteria. Occupation, age and marital status of the household head were common examples of this. Criteria generated via typical wealth ranking exercises may be incorrectly recorded as criteria of wealth rather than a simple 'identifier'. In most situations it is almost impossible to distinguish the intended usage. This was observed in all ten study sites, and I believe is an important factor in the generation of possibly invalid criteria.

Given a substantial number of factors that could influence the articulation of criteria, it would seem reasonable to expect that there is a Linguistic Construct *and* Cognitive Construct. But the questions still remains, how different is it, and what construct do the articulated criteria reflect? Where are the differences and similarities, and are these, in practical terms, important for specific purposes? These issues are quite relevant questions for the application of wealth ranking generally. To the author's knowledge, neither the theoretical relationship between the community construct and what people actually give as criteria or indicators of SES, nor the implications of a weak relationship have been discussed sufficiently in the literature. These issues are particularly relevant for those investigators that plan to use the criteria that communities give to study the dynamics of poverty or purposes related to the design of intervention programmes designed to assist the poor. Efforts to develop a theoretical base and accumulate empirical evidence for this relationship would seem justified.

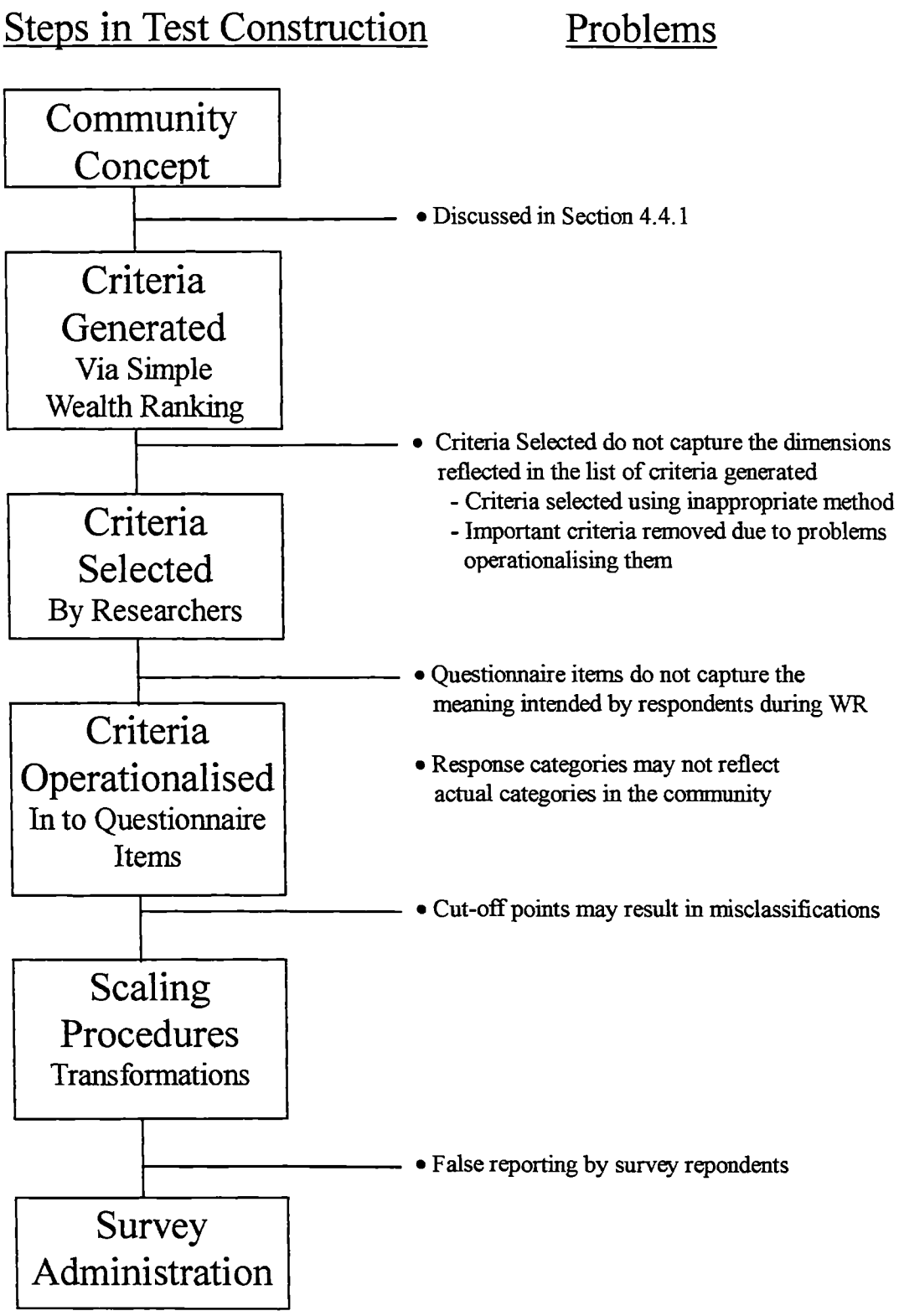
I have argued that there are a number of factors which could result in a cognitive and linguistic construct which are distinct, but related constructs. Lack of construct validity is not the only possible interpretation of the results.

4.4.2 Interpretation 2: There was Lack of Construct Validity or Unreliability of Some Other Variable in the Analysis

Even if we take the community generated index to be measuring a linguistic construct of SES, there are a number of potential problems noted with the index which should be clearly described. It is only through a thorough examination of these issues that we can improve upon the approach taken in this study, and more generally, the development of valuable indices of a community conceptualisation of SES.

Let us assume that we have extracted a list of criteria generated from the community that captures the linguistic concept. We now must construct an index that selects appropriate items, operationalizes them, and produces a questionnaire that is then carried out according to the normal principles of survey administration (e.g. Abramson, 1990). There are several critical points in this process where breakdown can occur. Figure 4.3 highlights these for the typical steps involved in test construction. I have already discussed in some detail the breakdown between the cognitive construct and the linguistic construct where the criteria are identified.

Figure 4.3 Problems in the Process of Test Construction



After the criteria have been generated via wealth ranking, a decision must be made as to which criteria will be included in the index of SES. In more typical test construction, items are selected using two main approaches: assessing the face validity, and conducting item analysis using a number of statistical techniques based on inter-item correlations (e.g. Streiner and Norman, 1989). Assessment of face validity is not truly relevant where we are attempting to develop an index that is intended to reflect a community construct of SES, not that of the outsider. As another fundamental aim of this study was to develop a procedure that could be adopted by typical peripheral social sector workers, inter-item correlation and factor analysis techniques were not appropriate. Therefore, the approach taken in this work was to select items based on two criteria. Firstly, that items could be operationalized in a questionnaire, and secondly, based on their frequency of use (interpreted as their relative importance).

These criteria for the selection of items may have resulted in poor items included in the index, and ultimately an index that measured neither the community construct, nor the hypothesised linguistic construct. It is possible that the criteria generated are an unsuitable representation of the categorisation task performed (wealth ranking), but using the 'frequency of use' may have magnified the importance of items with poor construct validity. 'Frequency of use' as a selection criteria may have had even less value if it is true that many respondents may be simply reusing criteria seen to be acceptable to the facilitator during the session.

As described in the methods section, some items were removed as they could not be operationalized. For example, work ethic or ability to plan was mentioned on a number of occasions as an indicator of wealth. One could argue that these are actually causes of poverty or wealth and not indicators *per se*. However, if they are taken to be criteria, it was not possible to develop a practical question line to capture this concept, certainly not in a way that would be inoffensive or possible for a household respondent to answer. Perhaps the rejection of certain items based on ease of operationalization, eliminated important dimensions of a community construct of SES. In fact this brings us to a frequent criticism of traditional indices. According to Jodha (1985):

> "The first step is to measure whatever can be easily measured....the second step is to disregard what can't be measured....the third step is to presume that what cannot be measured easily is not very important....the fourth step is to say that what cannot be easily measured really does not exist" (Jodha, 1985 cited in Kabeer, 1991, p. 3)

I have used this statement not to support an argument that all indicators *must* be operationalized to ensure that the community construct is adequately captured, but to highlight the implications of removing even a small number of items identified by community respondents. There are times when it is impossible to transfer a community notion of wealth and poverty into a simple questionnaire item that can be reliably and validly answered by respondents. This does beg the question however, of the value of attempting to use survey indices to capture a broad community construct of SES. It might be more realistic to reduce the scope of the construct substantially to meet the specific purposes of the measure. For example, rather than measuring SES, we could reduce the construct to, say, "the ability to pay for a basic consultation and drugs at a village health clinic".

Of course item selection could be improved substantially by using accepted techniques such as inter-item correlations and factor analyses. However, these techniques would move index construction into an arena with investigators who possess more advanced research skills, and out of reach of typical district level health service managers.

The next possible area of concern was that the questionnaire items did not successfully capture the intention of the criteria as given by the community. There was a clear and consistent problem with two items in particular. In all four validation sites, the community made a distinction during the wealth ranking sessions between being dependent on others for survival, and receiving assistance from others to improve the general SES of the household. 'Dependency on Others' means that without the assistance of a child or other relative, the household would not have sufficient resources to house, feed and cloth the household members; that is, meet some unquestionably basic needs. Alternatively, 'Receives Assistance', means that assistance was not actually required for survival, but was received, perhaps as part of the extended family support network. It may have been that a

household that would otherwise have been categorised as poor (SEG3 of 4) was ranked as comfortable (SEG2 of 4). Despite extensive re-writes of the questions, back translations and extended sessions with enumerators, this distinction was never well understood - by either the enumerators or the survey respondents. This may have had an impact on the success of the index as dependency is a well documented experience of the very poor (e.g. Jodha, 1988).

The technique used in this study to develop appropriate response categories for questionnaire items was more in-depth in nature than the criteria generation during wealth ranking. Using focus group discussions, the facilitator spent some time determining the appropriate categories for each criterion. This was explored for the village as a whole and not simply the households being ranked. I do not expect that response category formulation contributed to the invalidity of the index, but it should be considered when developing any scale or measure.

Two approaches to determining the cut-off points were used in this study, and although the indices did not demonstrate strong construct validity, the indices using proportional cut-off points were slightly better. I do not consider this had as important an effect on the validity of the measures as the other factors discussed, as the proportions were determined using the same proportions of households in each socio-economic group as the wealth ranking exercises.

False reporting by respondents should be of concern when assessing the quality of information collected on household SES. Of course it is not possible to account for the level of false reporting in this study, and impressions during data collection are insufficient evidence to claim low levels of false reporting. However, as I was present for all surveys, and accompanied enumerators throughout the survey period, I was able to observe and gain some impression of false reporting. Only flagrant false reporting can be detected through observation of the interviews. For example, on some occasions respondents said that they did not have certain material possessions that I observed in adjoining rooms, or in the compound on arrival or departure. Or, where households stated that they had small land holdings where we had learned during the wealth ranking sessions from a number of respondent groups that the household in fact had large land holdings. Overall, I did not detect a high

level of false reporting and I do not believe that this accounted for the invalidity of the index.

As described, there were several points in index construction (and to a lesser extent administration) that may have accounted for poor validity of the final index. Not only with the cognitive concept, but with a linguistic construct as well. Although I have listed a number of potential problems that could be encountered by other researchers, I believe that the most likely problem areas in this study were during item selection and to a lesser extent the operationalization of the items for the questionnaire. The implication of this is that there may be a strong relationship between the linguistic and cognitive construct, but because of poor index construction the linguistic construct was poorly measured.

A second area for debate about the validity of other variables in the study revolves around the issue of the suitability of the alternative measure of the community construct of SES; that is, the allocated wealth rank during the verification session. The MTMM assumes that the alternative measure is an accepted measure of the construct, but as we had seen in chapter two, validity or acceptability is assessed on a spectrum. There is no definitive point where we can say a measure is valid. Acceptance is based on a number of studies conducted in a number of settings, and we develop confidence over time. As discussed, wealth ranking itself has not been thoroughly examined to date. We have a small number of studies that point to different aspects of value, ranging from its practical application, to levels of agreement among respondents or respondent groups, and finally, a small number of studies confronting validity issues. To contribute to this literature, the current study was able to demonstrate a high level of reliability in a large study conducted over a broad range of geographic and cultural locations. This is further evidence about the *potential* value of the tool.

However this is insufficient evidence to assert that the allocated wealth rank during the verification session was a valid measure of the community construct. Therefore, poor validity of this measure cannot be ruled out as a possible explanation for poor validity of the community generated index.

4.4.3 Interpretation 3: The Method or Procedure Used to Test the Hypothesis is Faulty or Inappropriate

Common causes of the selection of inappropriate measures are related to the procedures selected to validate the proposed measures, or the statistics applied in the manipulation of the data. This study applied the multitrait, multimethod approach to assess construct validity.

The MTMM method was recommended and supported in the health (Streiner and Norman, 1989) psychology (Sullivan and Feldman, 1994) and sociology (Ghiselli, 1969) literature concerned with construct validity. However a thorough critique of the method was carried out by Sullivan and Feldman (1994), and this should be examined in relation to the current study to assess the implications for interpretation of the results.

Critique of the Three Criteria Used to Assess Validity on the MTMM matrices

Using path analysis, Sullivan and Feldman (1994) provide a critique the MTMM matrix, and base their critique on the three main criteria that are used to assess convergent and discriminate validity.

- *The validity coefficients should be large and statistically significant.*

Sullivan and Feldman show that the validity coefficient can be large either because correlations do in fact link the true variable with the measure, or, because the method effects are strong and the methods are in fact correlated. In the case of the present study, we can be reasonably assured that the methods are not likely to be correlated due to the contrasting nature of the method (household survey versus wealth ranking) and the respondents (individual self reporting for survey versus groups discussing other people in wealth ranking). Therefore, we can be more confident that the validity coefficient is not overly exaggerated by methods variance.

- *When comparing the validity coefficients and the different trait-different methods correlations in the same row and columns as that validity coefficient, correlations across methods should be higher for the same traits than different traits.*

In this case, a large difference between the correlations should indicate that the two measures of SES have more in common across methods than correlations of SES and Skin Tone. Therefore, the difference should be a function of the size of the correlation between SES and Skin Tone. However, according to Sullivan and Feldman (1994), the differences could be due to small methods effects (uncorrelated methods), or the method effects may be large, but approximately equal. In that case, there could be a large difference in the correlations even in the presence of substantial method effects. The implication of this again, is to assume strong method factors are not operating due to the disparate methods.

- *The correlations of the two measures of the same trait across methods should be higher than the correlations of measures of different traits using a single method. If the second correlation is larger then we would assume greater method variance than trait variance.*

Using causal models and associated equations, Sullivan and Feldman (1994) were able to show that if the method variance was large with respect to trait variance, the third criterion would be rejected, which was Campbell and Fiske's original intention. Although this criterion does hold up well under closer examination, it is still necessary to assume that the methods influence the measures in the same way; and for the comparison to be sensitive to method variance, the correlation between SES and Skin Tone must be quite small. In summary, if the two traits are "fairly highly" correlated, then even a small amount of methods variance will be overestimated.

Reactivity

Another issue identified by Sullivan and Feldman concerns the problem of reactivity whereby the first measure of the trait will have an impact on the second measure. The design of this study allows us to dismiss this concern as the methods not only are conducted in different ways (ranking versus survey), but they involve different respondents which should almost eliminate any reactivity effects. Even where a

respondent in a ranking sessions is discussing his own household *after* he has been surveyed, he must contend with the other four or five members of the respondent group which may be calling for a different allocation of rank.

Other Factors

The final concern with the MTMM matrix is in the situation whereby a second and distinct factor aside from the hypothesised trait, accounts for part of the correlation between measures. Campbell and Fiske's MTMM matrix cannot identify this situation, and the existence of such a factor cannot be rejected for this study.

The result of this discussion is that we must be prepared to make some assumptions about the relationship between methods, and the way the methods influence the measurement of the trait. For this study, I have shown that these assumptions can be made with a reasonable degree of confidence.

Finally, the correlation coefficient for the MTMM was chosen based on the nature of ordinal variables. Kendal's Tau was therefore applied, and statistical significance of the findings calculated.

I consider the alternative explanation, that the findings may be due to methods or procedures employed, as the least likely explanation. All procedures are clearly described to enable the reader to make an independent assessment.

4.5 Substitutability of the Community Generated Index for Wealth Ranking

The use of criterion-related validity as an appropriate choice to determine validity is debated in the literature, and was outlined in chapter one. According to Ghiselli (1969) care must be taken to use a criterion where no generally accepted measure of the construct exists, especially with constructs used in the social sciences. Later, others have argued (Messick, 1980) that the appropriateness of criterion validity is simply in deciding whether one measure is a suitable substitute for another, and no more. That is the interpretation used in this study. The sensitivity and specificity tests are only performed to assess whether or not the community generated index is a suitable substitute for the allocated wealth rank during the verification session.

This was desired as the wealth ranking technique cannot be applied in larger populations. In addition, sensitivity and specificity tests permit a discussion not only on the *extent* of the poor being misclassified, but also on the extent of misclassifications of the better-off and the implications for efficiency in health service management.

The findings of the sensitivity and specificity tests could not show criterion-relatedness for either the community generated or tradition indices. One community generated index achieved a satisfactory balance between sensitivity and specificity in only two of the four sites, and two traditional indices on one site each. An in-depth discussion comparing the implications for the protection of poor households, and wastage of scarce resources in the health service by incorrectly categorising the better-off as poor, is not indicated. Neither index could be considered a reasonable substitute for wealth ranking, and therefore application of the indices is not recommended. Issues concerning the validity of the criterion have already been identified as an alternative explanation for the findings.

4.6 Conclusions: Appropriate Applications of the Wealth Ranking Technique

In conclusion, the appropriate applications of wealth ranking is discussed as a result of this study. The following discussion will assess wealth ranking from two perspectives. Firstly, simple wealth ranking where households are simply allocated a rank of socio-economic status. Secondly, as a tool to generate indicators or criteria of SES for a range of purposes which might range from exploratory studies of poverty dynamics to selecting indicators for use in household surveys.

Simple Wealth Ranking as Applied in this Study

Simple wealth ranking has been shown to be a highly reliable method of identifying the SES of small groups of households. Previous studies have shown preliminary evidence to support a claim of validity for the allocated wealth rank. However, our knowledge of the validity of the allocated wealth rank as a measure of wealth is still not well developed. Despite the growing body of literature advocating wealth

ranking, there have been insufficient empirical studies carried out to be able to promote the technique with more confidence. Issues of geographical validity have not been well addressed so that we have only a limited knowledge about its value across a range of countries and cultures. Hughes *et al* (1993) argues that task equivalence is an important factor in the appropriate transfer of techniques across cultural groups. As such, different groups may interpret the wealth ranking task demands quite differently, and therefore validation in one site does not guarantee generalizability to other settings.

Promoters of simple wealth ranking have a responsibility to either provide empirical evidence of the relationship between the allocated rank and a construct of wealth, or, to highlight the gaps in theoretical knowledge to potential users of the technique. Many of the organisations who adopt new techniques do not have the skills or resources required to assess the value of techniques promoted by those considered to be experts in the field.

With a strengthened empirical base, wealth ranking may have potential for targeting purposes in small projects where resources and skills are limited. It may have also have potential as a tool to conduct assessments of criterion-related validity or as Messick (1980) has labelled the concept, 'substitutability'. Simple wealth ranking may be used to contribute more to the evidence of convergent and discriminant construct validity of any proposed measures of SES. There are a great many projects and research workers in partially subsistent economies that are currently using measures of SES that have never been assessed in terms of their validity. Assuming that the validity of the allocated wealth rank can be shown with more confidence, non-specialists are capable of using wealth ranking to provide preliminary evidence of the value of their current or proposed measures of SES. Simple statistical calculations are used which require only simple arithmetic. Of course, where estimates of significance are required, more sophisticated skills would be required. Where health service or epidemiological studies require validation of proposed measures without conducting a lengthy or expensive sub-study, wealth ranking may provide an effective and efficient tool.

Wealth ranking proved to be an excellent method for the MTMM approach to assessing construct validity, as it possessed the following qualities:

- Simple wealth ranking is simple to carry out
- Simple wealth ranking lends itself to identifying two traits at community level
- Simple wealth ranking is vastly different from the survey method - a basic requirement for using the MTMM, but one not always easy to meet.

The MTMM technique itself is also highly recommended for researchers without advanced skills, however some basic statistical skills are required to calculate the correlation coefficients, and computers are ideal. It is also a rapid and low cost method to conduct validation assessments of indices proposed for larger studies.

In terms of the development of a theoretical and empirical base, there are a number of specific areas that require investigation:

- *It is necessary to establish construct validity with more confidence*

Investigators with advanced research skills can apply ethnographic techniques to explore and describe the community construct more clearly. Alternatively, the allocated wealth rank can be correlated with a number of constructs known to be related to SES. As more confidence is developed in the measure, discriminant construct validity can be challenged with more plausible alternating hypothesis - that is with traits that are less obviously unrelated to SES as skin tone - thus refining the construct further.

- *It is necessary to determine the most appropriate was to summarise ranks allocated by a number of respondent groups*

Given that a range of respondent groups is generally considered to be ideal, what is the most appropriate way of summarising the ranks allocated by a number of respondent groups? Does an additional verification session with an in-depth discussion provide a more valid measure, or are more simple averages sufficient?

Wealth Ranking as a Tool to Generate Criteria of SES

This study proposed numerous weaknesses in the generation of criteria that were intended to capture a community construct of SES. I believe it is in this area that wealth ranking requires substantial investigation and development before it can be

promoted as a technique with high utility. An obvious area for concern, as a result of this research, is the advocation of wealth ranking as a technique to identify indicators of SES for survey purposes.

Wealth ranking is often promoted on the basis of the rapidity with which it can be carried out, and the simplicity of the technique. It is therefore assumed to be suitable for non-specialists. However, when wealth ranking is adapted to generate criteria of SES, more complex skills are required to extract the criteria, and assess how well they might capture a community construct of SES. The present study identified the possible existence of a cognitive and a linguistic construct, and suggested that these constructs may have a poor relationship in communities where respondents cannot easily articulate their categorisation process, or where there is insufficient time given to analyse the process. On a more fundamental level, one might question the appropriateness of wealth ranking, as a rapid technique, to explore these categorisation processes at all. Until these issues are investigated more comprehensively, it is suggested that wealth ranking may not be the most inappropriate tool for exploring a highly complex construct such as wealth. It's greatest limitation perhaps, lies with the speed in which the respondents are expected to identify the criteria they used to categorise households' SES. Indeed, one would expect that advanced research skills, an appreciation of ethnographic methods and analytical techniques, and a great deal of time are required to explore dimensions of wealth or dynamics of poverty adequately.

4.7 Combining qualitative and quantitative methods

The discussion of the results of this study draws us into questions relating to the choice of method to meet the research objectives. As discussed in detail in chapter 1 (section 1.5), this study chose to use wealth ranking as a tool to identify indicators of socio-economic status. This choice was made so that the concept and measures proposed would reflect the realities experienced by the communities for which indices of socio-economic status were to be constructed. This was considered desirable for a number of reasons, but largely because current measures may not reflect this reality at a district level. The project aimed to determine whether or not

the community generated indicators would be a more appropriate measure of socio-economic status than indicators identified by outsiders. Questions may arise relating to the appropriateness of using qualitative methods such as wealth ranking and focus groups discussions to generate indicators of socio-economic status that are translated into quantitative tools for survey purposes.

This section will provide a brief review of the epistemological foundations of qualitative and quantitative techniques, outline the debates concerning these techniques, and discuss the implications of these on the present study.

According to Reichart and Cook (1969) the quantitative world view is characterised as hypothetical-deductive, objective, and outcome-oriented. Hypothetical-deductive simply meaning that the research design is based on testing hypotheses, and infers particular instances from general laws rather than being inductive where a general law is inferred from a particular instance. These approaches are consistent with the positivist paradigm where one recognises an objective reality not dependant on the researcher (Olsen, no date). Objective research separates the researcher, or subject, from the respondent, or object, and the researcher focuses on the object in an effort to understand objective reality. A Positivist approach is typically characterised as being concerned with operational definitions, objectivity, replicability and causality. Surveys are seen as well suited to this approach because they can meet the needs of these concerns easily. Concepts can be operationalised, the distance between the observer and the observed maintains objectivity; replication is easily carried out by using the same tools in different contexts; and causality is addressed through statistical techniques such as path analysis. The labels of empiricism and positivism infer an epistemological view that the research is supported by a distinctive theory as what should pass as authorised knowledge (Bryman, 1984).

The positivist paradigm assumes truth consists of observable and verifiable facts, and not of internal conditions such as personal dispositions or values. This is where the fact-value dichotomy comes from. The positivist assumptions about reality are that there are social facts with a single objective reality, apart from individuals beliefs (Marcinkowski, 1993).

On the other hand, again according to Reichart and Cook (1969), the qualitative world view is characterised as social anthropological, inductive, holistic, subjective, and process-oriented. These types of researchers are often called phenomenologists and adhere to an interpretive paradigm. This views reality as subjective and socially constructed. The researcher/subject is placed within the context of the situation to understand it. The separation of the subject and object is reduced with the object becoming an active participant in the knowing process. The subjective researcher seeks to understand the world through the eyes of the respondent (Olsen, no date). In qualitative research there is a preference for a contextual understanding so that behaviour is understood in the context of meaning systems that are used by a specific group or community (Mishler, 1979).

Context and meaning in qualitative approaches are important to explain social action. According to Fogelson and Spiro (1965) an early pioneer of psychological anthropology, Hallowell, believed that one needed to consider psychological variables of cognition, perception and motivation. Put more simply, answers to three questions were essential for an understanding of the social world. What do the actors *know* about the world? How do they *perceive* the world? What do they *want* from the world? He believed that behaviour is not responsive to the objective properties of a stimuli, rather to their meaning. For Hallowell, an actor does not know or perceive *the* world, rather he knows or perceives *his* world, a world which is mediated through culturally constituted symbols. Sayer (1992) also supports this view by asserting that all social phenomena are *concept-dependant*. Unlike non-social objects they are not impervious to the meanings ascribed to them. What the practices, institutions, rules, roles or relationships *are* depends on what they mean in society to its members. Unquestioning use of every-day categories like occupation or ethnic groups is simply setting on paper issues that are not set in reality.

The assumption about truth in the phenomenological paradigm is that "there is no objective reality, apart from the knower, and therefore truth consists of a complex of value-laden observations and interpretations. As a result there is no a priori fact-

value dichotomy." (Marcinkowski, 1993, p 39). The assumption about reality is that multiple realities can be constructed through social processes.

In summary, the basic theme in positivist approaches is that the researcher views the social world from the outside, an objective reality is assumed, and social variables are described and compared without reference to what the observation of variables may mean to the subject. In phenomenological approaches the actors view is taken as the empirical point of departure (Bryman, 1984), and there are multiple realities. The role of theory also arises from the difference in the subject/object relationship. In a subjective approach, theory may be generated by the evidence during the study, while objective researchers devise hypotheses or theses prior to the study (Olsen, no date).

A summary of the debates or critiques concerning these two general approaches is useful to identify issues that may emerge from the current study. Logical positivists argue that research will ultimately be value-free as a result of increased quantification. Mathematics is a universal language, and positivism is based on the assumption that data can be identified and organised through quantification. They argue that because humanity is full of prejudices, one must use quantification to rid oneself of these constraints. Direct access to facts is achieved through "algorithmic or stepwise instructions that do not require interpretation". Positivists believe that value-free facts act independently, but the nature of interpretation of facts prevents their discovery. Therefore interpretation will only provide adulterated data and should be moved to the periphery of research (Murphy and Longino, 1992 p 144).

Critics of the positivist stance in sociology have written far more profusely in recent times. They reject the positivist approach for reducing social life to variables and the interactions between them. According to Blumer, (1956) in a highly influential critique of what he calls "variable analysis", argues that variables are selected according to specious impressions on what is important. This is usually on the basis of "conventional usage, on the basis of what can be secured through a given instrument or technique, on the basis of the demands of some doctrine, or on the basis of imaginative ingenuity in devising a new term" (Blumer, 1956 p 683). Blumer identifies three problems with the persistence with variable analysis. Firstly,

that there is a complete lack of rules or guidelines to govern the selection of variables. Secondly, that there is a lack of variables that represent abstract categories (generic variables). Without generic variables in empirical science, variable analysis can only result in separate and disconnected findings. Finally, variable relations ignore the context of the 'here and now' relations.

Blumer finishes his critique by drawing attention to several limits of variable analysis. He asserts that variable analysis, when dealing with human social life which demands the process of interpretation, ignores these processes. In addition, there is a deficiency which stems from the tendency to work with "truncated factors" which results in a concealment of the actual operations in group life (Blumer, 1956 p 688).

Guba and Lincoln (1994) identify additional problems with the positivist paradigm such as the dilemma where the theories brought in to be tested by the outsider, may have little or no meaning within the emic view of the peoples studied. Also, Guba and Lincoln draw attention to the fact that the traditional approach denies the researcher the discovery dimension by the insistence on testing a priori hypotheses; and that given facts have been shown to be facts only within a given framework, then the assumption that there is an objective reality is undermined. Finally, Guba and Lincoln discuss the unavoidable interaction between the subject and the object in social research. They assert that findings are discovered through the interaction between the researcher and the phenomena and that an objective observation is not plausible.

Each of the epistemological stances have a tendency to chose specific techniques that lend themselves to that stance. For example, positivist social research *tends* to use surveys while phenomenology *tends* to apply participant observation, unstructured interviews, or perusal of documents, for example. Often the debate about qualitative versus quantitative research gets lost in debates over technique, rather than debates over epistemology (Olsen, no date). The question should not be whether surveys are better than participant observation, rather, are surveys an appropriate technique in terms of a particular set of epistemological premises (Bryman, 1984). That is, does the selection and application of technique emerge

from the epistemological stance? It is, of course, possible to select a technique from a qualitative or phenomenological paradigm and apply it in a positivist or empirical fashion.

Marcinkowski (1993) distinguishes between two notions of research design. One, pure or formal designs which are the "prototypical" designs discusses in textbooks. The other is the "strategic" designs which reflect the research problems and questions one works with. According to Marcinkowski, the literature describes strategists as pragmatists who see research design as primarily a matter of fitting techniques to a specific problem. And there are a great number of studies conducted using design features of both paradigms, and a great number of commentators committed to combining methods.

Nau (1995) suggests that assertion to marry the two methods are post-positivist, and worthwhile. Nau thinks that as the two methods represent two ways of knowing, research purists should listen to another voice which "may serve to broaden our understanding of ourselves in a way heretofore obscured." (Nau, 1995 p 1).

Becker (1996) says that both types of research try to see how society works, to describe social reality, to answer specific questions about specific instances of social reality. Some scientists are interested in very general descriptions, in the form of laws about whole classes of phenomena. Others are more interested in understanding specific cases, how those general statements worked out in this case. In these terms there is a lot of overlap. However, the emphases are different. Qualitative research focuses on the understanding of specific historical or ethnographic cases, whereas quantitative work focuses on general laws of social interaction. The two style imply one another: "Every analysis of a case rests, explicitly or implicitly on some general laws, and every general law would suppose that the investigation of specific cases would show the law at work."

Becker (1996) describes two circumstances that explain why we think there is a difference between the two approaches. One, that the methods typically raise different questions at the level of the data on the way to generalisations of social life. The survey worker looks for quantitative or numerical differences between groups of people differing in interesting ways along some dimension. Their

argument consists of an 'explanation' of something based on a logic of difference between groups with different traits. For example people with certain socio-economic traits will be able to afford health care. They are interested in the way some variables condition the relationship between other variables. On the other hand, the fieldworker or qualitative researcher makes you aware of the constructed character of the 'variables'.

The complexity of this debate is evidenced in the body of literature dedicated to discussions on epistemology, method and technique. And confusion is apparent at many points. Bryman (1984) points to common debates in the literature that demonstrate a confusion between technique and epistemology. He argues that these are misguided discussions. For example, the assertion that qualitative research is inherently exploratory and therefore is well suited to discovery of new hypotheses to be verified by quantitative research. This assumes that the overall framework is positivist. The idea that qualitative research provides ideas is a technical issue, not an epistemological one.

Another example is the suggestion that it is best to combine methods as both quantitative and qualitative methods are complementary. Arguments for triangulated strategies are technical ones. Bryman argues that the combination of methods may indeed produce superior research, but debates about quantitative methods are largely epistemological, and asks how we can agree about the distinctions on an philosophical level, yet still ask to mix them. On a technical level the methods may be compatible, but on a epistemological level, the methods are written about as if they are not.

It is beyond the scope of this work to explore these debates further, however, the discussion has shown the importance of stating the epistemological stance in this study, and to identify the methodological issues that have emerged in this case.

The design of this study was based on a positivist stance. Hypotheses were constructed for testing, with the researcher as the subject identifying objective signs of wealth in the object or community. There is an inherent assumption in this work that there is an objective reality and that the researcher will describe that reality. There is an additional assumption that the construction and use of indices of socio-

economic status generally are valuable and meaningful. Qualitative methods were applied within a positivist paradigm using content analysis for the focus group interviews and 'number of times mentioned' in the wealth ranking. The mixing of qualitative and quantitative *techniques* (including the analyses of data collected) are not in conflict with this positivist paradigm.

However, it is not quite so clear-cut and that there are overlapping areas and potential confusion between research paradigms when using wealth ranking and focus group interviews to construct survey indices. First of all, the study was designed to work with a community construct of socio-economic status, and to extract indicators of this construct from the community members themselves. Thus, it could be said that the study was 'seeing the world from the actors viewpoint'.

Was this a phenomenological approach after all? I would argue it was not. The inherent assumption of an objective reality prohibits this assertion. The design of the study was such that indicators of wealth, as identified by the community, would be used to construct an index of socio-economic status that would be validated by comparison with another measure, in this case the wealth rank as allocated by community respondents. Although we are comparing constructs identified by the actors themselves, we are doing this within a 'hypothetical-deductive, objective, and outcome-oriented' framework (Reichart and Cook, 1969). Other features of the study rule out the phenomenological approach. The study did not attempt to explore the meaning or context behind the indicators or symbols of wealth. Rather, it allocated importance to indicators during the wealth ranking by quantifying them according to the number of times mentioned. It did not seek to *understand* the construct of wealth, it sought to *describe* it using indicators cited by the community members. It assumed that these indicators would represent a generality of some sort, that the indicators would be useful at the district level. It did not attempt to understand the symbols and signs of wealth at the household or village level in order to develop a theory of wealth or poverty (e.g. causes or expressions of wealth/poverty). After examination of the results, and reflection on the epistemological foundations of the study, it is clear that there are some tensions between methods (technique and epistemology) applied in this study.

Assuming that the positivist approach is plausible to begin with, the selection of wealth ranking techniques and focus group discussions may not have been the most appropriate choice to identify indicators of wealth that could be useful at a district level. First of all, the logic of the technique of counting the number of people who refer to a particular theme (which is used to prevail over allegations of researchers only providing fragments of data their support their argument) may have a fault. We then must ask, how many people must refer to a theme (or indicator) before it is deemed important? And in a smallish qualitative study it is unlikely that the views of those few that are interviewed will reflect the views of the community at large (Wainwright, 1997). I have discussed other issues of respondents citing indicators earlier. Therefore, to translate these into variables for quantitative survey may be misguided. According to Wainwright (1997) in qualitative research it is the quality of the insights that are gained rather than the number of respondents who share it. The concern is with the explanatory power of the data rather than it representing a commonly held view. In this way, we have moved away from a positivist paradigm in that we have taken a technique that may have been more suited to developing an *understanding* of the concept of poverty and tried to apply it in a quantitative fashion – at the same time ignoring the demands of generalizability inherent in the positivist paradigm.

The fact that the community-based indices did not perform well on validation may lead us also to question the choice of method to generate indicators on epistemological grounds. A positivist criteria for validity as used in this study, may have been less appropriate for the types of knowledge that wealth ranking or focus group discussions were most probably originally designed for. According to Altheide and Johnson (1994) qualitative research is carried out in ways which are sensitive to the nature of human and social contexts, and is commonly guided by the ethic to remain loyal or true to the phenomena under study, rather than to any particular set of methodological techniques or principles. For qualitative research, the general approach to assessing the value of the work is where the process by which the work occurred is clearly delineated, including accounts of the interactions among context, researcher, methods, setting and actors. Then four criteria are used to judge the quality of the work: plausibility, credibility, relevance and importance

of the topic. So once again, there may be some tension between the choice to use techniques that are fundamentally designed for a phenomenological or interpretive paradigm and then judge their value using a positivist paradigm stressing reliability and validity.

All this should beg the question as to whether wealth ranking is suitable for such a 'distortion'. Perhaps its usefulness is only when applied with a truly subjective, inductive purpose, and there is limited use for application in a positivist, objective research paradigm. Yet, as we have seen in chapter one, wealth ranking has been applied within a positivist framework to identify indicators of wealth. A deeper understanding of the behaviour of the technique within different epistemological paradigms would be helpful.

For the purposes of studies such as this one, it may have been more valuable, in order to remain within a positivist paradigm, to use alternative methods to identify locally relevant indicators. Using objective techniques, identifying the signs and symbols (not meanings) of wealth within the community, then using correlation coefficients to select those indicators which are found more frequently in wealthier/poorer households. Although this would have resolved the epistemological tension in this work, it would have taken the development of locally-specific indicators outside the reach of district level health workers.

Although I subscribe to a strategic design approach to research, it is important to expose the tensions inherent in this approach, and to identify the implications in doing so. I am not convinced that an individual researcher should to take a firm subjective *or* objective stance for all research problems. Indeed, as Olsen (no date) states, there is no real way of proving that there is one knowable reality or whether there are multiple realities of which we can have some individual knowledge. These things boil down to faith. However, what is required is that we are clear about our epistemic stances, and *aim* for consistency of methods within our chosen paradigm.

For this study, questions remain about the value of using wealth ranking to contribute to a survey index of socio-economic status, or more generally, to be subjected to quantitative analyses. The findings led support to 'the recognition of the irreducibly qualitative nature of some of the data which precludes its transformation into quantitative proxies' (Shaffer, 1996, p.25)

4.8 Endpiece: The Social Consequences of Measures and the Ethical Burden on Test Users

My closing comments are concerned with the broader implications of issues related to the validity of measures of SES used in health service delivery and health or medical research. This discussion emerges from what is now considered by mainstream measurement theorists to be a routine aspect of validity assessment, but one often ignored. Messick (1988) highlights the need to consider the social consequences of tests or measures as part of routine validity assessments. He argues that it is insufficient to make decisions on test validity simply on empirical grounds, and that it is essential to evaluate the intended and unintended social consequences of test interpretation and use. In an earlier paper, Messick (1980) discusses in detail two fundamental questions that need to be asked when assessing a proposed test or measure. Firstly, whether or not the test is a good measure of the characteristics it is interpreted to assess. This is a scientific and technical question and related largely to its properties such as reliability and construct validity. Secondly, should the test be used for the proposed purpose in the proposed way. This is an ethical question and to answer this question, it is necessary to assess the measure in terms of its social values. This assessment involves direct costs and benefits as well as likely possible side effects of applying the test in a given situation. Messick cautions us that just because a test may have good construct validity, this does not provide satisfactory answers to the second, ethical question.

The importance of this aspect of validity assessment requires emphasis, since it is often passed over. This study was not able to demonstrate a convincing degree of construct validity for either the community generated or traditional indices, and I have already recommended that application of the indices developed in this study using the wealth ranking technique is not attempted in the locations for which they were developed. Therefore, it is tempting to leave the discussion here and avoid a seemingly futile discussion about the social consequences of applying the indices for health service delivery. I feel that it is still worth flagging the issue concerning the interpretation and social consequences of measures of SES in general, and the

implications for users of measures of SES. This is all the more necessary because in many circumstances measures of SES or class are developed and applied in health service delivery or research without efforts to address either the technical aspects of validity, or the social consequences of applying the selected measure. What are the direct costs and benefits to the health service; the individual users, or the potential users of the health service? And, finally, what are the possible positive and negative side effects of using, or even not using the proposed measure?

Measures of SES are used in the health sector for a variety of purposes, with differing levels of social consequence if measures are invalid. I will briefly discuss some of the more important uses of measures of SES.

Studying Inequities in Health

Studies of equity in health status have been carried out in almost all settings, in both industrial and developing countries. No matter which way we measure SES, we almost always find the same result. Inequities exist in health, whether we use mortality, morbidity, access to, or utilisation of services as our dependent variable. Inequalities between social groups exist whether we use occupation, education, income, or composite measures of SES. Even with the bluntest instruments we can demonstrate clear patterns of inequity. For these purposes, it would seem that even a poor measure would not give rise to serious social consequences. As there are numerous studies carried out that confirm inequalities exist, a single study with converse findings, resulting from invalid measures, would hardly set in motion changes in public policy - the ultimate use of most equity studies. This is not to say that users of measures of SES concerned with measuring equity have no responsibility to ensure their measures are valid. Rather, that the ethical burden is not as high as in the following alternative applications.

Monitoring the Impact of Health Reforms on Equity

If the measures are being used to monitor the impact of health sector reforms, we would be more concerned about the accuracy of our measure. Invalid interpretations caused by poor measures may be used to 'sell' certain reforms on the agendas of policy makers, which may have a negative impact on the poor. Alternatively, as a

result of changing fashions in the reform package, support for reforms that have had a positive social consequence may be withdrawn.

Epidemiological Studies

In epidemiological studies examining the determinants of disease, SES is often measured. Like general equity studies, it is fairly clear now that SES is a major determinant of health status. It is generally accepted that we need to target health services towards the poor, and we now expect greater returns per health dollar invested when we target health services to the poor. Single studies identifying the determinants of ill health that use invalid measures of SES are unlikely to result in serious and direct negative social consequences.

Reaching and Protecting the Poor

Should we want to apply a new measure to decide which *individuals* should be exempted from health care charges, or which *individuals* should be targeted for a particular health service, then we have a rather different application for our measure. There may be disastrous social consequences if the measure is invalid. These consequences should be considered in terms of the individual families, and at a population level where a particular health service is misclassifying large numbers of the poor who require exemption from health care charges, or access to specific services. Conversely, wasting precious resources meant for the poor by incorrectly targeting the better-off is undesirable.

In summary, this study has highlighted the very real practical difficulties faced when attempting to identify the poor. Even if we accept that measuring equity is not required at the district level, targeting health services and exempting the poor from user charges are still major priorities of the Health Sector Reform movement, for effective and efficient use of health care resources. Even though the literature demonstrates that we have been struggling with this for well over one hundred years, many of the policy statements from the Health Sector Reform movement justify major changes in health service management - such as health sector financing and reducing the burden of disease - by protecting and targeting the poor. How are we achieving this?

We have an ethical obligation as the users of measures to ensure that they are measuring what we say they are measuring, especially those measures whose use has substantial social consequences. There may be fewer negative social consequences if we do not try to measure SES at all, than to measure SES badly.

Can we expect to see some of the energy of the Health Sector Reform movement put into such areas as successful identification of the poor, or at least assessing current strategies that claim to do so? If the reforms can demonstrate improvements in equity then this would provide valuable support for the reform strategies. Health Sector Reforms continue to face resistance from many commentators sceptical about the impact of the reforms on the poorest members of society.

Appendices

APPENDIX 2.1

Introduction for Wealth Ranking Sessions to Respondent Group Participants

Major Points	Rationale
1. Introduce all the team, their names, titles and organisations (facilitated by local leader).	Identify the government departments and research institutions that are collaborating to develop confidence in the exercise and the people present. Reduce suspicion.
2. The objective of the study is to learn more about how poor people can be identified because the common indicators normally used by government departments may not be appropriate for all villages in the country.	Develop a commitment to the exercise
3. This village is being used as an example only, and there will be no direct benefits received by the village, nor poverty alleviation programmes implemented as a result of this study. The study aims to be of benefit to Tanzanians generally.	Eliminate expectation of future direct benefit and therefore disappointment with no apparent projects implemented
4. We are not directly interested in *who* is poor and rich in this village, rather *how* you know who is poor or rich. We do not keep the names of the households we discuss, but record them as numbers.	Reduce false reporting motivated by a desire to present oneself or ones relatives and friends as poor in order to receive some perceived benefit
5. Instructions for ranking households	Efficient and effective session
6. Instructions or 'rules' for encouraging participation by all in discussion	Efficient and Effective session
7. Clarification of the terms used to reflect the four socioeconomic groups the households will be ranked into. Starting with the term *uwezo* the respondent group identifies appropriate labels for the piles through a discussion with, and among, participants until agreement is reached (Note participants often attempted to describe the criteria for classification in a very general way; group labels were quite consistent across research sites)	To ensure that the general concept reflects the breadth desired by the use of the term *uwezo* but that socioeconomic groupings themselves can be labelled in any way the respondents identified Avoid intra-group confusion about concepts and the labels applied to them
8. Clarification of the term household	Avoid intra-group confusion about concepts and the labels applied to them
9. Questions and answer session	Reduce suspicion and encourage efficient and effective session with maximum quality of data collected

APPENDIX 2.2

Example Focus Group Discussion Schedule

Note: This schedule is an example only. The nature of this qualitative method is that it changes and develops as data is collected and a deeper understanding of the issues is gained. All schedules included these basic questions, but several other questions were posed in different sites that were late dropped from the analysis and consequent discussion of results. This schedule provides all those questions that were posed in all sites and are a part of the rest of this phase of the work. The language is also slightly stilted as this is a translation of a schedule written and conducted in Kiswahili.

Facilitator: *Once the Wealth Ranking activity is completed, the following discussion is conducted and recorded on tape.*

COMMON CHARACTERISTICS QUESTIONS

1. Let us describe these groups as piles, not as individual households. Let us start for example with the group of people who are very poor. What are the common characteristics that made you put them in this very poor pile? You can just mention what items or things they have in common. Or what are their lifestyles that made you put them in this group?

2-4. *Facilitator:* *Repeat for all socioeconomic groups (1-4)*

5. Now for example, are there other things which you have not yet mentioned? You have mentioned [list criteria]. Are there other things in this village that may be used to rank people according to their wealth? We have mentioned many other things for the people we have been discussing, but what about the whole village, are there other things which would identify somebody as rich or poor man? Are their other indicators you have not yet mentioned?

RESPONSE CATEGORY FORMATION

6. Now we have discussed about various indicators of wealth that is for very rich man, comfy man, the poor and the very poor. You have mentioned for example, [criteria]. Now let us talk about [criteria] for the very rich man, how is [criteria] for them?

7 –>. *Facilitator:* *Repeat for as many criteria listed as possible avoiding respondent fatigue. Start with the most important criteria and move down the list prepared during the WR session.*

APPENDIX 2.3

Community Criteria
Peri-urban: Rufiji

Example Only : First 10 Households

HH #	Leaders	Poorer Men	Richer Men	Poorer Women	Richer Women
1	4	3	3	4	3
	Sells foodstuffs in stall	Sells fish gets money	Stalls sells foodstuffs	House in bad condition	Struggling to get food
		to help in farming	Farmer	Sells fish in market	Buys and sells fish
		Thatched mud house	Mud and thatch house	Small income fr farm	Farmer
		Poor condition			
2	4	4	4	4	4
	Depends on farming	No wife	No business activities	No wife	No farming activity
	which earns income	Kibanda	Depends only on farm	1 roomed kibanda	No IGA
	only annually	Depends on casual lab	1/4-1/2 acre for the year	Cooks & sleeps there	Traditional Healer
	Depends on casual lab		No wife/kids to help	Doesn't cultivate crops	
	At risk if farming fails			Conman to get food	
3	3	3	3	4	2
	Farmer	Iron roofed house	Depends on carpentry	No job	Farm 2-3 bags paddy
	Carpenter work subsid-	Carpenter		Children not enough food	Temp. carpenter
	ises farming			Iron roofed banda is	Banda w/ 3 rooms
				sisters	Iron sheets and mud
				Farms crops	
				Cuts poles and sells to	
				get food	
4	4	3	3	3	2
	Farmer irreg income	Farms paddy	Small income from	Gets 10-30 bags paddy	Farmer no other busine
	No other income source	Iron roofed house	farming	Iron roofed house/mud	Thatched house
		Thatched house	Doesn't last the year	Depends on farming	Small iron roofed house
		Assistance from kids			15 bags paddy
5	4	4	4	4	4
	If farm doesn't generate	Poor condition kibanda	Farmer	Depends on casual lab	Cuts poles and sells
	income does casual lab	No wife	Low income	No good place to live	Depends on this
	Full suit house	Cuts poles and thatch	Poor condition house	Can't believe what he	Sleeps in brothers
	Cannot smear with mud			eats	banda in poor condition
					Not married
					Farmer
6	4	4	4	4	4
	Farmer	Sickly can't work	Sickly with asthma	Sick	Wife farms & feeds fam
	Sick with asthma	Family depends on wife	Cannot work	Can't cultivate land	Sick with asthma
	Wife cultivates land	Children still young	Depends on wife to	Family dependant on	No IGAs
	House built for him	Wife sells cooked buns	grow food	wife	Tin roofed house in
	Kibanda			House poor condition	poor condition
				Children still young	
7	2	2	3	2	2
	Cultivates food/cash cr	7 acres cashews	No capacity to run	3 acres cashews	Goats
	Can maintain life	Good yield	business	Many goats	Cashews
	throughout the year	Goats	Depends on cashews	Good income fr paddy	Paddy
	Cashew - good income	Cultivates paddy using	and paddy	14 bags yield	House not good
	Sickly	casual labour	House poor condition	House poor condition	Thatched
	Doesn't depend on food		Sickly		
	crops				
8	4	4	4	4	4
	Farmer food crops	Renting a room/no house	Fisherman low status	Bird has more respect	Fisherman
	No cash crops	New to village	Still young	Has wife and kids	Rents a room
	Fisherman poor tools	Depends on fishing	Rents a room	Depends on traditional	Small farming
	Depends on fishing		Irregular income	ceremonies for food	
	for daily income			Doesn't work hard	
				Dances in trad ceremon	
				Rents a room	
				Irregular fisherman	
9	3	4	3	3	4
	Works in coop farm	Casual labourer but	Employee	Works in coop farm	Small scale farming
	Earns salary	paid reguarly	Casual labourer	Not married	2 sacks food
		Cultivates own food crop	Cultivates paddy/maize	Rents a room	Rents room
					Earns very little

APPENDIX 2.4 Statistics Applied in this Study

1. Testing Reliability: The Kappa Statistic

Kappa statistics were used to test respondent group reliability in ranking households by socio-economic status, the reliability of all respondent groups to rank specific socio-economic groups, and the performance of each respondent group in terms of their level of agreement with the other respondent groups. In addition, it was used to measure the variation in criteria generated by the different respondent groups within and across all study sites.

The Kappa statistic is based on "a series of N subjects from a larger population of subjects, with m independent observations for subject i (i = 1,2,....,N) sampled from a population of possible observations for that subject. Each observation is a classification of the subject into one or more of K response categories" (Kraemer, 1980, p. 207). In this situation, this translated into 50 subjects (or households) with 5 independent observations (respondent groups or judges) for subject i (i = 1,2,....50) into four possible observations of rich, comfortable, poor, or very poor.

To calculate the agreement simply, it is possible to identify how many exact agreements there were as a proportion of the number of possible agreements. This is an inadequate measure because it does not show where in the table of frequencies the agreements were; and also, as we should expect some level of agreement by chance alone this measure does not tell us anything about the proportion's relationship to chance agreement. Therefore it is useful to take a further step to determine the expected cell frequencies in the table that would be found with a chance agreement, and then determine how much better our raters level of agreement was than this chance agreement (Altman, 1992). The details of the calculation applied in Crunch 4 are detailed below. Correlation coefficients such as Pearson's Product Moment Correlation, although occasionally applied, are unsuitable for ordinal data (Cramer, 1994)

Since the original Kappa developed (Cohen, 1960; cited in Kraemer, 1980), the Kappa Coefficient has been extended to consider a number of different needs such as weighting of the severity of the disagreements. These were applied in this study. And finally, as mentioned, the Kappa statistic is applied in this study to determine the measure of agreement on particular socio-economic categories in the coding scheme. Kappas for multiple categories are expressed as "a weighted combination of the measures of agreement in each category" (Posner *et al.*, 1990). Both these calculations are explained in more detail below.

There are some limitations of the Kappa statistics and their interpretation which should be briefly reviewed as to their effect on this study. The most commonly referred to problem is that the Kappa value will depend on the proportion of subjects in each category, because of the possibly very different expected frequencies. Therefore it can be misleading to *compare* Kappa values from different studies (Altman, 1992). With this limitation in mind for the current study it is not useful to compare the results from the 10 sites *statistically*, nor aggregate overall Kappa values, rather, to look at the individual Kappa values independently for each site. A second limitation is that the Kappa value can be effected by the number of categories used. A six category response value will be lower than a four category response value. This limitation does not effect this study in that there are fixed categories used, but it should be noted that if I had applied three in Phase One the results would be stronger. The third important problem with interpretation is concerned with the magnitude of the Kappa value.

Posner *et al.*, (1990) show that the earlier discussions on the Kappa value were concerned with the null variance and whether or not the value is greater than zero, but argues that when we are interested in inter-rater reliability in most circumstances where raters have been trained to perform the task such as in medical diagnostics, we should be more than surprised if values were close to zero or negative! This brings the problems of interpretation of the magnitude into play. The interpretations of the strength of agreements are detailed below, but this is an important point in this study. Most research applying the Kappa Coefficient, are concerned with raters such as physicians or advanced diagnostic procedures that are expected to have a high level of agreement. In this study, we are investigating the natural process of categorisation of households into socio-economic categories by their neighbours acting as judges. Not only have these judges not been trained in social stratification, but they were not given any criteria by which to allocate a rank. With this in mind, we should view the interpretation of strength of agreement with some level of conservatism, and expect lower Kappa values than found in many applications, especially in medical diagnostics.

I selected the software package Crunch Statistical Package Version 4 and used the Kappa Module. The benefit of this package is that it can handle inter-rater agreement for up to 100 judges, whereas many programmes handle only two. In addition the Crunch 4 Kappa Module was designed to carry out all tests as described within the one programme.

The following Kappa Statistics were calculated according to Crunch Version 4, Kappa Module Update (1994). All Kappas presented are weighted (disagreement not agreement weights) meaning that there is a higher penalty, or calculated level of disagreement, if a household is ranked by one judge as "rich" and another "very poor" compared to non weighting where a discrepancy carries the same weight whether it be a discrepancy across one or two socio-economic groups. Squared, rather than absolute deviations were used as recommended by Streiner and Norman (1989).

The Observed Percentage Agreement for the weighted analyses were calculated and, according to Crunch Version 4, can most accurately be described as:

$$100 \quad X \frac{(1 - \text{Total Observed Disagreement})}{\text{Maximum possible total disagreement}}$$

Maximum possible total disagreement is the total disagreement resulting when each pair of judges rates a target in a way that resulted in the largest possible disagreement weight.

Chance percent agreement is the value that is at the heart of the generalised Kappa computation and is based on calculations used by O'Connel and Dobson (1984).

The final *Generalised Kappa* is then computed as:

$$\frac{(\text{Obs prop agreement} - \text{Chance prop agreement})}{(1 - \text{Chance prop agreement})}$$

The P Value: there is an asymptotic test based on referring kappa/standard error of kappa to a normal distribution. Note that it is often appropriate to report a one-tailed P-value when the alternative hypothesis is: agreement is better than chance.

In accordance with the suggested labels for Kappa magnitudes from Landis and Koch (1977) and also discussed by Posner *et al* (1990) the following labels were applied:

Kappa < 0	Poor
0.0 <=kappa <0.2	Slight
0.21 <=kappa <0.4	Fair
0.41 <=kappa <0.6	Moderate
0.61 <=kappa <0.8	Substantial
0.81 <=kappa <1.0	Almost/perfect

The Individual Judge Scores were examined for the five respondent groups to search for confirmation of certain judge groups being more reliable such as community leaders and women, and to see if any pattern of performance emerged over the 10 research sites. The Crunch software package calculates this following the guidelines of Williams (1976) as:

Judge *i* is singled out:

1) *Pi*, the proportion of times that judge *i* agrees with all the other judges; and,

2) *Qi*, the proportion of times the judges (other than *i*) agree with each other. The ratio of these two quantities is then an index of the particular judge's agreement level, relative to the other judges.

For each judge, the results will show *Pi*, *Pi Qi* (call Index) and approximate 95% confidence limits for the index. If the confidence limits are below 1, then this judge is clearly below the level of the others. Since it is of most interest to identify the worst judges, an upper 95% one-tailed limit is also displayed. These indices are computed for each judge and sorted best to worst.

2. Testing Convergent and Discriminant Validity of Indices: The Multitrait, Multimethod Matrices

This validation method is concerned with *convergent* and *discriminant* validity. These two notions of validity were introduced by Campbell and Fiske (1959) in an article that outlines the MTMM validation method. Convergent validity refers to the concept that if one attempts to measure a trait (such as socio-economic status) by methods which are very different, then we would expect that the results from the two methods to produce similar results. In other words, different approaches to measurement should converge when measuring the same trait. In terms of *discriminant* validity, one would expect that if these measures were valid, then they can discriminate between two different, unrelated traits. That is, when we are measuring two different traits by the same method we would not expect to see a high correlation which if present would be due to methodological contamination (Sullivan and Feldman, 1994). Campbell and Fiske (1959) also point out that more than one trait and more than one method is required to estimate the relative contributions of trait and method variance.

What we have seen is that two very different methods and two very different traits are required to conduct this validation procedure. For the ease of relating the following discussions to the current study, I will now outline the methods and traits applied. As stated, the methods used must be different, for example, they should not be two survey questionnaires with different items. So, I have chosen the wealth ranking method, where the community has ranked the study households into three socio-economic groups for the first method. The second method is the household survey which represents the study indices, both community generated and traditional. Therefore the analysis will construct a MTMM matrix for *each* index to be validated. For the traits, it was necessary to identify two traits that were unlikely to be correlated, and importantly, that could be measured by the two methods; wealth ranking and household survey. This initially posed a challenge as it was difficult to identify a trait that was expected to be *uncorrelated* with socio-economic status, the primary trait of interest. The reason for this difficulty was that the second trait of the household would need to be

known by respondent groups ranking households by that trait. After much failed deliberation, I decided to consult key informants within the study districts to identify a household trait that would be a) known by all respondents, and b) not related to socio-economic status. Community informants identified skin colour as an appropriate trait. On further enquiry it was found that people do in fact commonly use skin colour as a way of describing each other for purposes of identification, and, that three such skin tones are described using common Kiswahili vocabulary. These words are *mweupe* (white); *maji ya kunde* (brown, lit. bean water) and *mweusi* (black). This was chosen for it's ability to meet the two criteria, and because it was able to fall into the same three category ordinal variable with ease. It was also considered suitable as community informants identified this as a trait that would not cause sensitivity with respondent groups during WR sessions, nor survey respondents. Finally, the second trait was defined as the *skin tone of the head of the household.*

It is useful to show Campbell and Fiske's Matrix as applied to this study to facilitate further discussion. The data is fictional. The method "Household Survey" would represent *one* of the study indices to be validated.

Table 1 *Multitrait, Multimethod Matrix (Synthetic Data)*

		Community Rank		**Household Survey**	
	Traits	**SES**	**Skin Tone**	**SES**	**Skin Tone**
Community Rank	**SES**	(1.00)			
	Skin Tone	{0.19}	(1.00)		
Household Survey	**SES**	0.52	[0.16]	(1.00)	
	Skin Tone	[0.11]	0.57	{0.22}	(1.00)

Table 1 shows the two traits, each measured by two different methods, resulting in six correlations from the four separate variables, excluding the reliability coefficients. The brackets and shading assist in the identification of these four variables. In the curved brackets "( )" the reliabilities are located and can be best described as the "monotrait-monomethod values or correlations. Next, in the curly brackets "{ }" are the heterotrait-monomethod values. In the square brackets "[ ]" the heterotrait-heteromethod correlations are found, and in the shaded areas are the validity diagonals, or monotrait-heteromethod correlations. Before one can say that the new measure has potential, it is essential to meet the following three criteria as described by Campbell and Fiske (1959 p 82-83).

1) The values in the validity diagonals (shaded areas) should be sufficiently different from zero and large enough to encourage further examination of validity. The size of the validity values is a debatable point, but for this study, I have chosen $\geq .40$. In a discussion on these validity coefficients on a MTMM matrix, Sullivan and Feldman (1994) accept a the range of coefficients in a study with validity coefficients of .18 to 0.69, although they identified some reservations about the lower coefficients. My figure of 0.40, although not overly conservative, is based on the complex nature of the concept that we are measuring, and high values, although desirable, are not expected. In addition, the type of correlation coefficient employed

(Kendall's Tau) results in lower values than other coefficients available. A complete description of Kendall's Tau will be given later. Tests of significance were carried out on all values.

2) Each validity coefficient should be larger than *all* of the heterotrait, heteromethod correlations which are in the same rows or columns as the validity coefficient itself. For the above table, all shaded areas should be larger than the "[ ]" correlations.

3) Each validity coefficient should be larger than the heterotrait-monomethod correlations which involve the same variable as the validity coefficient. If this is not met, then it is suggestive that the *method* was more important than the *trait* being measured.

A fourth criteria is required where the same pattern of relationships should be seen in the heterotrait triangles (monomethod and heteromethod blocks) which can be found only in matrices with more than two traits and two methods. As can be seen, the example above has met the criteria required to suggest the measure has potential.

Kendall's Tau β for the MTMM Matrices

A correlation coefficient was required for the construction of the MTMM matrices. Kendall's Tau β is based on a comparison between the rank allocated by two variables to a number of subjects. It represents "a function of the minimal number of inversions or interchanges between neighbours which is required to turn one ranking into another" (Seigel, 1956, p. 215). The Kendall's Tau β was selected as it has a number of desirable properties that were appropriate for the data (Bailey, 1995). Firstly, the coefficient would not require specific numerical measurements, but was one that could deal with observations of a qualitative, subjective nature that could be arranged in an ordered series, in our case, the three socio-economic groups and three degrees of darkness of skin tone. Secondly, it is desirable that the coefficient will lie between +1 (complete association) and -1 (reverse) and 0 when SES and skin tone is distributed independently of the other.

In addition, Kendall's Tau β can also manage the situation where there are tied ranks, although the computation becomes far more complex (see Seigel, 1956, p. 217 for a discussion on calculations dealing with tied ranks). With studies where $n \geq 10$ this complication is reduced somewhat. In addition, Kendall's Tau does not require that the data be normally distributed. Shott (1990) also described how Kendall's Tau can successfully test the null hypothesis that there is no relationship between the two variables, whereas Spearman's Rank correlation is difficult to interpret when the null hypothesis is not true, and, is also subject to a number of more rigid assumptions.

3. Sensitivity and Specificity

These tests are based on the fundamental assumption that the community rank (using the wealth ranking technique) is a suitable "gold standard" or criterion by which to assess the accuracy of the community generated and traditional indices. For this validation procedure, I have decided to focus on the ability of the test to accurately identify the very poor households. Therefore, households were re-categorised into the "very poor" (SEG 3) and the "better-off" (SEG 1 and 2). Table 2 gives an example of how this procedure is carried out during the analysis stage. These tables were constructed separately for each of the study indices after transformation.

Table 2 *Sensitivity and Specificity (Synthetic data only)*

Index: TRAD 1	**COMMUNITY RANK**		**TOTAL**
	Very Poor	**Better-Off**	
Very Poor	29 (cell a)	5 (cell b)	34
Better-Off	2 (cell c)	11 (cell d)	13
TOTAL	31	16	47

According to Hennekens and Burning (1987), Sensitivity is calculated as $a/(a+c)$ and is interpreted as the probability of testing very poor if the household is "truly" very poor. As the sensitivity increases, the false negatives (ie: the number of houses incorrectly classified as better off), decreases. In this example, the Sensitivity would be: 29/(29+2) = 93.5%.

Specificity is calculated as $d/(b+d)$ and is interpreted as the probability of testing better-off if the household is "truly" better-off. A highly specific test will result fewer false positives (ie: a smaller number of better-off households being falsely classified as very poor). In this example, the Specificity would be: 11/(5+11) = 68.8%.

There is always some trade-off between sensitivity and specificity as the more able a measure is to identify the very poor correctly, some better-off households will be erroneously categorised as very poor. This is the result of a lack of clarity for some households in terms of actual status. For our purposes, it would seem reasonable to require a higher sensitivity as the implications are more serious if a household is incorrectly categorised as better-off when it is indeed very poor. Conversely, incorrectly identifying a household as very poor that is in fact better-off may be undesirable from an efficiency perspective, but less dire from a protection of the poor point of view.

APPENDIX 2.5

Criteria Generated from Wealth Ranking Exercises in Hai District

ITEM	Peri-urban Site	Rural Site	Total	Prop %
Work, Income, Capital				
Occupation	42	102	144	9.31
Employee in Household	18	6	24	1.55
Does not Farm	2	3	5	0.32
Income/able to afford *x*	12	10	22	1.42
Assets to generate income	4	4	8	0.52
Dependence on Casual Labour	30	50	80	5.17
Variety of Sources of Income	12	17	29	1.88
Regularity of Income	0	6	6	0.39
Housing Related				
Condition of House	27	3	30	1.94
Roofing Material	24	3	27	1.75
Wall material/Paint	32	13	45	2.91
Flooring Material	1	0	1	0.06
Quality of House	13	2	15	0.97
Number of Houses Owned	0	3	3	0.19
Ownership of House Plots	1	2	3	0.19
Tenure of House	21	83	104	6.73
Type of House	5	0	5	0.32
Size of House	9	0	9	0.58
Quality of Doors	1	0	1	0.06
Surroundings of House	3	0	3	0.19
Ownership of Rental Property	0	10	10	0.65
Size of House Plot	9	0	9	0.58
No. Houses in Compound	2	0	2	0.13
Permanency of Residence	0	2	2	0.13
Farm Related				
Crop Yield	16	1	17	1.20
Condition of Farm	17	0	17	1.20
Type of Crops	179	16	195	12.61
Size of Farm	110	16	126	8.15
Tenure of Farm Land	54	26	80	5.17
Progressivity of Farm Technique	3	0	3	0.19
Maturity of Crop	3	0	3	0.19
Farm Tools	0	1	1	0.06
Farm inputs or investments	0	1	1	0.06
Shop Related				
Ownership of a Shop	0	4	4	0.26
Shop Stock	0	1	1	0.06
Tenure of Premises	0	1	1	0.06

ITEM	Peri-urban Site	Rural Site	Total	Prop %
Family Composition				
Widow	9	1	10	0.65
Marital Status	16	3	19	1.23
Orphaned	2	1	3	0.19
Father Died	6	0	6	0.39
Female headed HH	11	7	18	1.16
Number of Children	10	10	20	1.29
Has Children	3	1	4	0.26
Size of Family	6	0	6	0.39
Lives Alone	5	0	5	0.32
Physical Status				
Age	43	38	81	5.24
Physical Status	13	15	28	1.81
Health Status	9	4	13	0.84
Disability	13	1	14	0.91
Alcoholic	1	0	1	0.06
Dependency Status				
Assistance from Children	18	12	30	1.94
Assistance from Others	4	8	12	0.78
Dependence on Children	1	5	6	0.39
Dependence on Others	15	6	21	1.36
House Built by a Relative	1	4	5	0.32
Living with Relatives/others	9	3	12	0.78
Ability to Assist Others	1	1	2	0.13
Dependant Relatives	7	3	10	0.65
Dependency of Children	15	5	20	1.29
Access to Utilities				
Toilet	1	0	1	0.06
Tap Water	1	3	4	0.26
Telephone	0	1	1	0.06
Electricity	0	4	4	0.06
Livestock Related				
Ownership of Livestock	23	4	27	1.75
Number of Livestock Owned	1	2	3	0.19
Type of Livestock Owned	15	10	25	1.62
Miscellaneous Criteria				
Clothing	3	0	3	0.19
Education	31	4	35	2.26
Time Living in Area	9	0	9	0.58
Material Possessions	2	0	2	0.13
Work Ethic	27	3	30	1.94
Ability to Plan Development	7	2	9	0.58
Food Security	7	4	11	0.71
TOTALS	**995**	**551**	**1546**	**100.00**

APPENDIX 2.6

Example of Focus Group Discussion Transcript Coded Statements
Questions relating to the Common Characteristics of the different SEGs

---

SITE 9: LEADERS RESPONDENT GROUP - Hai District, Rural Site

SEG 1 (Rich)

livestock 1
the people who are calm, they are not aggressive or trouble makers 1
there are those who are employed 1
able to maintain their farms very well 1
been able to invest here in the village from their salaries, like building houses or buying cars 1
children are well and are being educated 1
on the side of food crop, they also harvest a high crop yield 1
say coffee, they harvest a high crop yield 1
Some even have tractors for farming 1
the children in R households are healthy 1
They are educated people 1
They would have tapped water in their houses 1
worked hard and been able to build modern houses 1

SEG 2 (Comfortable)

The only thing which is lacking for them is the capital to start off whatever ventures they want to undertake 1
These are also people who work hard 1
They are people who seek for different opportunities 1

SEG 3 (Poor)

if they had capital, they could improve 2
It is not enough to make further progress, for example money which may be required for cultivating land or buying pesticide to spray the coffee 2
most of them subsidise with casual labour 2
On the side of income, they earn little 2
people with little education 2
They have a low income 2
They have very small land areas 2
usually borrow land to cultivate food crops 2

SEG 4 (Very Poor)

gets any help from the children, then they are satisfied with whatever they get 2
Some are physically disabled 2
the old people, they are not physically strong enough to work 2
they come to your house and ask for a portion to work for you 2
they don't think of tomorrow 2
would [not] have any income with which to make developments 2
[some are] those who don't care very much about their lives 2

APPENDIX 2.7
Hai District Questionnaire: English version

LSTM/ISS/AMREF/AMMP Validation Survey - Hai District

Name.....**Household ID No**..... **Balozi**.....

Interviewer..... **Date**..... **Time**.....

INTERVIEWER SECTION	DATA ENTRY SCORE CODES	
1. Occupation/Source of Income (Traditional Index: not analysed)		
What is the main souce of income of the head of household or of the person responsible for the wellbeing of this family ? *(indicate if respondent doesn't know)*	Enter Occupation Only	
2. Employees in Household (Community Index)		
How many people are employed in this household ? (including head of household) *Write the total number in the box.* 0 = none 99 = don't know ☐☐☐☐☐☐☐	0 through highest	= 2 = 1
3. Income Generating Activities (Community Index)		
How many other activities were done by the family to earn extra income for the past 12 months apart from main occupation of head of household or responsible person ? *Write total number in the box* 0 = nil 99 = don't know ☐☐☐☐☐☐☐	0 or 1 2 through highest	= 2 = 1
4. Dependence on Casual Labour (Community Index)		
How many months in the past year did your family depend on the casual labour for the neccesary requirements such as food, shelter and clothing ? *Write total number of months in the box* 0 = don't depend on casual labour 99 = don't know ☐☐☐☐☐☐☐	0 or 1 2 through highest	= 2 = 3

5. Tenure of House (Community and Traditional Index)

What is the ownership state of the house you are living in with your family ? *(Read aloud all options)*	Community	Traditional
1.My property (didn't inherit)	= 2	= 2
2.Inherited	= 2	= 2
3.Tenant(whole house)	= 2	= 2
4.Inherited (children's or other people's property)	= 3	= 2
5.Rent room(s)	= 3	= 3
6.Living with relatives or friends	= 3	= 3
7.No permanent shelter	= 3	= 3
8.Other.....	= system missing	
9.Don't know _____	= system missing	

6. Roofing materials (Community and Traditional Index)

What is the largest part, and state of the material used for roof thatching on your house ?	Community	Traditional
1.tiles	= 1	= 1
2.galvanized iron sheets in good condition	= 1	= 1
3.galvanized iron sheets in poor condition	= 2	= 1
4.coconut leaves in good condition	= 2	= 2
5.coconut leaves in poor condition	= 3	= 2
6.grass in good condition	= 2	= 3
7.grass in poor condition	= 3	= 3
8.no permanent shelter	= 3	= 3
9.other.....	= system missing	
10.don't know _____	= system missing	

7.Types of roofing material used
(interviewer's confirmation)

1.tiles
2.galvanized iron sheets in good condition
3.galvanized iron in poor condition
4.coconut leaves in good condition
5.coconut leaves in poor condition
6.grass in good condition
7.grass in poor condition
8.no permanent shelter
9.other.....
10.don't know _____

For cross-check Only

8. Materials used for erecting the walls (Community and Traditional Index)

What type of materials were used to build the walls of the house you are living in ?	Community	Traditional
1.cemented or burnt bricks	= 1	= 1
2.soil, pillars, sticks and cement/soil	= 2	= 2
3.soil, pillars, sticks, not cemented	= 2	= 2
4.grass/coconut leaves (full suit)	= 3	= 3
5.no permanent shelter	= 3	= 3
6.other.....	= system missing	
7.don't know _____	= system missing	

9. Condition of the house (Community Index)

Can you describe the general condition of your house ?
(by looking at the cracks or holes on the walls and floors)
(Read options below)

1. good, no need for renovating	= 1
2. needs some minor repairs	= 2
3. needs major repairs	= 3
4. no permanent shelter	= 3
5. don't know _____	= system missing

10. State of the house
(interviewer's confirmation)

1. good, no need for renovating
2. needs some renovation
3. needs major repairs
4. no permanent shelter
5. don't know _____

For cross-check Only

11. Size of the house (Traditional Index)

How many rooms does your house have?(excluding kitchen and bathroom)
(interviewer:incase of more than one house or other rooms occupied by family inlude all the rooms)
Write total number of rooms in the box
99 = don't know _____

Density Index Constructed With Following Question

12. Numer of People in Household (Traditional Index)

What is the size of this household (including servants)?
(people who stayed for two or more days in a week for the past three months in the house(s) owned by the head of household are to be included.)
Write the total number of people in the box

Density Index Constructed

_____

13. Water Availability (Traditional Index)

Where do you get water for domestic use?

1. tap water in the house	= 1
2. tap water in the village	= 2
3. tap water outside the village	= 2
4. Wells, in the farm	= 2
5. Wells, in the village	= 2
6. other sources, in the farm	= 2
7. other sources, in the village	= 3
8. other sources, outside the village	= system missing
9. don't know _____	= system missing

14. Toilet Use (Traditional Index)

What type of toilet(s) that are used in this household?

1. flush toilet(s) in the house	= 1
2. flush toilet(s) outside the house	= 1
3. pit latrines in the farm	= 2
4. pit latrines outside the farm	= 2
5. other.....	= system missing
6. no toilet	= 3
7. don't know _____	= system missing

15. Source of Energy for cooking (Traditional Index)

What type of energy source do you use for cooking ?

1. electricity	= 1
2. gas	= 1
3. kerosene	= 2
4. charcoal	= 2
5. firewood	= 3
6. other.....	= system missing
7. don't know _____	= system missing

16. Size of the farm (Traditional Index)

How big is your farm in acres? (include all plantations)
(write the number of acres in the box using deciminals for fractions eg: 1/2 acre = 0.5 acre) _____

5.0 through higest = 1
1.0 through 4.9 = 2
0 through 0.9 = 3

17. Types of crops grown (Community Index)

What type of crops are grown in your farm?
(cash crops may include food crops for the purpose of selling)

1. cash crops and food crops for sale	= 1
2. cash crops and food crops for own consumption	= 2
3. cash crops only	= 2
4. food crops only for sale	= 2
5. food crops for own consumption only	= 3
6. other.....	= system missing
7. don't know _____	= system missing

19. Condition of the farm (Community Index)

What was the condition of your farm during the last weeding season ?

1. clean and well maintained	= 2
2. small number of weeds, with little effect on productivity of the farm	= 2
3. large number of weeds reducing farm productivity	= 3
4. not attended, bad condition	= 3
5. don't cultivate	= system missing
6. other.....	= system missing
7. don't know _____	= system missing

20. Tenure of the farm (Community Index)

What is the tenure status of the land you farm on?

1. own the land	= 2
2. rent land for a fee (cash or in kind)	= 3
4. don't cultivate	= 3
5. other.....	= system missing
6. don't know _____	= system missing

21. Food Security (Community Index)

How many months in a year are you assured of food supply ?
Crop yields and income from other business are included
Write the total number of months in the box.

11 and 12 = 2
0 through 10 = 3
11 through 99 = s.m.
99 = don't know

22. Physical Capability (Community Index)

What is your opinion on the physical capability of the main person responsible for generating household income?

1. strong and physically fit = 2
2. not very strong but capable of doing small jobs for income generating = 2
3. weak and can't work = 3
4. don't know = system missing

23. Dependency on Others (Community Index)

Whom does your family or household depend upon for food, shelter and clothing ?

1. Self satisfied = 2
2. depend on father or mother (not living with the family) = 3
3. depend on brother or sister (not living with the family) = 3
4. depend on children (not living with the family) = 3
5. depend on relatives (not living with the family) = 3
6. depend on other people (not relatives and don't live with the family) = 3
7. don't know = system missing

24. Assistance from Others (Community Index)

From whom did this family received help during the past 12 months ? (exclude if *dependant* for basic necessities)

1. mother or father (not living with family) = 1
2. children (not living with family) = 1
3. brother or sister (not living with family) = 2
4. other relatives (not living with family) = 2
5. other people (not living with family) = 3
6. help not received = 2
7. don't know = system missing

25. Ownership of Material Possessions (Traditional Index)

Do you own any of the listed items below ?
Circle the correct answer

Sub-scale Construction

1. Car/Cars (and tractors)	Yes/No/Don't know
2. Motorcycle	Yes/No/Don't know
3. Freezer/Fridge	Yes/No/Don't know
4. Television/Video	Yes/No/Don't know
5. Sewing machine	Yes/No/Don't know
6. Sofa set (with cushion covers)	Yes/No/Don't know
7. Bed/Beds with matresses	Yes/No/Don't know
8. Cupboard (high quality timber)	Yes/No/Don't know
9. Radio/Radio cassette	Yes/No/Don't know
10. Bicycle	Yes/No/Don't know
11. Cupboard (low quality timber)	Yes/No/Don't know
12. Sofa set (without cushion covers)	Yes/No/Don't know
13. Pocket Radio	Yes/No/Don't know
14. Local beds	Yes/No/Don't know
15. Don't own any of the mentioned above	Yes/No/Don't know

26. Livestock (Community and Traditional Index)

Sub-scale Construction

(interviewer: put 0 = none, 999 = don't know)

Types of livestock owned	**Total**
Bulls (local type)	
Bulls (hybrid)	
Oxen	
Cows (local type)	
Cows (hybrid)	
Sheep	
Goats	
Pigs	
Chicken	
Ducks	
Donkeys	
Others	

27. Education Level (Community and Traditional Index)

What is the education level of the head of the household or of the person who provides income for the family?

Same for Both Indices

1. degree/diploma (completed) = 1
2. degree/diploma (still pursuing) = 1
3. degree/diploma (did not complete) = 1
4. secondary education (completed) = 1
5. secondary education (still pursuing) = 1
6. secondary education (did not complete) = 2
7. primary education (completed) = 2
8. primary education (still pursuing) = 2
9. primary education (did not complete) = 3
10. no formal education = 3
11. don't know _____ = system missing

28. Literary Status (Traditional Index)

Can the head of household or the responsible person read and write Kiswahili ?

1. Yes = 2
2. No = 3
3. Don't know _____ = system missing

33. Skin Tone (For Mult-method, multi-trait validation procedure only)

What is the skin colour of the head of the household ? (read parts)

1. black
2. brown
3. white
4. don't know _____

APPENDIX 3.1

Kappa Reliability Statistics for Respondent Groups: Sorted Best to Worst

Site 1: *Kisarawe District, Peri-Urban*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Women	0.813	1.09	1.04 - 1.14	1.13
Richer Men	0.800	1.06	1.01 - 1.11	1.10
Leaders	0.800	1.06	1.01 - 1.11	1.10
Poorer Women	0.738	0.93	0.86 - 0.99	0.98*
Poorer Men	0.717	0.88	0.82 - 0.95	0.94*

Site 2: *Kisarawe District, Rural*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Women	0.892	1.04	1.01 - 1.07	1.07
Richer Men	0.888	1.03	1.01 - 1.06	1.06
Poorer Men	0.873	1.00	0.96 - 1.04	1.04
Poorer Women	0.871	1.00	0.97 - 1.03	1.02
Leaders	0.831	0.93	0.89 - 0.96	0.96*

Site 3: *Rufiji District, Peri-urban*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Men	0.828	1.05	1.02 - 1.08	1.07
Leaders	0.812	1.01	0.98 - 1.05	1.05
Poorer Men	0.807	1.00	0.95 - 1.05	1.04
Poorer Women	0.803	0.99	0.95 - 1.04	1.03
Richer Women	0.778	0.94	0.90 - 0.99	0.99*

Site 4: *Rufiji District, Rural*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Leaders	0.881	1.03	1.01 - 1.06	1.06
Poorer Men	0.879	1.03	1.00 - 1.06	1.05
Richer Men	0.871	1.01	0.99 - 1.04	1.04
Richer Women	0.850	0.97	0.94 - 1.01	1.00
Poorer Women	0.838	0.95	0.92 - 0.98	0.98*

Site 5: *Mafia Island District, Rural*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Leaders	0.866	1.07	1.03 - 1.11	1.10
Poorer Men	0.861	1.06	1.02 - 1.09	1.08
Poorer Women	0.848	1.03	0.99 - 1.07	1.06
Richer Women	0.815	0.96	0.92 - 1.01	1.00
Richer Men	0.777	0.89	0.84 - 0.94	0.94*

Site 6: *Mafia Island District, Peri-urban*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Men	0.913	1.04	1.01 - 1.06	1.05
Poorer Women	0.902	1.01	0.98 - 1.04	1.04
Richer Women	0.888	0.99	0.96 - 1.02	1.01
Leaders	0.887	0.99	0.96 - 1.02	1.01
Poorer Men	0.883	0.98	0.95 - 1.01	1.01

Site 7: *Morogoro Rural District, Peri-urban*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Poorer Women	0.848	1.07	1.04 - 1.10	1.10
Richer Women	0.841	1.05	1.01 - 1.09	1.09
Richer Men	0.816	1.00	0.94 - 1.06	1.05
Leaders	0.806	0.98	0.95 - 1.02	1.01
Poorer Men	0.767	0.90	0.84 - 0.97	0.96*

Site 8: *Morogoro Rural District, Rural*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Women	0.852	1.02	0.98 - 1.05	1.05
Poorer Men	0.850	1.01	0.97 - 1.06	1.05
Poorer Women	0.845	1.00	0.97 - 1.04	1.03
Leaders	0.843	1.00	0.97 - 1.02	1.02
Richer Men	0.830	0.97	0.92 - 1.02	1.01

Site 9: *Hai Rural District, Rural*

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Richer Men	0.847	1.03	0.98 - 1.07	1.06
Richer Women	0.842	1.02	0.99 - 1.05	1.04
Poorer Women	0.833	1.00	0.97 - 1.03	1.03
Leaders	0.825	0.98	0.95 - 1.02	1.02
Poorer Men	0.820	0.97	0.92 - 1.02	1.02

Site 10: Hai Rural District, Peri-urban

Respondent Group	Observed Prop Agreement	Index	95% CL	One-Tailed Upper 95% CL
Poorer Women	0.958	1.05	1.02 - 1.07	1.07
Poorer Men	0.946	1.02	0.98 - 1.07	1.06
Richer Men	0.946	1.02	0.98 - 1.06	1.06
Leaders	0.933	1.00	0.96 - 1.04	1.03
Richer Women	0.883	0.91	0.87 - 0.96	0.95*

APPENDIX 3.2 Respondent Group Bias

All Sites

		POORMEN	RICHMEN	POORWOM	RICHWOM	LEADERS
SITE: K1	Over 3	0	0	0	0	0
N = 50	Over 2	8	0	9	0	0
	Over 1	19	4	17	4	1
	Same	17	36	19	37	39
	Under 1	1	8	3	6	10
	Under 2	0	1	0	0	0
	Under 3	0	0	0	0	0
	TOTAL	**45**	**49**	**48**	**47**	**50**
	Not Ranked	5	1	2	3	0

		POORMEN	RICHMEN	POORWOM	RICHWOM	LEADERS
SITE: K2	Over 3	0	0	0	0	0
N = 49	Over 2	0	1	0	0	0
	Over 1	2	5	15	3	27
	Same	39	41	28	37	19
	Under 1	4	2	1	6	3
	Under 2	3	0	0	0	0
	Under 3	0	0	0	0	0
	TOTAL	**48**	**49**	**44**	**46**	**49**
	Not Ranked	1	0	5	3	0

		LEADERS	RICHMEN	POORMEN	RICHWOM	POORWOM
SITE: R1	Over 3	0	0	0	0	1
N = 50	Over 2	0	0	0	1	0
	Over 1	3	1	15	21	5
	Same	33	33	30	23	26
	Under 1	14	13	4	3	12
	Under 2	0	3	0	1	3
	Under 3	0	0	0	0	2
	TOTAL	**50**	**50**	**49**	**49**	**49**
	Not Ranked	0	0	1	1	1

		LEADERS	RICHMEN	POORMEN	RICHWOM	POORWOM
SITE: R2	Over 3	0	0	0	0	0
N = 49	Over 2	0	0	0	0	0
	Over 1	3	5	4	2	0
	Same	46	40	39	30	25
	Under 1	0	4	6	17	24
	Under 2	0	0	0	0	0
	Under 3	0	0	0	0	0
	TOTAL	**49**	**49**	**49**	**49**	**49**
	Not Ranked	0	0	0	0	0

APPENDIX 3.2 Respondent Group Bias

All Sites

		LEADERS	RICHMEN	POORMEN	RICHWOM	POORWOM
SITE: M1	Over 3	0	0	0	0	0
N = 50	Over 2	0	4	0	0	1
	Over 1	3	15	3	5	7
	Same	39	28	37	32	31
	Under 1	8	3	10	7	9
	Under 2	0	0	0	2	1
	Under 3	0	0	0	0	0
	TOTAL	**50**	**50**	**50**	**46**	**49**
	Not Ranked	0	0	0	4	1

		LEADERS	POORMEN	RICHWOM	POORWOM	RICHMEN
SITE: M2	Over 3	0	0	0	0	0
N = 50	Over 2	0	0	0	0	0
	Over 1	3	4	5	2	1
	Same	27	41	36	42	36
	Under 1	20	5	9	6	12
	Under 2	0	0	0	0	1
	Under 3	0	0	0	0	0
	TOTAL	**50**	**50**	**50**	**50**	**50**
	Not Ranked	0	0	0	0	0

		LEADERS	RICHMEN	POORMEN	RICHWOM	POORWOM
SITE: MG1	Over 3	0	0	0	0	0
N = 47	Over 2	1	1	0	0	0
	Over 1	0	8	20	1	2
	Same	25	32	19	28	24
	Under 1	17	6	4	10	11
	Under 2	4	0	0	1	2
	Under 3	0	0	0	0	0
	TOTAL	**47**	**47**	**43**	**40**	**39**
	Not Ranked	0	0	4	7	8

		LEADERS	RICHMEN	RICHWOM	POORMEN	POORWOM
SITE: MG2	Over 3	0	0	0	0	0
N = 49	Over 2	0	3	0	0	0
	Over 1	4	16	18	23	3
	Same	37	27	24	25	42
	Under 1	8	2	4	1	2
	Under 2	0	1	0	0	0
	Under 3	0	0	0	0	0
	TOTAL	**49**	**49**	**46**	**49**	**47**
	Not Ranked	0	0	3	0	2

APPENDIX 3.2 Respondent Group Bias All Sites

		RICHMEN	LEADERS	POORMEN	RICHWOM	POORWOM
SITE: H1	Over 3	0	0	0	0	0
N = 50	Over 2	0	0	0	1	1
	Over 1	9	11	2	7	0
	Same	36	39	30	34	30
	Under 1	5	0	16	8	18
	Under 2	0	0	2	0	1
	Under 3	0	0	0	0	0
	TOTAL	**50**	**50**	**50**	**50**	**50**
	Not Ranked	0	0	0	0	0

		RICHMEN	POORMEN	LEADERS	RICHWOM	POORWOM
SITE: H2	Over 3	0	0	0	0	0
N = 44	Over 2	0	0	0	0	0
	Over 1	3	1	0	5	1
	Same	34	18	27	27	27
	Under 1	5	7	16	1	3
	Under 2	1	0	0	0	0
	Under 3	0	0	0	0	0
	TOTAL	**43**	**26**	**43**	**33**	**31**
	Not Ranked	1	18	1	11	13

ALL SITES TOTAL:	LEADERS	RICHMEN	POORMEN	RICHWOM	POORWOM
Over 3	0	0	0	0	1
Over 2	1	9	8	2	11
Over 1	55	67	93	71	52
Same	331	343	295	308	294
Under 1	96	60	58	71	89
Under 2	4	7	5	4	7
Under 3	0	0	0	0	2
Not Ranked	1	2	29	32	32
TOTAL	488	488	488	488	488

APPENDIX 3.3a

Intra Site Comparison of Criteria Generated

Kimanzichana (K1): Peri-urban site, Kisarawe District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.199	0.232	0.216	0.196	
Income (Level/Regularity)	0.000	0.008	0.069	0.014	
Assets to Generate Income	0.068	0.024	0.017	0.034	
Dependence on Casual Labour	0.006	0.016	0.000	0.000	
Condition or Quality of House	0.106	0.048	0.060	0.054	
Housing Materials	0.143	0.008	0.000	0.054	
Tenure of House	0.075	0.080	0.034	0.081	
Size of House	0.025	0.000	0.000	0.014	No
Number of Houses Owned	0.037	0.032	0.026	0.054	
Condition of Farm	0.000	0.024	0.017	0.020	Criteria
Type of Crop	0.230	0.048	0.017	0.155	
Size of Farm	0.006	0.096	0.078	0.027	
Tenure of Farm Land	0.006	0.064	0.069	0.041	Generated
Food Security	0.000	0.048	0.069	0.020	
Marital Status	0.000	0.008	0.034	0.014	
Size of Family	0.000	0.080	0.034	0.054	
Age of Children	0.000	0.008	0.009	0.000	
Physical and Health Status	0.000	0.032	0.052	0.007	
Receives Assistance	0.000	0.008	0.017	0.000	
Dependent Relatives	0.000	0.008	0.000	0.007	
Dependence on Others	0.012	0.016	0.017	0.014	
Material Possessions	0.031	0.040	0.052	0.041	
Shop Ownership	0.043	0.056	0.086	0.074	
Educational Status	0.000	0.000	0.009	0.007	
Other	0.012	0.016	0.017	0.020	

APPENDIX 3.3b

Intra Site Comparison of Criteria Generated

Kibuta (K2): Rural site, Kisarawe District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.190	0.116	0.180	0.081	0.241
Income (Level/Regularity)	0.029	0.023	0.022	0.010	0.027
Assets to Generate Income	0.007	0.017	0.067	0.030	0.018
Dependence on Casual Labour	0.029	0.017	0.000	0.030	0.018
Variety of Sources of Income	0.007	0.012	0.000	0.010	0.045
Condition or Quality of House	0.015	0.047	0.011	0.131	0.036
Housing Materials	0.102	0.198	0.000	0.030	0.018
Tenure of House	0.022	0.023	0.022	0.000	0.045
Size of House	0.000	0.035	0.000	0.000	0.027
Number of Houses Owned	0.000	0.006	0.011	0.010	0.000
Condition of Farm	0.000	0.012	0.022	0.000	0.009
Type of Crop	0.109	0.110	0.124	0.061	0.054
Size of Farm	0.000	0.099	0.045	0.020	0.000
Tenure of Farm Land	0.000	0.035	0.034	0.010	0.036
Food Security	0.022	0.006	0.101	0.040	0.018
Marital Status	0.058	0.041	0.011	0.051	0.027
Size of Family	0.015	0.006	0.000	0.020	0.018
Age of Children	0.029	0.017	0.000	0.030	0.036
Physical and Health Status	0.226	0.093	0.180	0.293	0.214
Receives Assistance	0.022	0.029	0.011	0.030	0.018
Dependent Relatives	0.000	0.006	0.000	0.000	0.027
Dependence on Others	0.051	0.017	0.034	0.051	0.045
Material Possessions	0.007	0.000	0.000	0.000	0.000
Shop Ownership	0.000	0.006	0.011	0.020	0.009
Other	0.058	0.029	0.112	0.040	0.018

APPENDIX 3.3c

Intra Site Comparison of Criteria Generated

Ikwiriri (R1): Peri-urban Site, Rufiji District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.348	0.298	0.193	0.160	0.111
Income (Level/Regularity)	0.036	0.088	0.012	0.017	0.037
Assets to Generate Income	0.036	0.000	0.012	0.022	0.005
Dependence on Casual Labour	0.036	0.026	0.018	0.006	0.026
Variety of Sources of Income	0.071	0.053	0.024	0.050	0.058
Ability to Employ Casual Labour	0.000	0.000	0.006	0.000	0.000
Condition or Quality of House	0.018	0.053	0.030	0.061	0.079
Housing Materials	0.036	0.053	0.301	0.227	0.143
Tenure of House	0.018	0.044	0.036	0.044	0.058
Size of House	0.000	0.000	0.000	0.033	0.037
Number of Houses Owned	0.009	0.000	0.006	0.000	0.000
Crop Yield	0.018	0.009	0.006	0.055	0.063
Condition of Farm	0.009	0.009	0.000	0.006	0.000
Type of Crop	0.080	0.053	0.090	0.044	0.026
Size of Farm	0.018	0.035	0.012	0.028	0.016
Tenure of Farm Land	0.000	0.000	0.000	0.006	0.000
Food Security	0.009	0.044	0.000	0.017	0.026
Marital Status	0.009	0.009	0.018	0.011	0.026
Size of Family	0.000	0.009	0.018	0.011	0.042
Age of Children	0.000	0.000	0.006	0.000	0.005
Physical and Health Status	0.107	0.114	0.078	0.050	0.101
Receives Assistance	0.000	0.000	0.018	0.006	0.011
Dependent Relatives	0.000	0.009	0.006	0.006	0.005
Dependence on Others	0.054	0.044	0.042	0.072	0.063
Material Possessions	0.000	0.000	0.006	0.006	0.000
Livestock Ownership	0.027	0.018	0.030	0.022	0.021
Shop Ownership	0.027	0.018	0.018	0.017	0.016
Educational Status	0.009	0.000	0.000	0.000	0.000
Other	0.027	0.018	0.012	0.028	0.021

APPENDIX 3.3d

Intra Site Comparison of Criteria Generated

Ngorongo (R2): Rural Site, Rufiji District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.224	0.198	0.198	0.434	0.261
Income (Level/Regularity)	0.042	0.043	0.056	0.060	0.035
Assets to Generate Income	0.036	0.009	0.025	0.012	0.000
Dependence on Casual Labour	0.006	0.103	0.093	0.024	0.026
Variety of Sources of Income	0.085	0.043	0.037	0.133	0.078
Ability to Employ Casual Labour	0.006	0.000	0.006	0.000	0.000
Condition or Quality of House	0.030	0.000	0.000	0.000	0.000
Housing Materials	0.079	0.009	0.012	0.000	0.000
Tenure of House	0.024	0.000	0.000	0.000	0.009
Size of House	0.006	0.000	0.000	0.000	0.000
Crop Yield	0.006	0.043	0.019	0.036	0.026
Condition of Farm	0.006	0.000	0.006	0.000	0.000
Type of Crop	0.048	0.017	0.037	0.024	0.000
Size of Farm	0.073	0.043	0.031	0.012	0.026
Food Security	0.000	0.034	0.025	0.012	0.000
Marital Status	0.018	0.026	0.025	0.036	0.017
Size of Family	0.109	0.017	0.043	0.012	0.096
Age of Children	0.024	0.009	0.012	0.012	0.061
Physical and Health Status	0.109	0.250	0.309	0.133	0.217
Receives Assistance	0.006	0.086	0.019	0.012	0.026
Ability to Assist Others	0.000	0.009	0.000	0.000	0.000
Dependent Relatives	0.012	0.000	0.000	0.000	0.052
Dependence on Others	0.024	0.052	0.037	0.036	0.061
Other	0.024	0.009	0.012	0.012	0.009

APPENDIX 3.3e

Intra Site Comparison of Criteria Generated

Baleni (M1): Rural Site, Mafia Island District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.043	0.043	0.049	0.049	0.107
Income (Level/Regularity)	0.009	0.000	0.000	0.000	0.007
Assets to Generate Income	0.034	0.011	0.018	0.049	0.029
Dependence on Casual Labour	0.034	0.043	0.073	0.024	0.021
Variety of Sources of Income	0.004	0.000	0.018	0.000	0.021
Condition or Quality of House	0.051	0.005	0.006	0.024	0.000
Housing Materials	0.068	0.016	0.049	0.000	0.036
Tenure of House	0.004	0.000	0.006	0.024	0.007
Size of House	0.034	0.000	0.000	0.000	0.000
Number of Houses Owned	0.009	0.005	0.000	0.000	0.014
Crop Yield	0.004	0.005	0.006	0.000	0.000
Condition of Farm	0.000	0.011	0.000	0.049	0.000
Type of Crop	0.141	0.198	0.244	0.220	0.279
Size of Farm	0.107	0.176	0.116	0.146	0.136
Tenure of Farm Land	0.017	0.021	0.012	0.000	0.021
Food Security	0.004	0.016	0.000	0.024	0.007
Marital Status	0.034	0.021	0.024	0.024	0.007
Size of Family	0.077	0.080	0.110	0.000	0.000
Age of Children	0.038	0.027	0.000	0.000	0.000
Physical and Health Status	0.081	0.139	0.110	0.146	0.086
Receives Assistance	0.004	0.005	0.006	0.000	0.014
Ability to Assist Others	0.004	0.000	0.000	0.000	0.000
Dependence on Others	0.034	0.000	0.030	0.098	0.021
Material Possessions	0.021	0.005	0.000	0.000	0.000
Livestock Ownership	0.103	0.112	0.104	0.098	0.150
Shop Ownership	0.004	0.011	0.006	0.000	0.021
Educational Status	0.000	0.027	0.000	0.000	0.000
Other	0.034	0.021	0.012	0.024	0.014

APPENDIX 3.3f

Intra Site Comparison of Criteria Generated

Kilondoni (M2): Peri-urban Site, Mafia Island District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.200	0.364	0.280	0.277	0.220
Income (Level/Regularity)	0.021	0.000	0.000	0.000	0.000
Assets to Generate Income	0.152	0.065	0.053	0.076	0.053
Dependence on Casual Labour	0.007	0.009	0.015	0.000	0.008
Variety of Sources of Income	0.028	0.009	0.061	0.000	0.030
Condition or Quality of House	0.000	0.000	0.008	0.000	0.008
Housing Materials	0.007	0.009	0.000	0.000	0.000
Tenure of House	0.021	0.037	0.038	0.118	0.121
Number of Houses Owned	0.076	0.037	0.076	0.101	0.076
Crop Yield	0.007	0.000	0.000	0.000	0.000
Condition of Farm	0.007	0.009	0.008	0.000	0.000
Type of Crop	0.076	0.056	0.053	0.042	0.030
Size of Farm	0.048	0.009	0.045	0.034	0.061
Tenure of Farm Land	0.028	0.047	0.068	0.126	0.106
Food Security	0.000	0.000	0.000	0.008	0.000
Marital Status	0.021	0.000	0.000	0.000	0.000
Size of Family	0.014	0.009	0.008	0.000	0.000
Age of Children	0.000	0.000	0.008	0.000	0.000
Physical and Health Status	0.062	0.121	0.015	0.008	0.045
Receives Assistance	0.062	0.047	0.008	0.000	0.000
Ability to Assist Others	0.000	0.000	0.008	0.000	0.008
Dependent Relatives	0.000	0.009	0.000	0.000	0.000
Dependence on Others	0.041	0.056	0.045	0.042	0.030
Material Possessions	0.028	0.028	0.068	0.050	0.068
Livestock Ownership	0.028	0.009	0.045	0.059	0.083
Shop Ownership	0.055	0.065	0.061	0.059	0.053
Educational Status	0.000	0.000	0.008	0.000	0.000
Other	0.014	0.000	0.023	0.000	0.000

APPENDIX 3.3g

Intra Site Comparison of Criteria Generated Kauzeni (MG1): Peri-urban Site, Morogoro Rural District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.063	0.239	0.117	0.223	0.079
Income (Level/Regularity)	0.047	0.051	0.061	0.045	0.009
Assets to Generate Income	0.024	0.000	0.006	0.006	0.000
Dependence on Casual Labour	0.039	0.036	0.028	0.006	0.000
Variety of Sources of Income	0.000	0.022	0.056	0.064	0.018
Condition or Quality of House	0.126	0.029	0.050	0.083	0.105
Housing Materials	0.157	0.000	0.117	0.045	0.202
Tenure of House	0.024	0.058	0.022	0.025	0.035
Size of House	0.000	0.000	0.006	0.006	0.018
Number of Houses Owned	0.016	0.000	0.006	0.000	0.009
Crop Yield	0.031	0.022	0.094	0.032	0.114
Condition of Farm	0.016	0.058	0.017	0.000	0.000
Type of Crop	0.008	0.000	0.056	0.045	0.044
Size of Farm	0.000	0.022	0.022	0.000	0.000
Tenure of Farm Land	0.000	0.022	0.017	0.000	0.000
Food Security	0.039	0.000	0.028	0.019	0.088
Marital Status	0.087	0.029	0.028	0.070	0.018
Size of Family	0.000	0.000	0.006	0.032	0.009
Age of Children	0.008	0.014	0.000	0.013	0.000
Physical and Health Status	0.150	0.239	0.106	0.115	0.140
Receives Assistance	0.039	0.007	0.017	0.045	0.000
Ability to Assist Others	0.008	0.000	0.000	0.000	0.000
Dependent Relatives	0.024	0.000	0.000	0.013	0.000
Dependence on Others	0.024	0.116	0.061	0.083	0.105
Material Possessions	0.008	0.000	0.006	0.006	0.000
Livestock Ownership	0.000	0.000	0.011	0.000	0.000
Educational Status	0.000	0.000	0.000	0.006	0.000
Other	0.063	0.036	0.067	0.019	0.009

APPENDIX 3.3h

Intra Site Comparison of Criteria Generated

Doma (MG2): Rural Site, Morogoro Rural District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.181	0.198	0.066	0.086	0.105
Income (Level/Regularity)	0.047	0.074	0.007	0.022	0.033
Assets to Generate Income	0.013	0.000	0.015	0.007	0.007
Dependence on Casual Labour	0.013	0.099	0.169	0.050	0.163
Variety of Sources of Income	0.027	0.008	0.015	0.014	0.007
Ability to Employ Casual Labour	0.000	0.041	0.162	0.144	0.098
Condition or Quality of House	0.000	0.000	0.000	0.000	0.020
Housing Materials	0.000	0.000	0.000	0.007	0.013
Tenure of House	0.000	0.041	0.000	0.022	0.000
Size of House	0.000	0.000	0.000	0.007	0.007
Crop Yield	0.168	0.099	0.265	0.223	0.111
Condition of Farm	0.000	0.000	0.000	0.000	0.007
Type of Crop	0.000	0.000	0.015	0.158	0.072
Size of Farm	0.034	0.000	0.051	0.029	0.052
Tenure of Farm Land	0.000	0.008	0.000	0.000	0.000
Food Security	0.027	0.083	0.059	0.014	0.065
Marital Status	0.013	0.008	0.000	0.036	0.026
Size of Family	0.054	0.017	0.007	0.014	0.072
Age of Children	0.054	0.008	0.015	0.000	0.007
Physical and Health Status	0.154	0.066	0.081	0.058	0.046
Receives Assistance	0.013	0.050	0.007	0.014	0.026
Ability to Assist Others	0.040	0.074	0.000	0.036	0.000
Dependent Relatives	0.007	0.000	0.007	0.000	0.007
Dependence on Others	0.067	0.091	0.037	0.022	0.033
Material Possessions	0.000	0.000	0.007	0.000	0.007
Livestock Ownership	0.007	0.017	0.007	0.007	0.007
Educational Status	0.007	0.000	0.000	0.000	0.000
Other	0.074	0.017	0.007	0.029	0.013

APPENDIX 3.3i

Intra Site Comparison of Criteria Generated

Mowa Njamu (H1): Rural Site, Hai District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.079	0.066	0.033	0.074	0.056
Income (Level/Regularity)	0.010	0.025	0.011	0.000	0.011
Assets to Generate Income	0.005	0.004	0.011	0.000	0.000
Dependence on Casual Labour	0.025	0.016	0.044	0.037	0.034
Variety of Sources of Income	0.005	0.008	0.027	0.011	0.011
Condition or Quality of House	0.050	0.021	0.011	0.090	0.034
Housing Materials	0.059	0.012	0.066	0.122	0.067
Tenure of House	0.030	0.016	0.016	0.037	0.006
Size of House	0.005	0.004	0.011	0.016	0.011
Number of Houses Owned	0.015	0.004	0.000	0.011	0.028
Crop Yield	0.025	0.016	0.011	0.000	0.028
Condition of Farm	0.020	0.012	0.044	0.000	0.011
Type of Crop	0.084	0.086	0.291	0.212	0.285
Size of Farm	0.084	0.111	0.187	0.053	0.123
Tenure of Farm Land	0.064	0.053	0.027	0.042	0.084
Food Security	0.015	0.008	0.000	0.005	0.006
Marital Status	0.054	0.033	0.027	0.063	0.028
Size of Family	0.005	0.037	0.005	0.037	0.006
Age of Children	0.015	0.037	0.000	0.005	0.011
Physical and Health Status	0.069	0.115	0.082	0.058	0.061
Receives Assistance	0.035	0.033	0.027	0.005	0.011
Ability to Assist Others	0.005	0.000	0.000	0.000	0.000
Dependent Relatives	0.015	0.012	0.005	0.000	0.000
Dependence on Others	0.040	0.045	0.011	0.016	0.006
Material Possessions	0.000	0.008	0.000	0.000	0.000
Livestock Ownership	0.035	0.037	0.027	0.042	0.056
Educational Status	0.054	0.049	0.000	0.042	0.000
Other	0.099	0.128	0.022	0.021	0.028

APPENDIX 3.3j

Intra Site Comparison of Criteria Generated

Sanya Juu (H2): Peri-urban Site, Hai District

Indicators or Criteria	Leaders	Rich Men	Poor Men	Rich Wom	Poor Wom
Occupation	0.193	0.180	0.130	0.297	0.186
Income (Level/Regularity)	0.071	0.019	0.029	0.009	0.000
Assets to Generate Income	0.021	0.025	0.043	0.018	0.029
Dependence on Casual Labour	0.107	0.075	0.130	0.045	0.129
Variety of Sources of Income	0.036	0.006	0.058	0.045	0.029
Condition or Quality of House	0.007	0.012	0.000	0.009	0.014
Housing Materials	0.029	0.068	0.000	0.000	0.043
Tenure of House	0.157	0.124	0.101	0.162	0.229
Number of Houses Owned	0.007	0.000	0.000	0.009	0.014
Crop Yield	0.007	0.000	0.000	0.000	0.000
Type of Crop	0.043	0.031	0.072	0.000	0.000
Size of Farm	0.021	0.031	0.072	0.018	0.014
Tenure of Farm Land	0.043	0.025	0.087	0.045	0.071
Food Security	0.007	0.006	0.014	0.009	0.000
Marital Status	0.007	0.037	0.000	0.027	0.014
Size of Family	0.021	0.025	0.029	0.018	0.000
Age of Children	0.000	0.006	0.029	0.009	0.014
Physical and Health Status	0.050	0.155	0.116	0.108	0.086
Receives Assistance	0.036	0.043	0.058	0.063	0.014
Ability to Assist Others	0.000	0.006	0.000	0.000	0.000
Dependent Relatives	0.000	0.006	0.014	0.009	0.000
Dependence on Others	0.007	0.037	0.014	0.036	0.029
Livestock Ownership	0.036	0.043	0.000	0.018	0.029
Shop Ownership	0.014	0.006	0.000	0.009	0.029
Educational Status	0.007	0.006	0.000	0.018	0.000
Other	0.071	0.025	0.000	0.018	0.029

APPENDIX 3.4

Cross District Comparison of Criteria Generated

Indicators or Criteria	Kisarawe	Rufiji	Mafia Is	Morogoro	Hai
Occupation	0.183	0.223	0.150	0.137	0.112
Income (Level/Regularity)	0.022	0.040	0.004	0.040	0.018
Assets to Generate Income	0.031	0.016	0.051	0.008	0.012
Dependence on Casual Labour	0.013	0.035	0.026	0.060	0.052
Variety of Sources of Income	0.008	0.059	0.017	0.025	0.019
Ability to Employ Casual Labour	0.000	0.002	0.000	0.044	0.000
Condition or Quality of House	0.057	0.031	0.012	0.040	0.029
Housing Materials	0.073	0.103	0.024	0.052	0.052
Tenure of House	0.045	0.026	0.033	0.022	0.067
Size of House	0.013	0.010	0.006	0.004	0.006
Number of Houses Owned	0.021	0.001	0.037	0.003	0.009
Crop Yield	0.000	0.029	0.003	0.115	0.011
Condition of Farm	0.011	0.004	0.005	0.010	0.011
Type of Crop	0.108	0.043	0.136	0.041	0.128
Size of Farm	0.042	0.030	0.091	0.022	0.082
Tenure of Farm Land	0.032	0.001	0.043	0.005	0.052
Food Security	0.031	0.016	0.005	0.040	0.007
Marital Status	0.027	0.019	0.015	0.032	0.034
Size of Family	0.025	0.038	0.039	0.022	0.019
Age of Children	0.014	0.012	0.011	0.012	0.013
Physical and Health Status	0.110	0.142	0.080	0.115	0.089
Receives Assistance	0.015	0.017	0.014	0.029	0.030
Ability to Assist Others	0.000	0.001	0.002	0.015	0.001
Dependent Relatives	0.005	0.009	0.001	0.006	0.006
Dependence on Others	0.027	0.049	0.034	0.056	0.025
Material Possessions	0.020	0.001	0.026	0.004	0.001
Livestock Ownership	0.000	0.013	0.083	0.006	0.036
Shop Ownership	0.035	0.010	0.031	0.000	0.004
Educational Status	0.002	0.001	0.004	0.001	0.023
Other	0.033	0.018	0.016	0.035	0.053

APPENDIX 3.5

Cross Site Comparison of Criteria Generated

All Sites Analysed Separately

DISTRICT	Kisarawe		Rufiji		Mafia Island		Morogoro Rural		Hai	
Indicators or Criteria	**Peri-urban**	**Rural**	**Peri-urban**	**Rural**	**Rural**	**Peri-urban**	**Peri-urban**	**Rural**	**Rural**	**Peri-urban**
Occupation	0.209	0.159	0.203	0.246	0.056	0.263	0.148	0.126	0.062	0.201
Income (Level/Regularity)	0.020	0.023	0.034	0.047	0.004	0.005	0.045	0.036	0.012	0.029
Assets to Generate Income	0.038	0.025	0.014	0.019	0.025	0.082	0.007	0.009	0.004	0.025
Dependence on Casual Labour	0.005	0.020	0.021	0.051	0.042	0.008	0.022	0.099	0.030	0.091
Variety of Sources of Income	0.000	0.015	0.050	0.070	0.009	0.027	0.035	0.014	0.012	0.031
Ability to Employ Casual Labour	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.089	0.000	0.000
Condition or Quality of House	0.069	0.046	0.051	0.008	0.020	0.003	0.075	0.004	0.040	0.009
Housing Materials	0.058	0.087	0.168	0.025	0.042	0.003	0.099	0.004	0.062	0.033
Tenure of House	0.069	0.023	0.042	0.008	0.005	0.066	0.032	0.011	0.021	0.151
Size of House	0.011	0.015	0.017	0.002	0.010	0.000	0.006	0.003	0.009	0.000
Number of Houses Owned	0.038	0.005	0.003	0.000	0.007	0.074	0.006	0.000	0.011	0.005
Crop Yield	0.000	0.000	0.034	0.023	0.004	0.002	0.059	0.173	0.016	0.002
Condition of Farm	0.015	0.008	0.004	0.003	0.005	0.005	0.018	0.001	0.017	0.000
Type of Crop	0.124	0.094	0.056	0.028	0.206	0.052	0.032	0.050	0.183	0.029
Size of Farm	0.047	0.038	0.021	0.041	0.133	0.041	0.010	0.034	0.111	0.029
Tenure of Farm Land	0.042	0.023	0.001	0.000	0.017	0.074	0.008	0.001	0.054	0.047
Food Security	0.031	0.031	0.018	0.014	0.008	0.002	0.032	0.049	0.007	0.007
Marital Status	0.013	0.039	0.016	0.023	0.023	0.005	0.046	0.017	0.041	0.020
Size of Family	0.040	0.011	0.018	0.061	0.067	0.006	0.010	0.034	0.019	0.020
Age of Children	0.004	0.023	0.003	0.023	0.018	0.002	0.007	0.017	0.015	0.009
Physical and Health Status	0.020	0.190	0.087	0.207	0.106	0.049	0.147	0.082	0.079	0.105
Receives Assistance	0.005	0.023	0.008	0.028	0.007	0.024	0.022	0.021	0.023	0.044
Ability to Assist Others	0.000	0.000	0.000	0.002	0.001	0.003	0.001	0.029	0.001	0.002
Dependent Relatives	0.004	0.007	0.005	0.012	0.000	0.002	0.007	0.004	0.007	0.005
Dependence on Others	0.015	0.038	0.056	0.041	0.026	0.043	0.077	0.049	0.025	0.025
Material Possessions	0.040	0.002	0.003	0.000	0.008	0.049	0.004	0.003	0.002	0.000
Livestock Ownership	0.000	0.000	0.024	0.000	0.114	0.046	0.003	0.009	0.039	0.029
Shop Ownership	0.064	0.008	0.018	0.000	0.009	0.058	0.000	0.000	0.000	0.011
Educational Status	0.004	0.000	0.001	0.000	0.007	0.002	0.001	0.001	0.031	0.007
Other	0.016	0.048	0.021	0.014	0.022	0.008	0.041	0.029	0.064	0.033
TOTAL	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

APPENDIX 3.6 Multitrait-Multimethod Matrices

Site: Kingulwira, Peri-urban, Morogoro Rural District N = 47 (4+ Respondent Group Agreement)

Index: COM1: Community Generated, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.2756 (.047)	-0.0394 (.774)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	0.1532 (.270)	(1.00)

Index: COM2: Community Generated, Weighted,Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.4471 (.000)	-0.1928 (.164)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.036 (.796)	(1.00)

Index: COM3: Community Generated, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.2686 (.540)	-0.0410 (.766)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	0.0736 (.597)	(1.00)

Index: COM4: Community Generated, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.4617 (0.001)	-0.1453 (.291)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.1698 (.223)	(1.00)

Index:TRAD1: Traditional, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.6726 (.000)	-0.1304 (.342)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.0511 (.713)	(1.00)

Index: TRAD2: Traditional, Weighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.2480 (.079)	-0.1885 (.176)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.1195 (.397)	(1.00)

Index: TRAD3: Traditional, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.4343 (.002)	-0.0359 (.794)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.0221 (.874)	(1.00)

Index: TRAD4: Traditional, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	.3887 (.007)	-0.2244 (.112)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	-0.0841 (.556)	(1.00)

Index: SELF RANK

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.2170 (.113)	(1.00)		
Household Survey	SES	0.1751 (.204)	0.1037 (.445)	(1.00)	
	Skin Tone	-0.1369 (.323)	0.6346 (.000)	.2112 (.125)	(1.00)

Site: Mikese Station, Rural, Morogoro Rural District N= 50 (4+ Respondent Group Agreement)

Index: COM1: Community Generated, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.2921 (.024)	-0.3358 (.011)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1816 (.169)	(1.00)

Index: COM2: Community Generated, Weighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	.0717 (.599)	-0.2799 (.043)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1406 (.308)	(1.00)

Index: COM3: Community Generated, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.4208 (.001)	-0.1973 (.134)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.3033 (.021)	(1.00)

Index: COM4: Community Generated, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.1584 (.239)	-0.2862 (.037)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1098 (.422)	(1.00)

Index:TRAD1: Traditional, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.6148 (.000)	-0.1070 (.422)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1388 (.296)	(1.00)

Index: TRAD2: Traditional, Weighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.2742 (.043)	-0.0711 (.608)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1750 (.204)	(1.00)

Index: TRAD3: Traditional, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.4857 (.000)	.0247 (.855)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1227 (.362)	(1.00)

Index: TRAD4: Traditional, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.1959 (.149)	-0.0854 (.538)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	-0.1232 (.372)	(1.00)

Index: SELF RANK

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.1063 (.420)	(1.00)		
Household Survey	SES	0.4252 (.001)	-0.1410 (.292)	(1.00)	
	Skin Tone	-0.2355 (.072)	0.3873 (.004)	0.0245 (.854)	(1.00)

Site: Lawate, Peri-Urban, Hai District N = 46 (4+ Respondent Group Agreement)

Index: COM1: Community Generated, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.3452 (.010)	-0.0946 (.485)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	0.0401 (.770)	(1.00)

Index: COM2: Community Generated, Weighted,Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.2511 (.075)	0.0975 (.494)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	0.3743 (.009)	(1.00)

Index: COM3: Community Generated, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	.4674 (.000)	-0.0246 (.856)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.0425 (.756)	(1.00)

Index: COM4: Community Generated, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	.3334 (.018)	-0.2406 (.091)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.1541 (.284)	(1.00)

Index:TRAD1: Traditional, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.3989 (.003)	-0.1731 (.200)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.1093 (.423)	(1.00)

Index: TRAD2: Traditional, Weighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.4060 (.003)	-0.2594 (.064)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.2025 (.152)	(1.00)

Index: TRAD3: Traditional, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.4258 (.001)	-0.1440 (.285)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.0720 (.597)	(1.00)

Index: TRAD4: Traditional, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.2665 (.056)	-0.1121 (.426)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.2477 (.082)	(1.00)

Index: SELF RANK

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.0136 (.920)	(1.00)		
Household Survey	SES	0.3767 (.006)	0.0410 (.767)	(1.00)	
	Skin Tone	-0.1868 (.169)	.4152 (.003)	-0.0065 (.963)	(1.00)

Site: Kashashe, Rural, Hai District N = 62 (4+ Respondent Group Agreement)

Index: COM1: Community Generated, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.3174 (.009)	-0.0361 (.759)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.0026 (.983)	(1.00)

Index: COM2: Community Generated, Weighted,Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.0057 (.964)	-0.0181 (.881)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	0.0095 (.939)	(1.00)

Index: COM3: Community Generated, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.5036 (.000)	-0.1387 (.235)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.0012 (.992)	(1.00)

Index: COM4: Community Generated, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.2948 (.019)	-0.1242 (.305)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.0177 (.887)	(1.00)

Index:TRAD1: Traditional, Weighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.3881 (.001)	-0.1716 (.144)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.1290 (.286)	(1.00)

Index: TRAD2: Traditional, Weighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.7389 (.000)	-0.2775 (.021)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.1202 (.331)	(1.00)

Index: TRAD3: Traditional, Unweighted, Proportional Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.4428 (.000)	-0.2044 (.081)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.0037 (.975)	(1.00)

Index: TRAD4: Traditional, Unweighted, Midpoint Cut-off Points

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.6599 (.000)	-0.2343 (.051)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.1310 (.290)	(1.00)

Index: SELF RANK

		Community Rank		Household Survey	
	Traits	SES	Skin Tone	SES	Skin Tone
Community Rank	SES	(1.00)			
	Skin Tone	-0.3687 (.002)	(1.00)		
Household Survey	SES	0.3373 (.007)	-0.1475 (.217)	(1.00)	
	Skin Tone	-0.0886 (.470)	0.2144 (.069)	-0.2803 (.023)	(1.00)

APPENDIX 3.7

Multitrait, Multimethod Criteria for Each Index: All Sites

SITE: Kingulwira, Peri-Urban, Morogoro Rural District (N = 47)

Study Index Column 1	Index Description Column 2	Validity Coeffic. [p value] Column 3	Criteria 1 *VC 0.40 *p < 0.05 Column 4	Criteria 2 VCs > DT/DM Column 5	Criteria 3 VCs > DT/SM Column 6
Community COM 1	Prop. Cutoff Points Weighted	0.2756 [.047]	No No	Yes	Yes
Community COM 2	Midpoint Cutoff Points Weighted	0.4471 [.000]	Yes Yes	Yes	Yes
Community COM 3	Prop. Cutoff Points Unweighted	0.2686 [.540]	No No	Yes	Yes
Community COM 4	Midpoint Cutoff Points Unweighted	0.4617 [.000]	Yes Yes	Yes	Yes
Traditional TRAD 1	Prop. Cutoff Points Weighted	0.6727 [.000]	Yes Yes	Yes	Yes
Traditional TRAD 2	Midpoint Cutoff Points Weighted	0.2480 [.079]	No No	Yes	Yes
Traditional TRAD 3	Prop. Cutoff Points Unweighted	0.4343 [.002]	Yes Yes	Yes	Yes
Traditional TRAD 4	Midpoint Cutoff Points Unweighted	0.3887 [.007]	No Yes	Yes	Yes
SELF RANK	Survey Respondent Ranks the Household	0.1751 [.204]	No No	Yes	No

SITE: Mikese Station, Rural, Morogoro Rural District (N = 50)

Study Index	Index Description	Validity Coeffic. [p value]	Criteria 1 *VC 0.40 *p < 0.05	Criteria 2 VCs > DT/DM	Criteria 3 VCs > DT/SM
Community COM 1	Prop. Cutoff Points Weighted	0.2921 [.024]	No Yes	Yes	Yes
Community COM 2	Midpoint Cutoff Points Weighted	0.0717 [.599]	No No	No	No
Community COM 3	Prop. Cutoff Points Unweighted	0.4208 [.001]	Yes Yes	Yes	Yes
Community COM 4	Midpoint Cutoff Points Unweighted	0.1584 [.239]	No No	No	Yes
Traditional TRAD 1	Prop. Cutoff Points Weighted	0.6148 [.000]	Yes Yes	Yes	Yes
Traditional TRAD 2	Midpoint Cutoff Points Weighted	0.2742 [.043]	No Yes	Yes	Yes
Traditional TRAD 3	Prop. Cutoff Points Unweighted	0.4857 [.000]	Yes Yes	Yes	Yes
Traditional TRAD 4	Midpoint Cutoff Points Unweighted	0.1959 [.149]	No No	No	Yes
SELF RANK	Survey Respondent Ranks the Household	0.4252 [.001]	Yes Yes	Yes	Yes

SITE: Lawate, Peri-Urban, Hai District (N = 46)

Study Index	Index Description	Validity Coeffic. [p value]	Criteria 1 *VC 0.40 *p < 0.05	Criteria 2 VCs > DT/DM	Criteria 3 VCs > DT/SM
Community COM 1	Prop. Cutoff Points Weighted	0.3452 [.010]	No Yes	Yes	Yes
Community COM 2	Midpoint Cutoff Points Weighted	0.2511 [.075]	No No	Yes	No
Community COM 3	Prop. Cutoff Points Unweighted	0.4674 [.000]	Yes Yes	Yes	Yes
Community COM 4	Midpoint Cutoff Points Unweighted	0.3334 [.018]	No Yes	Yes	Yes
Traditional TRAD 1	Prop. Cutoff Points Weighted	0.3898 [.003]	No Yes	Yes	Yes
Traditional TRAD 2	Midpoint Cutoff Points Weighted	0.4060 [.003]	Yes Yes	Yes	Yes
Traditional TRAD 3	Prop. Cutoff Points Unweighted	0.4258 [.001]	Yes Yes	Yes	Yes
Traditional TRAD 4	Midpoint Cutoff Points Unweighted	0.2665 [.056]	No No	Yes	Yes
SELF RANK	Survey Respondent Ranks the Household	0.3767 [.006]	No Yes	Yes	Yes

SITE: Kashashe, Rural, Hai District (N = 62)

Study Index	Index Description	Validity Coeffic. [p value]	Criteria 1 *VC 0.40 *p < 0.05	Criteria 2 VCs > DT/DM	Criteria 3 VCs > DT/SM
Community COM 1	Prop. Cutoff Points Weighted	0.3174 [.009]	No Yes	Yes	No
Community COM 2	Midpoint Cutoff Points Weighted	0.0057 [.964]	No No	No	No
Community COM 3	Prop. Cutoff Points Unweighted	0.5036 [.000]	Yes Yes	Yes	Yes
Community COM 4	Midpoint Cutoff Points Unweighted	0.2948 [.019]	No Yes	Yes	No
Traditional TRAD 1	Prop. Cutoff Points Weighted	0.3881 [.001]	No Yes	Yes	Yes
Traditional TRAD 2	Midpoint Cutoff Points Weighted	0.7389 [.000]	Yes	Yes	Yes
Traditional TRAD 3	Prop. Cutoff Points Unweighted	0.4428 [.000]	Yes Yes	Yes	Yes
Traditional TRAD 4	Midpoint Cutoff Points Unweighted	0.6599 [.000]	Yes Yes	Yes	Yes
SELF RANK	Survey Respondent Ranks the Household	0.3373 [.007]	No Yes	Yes	Yes

References

Abramson JH (1990) *Survey Methods on Community Medicine.* Churchill Livingston. London (4th edition).

Abramson JH., Gofin R., Habib J., *et al*, (1982) "Indicators of Social Class: a comparative appraisal of measures for use in epidemiological studies" *Social Science and Medicine* Vol 16(20): 1739-46.

Altman D., (1991) *Practical Statistics for Medical Research* Chapman and Hall. London.

Amis P. and Rakodi C., "Urban Poverty: Issues for Research and Policy". *Journal of International Development* Vol 6(5): 627-634.

Angoff WH., (1988) "Validity: An Evolving Concept". In Wainer H. and Braun H. (eds.) *Test Validity.* Lawrence Erlbaum & Assoc. New Jersey. Chapter 2.

Artz RD., Curtis RF., Fairbank DT., Jackson EF., (1971) "Community Rank Stratification: A Factor Analysis". *American Sociological Review* Vol 16 (Dec): 985-1002.

Atkinson AB., (1989) *Poverty and Social Security.* Harvester Wheatsheaf. New York.

Bailey NTJ., (1995) *Statistical Methods in Biology* Cambridge University Press. Cambridge.

Beck T., (1989) "Survival Strategies and Power Amongst the Poorest in a West Bengal Village" *IDS Bulletin* Vol 20 (2): 23-32.

Becker HS., (1996) 'The Epistemology of Qualitative Research' In Jessor R and Colby A., (eds.) *Essays on Ethnography and Human Development.* University of Chicago Press: Chicago.

Belcher JC., (1951). "Evaluation and Restandardization of Sewell's Socio-economic Scale". *Rural Sociology.* Vol 16: 246-255.

Benoit-Smullyan E., (1944). "Status, Status Types, and Status Interrelations". *American Sociological Review.* Vol 9: 151-161.

Berger J., Cohen BP., Zelditch M., (1972) "Status characteristics and social interaction" *American Sociological Review* Vol 37: 241-255.

Béteille A., (1969) "Ideas and interests: some conceptual problems in the study of social stratification in rural India". *International Social Science Journal.* Vol 21: 219-234.

Bills DB., Godfrey D., Haller A., (1985) "A Scale to Measure the Socio-economic Status of Occupations in Brazil" *Rural Sociology* Vol 50(2): 225-250.

Blalock HM., (1982). *Conceptualisation and Measurement in the Social Sciences.* Sage. Beverly Hills.

Bland R. (1979). "Measuring 'Social Class'". *Sociology* Vol 13: 283-291.

Blau PM., and Duncan OD., (1967). *The American Occupational Structure.* Wiley: New York.

Boston G., (1980). "Classification of occupations" *Population Trends* Vol 20: 9-11.

Bowden AO., Caldwell FF., West GA., (1934). "A study in prestige" *American Journal of Sociology* Vol 41: 193-204.

Bruce J., (1989). "Homes Divided" *World Development* Vol 17 (7): 979-991.

Bryman A., (1984) 'The debate about quantitative and qualitative research: a question of method or epistemology?' *The British Journal of Sociology* Vol 35 (1): 75-92.

Campbell DT. and Fiske DW., (1959). "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix" *Psychological Bulletin* Vol 56 (2): 81-105.

Carmines E., and Zeller R., (1994) "Reliability and Validity Assessment". In Lewis-Beck M (ed.) *Basic Measurement* Sage/Toppan. London.

Carter SE., Chidiamassamba A., Jeranyama P., *et al.* (1993). "Some Observations on Wealth Ranking After an RRA Looking at Soil Fertility Management in Northeastern Zimbabwe". *RRA Notes* No 18: 47-52.

Cattell R., (1942). "The Concept of Social Status" *The Journal of Social Psychology* Vol 15: 293-308.

Centers R (1949) *The Psychology of Social Classes: A study of class consciousness.* Princeton, NJ. Princeton University (Reissued by Russell and Russell, 1961)

Cernea MM., (1985) *Putting People First: Sociological Variables in Rural Development* World Bank, OUP: New York.

Chadwick MT. and Seeley JA.,(1996). *Indigenous Soil Classification: An Investigation into Vernacular Soil Typology in Nepal.* FORESC Occasional Paper 3/96, Ministry of Forests and Soil Conservation, Kathmandu.

Chamber R., (1992) *Rural Appraisal: Rapid, Relaxed and Participatory* IDS Discussion Paper no 311. IDS: Sussex.

Chambes R., (1989) 'Editorial Introduction: Vulnerability, Coping and Policy' *IDS Bulletin* Vol 20(2): 1-7.

Chambers R., (1995) *Poverty and Livelihoods: Whose Reality Counts?* Institute of Development Studies. Discussion Paper No. 347, Brighton.

Chambers R., (1997) *Whose Reality Counts?: Putting the First Last.* Intermediate Technology Publications: London.

Chapin FS., (1935). *Contemporary American Institutions: A Sociological Analysis.* Harper's Social Sciences Series. London.

Clignet R., (1977). "Educational and Occupational Differentiation in a New Country: The Case of Cameroun" *Economic Development and Cultural Change* Vol 25(4): 731-745.

Cohen J., (1960). "A Coefficient of Agreement for Nominal Scales". *Educational and Psychological Measurement.* Vol 20: 37-46. (Cited in Kraemer, 1980)

Cole M. and Scribner S., (1974) *Culture and Thought* Wiley. New York

Coleman J., (1966). "The Hypothesis of Intersocietal Similarity in Occupational Prestige Hierarchies" *American Journal of Sociology* Vol 72 (Sept): 210-216.

Collier P., Radwan S., Wangwe A., (1986). *Labour and Poverty in Rural Tanzania: Ujamaa and Rural Development in the United Republic of Tanzania.* Clarenden Press: Oxford.

Cooksey B., (1994). "Who's Poor in Tanzania? A Review of Recent Poverty Research" In Bagachwa MSD., (ed.). *Poverty Alleviation in Tanzania: Recent Research Issues.* Dar es Salaam University Press. Dar es Salaam.

Cornwall A. and Jewkes R., (1995) "What is Participatory Research?" *Social Science and Medicine* Vol 41 (12): 1667-1676.

Cramer D., (1994). *Introducing Statistics for Social Research.* Routledge. London. (Chapter 10)

Crompton R., (1993) *Class and Stratification: An Introduction to Current Debates.* Polity Press. Cambridge.

Cronbach LJ., (1951) "Coefficient alpha and the internal structure of tests" *Psychometrika* Vol 16: 297-334.

Cronbach LJ., and Meehl PE., (1955) "Construct Validity in Psychological Tests" *Psychological Bulletin* Vol 52: 281-302.

Cronbach LJ., (1971) "Test Validation" pp 443-507 in RL Thorndike (ed.) *Educational Measurement.* American Council on Education. Washington, DC.

Dahrendorf R., (1972). *Class and Conflict in Industrial Society.* Lowe and Brydome, London. (First published 1959)

Davis K. and Moore W., (1945) "Some principles of stratification". *American Sociological Review.* Vol 10 (2): 242-249.

Dawson SG., Manderson L., Tallo V., (1993). *A Manual for the Use of Focus Group Discussion: Methods for Social Research in Disease.* International Nutrition Foundation with WHO and UNDP, Boston. Mass.

Dawson SG., (1997) *The Effectiveness of Exemption Mechanisms: A Review of the Evidence from Developing Countries.* Unpublished Report. Liverpool School of Tropical Medicine, UK.

Day DD., (1940) "Discussion" *Rural Sociology* Vol V (March): 83-85 (Critique of Schuler's study "Social and Economic Status in a Louisiana Hills Community" Same volume).

Deaton A., (1982). "Inequality and needs: Some experimental results from Sri Lanka." *Population and Development Review.* (8) Supp: 35-49.

Drèze J., Lanjouw P, and Stern N., (1992) "Economic mobility and Agricultural Labour in Rural India: A Case Study, STICERD, London School of Economics, mimeo No.35.

Dumont L., (1972). *Homo Hierarchius.* Paladin. London.

Durkin MS., Islam S., Hasan ZM., Zaman SS,. (1994). "Measures of Socioeconomic Status for Child Health Research: Comparative Results from Bangladesh and Pakistan". *Social Science and Medicine* Vol 38(9): 1289-1297.

Edwards M., (1987) 'The irrelevance of development studies' *Third World Quarterly* Vol11 (1):116-135.

Ekehammar B., Sidanius J., Nilson I., (1987) "Social Status: Construct and External Validity" *Journal of Social Psychology* Vol 127(5): 473-481.

Erikson R., (1984) "Social Class of Men, Women and Families" *Sociology* Vol 18(4): 500-514.

Faia MA., (1981). "Selection by certification: a neglected variable in stratification research". *American Journal of Sociology* Vol 86:1093-1111.

Faunce WA., and Smucker MJ., (1966) "Industrialisation and community status structure" *American Sociological Review* Vol 31 (June): 390-399.

Fields G., (1994). "Poverty and Income Distribution: Data for measuring poverty and inequality changes in the developing countries". *Journal of Development Economics.* Vol 44: 87-102.

Firth R., (1968)."The Social Framework of Economic Organisations" In (eds.) LeClair E., Schneider H., *Economic Anthropology: Readings in Theory and Analysis.* Holt. NY.

Fischer C.,(1987) "The quality of qualitative research" *Theoretical and Philosophical Psychology* 7: 2-11.

Fogelson RD., Spiro ME., (1965) 'Introduction' In *Context and Meaning in Social Anthropology* Free Press: New York.

Food Studies Group/Sokoine University. (1992). *Agricutural Diversification and Intensification Study.* International Development Centre, Oxford.

Form WH., and Stone GP., (1957) "Urbanism, anonymity and status symbolism" *American Journal of Sociology.* Vol 62 (March): 504-514.

Frank RH., (1989) "Book Review: The Standard of Living: The Tanner Lectures by Amartya Sen". *Journal of Economic Literature* Vol XXVII (June): 664-666.

Gaiha R., and Deolalikar A., (1993) "Persistent, Expected and Innate Poverty: Estimates for Semi-Arid Rural South India, 1975-1984. *Cambridge Journal of Economics* Vol 17(4): 409-421.

Ghiselli EE., (1964). *Theory of Psychological Measurement.* McGraw-Hill. NY.

Goldthorpe JH., (1980). *Social Mobility and Class Structure in Modern Britain* Clarendon Press. Oxford (2nd edition, 1987)

Goldthorpe JH., (1983). "Women and Class Analysis: in Defense of the Conventional view" *Sociology* Vol 17: 465-488.

Grandin B., (1988), *Wealth Ranking in Small Holder Communities: A field manual.* Intermediate Technology Publications, Russell Press. Nottingham.

Green LW., George MA., Daniel M., *et al.*, (1995) *Study of Participatory Research in Health Promotion.* The Royal Society of Canada, BC.

Grice B., (1975) "Logic and Conversation" in Cole P., and Morgan J., (eds.) *Syntax and Semantics* Vol 3. Academic Press. New York. Cited in Taylor I., and Taylor M., (1990) *Psycholinguistics* Prentice-Hill. New Jersey.

Gross N., (1953). "Social class identification in the urban community". *American Sociological Review* Vol 18: 398-404.

Grosvenor-Alsop R., (1991) "Wealth Ranking in a Caste Area of India" Paper presented at Participatory Rapid Rural Appraisal: Introductory Course. IIED and the Agricultural Education and Training Unit, Wolverhampton Ploytechnic, UK.

Groverman V., (1990). "Wealth Ranking in Swaziland: a method to identify the poorest". *RRA Notes* No. 9: 6-11.

Guba EG. and Lincoln YS., (1994) 'Competing Paradigms in Qualitative Research' In Norman K. and Denzin Y. (eds.) *Handbook of Qualitative Research.* Sage: Thousand Oaks (pp 105-117).

Guijt I and Cornwall A., (1995) "Critical reflections on the practice of PRA" *PLA Notes* No. 24: 2-7.

Guion RM., (1976) "Recruiting, selection, and job placement" In Dunnette MD (Ed.) *Handbook of Industrial and Organizational Psychology.* Rand-McNally: Chicago. Cited in Messick (1988).

Haer JL., (1957). "Predictive utility of five indices of social stratification". *American Sociological Review* Vol 22 (Oct): 541-546.

Haller AO., Holsinger DB., and Saraiva HU., (1972). "Variations in Occupational Prestige Hierachies: Brazilian Data". *American Journal of Sociology.* Vol 77 (March): 941-56.

Haller AO., Lewis DM., (1966) "The Hypothesis of Intersocietal Similarity in Occupational Prestige Hierachies". *American Journal of Sociology* Vol 72 (Sept): 210-16.

Hauser RM., (1994). "Measuring socioeconomic status in studies of child developement". *Child Development* Vol 65: 1541-1545.

Hennekens CH., Burning JE., (1987) *Epidemiology in Medicine* Little, Brown and Co. Boston.
Hertfordshire.

Heyer J., (1989). "Landless Agricultural Labourers' Asset Strategies". *IDS Bulletin.* Vol 20(2): 33-40.

Hinton R., (1995) "Trades in different worlds: Listening to refugee voices". *PLA Notes* No.24: 21-26.

Hollingshead AB., Redlich FC., (1958). *Social Class and Mental Illness.* Wiley and Sons: New York (Appendix 1).

House W., Phillips-Howard KD., (1990) "Socioeconomic Differentiation Among African Peasants: Evidence from Acholi, Southern Sudan" *Journal of International Development.* Vol 2(1): 77-109.

Howes M. and Chambers R., (1979) 'Indigenous Technical Knowledge: Analysis, Implications and Issues.' *IDS Bulletin* Vol 10(2):5-11.
http://www.worldbank.org/html/edi/sourcebook/

Hughes D., Seidman E., Williams N., (1993). "Cultural Phenomena and the Research Enterprise: Toward a Culturally Anchored Methodology". *American Journal of Community Psychology* Vol 21 (6): 687-703.

Hymes DH., (1964)."Discussion on Burling's paper ("Cognition and componential analysis: God's Truth or Hocus Pocus?") *American Anthropologist.* Vol 66 (1):116-119.

Illife J., (1987). *The African Poor.* Cambridge University Press. Cambridge. Cited in: Amis and Rakodi (1994).

Jackson EF., Curtis RF., (1968). Conceptualization and Measurement in the Study of Social Stratification". In Blalock H., and Blalock A., (eds) *Methodology in Social Research.* McGraw-Hill.New York.

James SA., Strogatz DS., Ng SB., Ramsey OL., (1987) "Socioeconomic status, John Henryism and Hypertension in Blacks and Whites" *American Journal of Epidemiology* Vol 12: 465-481.

Jodha NS., (1985). "Social Science Research on Rural Change: Some Gaps". Paper presented at Conference on Rural Economic Change in South Asia: Differences in Approach and in Results between large-scale surveys and intensive micro-studies, Banglore, India 5-8 August. Cited in Kabeer N., (1991). *Gender, Production and Well-being: Rethinking the Household Economy* Institute of Development Studies, Discussion paper No: 288. Sussex (p. 3).

Jodha NS., (1988). "Poverty debate in India: a minority view" *Economic and Political Weekly* Special Number: 2421-2428, November.

Jones IG., Cameron D., (1983) "Social class analysis - an embarrassment to epidemiology" *Community Medicine* Vol 5: 116-24.

Kabeer N., (1989). *Monitoring Poverty as if Gender Mattered: methodology for Rural Bangladesh.* Institute of Development Studies. Discussion Paper no: 255, Brighton.

Kahl JA., Davis JA., (1955). "A Comparison of Indexes of Socioeconomic Status". *American Sociological Review* Vol 20: 317-325.

Karim MB., (1990) "Measuring Socioeconomic Status in Rural Households in Third World Countries: The case of Bangladesh" *International Journal of Contemporary Sociology* Vol 27: 3-4.

Kaufman HF., Duncan OD., Gross N., Sewell WH., (1953). "Problems of theory and method in the study of social stratification in rural society" *Rural Sociology* Vol 18: 12-24.

Kim JO. and Mueller C., (1978). *Introduction to Factor Analysis: What it is and how to do it".* Quantitative Applications in the Social Sciences Series No. 13. Sage University Paper. London.

Knox PL., (1974). "Social indicators and the concept of level of living". *The Sociological Review (New Series)* Vol 22: 249-257.

Kraemer HC., (1980). "Extension of the Kappa Coefficient". *Biometrics* Vol 36: 207-216.

Kroeger A. (1983) Health Interview Surveys in Developing Countries: A Review of the Methods and Results *International Journal of Epidemiology* Vol 12(4): 465-481.

Krueger RA., (1994) *Focus Groups: A Practical Guide for Applied Research* Sage. London.

Landis JR. and Koch GG., (1977) The measurement of observer agreement for categorical data. *Biometrics* Vol 33, 159-174.

Lasswell TE (1959) "Social class and size of community" *American Journal of Sociology* Vol 64 (March): 505-508.

Lenski GE., (1966). *Power and Privilege.* McGraw-Hill. NY.

Liberatos P., Link B., Kelsey J., (1988) "The Measurement of Social Class in Epidemiology" *Epidemiological Reviews* Vol 10: 87-121.

Lockwood D., (1958). *The Black Coated Worker.* Allen and Unwin. London. (Second edition, 1989, Clarendon Press. Oxford)

MacIntryre S. and West P., (1991) "Lack of class variation in health in adolescence: an artifact of an occupational measure of social class?" *Social Science and Medicine* Vol 32(4): 395-402.

Mack J. and Lansley S., (1985) *Poor Britain.* George Allen and Unwin. London.
Maliyamkono TL., and Bagachwa MSD., (1990) *The Second Economy in Tanzania.* East African Studies. London.

Maquet J., (1971). *Power and Society in Africa.* World University Library. London.

Marcinkowski T., (1993) 'A Contextual Review of the Qualitative paradigm in EE Research.' In Mrazek R., (ed.) *Alternative Paradigms in Environmental Education Research.* Monographs in Environmental Education and Environmental Studies Vol VIII (pp. 29-79). Troy, OH: NAAEE.

Marx K., (1967) "A Note on Classes" In Bendix R., Lipset S., (eds) *Class, Structure and Power: Social Stratification in Comparative Perspective.* 2nd edition. Routledge and Kegan Paul: London.(p 5-6)

Mayer K. and Buckley W. (1970). *Class and Society.* Random House: New York (3rd edition)

Mbughuni P., (1994) "Gender and Poverty Alleviation in Tanzania" In Bagachwa MSD., (ed.). *Poverty Alleviation in Tanzania: Recent Research Issues.* Dar es Salaam University Press. Dar es Salaam.

McLachlan H., (1986). "Concepts of Poverty and Deprivation" *Journal of Social Policy* Vol 15 (4): 499-501.

McNee A,. Khan N., Dawson S., *et al* (1995) "Responding to Cough: Boholano Illness Classification and Resort to Care in Response to Childhood ARI". *Social Science and Medicine* Vol 40(9):1279-1289.

Mearns R., Shombon G., *et al* (1992). "Direct and indirect uses of wealth ranking in Mongolia" *RRA Notes.* 15:29-38.

Messick S., (1980). "Test Validity and the ethics of assessment". *American Psychologist* Vol 35: 1012-1027.

Messick S., (1988) "The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement". In Wainer H. and Braun H., (eds.) *Test Validity.* Lawrence Erlbaum and Assoc. New Jersey. Chapter 3.

Mishler EG., (1979) 'Meaning in Context: is there any other kind?' *Harvard Educational Review* Vol 49(1): 2

Morgan M., (1983). "Measuring social inequality: Occupational classifications and their alternatives" *Community Medicine* Vol 5: 116-124.

Moser CA. and Kalton G., (1971) *Survey Methods in Social Investigation* Dartmouth. Aldershot (2nd edition).

Mosse D., (1995) "Social Analysis in Participatory Rural Development" *PLA Notes* No.24: 27-33.

Mueller C., and Parcel T., (1981). "Measures of socioeconomic status: Alternatives and Recommendations". *Child Development* Vol 52: 13-30.

Mukherjee N., (1992). "Villagers' Perceptions of Rural Poverty Through the Mapping Methods of PRA". *RRA Notes* No 15: 21-26.

Murphy JW., Longino Jr. CF., (1992) 'What is the justification for a qualitative approach to ageing studies?' *Ageing and Society* Vol 12: 143-156.

Nau DS., (1995) 'Mixing methodologies: Can bimodal research be a viable post-positivist tool?' *The Qualitative Report* Vol 2 (3).
[http://www.nova.edu/ssss/QR/QR2-3/nau.html.]

Nunnally JC., (1967) *Psychometric Theory.* McGraw-Hill. New York.

O'Connell DL., and Dobson AJ., (1984) General observer-agreement measures on individual subjects and groups of subjects. *Biometrics,* 40, 973-983.

Olsen H., (no date) 'Quantitative "versus" Qualitative Research: The Wrong Question. Web document available at: http://www.nova.edu/ssss/QR/text.html.

Orshansky M., (1965). "Counting the Poor: Another Look at the Poverty Profile". *Social Security Bulletin.* Vol XXXVII (1): 3-29.

Osborne A., and Morris TC., (1979) "The rationale for a composite index of social class and its evaluation". *British Journal of Sociology* Vol 30(1): 39-60.

Oxford Reference for Windows (1993). Oxford University Press. Oxford.

Øyen E., (1992) "Some basic issues in comparative poverty research". *International Social Science Journal* Vol 134: 615-626.

Piachaud D., (1981) "Peter Townsend and the Holy Grail". *New Society.* Vol 10 (Sept) 419-421.

Piachaud D., (1987) "Problems in the definition and measurement of poverty". *Journal of Social Policy* Vol. 16(2): 147-64.

Posner KL., Sampson PD., Caplan A., Ward RJ., Cheney FW., (1990) Measuring inter-rater reliability among multiple raters; an example of methods for nominal data. *Statistics in Medicine* 9, 1103-1115.

Pretty J., Subramanian S., *et al.* (1992). "Finding the poorest in a Tamil Nadu village: a sequence of mapping and wealth ranking" *RRA Notes.* 15: 39-42.

Rajaratnam J., Ganesan C., Thasian H., *et al.* (1992). *Validating Wealth Ranking of PRA and Formal Survey in Identifying the Rural Poor.* Project Report. Christian Medical College and Hospital, Tamil Nadu. South India.

Reason R., (1994) 'Three Approaches to Participative Inquiry' In Norman K. and Denzin Y. (eds.) *Handbook of Qualitative Research.* Sage: Thousand Oaks (pp 324-339).

Reichart CS., Cook TD., (1969) 'Beyond Qualitative versus Quantitative Methods'. In Reichart CS and Cook TD (eds.) *Qualitative and Quantitative Methods in Evaluation Research.* Sage: Beverly Hills.

Rein M., (1977) "Problems in the Definition and Measurement of Poverty" in Townsend P., (ed.) *The Concept of Poverty.* Heinmann Educational Books. London. (Chapter 2): pp 46-63.

Roll J., (1992). *Understanding Poverty: A Guide to Concepts and Measures.* Occasional Paper No. 15. Family Policy Studies Centre. London.

Rowntree BS., (1908). *Poverty: A study of town life.* McMillan. London.

Sanbrook R., (1982). *The Politics of Basic Needs: Urban Aspects of Assaulting Poverty in Africa.* Heinnemann Educational Books. London.

Sarch ML., (1992). "Wealth Ranking in The Gambia: Which Households Participated in the FITT Programme?" *RRA Notes* No 15: 14-20.

Sarris A., and van Den Brink R., (1993) *Economic Policy and Household Welfare during Crisis and Adjustment in Tanzania.* Cornell University Food and Nutrition Policy Program. University Press. NY.

Scase R., (1992). *Class.* Open University Press. Buckingham.

Schaefer S., (1992). "The Beans-Game: Experiences with a Variation of Wealth Ranking in the Kivu Region, Eastern Zaire" *RRA Notes* No 15: 27-28.

Schnore LF., and Varley DW., (1955) "Some Concomitants of Metropolitan Size" *American Sociological Review.* Vol 20 (August): 408-414.

Schrekenberg K ., (1995) "The respective merits of RRA and conventional methods for longer-term research" *PLA Notes* No.24:74-77.

Schuler EA., (1940). "Social and Economic Status in a Louisiana Hills Community". *Rural Sociology* Vol V (March): 69-83.

Scoones I., (1991). "Learning about wealth: An example from Zimbabwe". Paper presented at: Participatory Rapid Rural Appraisal: Introductory Course. IIED and the Agricultural Education and Training Unit, Wolverhampton Ploytechnic, UK.

Scoones I., (1995). "Investigating Difference: Applications of Wealth Ranking and Household Survey Approaches among Farming Households in Southern Zimbabwe" *Development and Change.* Vol 26(1):67-88.

Seeley J., Kajura E., Mulder D., (1995). "Methods used to study rural household coping strategies in rural South West Uganda". *Health Policy and Planning.* Vol 10(1): 79-88.

Seigel S., (1956) *Nonparametric Statistics for the Behavioural Sciences* McGraw-Hill and Kagakusha: Tokyo.

Seligson MA., (1977) "Prestige among Peasants: A Multidimensional Analysis of Preference Data". *American Journal of Sociology.* Vol 83 (3): 632-652.

Sen A., (1979). "Issues in the measurement of poverty" *Scandinavian Journal of Economics* Vol 81 : 285-307.

Sen A., (1983). "Poor, relatively speaking". *Oxford Economic Papers* Vol 35: 153-169.

Sewell WH., (1943) "A Short Form of The Farm Family Socioeconomic Status Scale". *Rural Sociology* Vol 8: 161-170.

Shaffer P., (1996) 'Beneath the Poverty Debate' *IDS Bulletin* Vol 27(1):23-35.

Sharrock G., Waldie JK., Joshi YR., (1993). "Wealth Ranking for Agricultural Research Purposes in the Eastern Hills of Nepal". *RRA Notes* No. 18: 53-59.

Shott S., (1990) *Statistics for Health Professionals* Saunders. Philadelphia.

Silverman SF., (1966). "An Ethnographic Approach to Social Stratification: Prestige in a Central Italian Community." *American Anthropologist.* Vol 68 (4): 899-921.

Slesnick DT., (1993) "Gaining Ground: Poverty in the Postwar United States". *Journal of Political Economy.* Vol 101(11): 1-38.

Spector PE., (1992) *Summated Rating Scale Construction: An Introduction.* Sage University Paper No. 82. Sage. Newbury Park.

Stanworth M., (1984) Women and class analysis: a reply to John Goldthorpe. *Sociology* Vol 18: 159-170.

Steeves RH, (1992): "A typology of data collected in naturalistic interviews". *Western Journal of Nursing Research* 14(4): 532-536

Steiner DL. and Norman GR., (1989). *Health Measurement Scales: A practical guide to their development and use.* Oxford Medical Publications. Oxford.

Streeten P., (1990). "Poverty Concepts and Measurement". *The Bangladesh Development Studies* Vol XVIII (3): 1-18.

Sullivan JL. and Feldman S., (1994) " Multiple Indicators: An Introduction" in Lewis-Beck MS (ed.)., *Basic Measurement* Part II: p 59-137 Sage: London

Tain H., (1874). *Notes in England.* p. 300-302. Cited in: Rose M., (1971) *The English Poor Laws 1780-1930.* David and Charles. Devon.

Thompson J., (1996) 'Moving the Indigenous Knowledge Debate Forward' *Development Policy Review* Vol 14: 105-112.

Townsend P., (1977). "Measures of Income and Expenditure as Criteria of Poverty" in

Townsend P., (1979). *Poverty in the United Kingdom.* Penguin Books. Hammonsworth.

Townsend P., (1987). "Deprivation". *Journal of Social Policy* Vol 16 (2): 125-146.

Townsend P., (1994). *The International Analysis of Poverty.* Harvester Wheatsheaf.

Townsend P., (ed.) *The Concept of Poverty.* Heinmann Educational Books. London. (Chapter 6): pp 100-112.

Tuden A., Plotnicov L (eds)., (1970). *Social Stratification in Africa.* Free Press. New York.

Tung L., and Baliña FT., (1993) "Farmer's Experiences on Wealth Ranking in The Philippines: Different farmers have different needs". *RRA Notes* No. 18: 60-65.

Turk C., (1995). "Identifying and Tackling Poverty" *PLA Notes* June: 37-41.
van der Hoeven R., (1988). *Planning for Basic Needs: A Soft Option or a Solid Policy? A Basic Needs Simulation Model Applied to Kenya.* Gower/ILO. Aldershot.

Victoria CG., Huttly SRA., Barros FC., *et al.*,(1992). "Maternal education in relation to early and late child health outcomes: Findings from a Brazilian Cohort Study". *Social Science and Medicine* Vol 34 (8): 899-905.

Viet-Wilson J., (1987) "Consensual Approaches to Poverty Lines and Social Security". *Journal of Social Policy.* Vol 16 (2): 183-211.

Wainer H. and Baun H., (1988) *Test Validity.* Lawrence Elbaun and Assoc. New Jersey.

Walker R. (1981). *Attitudes to Public Expenditure.* Paper given at the Annual Conference of the Social Administration Association, Cambridge, July. Cited in Walker R., (1987).

Walker R., (1987). "Consensual Approaches to the Definition of Poverty: Towards an Alternative Methodology". *Journal of Social Policy* Vol 16 (2): 213-226.

Weber M., (1965). *The Theory of Social and Economic Organisations.* Free Press. New York.

Weber M., (1966). "Class, status and party" In Bendix R., and Lipset S., (eds) *Class, Status and Power*, Routledge and Kegan. London.

Welbourn A., (1992) "A note on the use of disease problem ranking with relation to socio-economic well-being: an example from Sierra Leone" *RRA Notes* 16: 86-87.

Williams GW., (1976) Comparing the joint agreement of several raters with another rater. *Biometrics* 32, 619-627.

World Bank (1993a). *World Development Report 1993*: Washington DC.

World Bank (1993b). *Tanzania: A Poverty Profile.* Population and Human Resource Division, East Africa Department. Report No: 12298-TA (Questionnaire provided with report).

World Bank (1995) Poverty Reduction and the World Bank: Progress and Challenges in the 1990s. WB: Washington.

World Bank (1996) Participation Source Book. Document posted at:
Wright E. and Perrone L., (1977). "Marxist Class Categories and Income Inequality" *American Sociological Review.* Vol 42 (Feb): 32-55.

Zeller RA., Carmines EG., (1980) *Measurement in the social sciences: the link between theory and data.* Cambridge University Press, NY.