

Learning toward Transformation: Evaluating Material, Social, and Spiritual Impacts in Western Kenya

Jeffrey R. Bloem

The University of Minnesota*

Abstract: Transformational development—integrating the three goals of positive material, social, and spiritual change—is a popular concept among organizations and individuals who work in faith-based development. Despite broad agreement about the theory of transformational development, very little research empirically investigates how to best achieve these goals. By using a difference-in-differences empirical strategy, this study provides an example of a simple yet rigorous evaluation approach for measuring real progress toward meeting the goals of material, social, and spiritual change. This study finds effects after one year that are only marginal in their statistical significance but relatively large in their average impact. Thus, while economically meaningful impacts may be present, this study encounters difficulty with statistical identification due to a relatively small sample size. Nevertheless, this paper provides a useful example, with suggestions for improvement, for future research evaluating transformational development programs. JEL codes: O12, O20, Q01, and L31

*The author would like to thank the following people who have made this project possible and have improved the quality of this research: Renita Reed, Phil Walker, Alfred Kibairu, Elly Kisala, Leonard Ashavaga, and the faculty and staff at the Africa Theological Seminary for assistance and guidance in performing the field work for this paper. The author also thanks two anonymous referees for invaluable comments and feedback on an initial draft of this paper. Please direct all comments regarding this paper to bloem.jeff@gmail.com.

1 Introduction: The False Dichotomy

Ever since Wayne Bragg coined the term “transformational development,” the three goals of positive material, social, and spiritual change have been combined to form a unified theory (Bragg, 1984). Although the concept of transformational development is well developed (Myers, 2011), practical tension persists between desires to love God and love neighbor (Matthew 22:37–39; Mark 12:30–31; Luke 10:27; John 13:34). It is perhaps natural to view the world as two distinct and mutually exclusive parts: the spiritual world and the material world. The spiritual world is a private and personal place, concerned with a relationship with God, and strictly in the domain of the church; the material world is a public place, concerned with facts and evidence, and is the domain of business and the government.

In his classic book, *Walking with the Poor*, Bryant Myers explains how this separation between the spiritual and the material, or more specifically between Christian witness and social action, are at odds with the goals of transformational development (Myers, 2011). He points out that most modern governments and secular development organizations intentionally separate religion from development. This false dichotomy falls short of the richness of a theologically integrated public life, and suggests that God’s redemptive work takes place only in the spiritual world. Myers goes on to write:

The Christian development agency is not immune to [this] phenomenon. We express our captivity to a modern worldview when we say that holistic ministry means combining evangelism (meeting spiritual need) with relief and development (meeting physical need) as if these were divisible realms and activities. Then we make it worse by insisting that the church or the evangelism part of our organization do the former, while the development agency does the latter. A number of serious and thoughtful Christian groups are organized this way. By so doing we declare development independent of religion, something most of us do not really believe. (p. 7)

Most striking, from the perspective of an empirical development economist, is the lack of rigorous evaluation of programs within these organizations. Many Christian development organizations do not regu-

larly engage in externally-conducted impact evaluations, but more often perform less rigorous program evaluations internally.¹ An unfortunate consequence of this lack of commitment to collecting rigorous evidence is the underdevelopment of the practice of transformational development itself. If development is most effective and transformative when the spiritual world is integrated with the material world, like many believe, then Christian development agencies ought to be showing the rigorous evidence of this reality. Alas, this is hardly the case today.

A small group of faith-based development practitioners designed the Developing Marketplace Leaders (DML) program to explicitly expose, and experiment within, this false dichotomy in faith-based development practice. For many years faith-based NGOs have implemented faith-inspired business skills training programs that not only teach basic business skills but also emphasize themes such as social responsibility, human dignity, and the Biblical role of business people. After several years working first-hand with such programs, a small group of practitioners started to wonder if they could do better. Perhaps if spiritual integration and transformation was key to the theory of change of their program, then maybe local churches should be involved in some more meaningful manner.

In this paper, we² present empirical results from an impact evaluation of the DML program, a church-based business skills training program that aimed to make a transformational impact (e.g., positive social, material, and spiritual changes) in Western Kenya. We implement a difference-in-differences (DID) empirical strategy to identify the impact of the DML program across a variety of outcomes spanning the social, material, and spiritual realms of life. To perform this analysis, we collected baseline and endline data on both program participants and a comparison group. It is important to note at the outset that although the terms treatment and comparison group are used frequently throughout this paper, this study is not a randomized control trial (RCT). Proper randomization of participation in the DML program was not feasible due to several limiting factors at the time of program implementation.

The analysis presented in this paper finds several relatively large impact estimates. In particular, a year after the program began DML program participants reported an increase in household income and enterprise assets of roughly 21% and 18%, respectively, but a decrease in reported enterprise sales and profits of 38% and 40%, respectively. In addition, we find that, on average, participants report more spiritu-

al integration into their business activities. These findings should be accompanied by an important caveat regarding statistical inference. We do not find statistically significant effects, at the typical 5% level, across any outcome variables. We do, however, find several effects that are marginally significant. This reality is largely due to an underpowered research design caused by resource constraints and a relatively small sample size.

This paper aims to encourage Christian aid and development organizations to undertake rigorous evaluations. We present two key lessons on this objective. First, empirically evaluating spiritual change is possible, albeit in an approximate manner. We contend that it is no longer acceptable for Christian aid and development organizations to refrain from rigorous evaluation because “spiritual outcomes are immeasurable”. Although this statement is axiomatically true in a broad sense, this paper presents methods for approximating spiritual changes that are attributable to a transformational development program. Second, this paper highlights the need for a priori power calculations in order to understand the necessary sample size needed to detect effects. DML program administrators felt a need to produce rigorous lessons from a relatively unknown program design. This motivated a desire to engage in a more rigorous impact evaluation project. Although we applaud this general attitude and perspective, it must be accompanied with adequate funding, as more rigorous evaluations are likely more expensive than less rigorous evaluations due to increased demands for data collection and analysis. As will be recognized throughout, this study is limited by a relatively small sample size.

The next section provides additional background information about the conception of the DML program and a brief discussion of the present challenges in evaluating the impacts of business skills training programs in developing countries. Section three explains the research methodology and empirical strategy of this study. Section four presents and discusses the results of the estimated impacts of the program. Section five discusses the limitations of this research and provides suggestions for how to implement future work evaluating transformational development. Finally, section six concludes.

2 Training Business Skills: Beyond Development to Transformation

All over the developing world, a large fraction of the population runs

their own business. In Peru, Pakistan, and Nicaragua, for example, between 47 and 69 percent of households who live on less than \$2 per day operate a nonagricultural business (Banerjee and Duflo, 2007). Most of these business owners, however, do not practice the skills or habits commonly found among small business owners in more developed countries. So while many people own businesses in developing countries, very few of these businesses are actually owned and operated by what would commonly be characterized as a true entrepreneur. Thus, many development organizations offer trainings in basic business skills, such as accounting methods, management techniques, and marketing.

There are two broad schools of thought in the design of business skills training programs. The first (and more popular) approach is to teach specific business skills that business owners can implement in their own business. The assumption here is that the key constraint for business owners is simply a lack of managerial capital, and that everyone who completes a training session actually desires to grow and improve his or her business (Bruhn et al., 2010; Bloom and Van Reenen, 2010). Despite the popularity of this approach, it is not difficult to think of examples suggesting that this assumption may not hold in many contexts throughout the developing world.

A second school of thought is highlighted in Abhijit Banerjee and Esther Duflo's book, *Poor Economics*. The authors characterize many of those who operate business in developing countries as "reluctant entrepreneurs" (Banerjee and Duflo, 2011). In this conceptualization, some people are true entrepreneurs, but many own and operate their business primarily due to the lack of other viable economic alternatives, rather than an impassioned desire to succeed in owning their own enterprise. The assumption here is that attitudes and desires are of equal, or perhaps even greater, importance as skills. It is reasoned that even the most well-trained and knowledgeable business owners may still struggle to grow their businesses if they do not possess the attitude or desire to grow. In this vein, some studies evaluate programs that attempt to change entrepreneurial attitudes or desires (Glaub and Frese, 2011; Glaub et al., 2012).

The "Developing Marketplace Leaders" Program

In 2012, International Christian Ministries (ICM) developed a business skills training program. Within the ICM program design, the local

30 FAITH & ECONOMICS

church played a central role in program administration and dissemination. Rather than simply implement a faith-inspired business skills training curriculum, ICM aimed to utilize local churches and create so-called Marketplace Ministries within various local churches throughout Western Kenya. In 2013, ICM launched a pilot project for the DML program at the Africa Theological Seminary (ATS) in Kitale, Kenya, an ICM-affiliated academic institution. The design of the pilot project aimed to gather feedback and understand how to best implement a business skills training program with hope of achieving the integrated goals of transformational development.³

The central hypothesis of the DML program considers local churches as uniquely equipped to inform business skills and also to transform desires and attitudes. After the pastor of a local church completes an introductory course at ATS, he or she gives several sermons on the subject of the theology of work and Biblical stories of God working through businesspeople for His divine purpose. The local church then facilitates a business skills training class for the members of the church who run their own business. The training meets once a week for three hours over the span of four months, for a total of 36 hours. In total, there are twelve sessions, including such topics as theological foundations for work, book-keeping and accounting, marketing, and management. After completing the training course, participants form mentor groups within the church congregation for the purposes of providing encouragement and feedback while applying the lessons from the course.

The DML program aims to achieve the three impacts of transformational development: social, material, and spiritual change. This program is thus squarely rooted in the second school of thought on the design of business skills training programs. The DML program seeks not only to teach participants specific business skills, but also to transform the desires and attitudes of participants. The hypothesized theory of change of the DML program is similar to Max Weber's theory of the Protestant work ethic (Weber, 1930).⁴ In this sense, participants in the DML program view their work in the marketplace, the so-called secular world, as being related to worship to God. Coupled with increased business skills and knowledge, this theory suggests that a transformed attitude toward faith and work leads to social changes within participant's households, material changes via impacts within their own business, and spiritual changes in attitude and real behavior, such as support to their local church.

The final goal of the DML program is for the local church to establish a “Marketplace Ministry” within their local church organization. In the same way that churches have specific ministries for youth, men, women, and the hospitalized, the DML program seeks to establish sustainable ministries within local churches to disciple and encourage Christ-like practices and behavior in the marketplace. Although this final aspect of the DML program is important, it is untested within this pilot project evaluation because of the small sample size.⁵

Existing Evidence on Training in Business Skills

Although there is a notable literature on the impacts of business skills training programs, it is still premature to draw any general conclusions. Several factors limit the ability to generalize from the literature to date. First, the topics covered by these programs and the duration of lessons differ quite considerably across contexts. The differences in training curriculum and length of the particular training likely carry important implications regarding the measured impacts and dose effects of business skills training programs. Second, trainings also differ as to who participates. Some aim at training specific genders, some at existing business owners, and some at aspiring business owners. The specific characteristics of participants in the training likely determine the magnitude of measured outcomes. These two key challenges effectively restrict any sort of rigorous meta-analysis of business skills training in general, and leave many questions unanswered.

The impacts of business training programs remain an important empirical issue in development economics, as programs that aim to inform business skills and knowledge are relatively popular and widespread around the developing world. One of the most popular business skills training programs is the Start and Improve Your Business (SIYB) program, sponsored by the International Labor Organization (ILO), which has trained nearly 4.5 million people in over 100 countries and in over 40 languages (ILO, 2012). Despite this popularity, however, many rigorous evaluations of business skills training programs fail to find economically worthwhile and statistically significant impacts (McKenzie and Woodruff, 2014). These programs also tend to be relatively expensive. From 2002 to 2012, the World Bank invested \$9 billion dollars across 93 skills training programs (Sanchez et al., 2015). That is an average of roughly \$100 million dollars per program!

32 FAITH & ECONOMICS

In a focused review of the literature, McKenzie and Woodruff (2014) review several of the existing randomized experiments evaluating the impact of business skills training programs and summarize some of the major challenges. No impact evaluation to date reports positive impacts across all measured outcome variables. In terms of increased profits and revenues, only a couple of studies (Calderon et al., 2012; de Mel et al., 2012) found statistically significant increases in both categories. It should be noted, however, that de Mel et al. (2012) only found significant increases for potential entrepreneurs and not for existing business owners. Reporting the results of take-up of specific business practices covered in trainings presents a challenge, as there is large variation in the number of practices taught and the method of evaluating take-up. In general, studies find the impact of take-up on business practices to be positive, but statistically insignificant and very small. For example, Bruhn and Zia (2013) find a statistically significant positive impact on business practices but only for, on average, one out of three practices. The impact on business survival is again very minimal and generally statistically insignificant (Gine and Mansuri, 2014; Mano et al., 2012). Additionally, they find meager impacts on business startup rates. The largest impact estimate only predicts about half of the business training participants actually start a business post-training (Klinger and Schundeln, 2011). Most discouraging is the evidence that finds negative impacts on various measured outcomes due to business skills training (Valdivia, 2012).

Although there is little evidence of business training programs providing a reliable and cost-effective pathway out of poverty, it remains widely used. Nevertheless, it remains at least plausible that, packaged with the correct mix of other services and with the correct program design, business skills training programs may be a worthwhile aspect of effective poverty alleviation and development policies.

3 Research Methodology

This study uses a difference-in-differences (DID) empirical strategy to measure the impacts of a business skills training program, implemented through local churches, in three cities in Western Kenya. In formal expected outcomes notation, this calculation can be written as follows:

$$\text{DID} = (\bar{Y}_{T,1} - \bar{Y}_{T,0}) - (\bar{Y}_{C,1} - \bar{Y}_{C,0}),$$

where \bar{Y} represents an average of some outcome of interest, the subscripts T and C represent treatment and comparison group, respectively, the subscripts 1 and 0 represent the time period (post-treatment and pre-treatment, respectively). The DID framework calculates the average treatment effect on the treated (ATT) by subtracting the pre-treatment difference in outcomes from the post-treatment difference in outcomes. This research design improves upon simple before / after and participant / non-participant comparisons by subtracting away any time invariant factors that may have influenced the measured outcomes. Still of concern, however, are time variant factors that could change over time, perhaps differently between treatment and comparison groups. Indeed, the key weakness of a DID methodology is that one must assume that time variant factors do not influence the outcomes under consideration. The DID framework is best suited for the analysis in this study because we are unable to randomly assign participation in the DML program, but we were able to collect data on participants and non-participants both before and after the program. Random assignment was impossible because specific churches were chosen to participate in the DML pilot project precisely due to their perceived capacity to handle the administrative and logistical demands of facilitating the DML program.

Construction of the Counterfactual

Measuring impacts of a program in any setting requires the construction of a valid counterfactual. In the DID framework, the validity of the counterfactual rests on the critical “parallel trends” assumption. This is the assumption that, absent of the program, the treatment and comparison groups would follow parallel trends in outcomes over time. In other words, even if observable outcomes of the treatment and comparison groups in the sample are at different levels at baseline, the changes over time in their outcomes will progress at the same rate. If this assumption holds, the DID framework will be able to estimate unbiased program impacts, since there is no other factor that could cause a divergence in the change in outcome variables between the treatment and comparison groups other than the program itself.

Another concern relating to the validity of the counterfactual is selection bias. It is possible that those who participate in the program willingly signed up for the program, making them systematically different than those who do not sign up for the program. For example,

34 FAITH & ECONOMICS

it could be that business owners who have more motivation to grow their business sign up for a training program, while those who are less enthusiastic about their business do not sign up for the program. In this situation, any measured impact will be confounded by our inability to disentangle whether the business skills training program or the relative level of motivation (among other things) caused the observed differences. We can mitigate this concern by a critical detail of the research design. By rolling out this program through churches, we asked individuals included in both the treatment (participants) and comparison (non-participants) groups to sign up for the program after a brief announcement made during a church service. The treatment group churches mentioned the DML program would start in three to four weeks, while comparison group churches mentioned the DML program would start in about a year. Thus, individuals included in both the treatment group and comparison groups both willingly signed up to participate in the DML program.

A final concern about the validity of the counterfactual in this evaluation is the possibility of spillovers between treatment and comparison groups. In the presence of spillover effects, such that there is some sort of meaningful interaction between these two groups, the measured impacts may underestimate the true impact of the DML program. Spillovers could occur through word-of-mouth contact between friends in different church congregations. This sort of spillover is difficult to detect; to the extent that spillovers of this sort exist, we have estimated the lower bounds of the true impact of the DML program.

The DML evaluation treatment group included the congregations from the Friends (Quakers) Church in Kitale, the Deliverance Church in Kakamega, and the Anglican Church of Kenya (ACK) in Eldoret. These selections allow for several improvements in the external validity of the evaluation. First, the three locations are geographically diverse. Kitale, the smallest location in terms of population, is a bustling agricultural town close to Kenya's western border with Uganda. Kakamega is a mid-size city that has grown in recent years, due in large part to the operations of several universities in and around the city. Eldoret, the third or fourth largest city in Kenya (depending on how it is measured), is large enough to warrant its own international airport. All three cities are in Western Kenya but each is a member of its own county in the recently-minted Kenyan constitution. Second, each location engaged a different church

denomination. The Friends (Quakers), the Deliverance Church, and the ACK are all mainstream churches in Kenya. This is important from both a research and organizational perspective. From the research perspective, it was important that the impacts of the pilot project be due to the program rather than some unique denominational characteristic. From the organizational perspective, it was strategic to work with several large and nationwide church denominations in Kenya, to aid potential scaling up of the program in the future.

Within the DID framework, the comparison group must be comprised of individuals who are similar to (and who follow trends over time similar to) individuals in the treatment group. This involves a challenge, as church denominations are often closely tied to a host of different factors, including ethnicity, occupation, socioeconomic status, education, language, theology, beliefs, and tradition. Each church congregation serves a different set of individuals and families based on a seemingly infinite number of variables; no two churches are exactly the same.

It is here that working with mainstream churches paid its largest dividend. Each location hosting a treatment church also housed several congregations of the other two church denominations. This allowed us to select comparison church congregations that matched in as many observational characteristics as possible but varied in denomination. Therefore, in Kitale, where the Friends (Quakers) Church participated in the program, the Kitale Deliverance Church acted as a comparison church. In Kakamega, where the Deliverance Church participated in the program, the ACK church acted as a comparison church. In Eldoret, where the ACK church participated in the program, the Friends (Quakers) Church acted as a comparison church. This rotational matching method worked well as we matched churches with the same relative size and congregational make-up within each geographical location.

Although the critical parallel trends assumption cannot be tested and validated directly, we show through balancing tests in Table 1 that the treatment and the comparison groups are relatively similar at baseline. Although this study does not utilize random assignment to participate in the DML program, balancing tests remain useful to measure the comparison and treatment groups at baseline. Within most observed characteristics, we find no statistically significant difference between the control group and treatment group. Additionally, a joint test of significance of all covariates reveals that there is likely no systematic difference among

36 FAITH & ECONOMICS

observable characteristics at baseline. Although the fact that treatment and control groups are similar does not necessarily imply that the parallel trend assumption holds, it does make it more plausible.

On a descriptive level, the observed population includes individuals who live on between \$2 per day and \$20 per day. This is a relatively wide margin. It is not hard to imagine how vastly different lives are at the two ends of this distribution. Nonetheless, our population is not made up of many “extremely poor” individuals who live on less than \$1.90 per day. This is likely the case because one of the only requirements for participation in the DML program was that the participant already owned a business.

Data

The data for this analysis come from two waves of business- and household-level surveys. The first wave took place in late November 2013 through early February 2014, and the second wave took place exactly a year later. As previously discussed, program administrators implemented the DML pilot project in three locations across Western Kenya: Eldoret, Kakamega, and Kitale. Each location is distinct from the other in important ways, and each location included a treatment and comparison church congregation.

Given this structure, the sample is clustered into six groups, two in each geographical area. This creates issues relating to inter-cluster correlation and, in particular, issues associated with “too few” clusters in the sample. Although this problem could have been corrected by performing *ex ante* power calculations, choices about the structure and size of the pilot project evaluation were largely determined by the limited resources available to the DML program’s administration staff. Similar to most pilot projects, this pilot project began relatively small and in very few locations. Thus, while a larger study with more clusters would have allowed for impact estimates with greater statistical efficiency, such a study was outside of the budget constraint of the DML program at the time of the pilot project. This reality creates challenges when we begin to analyze this data, which will be discussed in detail later in the paper.

Identification Strategy

Estimation of the causal impact of the DML program uses basic ordinary least squares (OLS) regression analysis. The general equation for esti-

imating the average treatment effect through a DID framework is identified by the following econometric model:

$$y_{it} = \alpha + \beta \text{Treat} + \gamma \text{Time} + \delta (\text{Treat} * \text{Time}) + \phi' \mathbf{Z} + \varepsilon$$

In this model, y is an outcome variable of individual i in time t . The coefficient β measures differences of the outcome variable between participants and non-participants of the DML program. The coefficient γ measures difference in outcomes compared before and after the DML program. The coefficient on the interaction between treatment and time, δ , is the coefficient of interest, representing the ATT and the estimated impact of the DML program on the outcome variable y . The vector ϕ measures the influence of \mathbf{Z} control variables, and ε is an error term. Control variables included age, gender, educational attainment, number of children, number of direct dependents, number of months of experience in business, whether the respondent had received a business loan, and location level dummy variables.⁶

In our empirical model, observations are grouped into clusters based on their geographic region and their church congregation. Although error terms of observations are assumed to be uncorrelated across clusters, it is often the case that error terms are correlated within clusters. Thus, it is likely that our model has heteroskedastic error terms. Failure to account for inter-cluster correlation can lead to misleadingly small standard errors, overly narrow confidence intervals, t-statistics that are too high, and p-values that are too low (Cameron and Miller, 2015). This scenario increases the possibility of falsely rejecting the null hypothesis. “Cluster-robust” standard errors (White, 1984; Rogers, 1993; Bertrand, Dufo, and Mullainathan, 2004) address this issue, but critically assume that the number of clusters is large, approaching infinity. Thus, having “too few” clusters is a complicating factor for this study.⁷

We have one cluster per church congregation in the observed sample—only six clusters. In situations with small sample sizes and even smaller numbers of clusters, an increasingly common suggestion is the “wild” bootstrap approach of Wu (1986) and Liu (1988) (Cameron, Gelbach, and Miller, 2008; Cameron and Miller, 2015). We report estimates using his procedure, with 400 resamples, in Tables 2 and 3. The wild bootstrap technique, however, works best with an intermediate number of clusters. With less than roughly 11 clusters it results in p-values that are not point identified, and can still underestimate standard errors (Webb,

2014; Cameron and Miller, 2015). Webb's (2014) procedure uses a six-point distribution for weighting that greatly improves upon the discreteness of p-values calculated by the wild bootstrap approach with extremely small numbers of clusters. Table 4 reports the wild bootstrapped standard errors using Webb's weighting procedure with 1000 resamples, along with 95% confidence intervals for each of the impact estimate coefficients examined in this paper.

Finally, we run all of the analysis in this paper using OLS regression models. Although OLS is the workhorse model in econometrics, this choice deserves several brief comments. First, the regressions reported in Table 2 all have continuous dependent variables and thus OLS is a well-suited model for this estimation. Although the error terms of the explanatory variables are heteroskedastic, in a linear regression model OLS still produces unbiased and consistent coefficient estimates. (We address incorrect statistical inference, due to heteroskedastic errors caused by inter-cluster correlation within multiple clusters, by using the wild bootstrapping procedures discussed above.) Second, the regressions reported in Table 3 might normally be best estimated using a non-linear econometric model, because of the nature of the dependent variables: Regressions 1 and 2 use an ordered, discrete dependent variable, and regressions 3 and 4 use a dichotomous dependent variable. However, in these regressions inter-cluster correlation yields heteroskedastic errors, so using a non-linear model would produce inefficient, biased, and inconsistent estimates (Greene, 2012). Thus, we estimate the DID average treatment effect with the standard linear OLS regression model (Ai & Norton, 2003). Third, the dependent variable used for regressions 1 and 2 (in Table 3) is a one-through-four indexed score relating to an individual's integration of faith in business. OLS becomes more problematic at the extreme values of this index due to non-linearity. Figure 1 shows that, at baseline, the vast majority of the observed population had an index score of either 1 or 2. This, coupled with a positive relationship between DML participation and the index score, suggests this conceptual issue is unlikely to meaningfully influence our estimation.

4 Discussion of Results

In this study, we measure impacts of the DML program across three broad areas: Social, material, and spiritual changes. These three spheres

of life are clearly interconnected, and the boundaries between them are admittedly ambiguous. In this paper, we define the social sphere as outcomes realized at the household level rather than the business level. Thus, social outcomes include household income and education attainment of children. We define the material sphere as outcomes relating to an individual's business activities. In this study, material or enterprise outcomes specifically measure enterprise profits, sales, asset growth, and employment growth. Finally, the spiritual changes are framed by DML administrators' expectations of the program's potential effects on attitudes and behaviors relating to faith integration in business activities. We measure spiritual outcomes by using an index of faith integration in business, and by comment-coding several open-ended questions regarding commitment to faith while performing business practices. We also tracked self-reported tithes and offerings given to the local church.

Social Change

The first characteristic of interest is household income.⁸ Figure 2 shows the mean household incomes at baseline and endline within the treatment group and the comparison group. At baseline, the average monthly household income for those in the comparison group and the treatment group is relatively similar: 40,000 Kshs (\$476) for the comparison group and 35,000 Kshs (\$416) for the treatment group. At endline, the average monthly household income is no longer similar between the comparison and treatment groups. Those in the treatment group had an average monthly household income of roughly 65,000 Kshs (\$773) while those in the comparison group had an average monthly household income of roughly 44,000 Kshs (\$523).

The DID impact estimates, reported in Table 2, show that the average participant in the DML program reported monthly household incomes that were roughly 21% higher (27% without control variables) than those in the comparison group. This effect represents a potentially meaningful change in economic livelihood. Although this relationship between household income and DML program participation is both positive and large, it fails to pass the typical 5% statistical significance test. The effect is significant at the 15% level, which could be considered marginally significant. In Table 4, however, when using Webb's (2014) weights while performing the wild bootstrap, the standard error doubles in size and is actually larger than the impact estimate itself. The average

impact of participating in the DML program on household income may be relatively large and economically meaningful, but the present analysis is unable to confirm its statistical significance.

Program administrators were also interested in learning about effects on the education of children. At baseline, however, almost every respondent reported that all children of eligible age were already attending school. This level of school attendance remained constant in the endline survey as well. It is unclear if these self-reported educational attendance data are in fact true. Given the data, it is impossible to analyze the effect of DML participation on the education of children since there was no reported change.

Material Change

For the analysis of material change, the questionnaire asked respondents about the number of people their business employs, enterprise assets, monthly enterprise profits, and monthly enterprise sales. Figure 3 shows the 95% confidence interval for percentage changes of each of the enterprise outcomes relative to the comparison group. Importantly, each interval includes zero. Table 2 reports the DID impact estimates of outcome variables relating to enterprise growth. In particular, we find that enterprise sales and profits decreased by roughly 38% and 40%, respectively. These effects are significant at the 5% and 10% levels without control variables. With controls, however, the standard errors increase and the significance levels drop considerably. Additionally, Table 4 shows when using Webb's weights the standard errors increase even more. Thus, although these point estimates are again quite large, this study is unable to rule out the possibility that the effects are statistically zero. A similar story holds for the effect of the DML program on enterprise assets. We calculate that enterprise assets increased by 18%, but this relatively large and economically meaningful effect is not statistically significant at the 5% level. Additionally, when we include control variables and bootstrap with Webb's weights, the standard error is actually larger than the estimated coefficient. We are unable to determine if the impact of the DML program is statistically positive or negative.

A particular concern exists in regard to the measurement of these enterprise outcome variables. Those who participated in the DML program may calculate these variables differently after the training than they did before. Thus, a measured decrease in sales and profits or an increase in

assets may be due to a change in the respondents understanding of their own reality through increased skill in accounting practices. Several existing studies support this concern. One finds that business skills training reduces the number of errors in respondent's reporting of profit data (Drexler et al., 2014), and another finds that business skills training increases the difference between self-reported profits and profits calculated by subtracting reported expenses from reported profits (Berge et al., 2011). Based on the guidance of de Mel et al. (2009), however, the present study simply asked about sales and profits directly, without specifically breaking down revenues and costs. Future studies should note that asking specifically about revenues and costs may reduce the possibility that participating in a business skills training program impacts the way participants respond to these questions.

In summary, we find very little statistical evidence that participants in the DML program are much better off, in regard to enterprise outcomes, compared to those in the comparison group. Average impacts are quite large but in varying directions. We find decreases in sales and profits on average, but increases in assets. Several changes could be occurring among the enterprises who participated in the DML program. Although data availability makes it impossible to tell a definitive story, several possibilities exist. First, as mentioned above, it could be that some DML participants more accurately report business-level factors after the training than before; the estimated effects could be picking up the change in participants' perception of their business situation rather than a real-life change in sales, profits, or assets. Second, if we take the data literally, it could be that DML participants take time away from selling their products, perhaps in an attempt to scale up their enterprise through the accumulation of assets. Third, it is also possible that DML participants shift to other business activities after the training. In this case it would be very hard to measure any impacts on the business itself, since the attention of the business owner shifted onto an alternative income-generating activity. At the end of the day, however, these impact estimates remain tenuous as we fail to find statistical significance across any of these effects. To be clear, however, this study's failure to identify statistically significant effects is not evidence that any business was systematically made worse due to participating in the DML program.

Spiritual Change

Increased spirituality is an important aspect of the theory of change for the DML program. While increased spirituality may have important value in and of itself, it is hypothesized that a more holistic conception of worship—occurring not only on Sunday, but also Monday through Saturday in the marketplace—will lead to an ethic for entrepreneurship and a transformed attitude for business activity. Additionally, of particular concern to DML program administrators is the possibility that a business skills training program could make participants profoundly less spiritual. They feared, for example, that training skills such as bookkeeping and business management might make individuals more individualistic and less generous. We therefore spent considerable effort devising a method for approximately measuring the immeasurable.

The first variable is a stylized index of the Christian in the marketplace (Silvoso, 2009). This index consists of four levels of a Christian's spiritual involvement in the marketplace: Level 1 is "A Christian in the marketplace." Level 2 is "A Christian who is applying Biblical principles to business." Level 3 is "A Christian who follows the leading of the Holy Spirit in the marketplace." Level 4 is "A Christian committed to the spiritual transformation of the marketplace." A score on this index was assigned based on the respondent's answer to the open-ended question, "Could you please describe how you integrate your faith in your business?" At baseline, both treatment and control groups gave responses that suggested an average index score between level 1 and level 2—somewhere between "A Christian who happens to be in the marketplace" and "A Christian who is striving to apply Biblical principles to business." At endline, those in the control group largely remained at this level, while those in the treatment group provided statements that suggested an average index score between level 2 and level 3—somewhere between "A Christian who is applying Biblical principles to business" and "A Christian who follows the leading of the Holy Spirit" in the marketplace. Figure 4 shows the average responses broken down between treatment and control groups at baseline and endline.

Regressions 1 and 2 in Table 3 report the DID impact estimates of participation in the DML program, and show a positive and statistically significant result at the 5% level. This result seems to indicate that participating in the DML program causes an average increase in the Silvoso index of slightly less than one level, or 0.77.⁹ This result is statistically

significant at the 5% level, both with and without control variables. When the wild bootstrap method is performed using Webb's weights, as shown in Table 4, the statistical significance of this result falls to being significant at the 10% level with the inclusion of control variables. This result, however, persists as one of the strongest—in the sense of statistical significance—in this entire study. It may be reasonable to suggest that the DML program is influencing the way in which participants view integration of faith in their daily work in the marketplace.

Next, the questionnaire asked respondents what they viewed as more spiritual: their church, their business, or if they were equal. This question was converted into a binary response (e.g., is the church more spiritual or is your business and your church equal in spirituality), as nobody reported that their business was *more* spiritual than their church. We coded responses in the form of a binary dummy variable where 1 indicates a belief that church and business are equal in spirituality and 0 indicates a belief that the church is more spiritual than business. Figure 5 shows that participants in the DML program were slightly more likely than those in the comparison group to see their work in the private sector as of equal spirituality to their church.

Regressions 3 and 4 in Table 3 report the DID impact estimates, including a positive relationship between DML participation and perceived spirituality of business, significant at the 10% level without control variables. When control variables are included, the effect falls by roughly a third, and the statistical significance drops, though the effect is still marginally significant. In Table 4, when using Webb's weights while calculating wild bootstrapped standard errors, the significance level of this effect falls even more. Taken together, the results suggest that participants in the DML program may be more likely to view their business and their church as equally spiritual than those in the comparison group.

Finally, the last spiritual outcome of interest is household giving to the church through tithes and offerings. The outcomes between the treatment and comparison groups at baseline and endline are shown in Figure 6. DML program participants reported much higher levels of giving compared to non-participants at endline. DML program participants, however, also reported much higher levels of giving at baseline. When calculating the DID impact estimates, shown in Table 2 and Table 4, we find the impact on household giving to local churches from participating in the DML to be slightly negative and far from statistically significant. Giv-

44 FAITH & ECONOMICS

ing to the church in proportion to household income actually decreased slightly among DML participants on average. At baseline, the comparison group reported giving an average of 3% of their household income to the church, while the treatment group reported giving an average of 8%. At endline, the comparison group reported giving an average of 4% of their household income to the church, while the treatment group reported giving 6% of their household income. These changes, however, are both realistically and statistically insignificant.

Taken together, it remains plausible that the DML program inspired spiritual change among its participants, but our research design lacks the statistical power to detect these effects at typical levels of statistical significance. Impact estimates, however, suggest that increased measures of spirituality due to DML participation are relatively meaningful on average and are marginally significant in a statistical sense. Either via an open-ended response to a question regarding faith integration in daily business activities or a closed response to a question about the spirituality of business relative to their church, participants report increased faith integration in their business lives on average. Finally, even though we fail to find an increase in household giving on average, we also fail to find a significant decrease in household giving to the church. Thus, we do not find evidence that suggests that participants became less spiritual after the DML program.

Although this study is unable to find evidence of a statistically significant impact of the DML program—positively or negatively—across a variety of outcomes, this is one case in which the absence of evidence does not imply evidence of absence (of meaningful impact). It could very well be true that the DML program did cause meaningful changes in the lives of those who participated; however, this study is unable to establish statistical significance for these changes.

5 Discussion of Limitations and Future Priorities

Several limitations of this research require further exposition, as they pertain to and inform future work evaluating transformational development initiatives. This paper seeks to provide an example of one possible approach, of which there are many, to evaluating transformational development in an empirically rigorous manner. There are a number of details that limit this evaluation and policy-relevant lessons. In this section, we dis-

cuss these limitations in the hope that such issues do not limit the scope of future empirical evaluations of transformational development initiatives.

First, as has been evident throughout the discussion of results, low statistical power driven by an insufficient sample size limits the explanatory capabilities of this study. The statistical power of a research design represents the probability that the null hypothesis will be rejected given that the null hypothesis is indeed false. The issue of low statistical power is not a new problem for empirical studies, and as McKenzie and Woodruff (2014) point out, it is a common problem limiting the measurement of impacts of business skills training programs in developing countries. Several key details of most business skills training programs inherently reduce the statistical power of an empirical evaluation. Power is influenced by sample size and whether treatment is administered at the individual or group level. When development programs are administered at the group level, such as business skills training programs being administered to a class, the statistical power for a given sample size is diminished. Additionally, as heterogeneity within a sample population increases, statistical power decreases. The sample studied in this paper was made up of individuals who, on average, live on between \$2 and \$20 per day—a heterogeneous group. Finally, low take-up rates and survey attrition weaken the precision of the estimates of the treatment effects and therefore reduce statistical power. Many business skills training programs are time-intensive, and thus a relatively large share of an initial class end up dropping out of the program. All of these factors limit the ability of the present study, and other empirical evaluations of business skills training programs, to detect statistically significant effects.

In the present study, the minimum detectable effects of the DML program are often very large, especially considering that they were measured within one year of the training program itself. Although the measured average effects, shown in Tables 2, 3, and 4, are often quite large and economically meaningful, the minimum detectable effects are often even larger. Statistical significance is elusive, due to relatively large standard error measurements calculated across most outcome variables, primarily caused by the small number of clusters and a relatively small sample size. It would have been helpful to have performed an *ex ante* power calculation in order to understand the necessary sample size and number of clusters needed to measure sufficiently precise impact estimates (see Duflo, Glennerster, and Kremer, 2008).¹⁰

A second limitation of this research is the potential for “desirability bias,” the tendency for respondents to provide answers that they think interviewers would like to hear. The concern here is that respondents may claim to behave in ways that align with what they learned during the program, but may not actually behave in these ways. The potential for desirability bias is particularly strong among the spiritual outcomes. During the course of the DML program, participants were told numerous times about the importance of faith integration into their daily business activities. The use of a DID empirical strategy is not a sufficient solution to this problem, as this strategy only subtracts away time-invariant differences between the treatment and the comparison group. The problem lies in the fact that participation in the DML program may change respondents’ likelihood of reporting spiritual integration in business activities.

The problem of desirability bias cannot be completely avoided in any empirical work involving self-reported survey data of subjective concepts like “spirituality.” One way to minimize this issue, utilized in the present study, is asking an open-ended question that invites the respondent to provide a concrete example. The questionnaire asked all respondents: “Could you please describe how you integrate your faith in your business?” The answer to this question was then scored according to the index and rubric developed in Silvos (2009). Prompting respondents to provide a concrete example in answering this open-ended question potentially diminishes the possibility of desirability bias influencing the analysis of this study. This assumes that those who are not actually acting on the lessons of the DML program will have a more difficult time reporting a concrete example of faith integration in their business. Although this procedure carries the potential for diminishing desirability bias, the possibility of this limiting factor cannot be sufficiently ruled out. Readers must keep this limitation in mind when reflecting on this research and when designing future evaluation studies using self-reported survey data.

Future work measuring spiritual change driven by transformational development programs could implement a qualitative method for “measuring the measurement error.” In Blattman et al. (2016), the authors describe a simple, yet intensive, method for validating self-reported survey data in the field through the use of open-ended qualitative questioning and participant observation. This method sends so-called validators into the field to embed themselves in the daily lives of a random subset of a sample population. Validators visit survey respondents four times over

the course of a ten-day period. Through this process, relationships and trust are built to an extent that validators are presumably able to elicit observations that are closer to the truth than observations from a rather coarse household survey.

Finally, this paper provides one possible approach to evaluating transformational development in an empirically rigorous manner, but a DID research design may not be the best evaluation technique for every transformational development program. Over the past two decades, the field of development economics has passed through the so-called credibility revolution (Angrist and Pischke, 2010). Some development organizations now feel increased pressure to show that their programs make a positive impact in the world.¹¹ This pressure, in our view, is positive as it strengthens our ability to effectively respond to important issues of human suffering in our world and provides insights into the responsible stewardship of resources. It is our feeling that Christian aid and development organizations should follow this trend and focus on whether their programs generate positive material, social, and spiritual impacts. It is not our aim, however, to suggest that every transformational development program should implement a DID evaluation methodology, as the DID approach may not be best suited for every situation.¹²

The DID approach greatly improves upon common approaches to measuring outcomes before and after some intervention, or differences between beneficiaries and non-beneficiaries. In fact, the DID approach works precisely by combining these two approaches and measuring impact by taking the differences between before and after some intervention in the differences between the beneficiaries and non-beneficiaries. Although this approach is relatively straightforward, other approaches should be considered in other contexts. For example, when it is possible to randomize the assignment to treatment and comparison groups, a randomized control trial (RCT) may be a better evaluation approach. In the case of an RCT, assumptions like “parallel trends” are unnecessary, as random assignment—properly implemented—will lead to an unbiased impact estimate.

We previously mentioned that effective rigorous evaluations might be more expensive than less rigorous evaluations due to increased demands for data collection and analysis. It also may be the case that more rigorous evaluations are more beneficial to policymakers, program administrators, and ultimately the end beneficiaries than less rigorous eval-

uations, because policy-relevant lessons are less likely to be confounded by additional factors. This being the case, Christian aid and development organizations should put forth effort to weigh the relative costs and benefits of rigorously answering policy-relevant questions when considering a specific evaluation approach.¹³

6 Conclusion

This paper seeks to build on the theory of transformational development (Bragg, 1984; Myers, 2011) by providing a detailed example of an empirical evaluation that measured results across material, social, and spiritual dimensions of life. Using a difference-in-differences (DID) empirical strategy to measure the impacts of a church-based business skills training program in Western Kenya, this paper showcases the key steps in designing and implementing a rigorous impact evaluation. Although the methods used in this paper are not the only ways to design a credible impact evaluation, this evaluation may guide faith-based organizations interested in producing evidence about what works best in achieving the three goals of transformational development.

We do not find any statistically significant effects, at the typical 5% significance level, largely due to a small sample size and, relatedly, “too few” clusters. This drives our standard errors upward, which raises the threshold for the minimum detectable effect. Some of the measured effects, however, are quite large and quite possibly economically meaningful. In particular, a year after the DML program began participants reported an increase in household income and enterprise assets of roughly 21% and 18%, respectively. Respondents also reported a decrease in enterprise sales and profits of 38% and 40%, respectively. Thus, the impacts of the DML program may be quite large and economically meaningful, but the present study does not possess sufficient statistical power to ascribe statistical significance to these impacts.

The impacts of the DML program on various spiritual outcomes are the strongest—in the sense of statistical significance—in this entire paper. We find that participation in the DML program plausibly causes a positive increase on a scale of faith-integration in business activities. In particular, we find a positive impact on Silvosó’s (2009) scale, and an increased likelihood that respondent report their business is just as spiritual as their church. Additionally, we also do not find evidence that the

DML program made anyone less generous to their local church, measured through self-reported tithes and offerings.

Two final points serve as key lessons of this study. First, empirically evaluating spiritual change is possible. Many Christian aid and development organizations refrain from rigorous evaluation because a key portion of their programming aims to generate spiritual impacts. These impacts, some organizations contend, are immeasurable. This paper aims to strengthen the rigorous evidence base upon which Christian aid and development organizations operate. To do this, we present methods for approximately assessing spiritual changes that are attributable to a transformational development program. Second, this paper highlights the need for a priori power calculations in order to understand the necessary sample size needed to detect effects. There is increasing interest among both Christian and secular aid and development organizations in performing more rigorous impact evaluations. We applaud this broad shift, but argue it must be accompanied by adequate funding, since more rigorous evaluations are likely more expensive than less rigorous evaluations. Choosing the right evaluation approach is also consequential, as there is no single research design best suited to all contexts. It is our hope that this paper assists aid and development organizations to make more thoughtful decisions about the best size and design of their evaluations in the process of learning toward transformation.

Endnotes

- 1 An exception is Wydick, Glewwe, and Rutledge (2013) evaluating the impacts of Compassion International's child sponsorship program.
- 2 Throughout this paper, I will use the terms "we" and "our," et cetera. Although this is a single-authored paper, the project involved a small team of excellent colleagues who helped collect data and facilitate this project.
- 3 The relationship between "Transformational Development" and the "Prosperity Gospel" requires a brief comment. While both ideas promote material success, the latter suggests that material success can be earned explicitly through spiritual righteousness, while the former builds a theory of the multidimensional nature (i.e., material, social, and spiritual) of poverty and human development. For more on this, see the work of Bryant Myers (2011, 2015).

50 FAITH & ECONOMICS

- 4 We do not aim to test or make any statement relating to the validity of Weber's hypothesis in this paper. It is merely mentioned here to explain a potential mechanism of the DML program. For empirical research on Weber's hypothesis, see Becker and Woessmann (2009) and Cantonie (2014).
- 5 Only six churches were involved in the evaluation of this pilot project.
- 6 See Bruhn and McKenzie (2009) for a justification for adding strata dummy variables.
- 7 In practice, there is no canonical definition for "too few." Cameron and Miller (2015) state that, depending on the situation, "too few" may range from between 20 to 50. In general, more is better.
- 8 It may seem odd to categorize household income as a social change. Indeed income is a material outcome. We make this distinction, however, based on outcomes that are realized inside the business and outcomes that are realized inside the home. As we will discuss briefly later in the paper, household income could come from additional sources other than the business under observation in this study.
- 9 We acknowledge that this interpretation requires an assumption of cardinality of Silvano's index. See Schroder and Yitzhaki (2017) for a discussion about this assumption.
- 10 The Abdul Latif Jameel Poverty Action Lab has an excellent resource for performing ex ante power calculations: "How to Do Power Calculations in Optimal Design Software." It is freely available online: https://www.povertyactionlab.org/sites/default/files/resources/ExerciseC_Participant.powercalc.TA_.pdf
- 11 Here "impact" is understood as the difference, all else equal, between a treatment group and the counterfactual.
- 12 David Evans, of the World Bank, and Bruce Wydick, of the University of San Francisco, have written an excellent blog post on the World Bank's *Development Impact blog*, entitled "Is My NGO Having a Positive Impact?"
- 13 Innovations for Poverty Action (IPA) have developed a useful framework, their "Goldilocks Framework," with numerous resources available to help NGOs identify and implement a "right-fit" evaluation approach. Available online <http://www.poverty-action.org/goldilocks/principles>

References

- Ai, C., & Norton, E.** (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123–129.
- Angrist, J., & Pischke, J. S.** (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30.
- Banerjee, A., & Duflo, E.** (2007). The economic lives of the poor. *Journal of Economic Perspectives*, 21(1), 141–168.
- Banerjee, A., & Duflo, E.** (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.
- Becker, S. O., & Woessmann, L.** (2009). Was Weber wrong? A human capital theory of Protestant economic history. *The Quarterly Journal of Economics*, 124(2), 531–96.
- Berge, L. I., Oppedal, K. B., & Tungodden, B.** (2011). Human and financial capital for microenterprise development: Evidence from a field and lab experiment. *NHH Discussion Paper Sam 1*.
- Bertrand, M., Duflo, E., & Mullainathan, S.** (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Blattman, C., Jaminson, J., Koroknay-Palicz, T., Rodrigues, K., & Sheridan, M.** (2016). Measuring the measurement error: A method to qualitatively validate survey data. *Journal of Development Economics*, 120, 99–112.
- Bloom, N., & van Reenen, J.** (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24(1), 203–24.
- Bragg, W.** (1984). Beyond development to transformation. *International Review of Mission*, 73(290), 153–165.
- Bruhn, M., Karlan, D., & Schoar, A.** (2010). What capital is missing in developing countries? *American Economic Review Papers and Proceedings*, 100(2), 629–33.
- Bruhn, M., & Zia, B.** (2013). Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs. *Journal of Development Effectiveness*, 5(2), 232–66.
- Bruhn, M., & McKenzie, D.** (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4), 200–232.

- Calderon, G., Cunha, J., & de Giorgi, G.** (2012.) Business literacy and development: Evidence from a randomized trial in rural Mexico. Mimeo. Stanford University
- Cameron, A. C., & Miller, D.** (2015.) A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–373.
- Cameron, A. C., Gelbach, J., & Miller, D.** (2008.) Bootstrap-based improvements for inference with clustered errors. With Gelbach, J., & Miller, D. *Review of Economics and Statistics*, 90, 414–427.
- Cantonie, D.** (2014.) The economic effects of the Protestant Reformation: Testing the Weber hypothesis in the German lands. *Journal of the European Economic Association*, 13(4), 561–598.
- de Mel, S., McKenzie, D., & Woodruff, C.** (2009.) Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, 88(1), 19–31.
- de Mel, S., McKenzie, D., & Woodruff, C.** (2012.) Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. Mimeo. World Bank.
- Drexler, A., Fischer, G., & Schoar, A.** (2014.) Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2), 1–31.
- Duflo, Esther, Glennerster, R., & Kremer, M.** (2008.) Using randomization in development economics research: A toolkit. *The Handbook of Development Economics, Vol 4*. Schultz, T. P. & Strauss, J. A. (Eds.). Amsterdam: North Holland. 3894–3895.
- Giné, X., & Mansuri, G.** (2014.) Money or ideas? A field experiment on constraints to entrepreneurship in rural Pakistan. *SSRN Scholarly Paper ID 2461015*. Rochester, NY: Social Science Research Network.
- Glaub, M., & Frese, M.** (2011.) A critical review of the effects of entrepreneurship training in developing countries. *Enterprise development and microfinance*, 22(4), 335–353.
- Glaub, M., Frese, M., Fischer, S., & Hoppe, M.** (2012.) A psychological personal initiative training enhances business success of African business owners. Mimeo. National University of Singapore Business School.
- Greene, W. H.** (2012.) *Econometric Analysis*, Seventh Edition. Boston: Prentice Hall.
- International Labor Organization.** (2012.) The SIYB Program History. Available at: http://ilo.org/empent/areas/start-and-improve-your-business/WCMS_192060/lang--en/index.htm

- Klinger, B., & Schundeln, M.** (2011.) Can entrepreneurial activity be taught? Quasi-experimental evidence from Central America. *World Development*, 39(9), 1592–610.
- Liu, R.** (1988.) Bootstrap procedures under some non-i.i.d. models. *Annals of Statistics*, 16, 1696–1708.
- Mano, Y., Iddrisu, A., Yoshino, Y., & Sonobe, T.** (2012.) How can micro and small enterprises in sub-Saharan Africa become more productive? The impacts of experimental basic managerial training. *World Development*, 40(3), 458–68.
- McKenzie, D., & Woodruff, C.** (2014.) What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Research Observer*, 29(1), 48–82.
- Miller, M., Reichelstein, J., Salas, C., & Zia, B.** (2015.) Can you help someone become financially capable? A meta-analysis of the literature. *World Bank Research Observer*, 30(2), 220–246.
- Myers, B.** (2011.) *Walking with the poor: Principles and Practices of Transformational Development*. Revised Edition. Maryknoll, NY: Orbis Books.
- Myers, B.** (2015.) Progressive Pentecostalism, development, and Christian development NGOs: A challenge and an opportunity. *International Bulletin of Missionary Research*, 39(3), 115–120.
- Rogers, W.H.** (1993.) Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13, 19–23.
- Sanchez Puerta, M. L., Robalino, D. A., Strokova, V., Lord, N., & Perinet, M.** (2015.) *Skills and jobs: lessons learned and options for collaboration*. Washington, D.C.: World Bank Group. Available online: <http://documents.worldbank.org/curated/en/2015/05/24566584/skills-jobs-lessons-learned-options-collaboration>
- Schroder, C., & Yitzhaki, S.** (2017.) Revising the evidence for cardinal treatment of ordinal variables. *European Economic Review*, 92, 337–358.
- Silvoso, E.** (2009.) *Anointed for Business: How to Use Your Influence in the Marketplace to Change the World*. Minneapolis, Minnesota: Chosen Books.
- Valdivia, M.** (2012.) *Training or technical assistance for female entrepreneurship? Evidence from a field experiment in Peru*. GRADE working paper.
- Webb, M. D.** (2014.) Reworking wild bootstrap based inference for clustered errors. *Queens Economics Department Working Paper 1315*.

54 FAITH & ECONOMICS

- Weber, M.** (1992, 1930.) *The Protestant Ethic and the Spirit of Capitalism*. London and New York: Routledge.
- White, H.** (1984.) *Asymptotic Theory for Econometricians*. San Diego: Academic Press.
- Wu, C. F. G.** (1986.) Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261–1295.
- Wydick, B., Glewwe, P., & Rutledge, L.** (2013.) Does international child sponsorship work? A six-country study of impacts on adult life outcomes. *Journal of Political Economy*, 121(2), 393–426.

Table 1. Summary of Baseline Covariates - Balancing Test

	Comparison	Treatment	Difference
Age	39.3	42.2	-2.92**
# of Children	3.15	3.44	-0.29
# of Dependents	3.89	3.86	0.02
Education (Years)	11.2	11.7	-0.49
Sex (1=Male)	0.37	0.36	0.01
Marital Status:			
Single	0.15	0.08	0.06
Married	0.77	0.80	-0.02
Divorced	0.00	0.00	0.00
Widowed	0.06	0.10	-0.04
Experience in Business	10.3	12.8	-2.48*
Business Sector:			
Retail	0.72	0.63	0.08
Service	0.22	0.23	-0.01
Farming	0.04	0.07	-0.03
Manufacturing	0.00	0.01	-0.01
Non-Profit	0.00	0.02	-0.02*
% Taken a Loan	0.39	0.45	-0.06
Num. of Loans	2.36	2.03	0.32
Num. of Employees	1.15	2.50	-1.35**
Monthly Labor Costs	24,300	28,039	-3738
Monthly Sales	59,725	99,743	-40,018*
Monthly Profit	20,336	58,000	-37,664*
Total Assets	1,051,829	863,239	188,589
Annual Investment	437,619	190,967	246,651
Household Income	38,696	35,714	2,981
Household Savings	5,894	6,966	-1,071
Join Test (F-Statistic)			Treatment
All Covariates			0.0249

All monetary figures are in Kshs (\$1 US = 84 Kshs). Statistical significance is represented through p-values: ***p<0.01, **p<0.05, *p<0.1.

Figure 1. Histogram of Silvoso Responses at Baseline

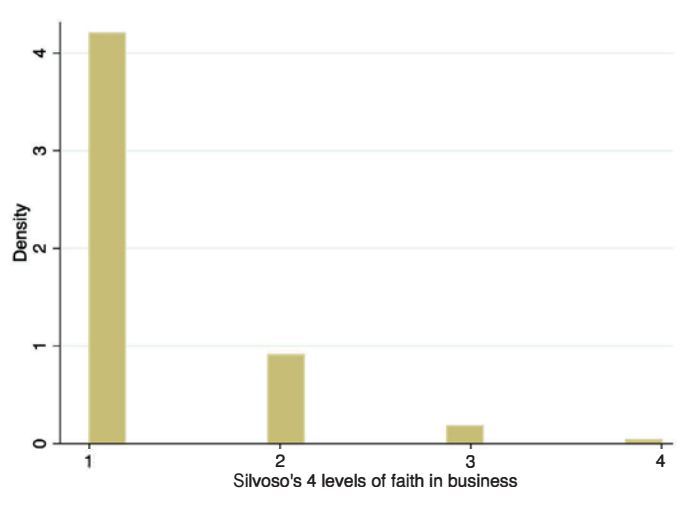


Figure 2. Household Income (Kshs; \$1 US = 84 Kshs)

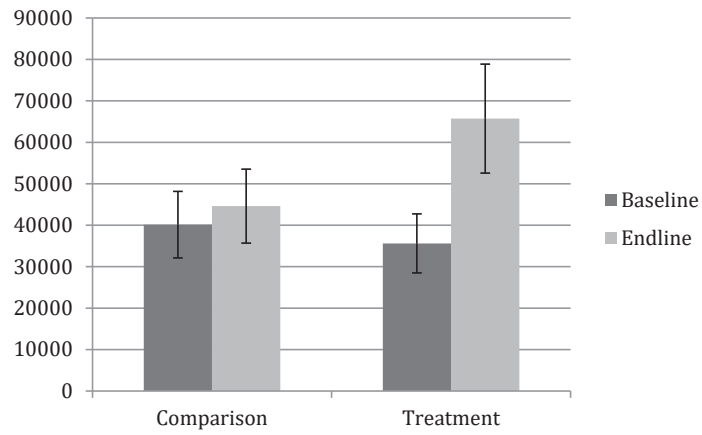


Figure 3. Range of (log-level) ATT Effects of Enterprise Outcomes (95% CI)

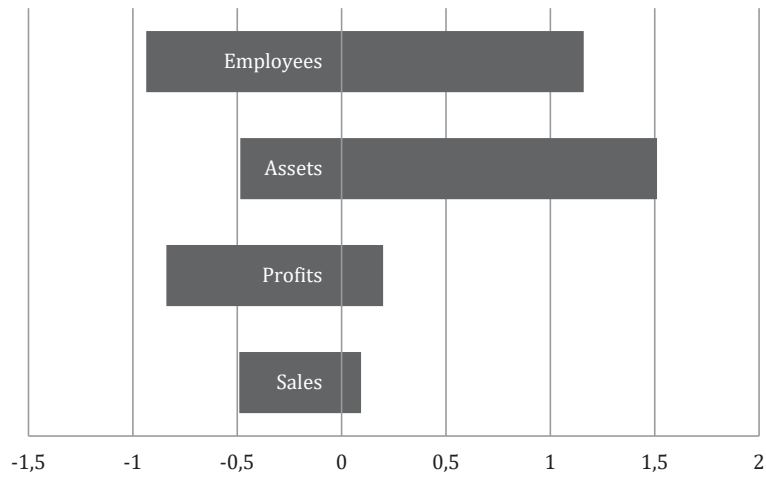


Figure 4. Silviso's Four Levels of the Christian in the Marketplace

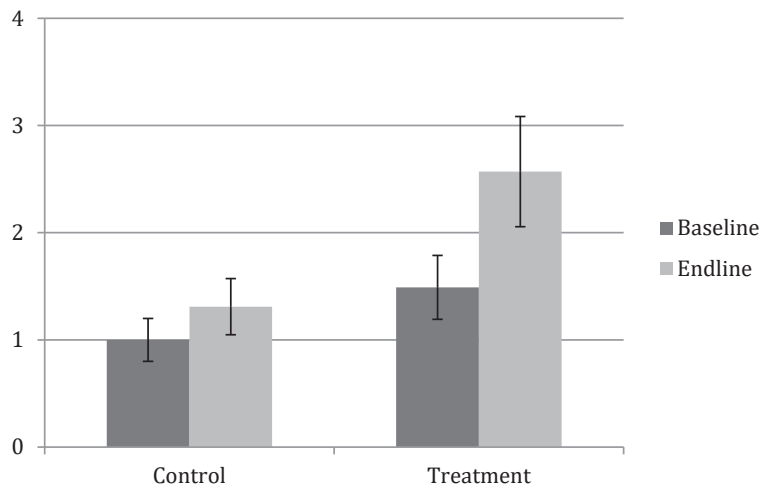


Figure 5. Spirituality of Business Relative to the Church

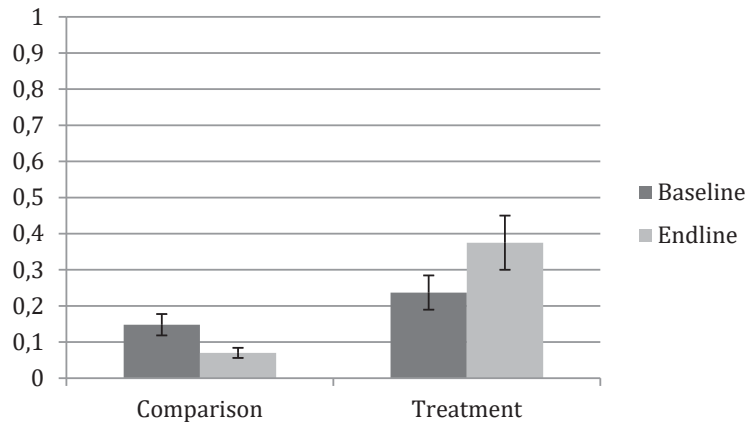


Figure 6. Household Giving through Tithes and Offerings (Kshs; \$1 US = 84 Kshs)

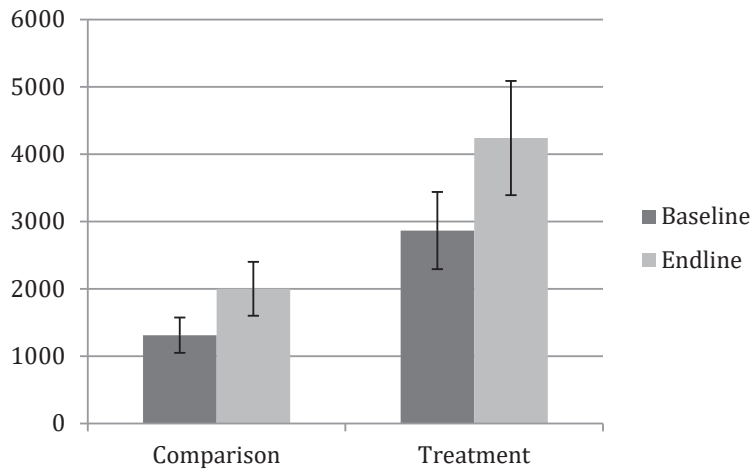


Table 2. Difference-in-Differences Impact Estimates (*facing page*)

Wild bootstrapped standard errors calculated with 400 resamples. Statistical significance is represented through p-values: ***p<0.01, **p<0.05, *p<0.1.

	Log Household income		Log Household Giving		Log Enterprise Sales		Log Enterprise Profits		Number of Employees		Log Enterprises Assets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment Dummy	0.1301 (0.1939)	0.0737 (0.1497)	0.4042 (0.3104)	0.2281 (0.3584)	0.3815 (0.2422)	0.4148 (0.3052)	0.5555** (0.2509)	0.4994 (0.3069)	0.1504* (0.1775)	0.0927 (0.2884)	0.6601 (0.2443)	0.5784 (0.1968)
Time Dummy	0.2947*** (0.0743)	0.4696*** (0.0871)	0.6245*** (0.0921)	0.6175 (0.1447)	0.6914*** (0.0608)	0.7146*** (0.1647)	0.5637*** (0.1170)	0.4694* (0.2162)	0.0992* (0.1358)	0.2122 (0.1236)	0.1797 (0.1961)	0.2766 (0.2267)
DID Impact Estimate	0.2731 (0.1895)	0.2195 (0.1647)	-0.0105 (0.1315)	-0.0829 (0.2009)	-0.2405** (0.1126)	-0.3761 (0.2052)	-0.3651* (0.1776)	-0.3991 (0.2482)	-0.0851 (0.2656)	-0.0876 (0.1870)	0.4718 (0.2701)	0.1817 (0.2016)
Age	-0.0232*** (0.0048)	-0.0186 (0.0057)	-0.0186 (0.0057)	-0.0186 (0.0057)	-0.0311 (0.0098)	-0.0311 (0.0098)	-0.0247** (0.0061)	-0.0247** (0.0061)	0.0135 (0.0070)	0.0135 (0.0070)	-0.0083 (0.0188)	-0.0083 (0.0188)
Education	0.0895*** (0.0129)	0.0895*** (0.0129)	0.1115*** (0.0132)	0.1115*** (0.0132)	0.0486 (0.0267)	0.0486 (0.0267)	0.0714* (0.0264)	0.0714* (0.0264)	0.0323 (0.0136)	0.0323 (0.0136)	0.1713** (0.0277)	0.1713** (0.0277)
Number of Children	0.0648** (0.0197)	0.0648** (0.0197)	0.0453 (0.0556)	0.0453 (0.0556)	0.0013 (0.0423)	0.0013 (0.0423)	0.0233 (0.0279)	0.0233 (0.0279)	0.0007 (0.0292)	0.0007 (0.0292)	0.0564 (0.0537)	0.0564 (0.0537)
Number of Dependents	0.0154 (0.0127)	0.0154 (0.0127)	0.0113 (0.0117)	0.0113 (0.0117)	0.0031 (0.0051)	0.0031 (0.0051)	-0.0225 (0.0123)	-0.0225 (0.0123)	0.0217 (0.0181)	0.0217 (0.0181)	-0.0441 (0.0438)	-0.0441 (0.0438)
Business Experience	0.0238*** (0.0027)	0.0238*** (0.0027)	0.0183*** (0.0030)	0.0183*** (0.0030)	0.0472*** (0.0050)	0.0472*** (0.0050)	0.0537*** (0.0058)	0.0537*** (0.0058)	0.0092 (0.0071)	0.0092 (0.0071)	0.0451*** (0.0088)	0.0451*** (0.0088)
Business Loan	0.2166 (0.1276)	0.2166 (0.1276)	0.2860 (0.1236)	0.2860 (0.1236)	0.6426*** (0.1181)	0.6426*** (0.1181)	0.4486*** (0.0664)	0.4486*** (0.0664)	-0.1548* (0.0445)	-0.1548* (0.0445)	0.7108*** (0.1248)	0.7108*** (0.1248)
Constant	9.9184*** (0.1250)	9.0862*** (0.2544)	6.5634*** (0.1707)	5.6545*** (0.4478)	10.1186*** (0.2038)	9.9303*** (0.4874)	9.0656*** (0.1862)	8.4429*** (0.4706)	0.7777*** (0.1353)	-0.2883 (0.2311)	11.6926*** (0.2154)	9.1202*** (0.7810)
Location Dummy?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N =	420	355	423	371	415	363	392	344	210	180	414	359
R-Squared	0.0635	0.2279	0.0920	0.1897	0.0614	0.2103	0.0549	0.2329	0.0052	0.0892	0.0757	0.2413

Table 3. Impacts on Spirituality

	Ordinary Least Squares <i>Silvoso Index</i>		Linear Probability <i>Spirituality of Business</i>	
	(1)	(2)	(3)	(4)
Treatment Dummy	0.4884 (0.1834)	0.4723 (0.2389)	0.0888 (0.0509)	0.1250 (0.0239)
Time Dummy	0.3030 (0.1938)	0.3211 (0.2003)	-0.0728*** (0.0543)	-0.0498*** (0.0364)
DID Impact Estimate	0.7790** (0.3422)	0.7796** (0.3400)	0.2105* (0.0995)	0.1406 (0.0758)
Age		-0.0088*** (0.0026)		0.0005 (0.0012)
Education		0.0210* (0.0097)		0.0183** (0.0044)
Number of Children		0.0271 (0.0117)		0.0023 (0.0086)
Number of Dependents		0.0048 (0.0097)		-0.0030 (0.0061)
Business Experience		0.0068 (0.0063)		0.0048 (0.0031)
Business Loan		-0.0276 (0.0251)		0.0057 (0.0170)
Constant	1.0078*** (0.0937)	0.9425*** (0.1790)	0.1484*** (0.0244)	-0.1552 (0.0894)
Location Dummy?	No	Yes	No	Yes
N =	485	403	469	388
R-Squared	0.4853	0.5060	0.073	0.1235

Wild bootstrapped standard errors calculated with 400 resamples. Statistical significance is represented through p-values: ***p<0.01, **p<0.05, *p<0.1.

Table 4. Wild Bootstrapped Standard Errors and 95% Confidence Intervals using Webb’s Weighting

	Log Household Income		Log Household Giving		Log Enterprise Sales		Log Enterprise Profits	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Impact Estimate	0.2731	0.2195	-0.0105	-0.0829	-0.2405*	-0.3761	-0.3651	-0.3991
Std. Error	(0.3534)	(0.2826)	(0.2378)	(0.5775)	(0.1487)	(0.2883)	(0.2648)	(0.4707)
95% C.I.	[-0.3327, 1.053]	[-0.3738, 0.7342]	[-0.5750, 0.3574]	[-0.7969, 1.467]	[-0.4899, 0.0930]	[-0.8407, 0.2898]	[-0.8393, 0.1991]	[-1.224, 0.6215]
Control Vector?	No	Yes	No	Yes	No	Yes	No	Yes
	Number of Employees		Log Enterprise Assets		Silvoso Index		Spirituality of Business	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Impact Estimate	-0.0851	-0.0876	0.4718	0.1817	0.7790**	0.7796*	0.2105*	0.1406
Std. Error	(0.5346)	(0.4345)	(0.5092)	(0.4197)	(0.3500)	(0.4335)	(0.1081)	(0.0925)
95% C.I.	[-0.9358, 1.16]	[-0.8996, 0.8037]	[-0.4851, 1.511]	[-0.5035, 1.142]	[0.1909, 1.563]	[0.2203, 1.92]	[-0.01424, 0.4099]	[-0.05471, 0.3081]
Control Vector?	No	Yes	No	Yes	No	Yes	No	Yes

Wild bootstrapped standard errors calculated with 1000 resamples, the null hypothesis imposed, while using Webb’s weights. Statistical significance is represented through p-values: ***p<0.01, **p<0.05, *p<0.1.

