A questionnaire-based efficacy index to measure the impact on wellbeing of development interventions in sub-Saharan Africa

Emma Kandelaars

Bachelor thesis Econometrics

July 2009

Supervisor: Dr. A. Koning

1 Introduction

Ever since its emergence, there has been fierce discussion about the best structure, design and implementation of development aid. The wide divergence of opinions within this discussion underscores the ambiguity of results obtained so far by development interventions. This lack of clarity makes the evaluation of development projects indispensable; it allows for the optimal adjustment of project design to local circumstances and can provide practical guidelines and support for the structuring and implementation of future projects. For an overview of the debate on aid effectiveness, the reader is referred to Easterly's 'The White Man's Burden' (Easterly, 2006).

Since the 1990s the focus in the development industry has redirected from the outputs of aid, to the impact and effects of that output (Conlin *et al.*, 2008; OECD, 2002). While the delivery of outputs could be measured using predefined objective indicators (such as the number of wells dug for a water project), the effects of these outputs on the wellbeing of beneficiaries are difficult to predict and impossible to generalize across development interventions. In order to draw conclusions about a programme's impact on beneficiaries, the effects should be evaluated with broadly defined, subjective instruments (for the water project this could be an assessment of the impact of the improved access to water on the lives of the people benefiting from the well). Currently, no evaluation techniques that take this approach exist in humanitarian aid.

The Logical Framework Approach (LFA) emerged as the aid industry's tool for project planning and appraisal (Cracknell, 2000), and is applied by numerous aid agencies (Sartorius, 1991). Field experience has shown that application of the LFA for evaluative purposes is inadequate and that this instrument lacks a proper effect measure (Crawford *et al.*, 2002). A less frequently used evaluation approach, specifically designed to measure the impact of aid, is the Most Significance Change (MSC) technique. This method assesses the effect of a project by systematically evaluating 'significant change stories' which are gathered from the beneficiaries of the project. This technique strategy may easily produce biased results (Dart *et al.*, 2003).

In health care, a commonly applied approach for the assessment of the effects of interventions is the evaluation of changes in quality of life of the participants. Using questionnaires to obtain quality of life measurements before and after the intervention, it is estimated how the intervention affected the wellbeing of its participants. I expect that a similar approach can be adopted for the evaluation of the effects and impact of development aid. Since aid projects attempt to improve their beneficiaries' wellbeing, a questionnaire should be developed to quantify and measure this concept. Repeated measurements within the beneficiaries of a project and within a control group should enable the project evaluator to make an assessment of the wellbeing effect caused by the aid project. With wellbeing as the single outcome measure, results of these evaluations can be compared to results obtained by other development interventions.

Extensive literature has evolved around effect measurement in health care and numerous questionnaires have been developed for this purpose. Direct application of these questionnaires to the development field would not be appropriate since the scope of these questionnaires is limited to health-related quality of life (Hagerty *et al.*, 2001). I will therefore attempt to develop a new questionnaire which can measure changes in wellbeing with a focus on a developmental context.

Since this questionnaire will be administered in varying circumstances, I attempt to develop a dynamic mechanism for converting the scores on the items of the questionnaire into one single index which allows for the optimization of the sensitivity of the technique. This mechanism will ensure that for any local condition an expression can be derived for the optimal determination of the changes in wellbeing.

The purpose of this paper is to develop a questionnaire and corresponding index to optimally measure the change over time in wellbeing of beneficiaries directly caused by a development project. In section 2 I will elaborate on the methods I use to construct the questionnaire and corresponding index, and to evaluate its quality. Section 3 gives insights in the data gathering process, and the results and a discussion of these results are presented in Section 4. I conclude in Section 5.

2 Methodology

In this methodology chapter a division is made in a section concerning the design of the questionnaire and a section concerning the calculation of the index. In the text on the questionnaire I elucidate on its construction process and I define a number of measurement properties that will be used to assess the quality of the questionnaire. The section on the index presents two methods that can be used to convert the answers on the questionnaire into a single index score. I start with a commonly used method that only utilizes the information obtained at one point in time and consequently, I give a detailed deduction of the method I developed which ensures the optimal expression of the change in wellbeing over time measured by the questionnaire.

2.1 Questionnaire

This first section of the Methodology describes both selection of the outcome measure of the questionnaire and the choice of questions and domains, together with the assessment of the quality of the resulting questionnaire.

2.1.1 Construction

2.1.1.1 Concept

As introduced in Section 1, I aim to develop an instrument that can optimally measure the effects of development programmes on the lives of their beneficiaries over time. Since all development programmes are set up with the aim to improve people's wellbeing, wellbeing is the concept needs to be quantified. I define wellbeing as follows:

Individuals' perceptions of their position in life in relation to their goals, expectations, standards and concerns, and in the context of the material and immaterial resources individuals have in their command to maintain and/or improve this position.

2.1.1.2 Domains

Extensive inspection of the existing literature on instruments assessing (health related) quality of life, has led to the identification of two instruments that are the closest approximation for this paper's aim of measuring wellbeing. The first instrument is the WHOQOL-100, which is the most broadly developed and commonly used instrument to assess quality of life (Miller *et al.*, 2008). The

WHOQOL-100 consists of six domains: physical health, psychological, level of independence, social relations, environment and spirituality/religion/personal beliefs. Developed by a medical organization, the main criticism on this questionnaire has been that health has been overemphasized. Domains that were previously found to be important facets of wellbeing (material wellbeing, productivity/employment) are missing from this questionnaire (Hagerty *et al.*, 2001).

The second closely related instrument is the Economist's Intelligence Unit's (EIU) national quality of life measure. By combining the subjective results of life satisfaction surveys with objective factors that have been shown to be associated with life satisfaction in the literature, it identifies the following determinants of wellbeing: material wellbeing, health, political stability and security, family life, community life, climate and geography, job security, political freedom and gender equality (Economist's Intelligence Unit, 2004).

Careful examination of the domains identified by the above instruments and consideration of their shortcomings and deficiencies, has led me to the conclusion that wellbeing could best be segregated into the following domains: material wellbeing, mental and physical health, work and education, social relations, security, freedom and mobility, and overall life satisfaction.

2.1.1.3 Questions

The items in this questionnaire should be designed to jointly measure the concept of wellbeing. As wellbeing is defined as an individual's perception, it should be measured solely with subjective questions. In order to provide development projects with some form of reference frame and to facilitate the future identification of objective factors that are likely to influence wellbeing, this questionnaire will however also contain several objective questions. These objective questions are formulated to each correspond to at least one of the domains in this questionnaire. I emphasize that these objective questions are not part of the subjective wellbeing assessment, but should only be used for reference purposes.

For each domain subjective questions were formulated asking for the respondent's satisfaction with various facets of life. Questions were based on items from the WHOQOL-100 (WHO QOL group, 1995), the Manchester Short Assessment of quality of life (MANSA) (Priebe *et al.*, 1999) and the 2004 UNDP/World Bank poverty assessment of the Maldives (De Kruijk *et al.*, 2004), together with discussions with practitioners from the development field and researchers experienced in questionnaire construction. This procedure resulted in a final selection of: four questions for the health, social relationships and overall life satisfaction domains, three questions for the domains material wellbeing, and freedom and mobility, two questions for work and education and one question for the security domain.

Response categories for all subjective and part of the objective questions follow a five point Likert scale. The remainder of the objective questions are of the yes/no-type.

2.1.1.4 Target population

The target population for this evaluation instrument are beneficiaries with the age of 16 or older of development projects in sub-Sahara Africa.

2.1.1.5 Time frame

This questionnaire should be administered at least twice: before respondents enter into the aid project and after the project has ended. For projects that have no clearly defined end, the second questionnaires should be administered after a pre-defined period of time, which could range from six months to several years, depending on the character of the project. When interested in a more continuous monitoring of the impact of the project, it can of course also be decided to perform additional measurements in between.

2.1.2 Measurement properties

Since the aim of this paper is to measure the unobservable concept of wellbeing by using multiple items that can be quantified, it is important that these individual items are indeed measuring the same concept. This is also referred to as a questionnaire's internal consistency and should be assessed using Factor analysis and Cronbach's alpha (Terwee *et al.*, 2006).

2.1.2.1 Factor analysis

Factor analysis is a technique to describe the covariance relationships among many variables in terms of a few underlying, unobservable constructs. In the basic reasoning underlying factor analysis it is assumed that variables can be grouped by their correlations. Variables constituting a certain group have high correlations among themselves and relatively low correlations with variables in other groups. This makes it plausible that for each group a single underlying construct is responsible for the observed correlations. These constructs are called factors (Jonhson and Wichern, 2002).

In this paper factor analysis is applied to explore the concept of wellbeing and to identify any underlying constructs. Correspondence of any identified constructs to the from the literature derived domains constituting the concept of wellbeing, is assumed to be an indication of good internal consistency of the questionnaire (Terwee *et al.*, 2006).

Factor model

In a factor model it is assumed that each item on a questionnaire can be segregated into a set of underlying factors and an error term. The set of factors is identical for each item, but the factors have varying importance for the different questions. This varying importance is translated into the factor loadings, which indicate how much of the variation in an item is explained by a certain factor. This idea can be formulated with the following theoretical model, adopted from Johnson and Wichern (2002): the observable random vector $\mathbf{X}$, consisting of the p items $X_1, X_2, \ldots, X_p$, has mean μ and covariance matrix $\mathbf{\Sigma}$. It is assumed that $\mathbf{X}$ is linearly dependent upon severable common factors $F_1, F_2, \ldots, F_m$ and p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, called specific factors, where the ith specific factor is associated only with the ith item X_i. The factor analysis model therefore becomes:

$$
\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \ldots + \ell_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \ldots + \ell_{2m}F_m + \varepsilon_2 \\
&\vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots + \ell_{pm}F_m + \varepsilon_p
\end{aligned}
$$

or in matrix notation:

$$
\underset{(p\times 1)}{\mathbf{X}} - \boldsymbol{\mu} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}
$$

The coefficient ℓ_{ij} is called the loading of the ith item on the jth factor, so that the matrix $\mathbf{L}$ is the matrix of factor loadings. The unobservable random vector $\mathbf{F}$ is assumed to have mean $\mathbf{0}$ and covariance matrix $\mathbf{I}$. For the unknown random vector $\boldsymbol{\varepsilon}$ the assumption is made that it has mean $\mathbf{0}$ and diagonal covariance matrix $\boldsymbol{\Psi}$. Lastly, $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are assumed to be independent, so that their joint covariance matrix is equal to $\mathbf{0}$.

The above factor model implies a covariance structure for $\mathbf{X}$. Based on the assumptions on $\mathbf{F}$ and $\boldsymbol{\varepsilon}$, it can be derived that:

$$
\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}
$$

which could also be read as:

$$
\begin{aligned}
&\mathrm{Var}(X_i) = \ell_{i1}^2 + \ldots + \ell_{im}^2 + \psi_i \qquad \text{and} \\
&\mathrm{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \ldots + \ell_{im}\ell_{km}
\end{aligned}
$$

This shows that the variance in each item can partly be explained by the factors and partly by the error term. ψ_i is that portion of $\mathrm{Var}(X_i)$ due to this error term or specific factor. The sum of the squared loadings of the ith item on the m common factors, called the ith communality and denoted by h_i^2, shows how much of the variance in item i is explained by the m factors in the model. It can also be deducted that the matrix of factor loadings $\mathbf{L}$ actually represents he covariance relations between $\mathbf{X}$ and $\mathbf{F}$.

Administering the questionnaire with n respondents gives observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ on the p items of the questionnaire. When the questions show to be sufficiently correlated (that is $\boldsymbol{\Sigma}$ significantly deviating from a diagonal matrix), the factor loadings ℓ_{ij} and specific variances ψ_i can be estimated with various methods. Selection of the appropriate method is dependent on assumptions regarding the generalization of the obtained results and on whether the factor analysis is exploratory or intended for hypothesis testing (Tinsley *et al.*, 1987). Since I am merely interested in exploring the concept of wellbeing and wish to generalize any results obtained from the sample to a population, the maximum likelihood method is selected for factor extraction.

Factor extraction

The maximum likelihood method finds estimates for the factor loadings and specific variances by selecting those values that make the underlying data most 'likely'. For valid application of the

maximum likelihood method, $\mathbf{F}_k$ and $\boldsymbol{\varepsilon}_k$ are assumed to be jointly normal and the observations $\mathbf{X}_k - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_k + \boldsymbol{\varepsilon}_k$ are therefore also normal. The likelihood function then becomes:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right) tr\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{k=1}^{n}(\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]}$$

which depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$. In this model $\mathbf{L}$ can take numerous forms through orthogonal transformation, which makes that this likelihood is not yet well-defined. To ensure that the likelihood is well-defined a uniqueness condition is imposed.

The estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ are obtained by numerical maximization of the maximum likelihood function using the statistical software package SPSS. In order to correct for differences in measurement scales of the questionnaire, the observations are first standardized and the factor analysis is performed using the correlation matrix $\mathbf{R}$. This matrix is inserted in the maximum likelihood expression for the estimator of $\boldsymbol{\Sigma}$.

Factor rotation

Since interpretation of the obtained factor loadings can be difficult, the discrimination between factors can be improved by a technique called factor rotation. If a factor is seen as a classification axis along which variables can be plotted, then factor rotation rotates these factor axes such that variables are loaded maximally to one factor. Ideally I thus strive to obtain a solution where each variable loads highly onto one single factor and has relatively small loadings for all other factors (Field, 2009).

Different rotation methods can be applied and selection of the appropriate method is based on the dependency of the factors. On a theoretical level dependency between the different constructs constituting wellbeing is very likely and therefore oblique rotation of the original factor loadings is conducted in this paper.

The oblique rotation method Direct Oblimin amounts to finding a transformation matrix $\mathbf{T}$ $\left(\mathbf{T} = t_{gj} \text{ with } (j, g = 1,2,\ldots,m)\right)$ that minimizes the function $F\left(\mathbf{L}(\mathbf{T}')^{-1}\right)$ under the condition that $\text{diag}(\mathbf{T}'\mathbf{T}) = \mathbf{I}$. With:

$$F\left(\mathbf{L}(\mathbf{T}')^{-1}\right) = \sum_{j<g}^{m}\left(\sum_{i=1}^{p}\ell_{ij}^2\ell_{ig}^2 - \frac{\gamma}{p}\sum_{i=1}^{p}\ell_{ig}^2\sum_{i=1}^{p}\ell_{ig}^2\right)$$

where $\mathbf{L}$ is the original factor loading matrix and ℓ_{ij} is thus the loading of the ith variable on the jth factor. For practical purposes, the value of γ should be zero or negative. Minimization of $F\left(\mathbf{L}(\mathbf{T}')^{-1}\right)$ is attained by systematically performing rotations until $F\left(\mathbf{L}(\mathbf{T}')^{-1}\right)$ converges (Harman, 1976). This minimization routine is performed using the statistical software package SPSS.

2.1.2.2 Reliability analysis

For a scale it is important that all individual items in the questionnaire are consistently reflecting the same concept. If the separate questions are not consistently measuring the concept of interest, it

would not be justified to convert them into one final overall score. If the questionnaire is randomly split in half and respondents' scores on both halves of the questionnaire are highly correlated, it is likely that the individual questions are indeed reflecting the same concept (Field, 2009). This idea is approximated with Cronbach's alpha, which is defined in the following way:

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p} \mathrm{Var}(X_i)}{\mathrm{Var}(Y)}\right)$$

where p is the number of items X_i and $Y = \sum X_i$ (Cronbach, 1951). A low Cronbach's alpha thus indicates a lack of consistency between the items in the questionnaire, while too high values signify that some items in the scale could be redundant. It is generally accepted that a value between 0.7 and 0.95 indicates a good internal consistency of the questionnaire (Terwee *et al.*, 2006).

Because wellbeing is assumed to consist of several underlying constructs (to be multidimensional), the Cronbach's alpha should be determined separately for these different domains. The coefficient then indicates to what extent the individual items in a particular domain are consistently measuring that specific underlying construct (Field, 2009).

2.2 Index

Once the questionnaire has been constructed, it has to be decided how to convert the scores obtained on the items of the questionnaire into one final wellbeing index score. It could be assumed that all questions are equally important to the concept they are representing, which would amount to simply adding the scores on the questions to obtain the index score, but this does not seem to be a very realistic assumption. After all, it would be logical that a respondent feels that for example his level of satisfaction with his financial situation is more determining for his total level of wellbeing than his satisfaction with say, his access to media. From this reasoning follows that different weightings should be ascribed to each question, so that the weighted scores together give an expression for a respondent's total level of wellbeing.

The question specific weightings cannot be observed directly and therefore need to be estimated. Different methods can be used to perform this estimation and I start by describing a very commonly used method that determines the question weightings using data that is obtained at one point in time. Consequently I deduct a new method that utilizes information obtained at different points in time to optimally express any change in wellbeing measured.

2.2.1 Factor score coefficients

With the factor analysis described in Section 2.1.2 factor loadings for each question can be obtained. Although it might seem logical to use these factor loadings as question specific weightings, the fact that this approach would not correct for differences in measurement scales or variable variances, makes that this technique is rarely used (Field, 2009). Instead, the factor loadings are usually employed in a multivariate regression of the factors on the items. The coefficient estimates of this regression analysis produce weightings for each question, which could then be used to obtain factor scores for each respondent. This can be formulated as follows:

$$\hat{\mathbf{f}}_k = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_k - \bar{\mathbf{x}})$$

where $\hat{\mathbf{f}}_k$ is the kth factor score vector, $\hat{\mathbf{L}}$ is the maximum likelihood estimate for the factor loading matrix, $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix and $\mathbf{x}_k$ is a vector with observations. The question weightings are thus given by $\hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}$, where I used the original correlation matrix instead of $\hat{\boldsymbol{\Sigma}}$ to reduce the effects of a possibly incorrect determination of the number of factors (Johnson and Wichern, 2002).

Applying this regression method I calculate the question specific weightings for each domain and using the proportion of variance explained by each factor as a proxy for the factors' relative importance to the total level of wellbeing, the questions' index weightings can be obtained.

2.2.2 Optimal expression change

When performing measurements at multiple points in time, a respondent's score on a particular question could be segregated into several separate elements. Naturally, every question has an item specific grand mean, which is the average response to a specific question. For a particular respondent, the score on a question is obviously not only determined by this grand mean, but this mean is supplemented with some respondent specific effect. When measurements are performed at multiple points in time, the moment of measurement also starts playing a role and the grand mean and respondent specific effect are additionally complemented with a time element. Any remaining variation in scores is captured by an error element.

This intuitive separation of scores into different elements could be formulated into the following two-way layout MANOVA model: performing multiple measurements within n respondents $(k = 1,2,..,n)$ at times $t = 1,2,..,s$, the response of respondent k at time t is given by:

$$\mathbf{X}_{kt} = \boldsymbol{\mu} + \boldsymbol{\alpha}_k + \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_{kt}$$

where $\mathbf{X}_{kt}$, $\boldsymbol{\mu}$, $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_t$, $\boldsymbol{\varepsilon}_{kt}$ are $(p \times 1)$ vectors with values for the p items on the questionnaire $(i = 1,2,..,p)$ and where $\boldsymbol{\mu}$ is a vector with unknown grand means, $\boldsymbol{\alpha}_k$ is a vector with unknown respondent-dependent values, $\boldsymbol{\beta}_t$ a vector with unknown time-dependent values and $\boldsymbol{\varepsilon}_{kt}$ a random vector with values that are both respondent and time variant. $\boldsymbol{\varepsilon}_{kt}$ is assumed to follow a jointly normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

For each respondent it can be assumed that his or her level of wellbeing is a combination of the items in the questionnaire. The differences in relative importance of these items justify the assignment of different weightings to each item. Adding the weighted scores on all questions would then produce a respondent's total level of wellbeing. Since this questionnaire is used to measure the change in wellbeing over time, it could be effective to assign such weightings to the items on the scale that the change in wellbeing is optimally expressed. In order to acquire such weightings, wellbeing could be formulated in the following way:

$$Y_{kt} = \mathbf{a}'\mathbf{X}_{kt} = \mathbf{a}'\boldsymbol{\mu} + \mathbf{a}'\boldsymbol{\alpha}_k + \mathbf{a}'\boldsymbol{\beta}_t + \mathbf{a}'\boldsymbol{\varepsilon}_{kt}$$

where Y_{kt} is the wellbeing of respondent k at time t, $\mathbf{a}$ is a $(p \times 1)$ vector with question specific weightings, which are assumed to be constant over time, and $\boldsymbol{\mu}$, $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_t$ and $\boldsymbol{\varepsilon}_{kt}$ are as defined above. Y_{kt} could then also be written as:

$$Y_{kt} = \delta + \gamma_k + \theta_t + \delta_{kt}$$

where δ_{kt} is assumed to follow a normal distribution with mean $\mathbf{a}'\mathbf{0} = 0$ and measurement error variance $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \sigma^2$, which is assumed to be constant over time. Closer inspection shows that this model amounts to a two-way layout ANOVA model, where the wellbeing of respondent k at time t can again be segregated into a grand mean, a respondent effect, a time effect and an error term. When measurements are performed at two different points in time, a respondent's change in wellbeing will solely be determined by the time and error effect, all other elements are constant over time after all. The expected change in wellbeing over the period $t = 1$ to $t = 2$ for respondent k can therefore be given by $E(Y_{k2} - Y_{k1}) = E((\theta_2 - \theta_1) + (\varepsilon_{k2} - \varepsilon_{k1}))$, which equals $(\theta_2 - \theta_1)$ since $E((\varepsilon_{k2} - \varepsilon_{k1})) = 0$. The variance of this change in wellbeing is given by $\mathrm{Var}(Y_{k2}) + \mathrm{Var}(Y_{k1}) - 2\mathrm{Cov}(Y_{k2}, Y_{k1})$, which equals $\sigma^2 + \sigma^2 = 2\sigma^2$, because the covariance between Y_{2k} and Y_{1k} is 0. In the remainder of this paragraph I restrict myself to the situation of measurements performed at two different points in time, but the results obtained are readily extended to cases with more than two measurements.

Determination of the optimal weightings is attained by maximizing the on $\mathbf{a}$ dependent expected change in wellbeing over time, relative to the variance of the measurement error, which also depends on $\mathbf{a}$. This gives the following maximization statement:

$$\mathrm{Max}\left(\frac{\theta_2 - \theta_1}{\sqrt{2\sigma^2}}\right)^2 = \mathrm{Max}\left(\frac{\mathbf{a}'\boldsymbol{\beta}_2 - \mathbf{a}'\boldsymbol{\beta}_1}{\sqrt{2\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}\right)^2 = \mathrm{Max}\,\frac{\left(\mathbf{a}'(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)\right)^2}{2\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}$$

Using the Cauchy-Schwarz maximization lemma (see (2-50) Johnson and Wichern, 2002, p.81), it is determined that the above expression is maximized when:

$$\mathbf{a} = c(\boldsymbol{\Sigma})^{-1}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)$$

for any constant $c \neq 0$. This means that the optimal question specific weightings can be attained multiplying the inverse covariance matrix of the error term by the difference of the time effect. The optimal weightings are determined using Matlab.

In the preceding argument it is assumed that wellbeing is a combination of the items in the questionnaire, that is, wellbeing is said to be one-dimensional. Before I have argued however, that wellbeing is constituted by different domains and that these domains are represented by a combination of the questions, which amounts to wellbeing being a multidimensional concept. Therefore I first use the deducted method to obtain question weightings that express the items' relative importance to the domain they are constituting. Using these weights domain scores for each participant can be calculated and these are then used to attain domain weightings which express the importance of the domains to the total level of wellbeing. Multiplying the question specific weightings by the weighting of the relevant domain, gives the final multi-dimensional weightings.

3 Data

3.1 Collection data

As this research was subject to a very limited time frame, it was decided to solely collect data from projects that were believed to have an impact on the lives of their beneficiaries within 3 weeks. Discussion with experienced practitioners from the development field, led us to the conclusion that micro finance and income generating projects would best be suited for this purpose. Several of such aid organizations in Uganda and Ghana were contacted and three projects (Casud, Ghana (www.casud.nl), Adopteer een Geit, Uganda (www.adopteereengeit.nl) and SYPO microdevelopment, Uganda (www.sypo.nl)) were found willing to contribute to this research. Casud and SYPO microdevelopment provide small loans to groups of women in order to enable them to set up or expand their own business. Profits from these businesses are supposed to boost family income. Adopteer een Geit stimulates the development of single teen ache mothers, by giving away goats which provide them with an independent source of income.

Most respondents were unable to read English and in order to create identical administration conditions all questionnaires were administered by an interviewer. Each project was asked to select one interviewer to administer all the questionnaires. Since neither of the authors was fluent in the native language of Uganda or Ghana, the questionnaire was composed in English and had to be translated by the interviewers. The interviewers were asked to formulate a translation of the complete questionnaire before the start of the interviews and to exactly follow this translation with the each administration of the questionnaire.

The interviewers received extensive instructions to emphasize to the respondents that the results of this questionnaire will in no way influence the respondent's participation in the related aid project. Next to this, interviewers were asked to take respondents to a place with sufficient privacy and no distraction, and to make sure that the respondents felt comfortable to answer these private questions. It was also emphasized that the interviewer should only read out the questions and under no condition explain or comment on a question.

Respondents that were to benefit from a project, were interviewed before they were included in the project and again two or three weeks after inclusion in the project. Practically, this meant that the women were interviewed before they obtained their loan or received their goat and two to three weeks after reception of the goat or credit. A second group of respondents was not included into a project, but was also interviewed with a time interval of two or three weeks. A third group of respondents was interviewed only once.

After administering both rounds of questionnaires, the filled out forms were either scanned and sent by email or simply sent by post.

3.2 Descriptives data

In total 74 people were interviewed, of which 53 people were interviewed twice. Of these 74 respondents 51 were included in an aid project (15 Casud, 28 Adopteer een Geit and 7 SYPO microdevelopment). Respondents were between 16 and 77 years old, with an average age of 30 years. The majority of respondents was female, since all projects focussed on female beneficiaries; in

total 67 women and 7 men were interviewed. 50% Of the respondents was single, 39% was married, 8% was either separated or divorced and 3% was widowed.

No respondents were found to have obtained either the highest or lowest possible score, which indicates that no ceiling effects are present. It could therefore be assumed that the employed range of the scale allows for sufficient differentiation between respondents.

4 Results and discussion

This section presents the results obtained on the measurement properties of the questionnaire and describes the question weightings deducted through both methods.

4.1 Questionnaire

Initial factor and reliability analyses were conducted to explore the dataset. Any items that had either too little correlation with any other items on the scale (less than 15% of inter item correlations exceeding 0.3), did not load substantially on any of the identified factor (no loadings over 0.4), or showed to negatively affect scale reliability, were deleted from the dataset (criteria from Field, 2009). In this process the following 3 questions were discarded: 'How satisfied are you with your achieved level of education?', 'How satisfied are you with your mental health?' and 'How often do you have negative feelings such as blue mood, despair, anxiety, depression?'.

The Kaiser-Meyer-Olkin measure $(KMO = 0.63)$ for the final dataset indicated that the sample size was sufficient. The KMO values for the individual items were all above 0.5, except for 2 items. These items however had values only just below 0.5 (0.46 and 0.48). Barlett's test of sphericity $\chi^2(df = 171) = 455.20$ with $(p < 0.0001)$ indicated that correlations between questions were sufficiently large.

The final factor analysis was conducted on the remaining 18 items of the wellbeing scale using the maximum likelihood method. The initial factor loadings were rotated using the oblique rotation method Direct Oblimin. The goodness of fit test for 4 factors had a $\chi^2(df = 101) = 135.25$ with $(p = 0.013)$, which could justify the selection of 4 extracted factors. Our analysis however was bothered by a Heywood case, which could lead to inaccurate results. The factor loadings obtained after rotation are displayed in Table 1.

The questions that cluster on the same factors, suggest that factor 1 represents everyday life, factor 2 social relations, factor 3 freedom and relations, and factor 4 standard of living. For the items that are assumed to constitute a particular domain, the corresponding factor loadings are in bold. Items were assigned to a domain when the relevant absolute factor loading exceeded 0.4 (criteria Field, 2009), except in the case of the question 'How satisfied are you with your personal safety' for factor 1 and the question 'To what extend do you feel that you can accomplish what you want in life?' for factor 3. Those two questions showed to deplete the corresponding subscale reliabilities and it was therefore decided not to relate them to those two domains. The results of all reliability analyses of the sub scales composed by the questions corresponding to the factor loadings in bold, are also presented in Table 1. The division of the questions over the domains is used in subsequent analyses.

	Rotated factor loadings			
Item	**Everyday life**	**Social relations**	**Freedom and relations**	**Standard of living**
How much do you enjoy life?	**,967**	-,118	,121	,120
How satisfied are you with your work life or training or education as your main occupation? If unemployed or retired, how satisfied are you with being unemployed or retired?	**,573**	,209	,201	,085
How satisfied are you with your relationships with your friends?	,132	**,964**	,136	-,220
How satisfied are you with your personal safety?	-,411	**,683**	,229	,051
To what extent do you feel that your life is meaningful?	,079	**,599**	-,161	,017
How satisfied are you with your relationship with your family?	-,030	**,533**	,324	,312
How satisfied are you with your role in the community?	,065	**,408**	,187	,115
How satisfied are you with your sex life?	-,032	,114	**,854**	-,190
How satisfied are you with your access to media such as newspapers, radio, television and internet?	,048	,001	**,665**	,059
How satisfied are you with your romantic relationship or with the fact that you do not have one?	,289	,062	**,572**	-,044
To what extent do you feel that you can accomplish what you want in life?	,292	**,465**	-,471	,095
How satisfied are you with your access to transportation?	,110	,137	**,424**	,192
How satisfied are you with your home?	,098	-,125	,014	**,729**
How satisfied are you with your personal possessions?	-,011	-,129	,036	**,589**
How satisfied are you with your access to health services?	-,099	,300	,187	**,554**
How satisfied are you with your physical health?	,185	,099	,200	**,545**
To what extent do you have the opportunity for leisure activities?	-,039	-,032	-,227	**,524**
How satisfied are you with your financial situation?	,015	,105	,053	**,464**
How satisfied are you with your life as a whole?	,110	,114	-,191	**,455**
Eigenvalues	**3,769**	**1,883**	**1,932**	**2,245**
% of variance	**19,836**	**9,909**	**10,166**	**11,817**
Cronbach's α	**0,377**	**0,783**	**0,739**	**0,757**

Table 1: Summary of the exploratory factor analysis results for the wellbeing questionnaire.

4.2 Index

Using both the conventional factor scores coefficient method and the method deducted to optimally express the change in wellbeing measured over time, I determined the weightings for each question. These weightings are represented in Table 2, where the questions are clustered by domain.

Item	Item weightings	
	Factor scores coefficients	Optimal expression change
How much do you enjoy life?	0,38	-0,261
How satisfied are you with your work life or training or education as your main occupation? If unemployed or retired, how satisfied are you with being unemployed or retired?	0,00	-1,234
How satisfied are you with your relationships with your friends?	0,18	-0,568
How satisfied are you with your personal safety?	0,00	-0,198
To what extent do you feel that your life is meaningful?	0,00	-0,239
How satisfied are you with your relationship with your family?	0,01	0,504
How satisfied are you with your role in the community?	0,00	-0,185
To what extent do you feel that you can accomplish what you want in life?	0,11	-0,071
How satisfied are you with your sex life?	0,03	-0,199
How satisfied are you with your access to media such as newspapers, radio, television and internet?	0,02	0,497
How satisfied are you with your romantic relationship or with the fact that you do not have one?	-0,03	-0,055
How satisfied are you with your access to transportation?	0,01	0,003
How satisfied are you with your home?	0,05	0,239
How satisfied are you with your personal possessions?	0,03	-0,014
How satisfied are you with your access to health services?	0,05	-0,72
How satisfied are you with your physical health?	0,04	0,694
To what extent do you have the opportunity for leisure activities?	0,03	0,69
How satisfied are you with your financial situation?	0,02	0,445
How satisfied are you with your life as a whole?	0,02	-0,77

Table 2: Question weightings derived by the factor score coefficient method and the method that optimally expresses the change in wellbeing measured over time.

Using these weightings I determined the total wellbeing index scores of all respondents and calculated the percentage absolute change in scores measured for the three different projects. Using a dependent samples t-test I tested whether the methods yield significantly different results. This is presented in Table 3.

Project	Mean percentage absolute change in scores		t-statistic	p-value
	Factor score coefficients	Optimal expression change		
Casud	9,16	47,21	3,895	0,002
Adopteer een Geit	19,19	59,19	3,323	0,003
SYPO microdevelopment	6,97	28,99	1,486	0,188

Table 3: Overview of the mean percentage absolute changes per project as measured using the two different methods for the determination of question weightings. The t-statistics and corresponding p-values indicate to what extent the percentage absolute changes are significantly different for the two methods.

5 Conclusion

In order to provide aid projects with an instrument to assess the effects of their development project, I have developed a questionnaire and index that measure the change in wellbeing over time of the beneficiaries of aid projects.

Using the already existing literature on wellbeing, I defined the concept of wellbeing, the domains that constitute this concept and I formulated 21 subjective questions. These subjective questions formed the heart of the wellbeing assessment, but were in the questionnaire also accompanied by objective questions in order to provide projects with some sort of reference frame.

The internal consistency of the questionnaire was assessed using factor and reliability analysis. The factor analysis resulted in a division in factors that did not completely match the domains that were originally subtracted from the literature. This suggests that in future research the domains constituting wellbeing should be reconsidered. The reliability analysis showed that the items on the scales are consistency measuring the same concept, which indicates that the questions have been properly formulated.

To convert the answers given to the questionnaire into a single wellbeing score, I used both the conventional factor scores coefficient method and I developed a method to optimally express the change in wellbeing over time. Both methods delivered clearly differentiating results, where the newly derived method showed to yield a more optimal expression of the change in wellbeing over time.

Although subject to a very limited time frame, this research has provided a good first step on the way to proper effect measurement of aid projects. Because of the indisputable importance of proper effect measurement in development aid, this instrument can form a valuable contribution to the field. Future research on this instrument recording more respondents from a broader spectrum of projects over a longer time period should be performed to further strengthen and validate this tool, so as to ensure that it can be developed into the standard used by all parties involved in development cooperation.

6 References

Beckerman H., M.E. Roebroeck, G.J. Lankhorst, J.G. Becher, P.D. Bezemer and A.L.M. Verbeek. Smallest real difference, a link between reproducibility and responsiveness. Quality of Life Research 10: 571–578, 2001.

Björkman T, Svensson B. Quality of life in people with severe mental illness. Reliability and validity of the Manchester Short Assessment of Quality of Life (MANSA). Nord J Psychiatry (2005) 59:302_/306. Oslo. ISSN 0803-9488.

Conlin S. and R.L. Stirrat. Current Challenges in Development. Evaluation. Evaluation (2008) 14; 193

Cortina J.M. What Is Coefficient Alpha? An Examination of Theory and Applications. Journal of Applied Psychology. (1993) 78; 1,98-104.

Crawford P, P Bryce. Project monitoring and evaluation: a method for enhancing the efficiency and effectiveness of aid project implementation. International Journal of Project Management (2003) 21; 363–373.

Cronbach L.J. Coefficient alpha and the internal structure of tests. PSYCHOMETRIKA (1951) 16; 3.

Dart J. and R. Davies. A Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique. American Journal of Evaluation (2003) 24;2,137–155.

Easterly, William. The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good. Penguin Press HC, The (2006) ISBN 1594200378

Field, A. Discovering statistics using SPSS. Sage Publications Ltd, third edition (2009), ISBN 9781847879066.

Hagerty, M.R., R.A. Cummins, A.L. Ferriss and K. Land. Quality of life indexes for national policy: Review and agenda for research Social Indicators Research (2001) 55, 1.

Harman H.H., Modern Factor Analysis. The University of Chicago Press, Third Edition Revised (1976), ISBN 0226316521.

Jaeschke R. Ascertaining the Minimal Clinically Important Difference. Controlled Clinical Trials (1989) 10,407-415.

Johnson, R.A., Wichern D.W., Applied multivariate statistical analysis. Pearson Education Ltd., Fifth edition (2002), ISBN 0131219731.

Make B. How Can We Assess Outcomes of Clinical Trials: The MCID Approach. COPD: Journal of Chronic Obstructive Pulmonary Disease (2007) 4;3,191-194.

Miller S.M., F. Chan, J.M. Ferrin, C Ling and J.Y.C. Chan. Confirmatory Factor Analysis of the World Health Organization Quality of Life Questionnaire–Brief Version for Individuals With Spinal Cord Injury. Rehabil Couns Bull OnlineFirst (2008) 10.1177/0034355208316806.

Priebe S., P. Huxley, S. Knight and S. Evans. APPLICATION AND RESULTS OF THE MANCHESTER SHORT ASSESSMENT OF QUALITY OF LIFE (MANSA). International Journal of Social Psychiatry (1999) 45;1, 7-12.

Terwee C.B., F.W. Dekker, W.M. Wiersinga, M.F. Prummel and P.M.M. Bossuyt. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. Quality of Life Research (2003) 12, 349–362.

Terweea C.B., S.D.M. Bota, M.R. de Boera, Daniëlle A.W.M. van der Windta, D.L. Knola, J. Dekkera, L.M. Boutera and H.C.W. de Veta. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology (2007) 60;34-42.

Tinsley H.E. A. and D. J. Tinsley. Uses of Factor Analysis in Counseling Psychology Research. Journal of Counseling Psychology (1987) 34; 4, 414-424.

Appendix

Wellbeing questionnaire

Personal characteristics

Before we begin we would like to ask you to answer the following questions about yourself.

Name		
Name father		
Name mother		
Place of residence		
Date of birth (DD/MM/YYYY)		
Place of birth		
Sex	M	F
Marital status	Single	Separated
	Married	Divorced
	Living as married	Widowed

Questions

This questionnaire asks you how you feel about different aspects of your life. Please answer all the questions. If you are unsure about which response to give to a question, please choose the one that appears to be most appropriate. This can often be the answer that first comes to mind.

Please keep in mind your standards, hopes, pleasures and concerns. Listen to each question carefully, assess your feelings and give the answer that appears to be most appropriate for you. We ask that, unless indicated otherwise, you think about your life in the past four weeks.

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
1.	How satisfied are you with your life as a whole?	1	2	3	4	5
2.	How satisfied are you with your financial situation?	1	2	3	4	5

		0 – 12,500 UGX	12,501 – 37,500 UGX	37,501 – 125,000 UGX	125,001 – 312,500 UGX	More than 312,501 UGX
3.	What is your income per week? Take the average over the past four weeks.	1	2	3	4	5

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
4.	How satisfied are you with your home?	1	2	3	4	5
5.	How satisfied are you with your personal possessions?	1	2	3	4	5

		Tap in the house	Water tank near the house	Less than 20 minutes walk (back and forth)	Between 20 and 60 minutes walk (back and forth)	More than 60 minutes walk (back and forth)
6.	What is the water source nearest to your home?	5	4	3	2	1

		Yes	No
7.	Do you have close access to electricity?	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
8.	How satisfied are you with your achieved level of education?	1	2	3	4	5

		No education	Primary school	Secondary school	Subsequent diploma, other than university	University
9.	What is your highest completed level of education?	1	2	3	4	5

		Yes	No
10.	Can you read and write?	2	1

		Yes	No
11.	Do you know how to operate a computer?	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
12.	How satisfied are you with your physical health?	1	2	3	4	5
13.	How satisfied are you with your mental health?	1	2	3	4	5
14.	How satisfied are you with your access to health services?	1	2	3	4	5

		0-2 days	3-5 days	6-10 days	11-18 days	19-28 days
15.	During the past four weeks, how many days did poor physical or mental health keep you from doing your usual activities?	5	4	3	2	1

		Not at all	A little	A moderate amount	Very much	Extremely
16.	How much do you enjoy life?	1	2	3	4	5
17.	To what extent do you feel that your life is meaningful?	1	2	3	4	5
18.	To what extent do you feel that you can accomplish what you want in life?	1	2	3	4	5

		Never	Seldom	Quite often	Very often	Always
19.	How often do you have negative feelings such as blue mood, despair, anxiety, depression?	5	4	3	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
20.	How satisfied are you with your personal safety?	1	2	3	4	5

		Yes	No
21.	In the past year have you been a victim of physical violence?	1	2
22.	In the past year have you been accused of a crime?	1	2

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
23.	How satisfied are you with your relationship with your family?	1	2	3	4	5
24.	How satisfied are you with your relationships with your friends?	1	2	3	4	5
25.	How satisfied are you with your romantic relationship or with the fact that you do not have one?	1	2	3	4	5
26.	How satisfied are you with your sex life?	1	2	3	4	5

		Yes	No
27.	In the past week have you visited a friend, been visited by a friend or met a friend outside home or work?	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
28.	How satisfied are you with your role in the community?	1	2	3	4	5

		Yes	No
29.	Do you fulfil an official unpaid function in the community?	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very satisfied
30.	How satisfied are you with your work life or training or education as your main occupation? If unemployed or retired, how satisfied are you with being unemployed or retired?	1	2	3	4	5

		0 hours	1-10 hours	11-20 hours	21-30 hours	More than 30 hours
31.	How many hours per week do you perform paid work, also including farming and other income generating activities?	1	2	3	4	5

		Not at all	A little	Moderately	Mostly	Completely
32.	To what extent do you have the opportunity for leisure activities?	1	2	3	4	5

		Yes	No
33.	Do you have a hobby?	2	1

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very Satisfied
34.	How satisfied are you with your access to transportation?	1	2	3	4	5

		Very dissatisfied	Dissatisfied	Neither dissatisfied nor satisfied	Satisfied	Very Satisfied
35.	How satisfied are you with your access to media such as newspapers, radio, television and internet?	1	2	3	4	5

		Yes	No
36.	Do you, or does anyone in your household, own a mobile phone?	2	1

Do you have any comments about this questionnaire?

.....

.....

.....

.....

.....

THANK YOU VERY MUCH FOR YOUR HELP