Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

RANDOMIZING RELIGION: THE IMPACT OF PROTESTANT EVANGELISM ON ECONOMIC OUTCOMES*

GHARAD BRYAN
JAMES J. CHOI
DEAN KARLAN

We study the causal impact of religiosity through a randomized evaluation of an evangelical Protestant Christian values and theology education program delivered to thousands of ultrapoor Filipino households. Six months after the program ended, treated households have higher religiosity and income; no statistically significant differences in total labor supply, consumption, food security, or life satisfaction; and lower perceived relative economic status. Exploratory analysis suggests that the income treatment effect may operate through increasing grit. Thirty months after the program ended, significant differences in the intensity of religiosity disappear, but those in the treatment group are less likely to be Catholic and more likely to be Protestant, and there is some mixed evidence that their consumption and perceived relative economic status are higher. *JEL* Codes: D12, I30, O12.

*This study was registered, along with a preanalysis plan, in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0001060. Institutional Review Board approval by Innovations for Poverty Action (#1185), NBER (#17_116), and Yale University (#1410014779). For funding, we thank the Bill & Melinda Gates Foundation for funding-related and overlapping data collection on access to savings; Celia and Joseph Grenny, NIH grant P01AG005842, and the Yale University Economic Growth Center. We thank Manuel Victor Sapitula and Josephine Dionisio for qualitative field work that informed this research. We thank Sachet Bangia, Nate Barker, Leah Bridle, Samantha Horn, Rebecca Hughes, Marius Karabaczek, Sana Khan, Megan McGuire, Neil Mirochnick, Isabel Oñate, Nassreena Sampaco-Baddiri, Cornelius Saunders, Martin Sweeney, and Sneha Stephen from Northwestern University and Innovations for Poverty Action for research assistance and management support. We thank Peter Aronow, Latika Chaudhary, Gregory Cox, Dan Hungerman, Laurence Iannaccone, Aniceto Orbeta, Jared Rubin, and seminar audiences at the Asian Institute of Management, De La Salle University, the Philippine Institute for Development Studies, and numerous venues in the U.S. and Europe for helpful comments. We thank Lincoln Lau, David Sutherland, Peter Nitchke, Daniel Mayhugh, Zaldy Rodriguez, the ICM Metrics team, Danilo Mijares and the ICM Bacolod staff, Lilian Barinas and the ICM Dumaguete staff, Jonathan Sanchez and the ICM Koronadal staff, and Evren Managua and the ICM General Santos staff for their collaboration, patience, flexibility, and curiosity throughout the design and implementation of this study.

© The Author(s) 2020. Published by Oxford University Press on behalf of President and Fellows of Harvard College. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
The Quarterly Journal of Economics (2020), 1–88. doi:10.1093/qje/qjaa023.
Advance Access publication on June 22, 2020.

I. Introduction

A literature dating back at least to Adam Smith and Max Weber has argued that religiosity is associated with a set of characteristics that promote economic success, including diligence, thriftiness, trust, and cooperation (Iannaccone 1998; Iyer 2016). More recent research has linked religiosity to positive outcomes in domains such as physical health (Ellison 1991), crime rates (Freeman 1986), drug and alcohol use (Gruber and Hungerman 2008), income (Gruber 2005), and educational attainment (Freeman 1986; Gruber 2005). Other studies have argued for negative economic effects of some aspects of religiosity due to a focus on otherworldliness (Weber 1958 in his discussion of Catholicism) and substitution toward church attendance away from production (Barro and McCleary 2003). Despite extensive research, claims that religion causes outcomes remain controversial, in part because people choose their religion. Naturally occurring religious affiliation is likely to be correlated with unobserved personal characteristics, which may be the true drivers of the observed correlations. Iannaccone (1998, 1475) writes that "nothing short of a (probably unattainable) 'genuine experiment' will suffice to demonstrate religion's causal impact."

To study the causal impact of religiosity, we partnered with International Care Ministries (ICM), an evangelical Protestant antipoverty organization founded by Filipino pastors that operates through a local network of pastors in the Philippines, to conduct an evaluation that randomly assigned invitations to attend Christian theology and values training. There are 285 million evangelical Christians in the world, comprising 13% of Christians and 36% of Protestants (Hackett and Grim 2011).[1] ICM is representative of an important sector that attempts to generate religiosity while alleviating poverty.

ICM's program, called Transform, normally consists of three components—Protestant Christian theology, values, and character virtues (V), health behaviors (H), and livelihood (i.e.,

1. The National Association of Evangelicals lists four defining characteristics of evangelical Christians that have been identified by historian David Bebbington: "the belief that lives need to be transformed through a 'born-again' experience and a life long process of following Jesus," "the expression and demonstration of the gospel in missionary and social reform efforts," "a high regard for and obedience to the Bible as the ultimate authority," and "a stress on the sacrifice of Jesus Christ on the cross as making possible the redemption of humanity" (https://www.nae.net/what-is-an-evangelical/, accessed April 20, 2018).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

self-employment) skills (L)—taught over 15 weekly meetings (plus a 16th meeting for a graduation ceremony). Each meeting lasts 90 minutes, with 30 minutes allotted to each component. ICM's leadership believes that the Values curriculum lies firmly in the mainstream of evangelical belief. Between 2009 and 2017, 194,000 people participated in Transform. The basic structure of the program, using a set series of classes outside of a Sunday worship service to evangelize, is a common model. For example, over 24 million people in 169 countries have taken the evangelistic Alpha course since 1977 (Bell 2013), and Samaritan's Purse has enrolled 11 million children in about 100 countries in its evangelistic Greatest Journey course since 2010 (Samaritan's Purse 2017). Like Transform, these are courses of approximately a dozen sessions.

We randomly assigned 320 communities (from which we selected 7,999 households) to receive the full Transform curriculum (VHL), to receive only the Health and Livelihood components of the curriculum (HL), to receive only the Christian values component of the curriculum (V), or to be a no-curriculum control (C). (Our experimental design was such that the total amount of religious outreach done by ICM was unaffected during the course of our study. We discuss ethical considerations regarding our study in Section VI.) We identify the effect of religiosity by comparing invited households in VHL communities to invited households in HL communities, and invited households in V communities to households in C communities that would have been invited had that community been assigned to be treated.

We measure outcomes approximately 6 months and 30 months after the training sessions ended and analyze them in accordance with a preanalysis plan. At six months, we find that those who were invited to receive the V curriculum have significantly higher religiosity than those who did not receive the V curriculum, demonstrating that the treatment had its intended first-stage effect. Examining downstream economic outcomes while correcting for multiple hypothesis tests by controlling the false discovery rate (FDR), we find that the V curriculum increased household income by 9.2% but had no statistically significant effect on total labor supply, consumption of a subset of goods, food security, or life satisfaction, and it decreased perceptions of relative economic status within one's community by 0.11 points on a 10-point scale.[2] Post hoc analysis shows that the income

2. In post hoc analysis not contained in our preanalysis plan, we find that the treatment effects on religiosity and income remain statistically significant when

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

effect is strongly concentrated on the Transform invitee and is not significant for other household members' labor income, providing further support that the estimated income effect is not a Type I error.

Exploratory regressions suggest that the religiosity treatment effect operates by increasing grit (Duckworth et al. 2007)—specifically, the portion of grit associated with perseverance of effort (and in particular, agreement with the statements "I am a very hard worker," "I finish whatever I begin," and "Setbacks don't discourage me"). This mechanism accords with Weber's conception of the Protestant work ethic. We find no consistent movement in the other potential mechanisms that we measured: social capital, locus of control (other than the belief that God is in control, which increases), optimism, and self-control. Furthermore, post hoc analysis finds that the HL treatment had no statistically significant effects on income or perceived relative economic status at six months.[3] Because the HL treatment includes many of the nonreligious aspects of the V intervention (e.g., meeting in a group over a number of weeks), this null finding suggests that the six-month V curriculum treatment effect primarily captures the effect of altered religiosity.

By 30 months, there is no longer a statistically significant difference in the intensity of religiosity between the experimental groups. However, individuals who received the V curriculum are 3.6 percentage points less likely to identify as Catholic and 2.3 percentage points more likely to identify as Protestant. To put these changes in context, the control group at 30 months is 70% Catholic and 21% Protestant.

There is mixed evidence on the effects on downstream economic outcomes. Relative to the no-curriculum control, those who

we instead control the family-wise error rate (FWER). We discuss in Section IV.E the conceptual differences between controlling the FDR versus the FWER. We do not combine all of our outcomes into a single index and compute an unadjusted p-value for that index because the outcomes are not all proxies for a single concept. If we were to find that half of our outcomes had positive treatment effects and the other half had negative treatment effects of equivalent magnitude, we would not conclude that the treatment had zero effect. An F-test of the outcomes jointly equaling zero would tell us whether the treatment had any statistically significant effect, but it would not tell us which outcomes the treatment affected.

3. The p-value of the null hypothesis that receiving any HL curriculum has no effect is .299 for income (95% confidence interval = [−2.8%, 9.0%]) and .395 for perceived relative economic status (95% confidence interval = [−0.13, 0.05]).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

received only the V curriculum have a significantly higher perceived relative economic status (0.34 points on a 10-point scale) and marginally significantly higher consumption (7.5% of the control group mean, FDR q-value = .062). Exploration of the mechanisms responsible for these positive effects finds that V curriculum recipients are more optimistic, even though they do not have higher grit. On the other hand, we find no statistically significant effects on primary economic outcomes when combining the VHL versus HL and V versus control comparisons. This difference in findings is driven by the fact that the HL group appears better off than the VHL group at 30 months. Relative to the no-curriculum control, the HL group has significantly higher income and perceived relative economic status (in tests that do not adjust for multiple comparisons).

Interpreting these results requires an understanding of the context and details of the intervention. ICM operates in a setting where most people claim to be religious. In the six-month survey, only 2.4% of those who did not receive the V curriculum and 2.3% of those who did receive the V curriculum indicate that they are "not religious at all." Our experiment should therefore be understood as measuring the effects of strengthening preexisting religiosity or changing the emphasis of preexisting religious beliefs, rather than the effects of causing the completely irreligious to become religious. Arguably, these intensive margin effects are the most relevant ones, since 84% of the world's population is religious (Pew Research Center 2015). It is also important to note that ICM targets the ultrapoor within communities, and the communities in our study (including those in the no-curriculum control) are chosen by pastors who presumably believe that they would be able to run a successful program there. Most expansions by religious organizations into a community are probably based on a belief that the community would be receptive, so these are an externally relevant type of community. It is possible that the ultrapoor are more receptive to religious outreach than less impoverished individuals (Chen 2010), so ICM's outreach may be more effective than comparable outreach to higher-income populations.

In addition, religiosity is not a singular concept, and its causal impact will likely depend on many factors. Johnson, Tompkin, and Webb (2008) differentiate between "organic" exposure to religion over a prolonged period of time (e.g., through one's upbringing at home) and "intentional" exposure through participation in a specific program targeting a specific set of individuals. Both are

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

important channels of religious propagation, and the type of religiosity produced may depend on the channel. Our study is about intentionally generated religiosity of a specific kind (evangelical Protestant Christian), and a significant aim is to establish, in the context of a randomized controlled trial, that intentional exposure to a religious program can generate the critical first stage: an exogenous change in religiosity.

Our article contributes to a recent strand of literature arguing that noncognitive skills are important drivers of economic outcomes and can be improved through specific interventions (Duckworth et al. 2007; Kautz et al. 2014; Blattman, Jamison, and Sheridan 2017). This body of work raises the possibility that programs to improve noncognitive skills might have large positive effects on the lives of the most disadvantaged people, but three obstacles need to be overcome to meet this goal. First, with a few exceptions (e.g., Blattman, Jamison, and Sheridan 2017), existing studies concentrate on high-income countries, whereas most of the world's poorest people live in the developing world. Even if we can assume that noncognitive skills are similarly malleable in the developing world, it is not clear that the environment and market structures allow for economic gains. Second, much of the literature concentrates on children, and little is known about the ability to improve the noncognitive skills of adults, although Kautz et al. (2014) note that noncognitive skills are more malleable later in life than cognitive skills are. Finally, it is unclear whether interventions that create large improvements can be delivered in a cost-effective, scalable manner. Our results suggest that church-based programs might be a solution for building noncognitive skills. Church-based programs make use of a large existing infrastructure, teach a well-understood and developed set of values, and are often low cost because they leverage volunteer labor via the intrinsic motivation of church members.

Our work also relates to a growing number of publications that use instrumental variables or natural experiments to study the causal effect of religion on economic outcomes.[4] Clingingsmith, Khwaja, and Kremer (2009) find that winning a lottery for *hajj* visas changes beliefs, values, and religious practices. Barro and

4. Laboratory experiments that study religious effects by exogenously varying the salience of religion include Shariff and Norenzyan (2007), Mazar, Amir, and Ariely (2008), Hilary and Hui (2009), Horton, Rand, and Zeckhauser (2011), and Benjamin, Choi, and Fisher (2016). See Shariff et al. (2016) for a review of the laboratory literature.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

McCleary (2003) conduct a cross-country analysis of economic growth using the existence of a state religion, state regulation of religion, adherence shares for the major religions, and a religious pluralism index as instruments. They find that religious beliefs (believing) increase economic growth, whereas religious service attendance (belonging) decreases growth. Because our study does not induce independent exogenous variation in beliefs versus behaviors, we cannot add further evidence on this "believing versus belonging" hypothesis. Gruber (2005) uses local ancestral mix as an instrument and finds that religious participation in the United States (which is mostly Christian) increases education, income, and marriage rates and decreases disability and divorce rates. Gruber and Hungerman (2008) exploit the repeal of U.S. state laws prohibiting retail activity on Sundays and find that Christian religious participation decreases drinking and drug use. Bottan and Perez-Truglia (2015) study the decline in Catholic religious participation caused by clergy scandals and find evidence that religious participation increases charitable giving.

Becker and Woessmann (2009) and Cantoni (2015) use geographic distance from Wittenberg, where Martin Luther posted his Ninety-Five Theses, as an instrument for adoption of Protestantism. Becker and Woessmann (2009) conclude that Protestantism does increase income, but this can be entirely accounted for by its effect on literacy, whereas Cantoni (2015) finds no effect on economic growth. Woodberry (2012) argues that Protestants' desire for people to read the Bible fostered mass education, mass printing, and civil society, making it more likely that a country on the receiving end of high historical Protestant missionary activity is a democracy today. Basten and Betz (2013) and Spenkuch (2017) use different peace treaties signed 500 years ago as instruments for local Protestant versus Catholic share and find support for a Protestant work ethic.

The remainder of the article is structured as follows. Section II describes ICM's Transform program, and Section III describes the experimental design. Section IV covers our six-month survey. Section V discusses our 30-month survey and a survey of pastors conducted to examine whether the fading of the religiosity results at 30 months is due to the pastors engaging with the control group after Transform ended. We discuss ethical considerations regarding our study in Section VI, and Section VII concludes.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

II. The ICM Transform Program

Transform's Values curriculum begins by teaching participants to recognize the goodness of the material world and their own high worth as God's creation. The theme then shifts toward humanity's rebellion against God and its negative consequences, while contrasting that with the message that "believers of Jesus will discover joy in sorrow, strength in weakness, timely provision in time of poverty, and peace in the midst of problems and pain." (Transform does not teach prosperity theology—the belief that following God will guarantee economic prosperity and physical health.[5]) The Protestant doctrine of salvation by grace—people cannot earn their way into Heaven by performing good works, but can only be saved by putting their faith in Jesus, upon which God forgives their sins as a free act of grace—is taught. The proper response to God's grace is to do good works out of gratitude. The final section of the curriculum covers what such good works would be. They include not wasting money on gambling and drinking, saving money, treating everyday work as "a sacred ministry," and becoming active in a local church community. Participants are encouraged to find hope in the midst of disasters through faith and generally see that "life's trials and troubles" are "God's pruning knife" that will result in "more fruitfulness." In other words, the curriculum teaches students that their suffering has meaning and purpose, and aims to build the ability to persevere through setbacks. These curricular elements dovetail with the growing literature on noncognitive skills that emphasizes the importance of characteristics like conscientiousness, grit, resilience to adversity, self-esteem, and the ability to engage productively in society (Kautz et al. 2014).

The Health training focuses on building health knowledge and changing health and hygiene practices in the household. In addition, ICM staff identify participants experiencing malnourishment and common health issues such as diarrhea, tuberculosis, and skin problems. They then receive nutritional supplements (estimated to have market value of approximately US$5 per family per week), deworming pills, other medical treatments, and follow-up care.[6]

5. The teacher's manual for the Values curriculum says that "we also see ordinary and simple people who enthrone God as their Lord and Savior discover the deep satisfaction and contentment that make them happy even in their relative poverty."

6. For a small number of households (less than 1%), ICM also arranges treatment for serious medical needs.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

The Livelihood section of the program consists of training in small business management skills, training in one of several different livelihood options (e.g., an introduction to producing compost through vermiculture), and being invited to a savings group. Minor agricultural assistance is given in the form of small seed kits. These activities are intended to provide key tools for achieving a more sustainable income and smoothing economic shocks.

The Health and Livelihood components are led by two employees of ICM, while the religious training is led by a local pastor following an ICM-provided curriculum. The local pastor is not compensated by ICM but does receive training and support. Six lay volunteers from the pastor's church serve as counselors who offer support and encouragement to the participants.

The teacher's manuals used by ICM are available on the authors' websites.

III. Experimental Design

For the experiment, ICM recruited 160 pastors to each choose two communities in which they did not already minister and that were at least 10 kilometers away from each other. Selected communities were required to be predominantly Catholic or Protestant—which meant that Muslim-majority communities were excluded[7]—and not to have been previously contacted by ICM. In each community, the pastor created a list of 40 households that they considered to be the poorest and thus eligible for participation in Transform and interacted with these households to assess their willingness to participate in the program, should it be launched in their village. The pastor identified one member of the household—usually the female head of household or the female spouse of the male head of household—as the potential invitee to Transform. ICM staff then administered a poverty verification questionnaire, based on indicators such as the quality of a home's construction materials; access to electricity, clean water, and sanitation; and household income—most of which do not rely on self-reports. The previously identified individuals in the 30 households deemed poorest out of the 40 households were then invited to participate in the program if their community was selected for treatment.

7. There is only one ICM base (located in Mindanao) that is close to any communities that are predominantly Muslim.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

The randomization was a two-stage clustered design. In the first stage, the pastors were randomly assigned to either group VHL-C or group HL-V. In the second stage, pastors in group VHL-C had one of their communities randomly assigned to receive the full Transform program (VHL) and the other to be a no-treatment control (C). Pastors in group HL-V had one of their communities randomly assigned to receive only the Health and Livelihood component of Transform (HL), and the other to receive only the Christian values component of Transform (V).[8] We implemented this randomization scheme because each pastor had capacity to provide values training in only one community, and thus the scheme allowed every invited pastor to be involved in exactly one Transform implementation. The design also meant that the total amount of religious outreach done by ICM was not altered due to the study. Because the treatments were assigned at the community level, the estimated effect of the Values treatment on downstream economic outcomes should be interpreted as the effect of increasing religious engagement for a group of individuals in a community, rather than the effect for an isolated individual. We view this as a desirable feature, since religion is most often experienced and practiced in a communal context.

The four-month Transform program ran from February to May 2015. HL/VHL households on average attended 8.9 class sessions, and 83% attended at least 1.[9] Participants in the VHL and HL treatment arms received nutritional supplements as described in Section II. Participants in the V treatment arm received food assistance only for child malnutrition, and ICM estimates that there were fewer than five such cases. ICM arranged treatment for serious medical needs in the VHL, HL, and V arms (less than 1% of participants).

ICM carried out the experimental implementation, independent of the researchers, although the research team did the randomization. ICM covered the costs of the V and VHL treatments, but the researchers raised funds to cover the costs of implementing the HL curriculum, as ICM's unrestricted donations were typically raised with the understanding that they would be used for

8. Both HL and V communities were also assisted by six counselors recruited by the pastors prior to the random assignment.

9. ICM did not track attendance in the V group. If somebody was sent in the place of an invited individual, ICM recorded that individual as present. We cannot distinguish these substitute attendances from regular attendances.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

programs that included a religious component. Neither ICM nor its donors provided compensation to the researchers.

IV. Six-Month Survey

IV.A. Data Collection

Approximately six months after Transform ended (between August 12, 2015, and January 14, 2016), we sent surveyors to the poorest 25 households selected by the pastors in each community.[10] Respondents were compensated with 100 PHP (about US$2.5), irrespective of whether they completed the survey.

To reduce the correlation between treatment assignment and social desirability bias in survey responses, we used surveyors from Innovations for Poverty Action (IPA), a nonprofit research organization independent from ICM. Respondents were not told of any relationship between ICM and IPA, and the informed consent script introduced the survey as follows: “Hello, my name is _____ with the research organization Innovations for Poverty Action. I am working to learn about the economic and social conditions and well-being of families in the Philippines. You are being invited to be one of the participants in this study. We expect the results from this survey will help Filipino NGOs and international organizations to develop policies and procedures that improve the lives of people.”

As we will discuss in Section IV.B, we divide our outcomes into primary religious outcomes, primary economic outcomes, mechanisms, and secondary outcomes. All of the questions about primary economic outcomes came before the main religiosity questions. If these direct religiosity questions caused subjects to discern a link between ICM and IPA, only some of the secondary outcome and mechanism questions would have been affected. We did mention religion at three points before measuring primary economic outcomes. First, the script for obtaining informed consent said, “If you agree to participate in this study, we will ask you questions about your household’s economic, health, social, and religious status.” Second, when constructing the household roster, we asked about each household member’s relation to the head of the household, permanence of his or her residence in the home, gender,

10. We sampled the 25 poorest households, rather than the full 30 identified by ICM, because of budget constraints and the programmatic importance of measuring the impact on the poorer individuals within the sample.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

age, religious denomination, marital status, schooling, literacy, and work status. Third, we asked five list-randomized questions (described in Section IV.B), two of which measure religiosity in an obscured way. Given the many different characteristics in the informed consent script and measured in the household roster, the obscured nature of the list-randomized questions, and the fact that only two of the five list-randomized questions had religious content (which in turn was shown to only half the respondents; see Section IV.B), we think it is unlikely that respondents would have inferred a link to ICM when we were eliciting primary economic outcomes.

Surveyors attempted to interview, in descending order of preference, (i) the person previously identified as a potential Transform invitee, (ii) the female head of household if the head of household was female, (iii) the female spouse/partner of the male head of household, or (iv) the person reporting to be responsible for health and household expense decisions. Out of 7,999 households targeted for surveying, we successfully surveyed 6,507 (81%); in 88% of these households, the respondent was the potential/actual Transform invitee. Insurgent violence and political opposition prevented the field teams from surveying in six communities (150 households), and some households refused to be surveyed (60 households), could not be contacted (1,252 households), or suffered from survey data issues (30 households).

Management data and internal control checks identified five instances (out of the 157 pastors whose communities we surveyed) in which ICM and the pastor switched the assignments within a community pair, treating one with what the other was supposed to receive, and vice versa. Because of the paired randomization, we drop these five community pairs in our analysis without harming internal validity. There was also one community that was supposed to receive the V treatment but did not. We retain this community in our regressions (coded as a V community), because the compliance issue was not present in both communities in the pair.[11] Thus, we only use data from 6,276 households in our main analyses. Online Appendix Table 1 shows that the attrition rate and the number of days between program end and survey date do not differ significantly across the four experimental groups.

11. We show in Online Appendix Tables 3–5 the main six-month regressions including the five dropped pairs, using the assigned treatment status for each community. Online Appendix Tables 48–50 show analogous regressions for the 30-month survey.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Before the intervention, we intended to conduct a baseline survey of the 7,999 households. However, we underestimated the time this would take, and we were unable to delay the start of Transform to complete the baseline survey. This means that we have baseline data on only 2,634 of the households. Online Appendix Table 1 shows that the four experimental groups are well-balanced on characteristics measured in the six-month survey that are unlikely to have changed in response to the treatment. Online Appendix Table 2 shows that in the subsample of households we were able to survey at baseline and which are not in the excluded communities, household income and respondent age, education, income, and religiosity at baseline do not predict attrition from the six-month survey, but men are 4.8 percentage points less likely to be in the six-month survey.[12]

We filed a preanalysis plan with the American Economic Association RCT Registry before seeing any follow-up data. In accordance with our first filing, we examined the follow-up data blinded to treatment assignment and filed a supplement to the preanalysis plan.[13]

12. In unreported regressions, we find that when baseline characteristics are interacted with treatment assignment, these interactions jointly predict attrition at six months at $p < .05$ for household income and $p < .10$ for respondent income, education, and religiosity. The income interactions' significance is driven by higher income predicting less attrition in the HL group relative to the control group. The religiosity interaction's significance is driven by higher religiosity predicting more attrition from the VHL group relative to the control group. These would bias against our finding positive income and religiosity effects of the Values curriculum. The education interaction's significance is driven by higher education predicting more attrition in the VHL group relative to the control group, which would again bias against finding a positive Values curriculum income effect. However, this appears to have had a minor effect in practice because average education is well balanced between the VHL and control groups in the complete six-month survey sample ($p = .777$; see Online Appendix Table 1).

13. In accordance with the first phase of our preanalysis plan, we analyzed the data stripped of treatment status. We randomly generated treatment assignments and checked whether including control variables from the available baseline observations reduced the standard errors of the coefficients on the randomly generated treatment dummies. We did not find any efficiency gains, so we decided not to use the baseline survey in our final regressions. We do, however, include controls for demographic variables that were collected after the intervention and which were unlikely to be affected by the treatment. Online Appendix Tables 6–8 show the six-month treatment effect estimates on the primary outcomes, mechanisms, and secondary outcomes if we additionally control for baseline survey measurements and dummies for each of these baseline variable values being missing.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

IV.B. Outcome Variables

Our preanalysis plan divided outcomes into primary religious outcomes, primary economic outcomes, mechanisms, and secondary outcomes. Many of these outcome variables are indices, which we standardize so that the control group has zero mean and unit variance. If the index is found in previous academic literature, we use the construction method from that literature, which in our cases always involves simply summing the components (which are sometimes reverse-coded). If there is no preexisting index, we use the index construction methodology of Kling, Liebman, and Katz (2007). We sign all component variables such that higher values are telling a consistent story for each component of the index. Then we standardize each component by subtracting its control group mean and dividing by its control group standard deviation. We compute the sum of the standardized components and standardize the sum once again by the control group sum's standard deviation.[14] Appendix Table A.1 shows all of the questions that make up our variables. Online Appendix Tables 12–40 show the treatment effect estimates on each component of the outcome variables.[15]

The primary religious outcomes are the intrinsic religious orientation scale and the sum of the two extrinsic religious orientation scales of Gorsuch and McPherson (1989), a general religion index that consolidates responses to nine religious belief and practice questions and the average of two binary indicators for whether the respondent reports that “I have made a personal commitment to Jesus Christ that is still important to me today” and “I have read or listened to the Bible in the past week.” These last two binary indicators are elicited using list randomization, a technique for eliciting responses to sensitive questions that conceals any given individual's response from the interviewer (Droitcour et al. 1991; Karlan and Zinman 2012). We do this to minimize experimenter demand and social desirability effects. In a list-randomized elicitation, participants are randomly selected to receive either a list of n nonsensitive statements or these same n statements plus a sensitive statement. They are asked to answer how many of the statements are true without specifying which ones are true. The

14. For observations without information on one or more components of the index, we impute the missing component standardized values as the mean of the nonmissing components' standardized values for that individual/household.

15. We also include Online Appendix Table 41, which shows treatment effects on consumption of “temptation goods” (cigarettes and alcoholic beverages).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

difference in the average number of statements reported to be true between participants who received n statements and $n + 1$ statements is the estimated fraction of participants for whom the sensitive statement is true.[16]

After data collection, we discovered an issue with our measure of intrinsic religiosity. The indices for intrinsic and extrinsic religious orientation were measured using one 14-question block, with 8 questions constituting the intrinsic index and 6 constituting the extrinsic index. For each question, respondents were asked to state on a Likert scale a level of agreement with a statement. In 11 out of the 14 questions, stronger agreement corresponds to stronger religiosity. In the remaining three—all of which are part of the intrinsic index—weaker agreement corresponds to stronger religiosity. We believe that respondents did not perceive the subtle changes in the direction of the questions, causing them to use stronger agreement to express stronger religiosity even for the reversed questions.[17] Agreement levels are positively correlated across all seven intrinsic orientation statements, regardless of whether greater agreement corresponds to greater religiosity. Because of this, we have chosen to exclude the three reversed questions from the intrinsic index used for the main analysis.

16. An individual's answer about the sensitive statement can only be deduced if they answer 0 (implying falsity of the sensitive statement) or $n + 1$ (implying truth of the sensitive statement). An individual can answer truthfully about the longer list while being assured that their response to the sensitive statement is concealed if the number of nonsensitive questions that are true for them is not 0 or n. Among respondents who did not receive the sensitive statement, the fraction who did not give a boundary response was 73% for the list associated with the commitment to Jesus statement and 80% for the list associated with the Bible statement. The corresponding percentages are 82%, 83%, and 86% for the water treatment, hand washing, and domestic abuse questions, respectively. Therefore, the list-randomization questions concealed the truth about the sensitive statements for the majority of our respondents.

17. Thirty-three percent of respondents answered "agree" or "strongly agree" to all 14 questions, regardless of whether the question was reversed, whereas only 0.02% of respondents answered "agree" or "strongly agree" to all nonreversed questions and "disagree" or "strongly disagree" to all reversed questions. The finding that many subjects indiscriminately agree with statements to express a general support for religion goes back to the earliest research on intrinsic and extrinsic religious orientation. Allport and Ross (1967, 441) argue, "In responding to the religious items these individuals seem to take a superficial or 'hit and run' approach. Their mental set seems to be 'all religion is good.' 'My religious beliefs are what really lie behind my whole life'—Yes! 'Although I believe in my religion, I feel there are many more important things in my life'—Yes!" They classify such types as the "indiscriminately pro-religious" and find that they are likely to be less educated.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Our broad conclusions about the six-month treatment effect on religiosity are unchanged by this choice.[18]

The primary economic outcomes are household expenditure on a sample of consumption goods, a food security index, household income, total household adult labor supply in hours, an index of life satisfaction, and perceived relative economic status.

The mechanism outcomes are three measures of social capital (a general trust index, a strength of social safety net index, and a participation in community activities index), three measures of a sense that one has control over one's life (a perceived stress index, the Levenson (1981) Powerful Others index modified to apply to God's control of one's life, and a locus of control index that combines the internality and chance subscales of Levenson (1981) and the World Values Survey locus of control question), three measures of optimism (the Life Orientation Test—Revised index (Scheier, Carver, and Bridges 1994), an index of expectations about one's life satisfaction and relative economic status five years in the future, and a general optimism index), the Short Grit Scale (Duckworth and Quinn 2009), and a subset of the Brief Self-Control Scale (Tangney, Baumeister, and Boone 2004).

The secondary outcomes are an index of belief in the Protestant doctrine of salvation by grace (an outcome of interest to ICM because the doctrine is taught in the V curriculum, and the mechanism through which Weber (1958) hypothesized that Protestantism's encouragement of capitalistic activities operated), an asset index, a financial inclusion index, a health index, two hygienic practice variables, a home quality index, a migration and remittance index, an absence of domestic discord index, absence of domestic violence, child labor supply, and the number of children enrolled in school.

18. If we instead use the eight-question intrinsic measure, as stated in our preanalysis plan, the point estimate of the "Any-V" treatment effect on intrinsic religious orientation in the pooled regression specification is 0.04 standard deviations, and its q-value rises to .084. In the disaggregated regression specification, the point estimate of the V versus control effect on intrinsic religious orientation is 0.01 standard deviations ($q = .899$), and the point estimate of the VHL versus HL effect on intrinsic religious orientation is 0.074 standard deviations ($q = .330$). The q-values on the other religious outcomes are qualitatively similar regardless of whether we use the eight-question or five-question intrinsic measure. Therefore, even though the estimates of the V curriculum's effect on intrinsic religious orientation weaken when we use the eight-question measure, we still find robust first-stage effects on other measures of religiosity.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

IV.C. Prespecified Econometric Strategy

In this subsection, we discuss our prespecified econometric strategy. (In Section IV.E, we present several post hoc analyses.) Treatment effects are estimated using OLS regressions with the following explanatory variables: treatment indicator variables, an indicator variable for the respondent's gender, an indicator variable for the respondent being married, an indicator variable for the respondent being divorced or separated, the respondent's years of educational attainment,[19] the number of adults in the household (age $\geqslant 17$), the number of children in the household (age < 17), and the number of days between June 1, 2015, and the interview date.[20] We also include fixed effects for each pair of communities chosen by a given pastor (community-pair fixed effects) where possible, as discussed in detail below. We cluster standard errors by community (the unit of randomization).

We estimate the treatment effect on list-randomized variables by stacking the responses of those who did and did not receive the sensitive statement in a regression that controls for treatment assignment indicator variables, an indicator variable for whether the individual received the sensitive statement, the interaction between receiving the sensitive statement and each treatment indicator variable, and all the other nontreatment variable controls from the main specification. The coefficients on the interaction variables are the treatment effects of interest. We estimate the control mean by calculating within the control group the difference (without adjusting for covariates) in the mean response between those who got the sensitive statement and those who did

19. Preschool only is coded as 0.5 years, 1st grade only is coded as 1 year, 2nd grade only is coded as 2 years, . . . 11th grade only is coded as 11 years, some 12th grade without high school graduation is coded as 12 years, high school graduation is coded as 13 years, partial vocational education is coded as 14 years, complete vocational education is coded as 15 years, partial college is coded as 16 years, and college graduation is coded as 17 years. There are 27 observations for which the respondent's name is not in the household roster, and thus respondent demographic information is missing. We code the respondent demographic variables as equaling 0 for these 27 observations and control for an indicator variable equal to 1 if respondent demographic information is missing.

20. These control variables were measured at the same time as the outcome variables, but are unlikely to have been affected by the treatments. Online Appendix Tables 9–11 show the treatment effect estimates on the prespecified outcomes when the only explanatory variables are the treatment dummies and community pair or ICM base dummies.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

not. When two list-randomized variables are combined to form an outcome variable, we stack the responses for both variables into a single regression while retaining the same control variables as above. The coefficient on each interaction variable in this case is the treatment effect on the average of the two outcomes of interest.

We test for the effect of religiosity by comparing VHL to HL respondents and V to control respondents. We do not reject the hypothesis that the V and HL curricula have additive effects when testing jointly across all outcomes of interest; the p-values for this test are .344, .634, .890, and .234 when looking across religious primary outcomes, all primary outcomes, all primary outcomes and mechanisms, and all outcomes, respectively. Therefore—following our preanalysis plan—we also run a pooled specification that estimates the effect of being invited to receive any V curriculum, while controlling for whether the household was invited to receive any HL curriculum. This pooled specification gives consistent inference on the average of the V curriculum effect with and without a concurrent Health and Livelihood curriculum and has greater statistical power than a specification that separately estimates the VHL-versus-HL and V-versus-control effects.[21]

Because we conducted a matched-pair randomization, our pooled specification controls for the community-pair fixed effects previously mentioned. In our disaggregated specification, where we estimate VHL, HL, and V treatment effects separately, the estimation of the VHL treatment effect versus control also controls for community-pair fixed effects. However, we cannot include community-pair fixed effects when estimating the HL and V treatment effects versus control because pastors were assigned either to get one HL and one V community, or to get one VHL and one control community. No pastor who had one community assigned to control had the other assigned to HL or V. We therefore generate the disaggregated specification's treatment estimates from two independently estimated regressions: one to estimate the treatment effect for VHL relative to control with community-pair fixed effects, and a second to estimate the treatment effects for HL and V relative to control with fixed effects for the ICM base with which the community is associated.[22]

21. The fact that we cannot reject that the treatment effects are additive gives some confidence that this average effect is the same as the Values curriculum effect without a concurrent Health and Livelihoods curriculum.

22. There are four ICM bases. Our preanalysis plan stated that we would control for community-pair fixed effects in all regressions. We have deviated from

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Because of the multiple hypotheses tested, we follow Banerjee et al. (2015): for each primary test in our preanalysis plan, we calculate a q-value—the minimum FDR (i.e., the expected proportion of rejected null hypotheses that are actually true) at which the null hypothesis would be rejected for that test (Benjamini and Hochberg 1995; Anderson 2008), given the other tests run within the family.[23] For the purposes of this correction, and in accordance with our preanalysis plan, we consider the tests on primary religious outcomes to be one family (because they are a test of the study's first stage, a null result here would eliminate the justification for examining the nonreligious outcomes) and the tests on primary nonreligious outcomes to be another family. We implement adjustments once among the pooled specification regressions and separately among the disaggregated specification regressions. In other words, the tests run within the pooled specification do not affect the q-values from the disaggregated specification and vice versa. Following our preanalysis plan, we do not apply multiple hypothesis test corrections to our tests of hypothesized mechanisms and secondary outcomes because these analyses are exploratory.

IV.D. *Results of Prespecified Analyses*

The majority of our sample (69%) self-identifies as Catholic, and 21% as Protestant. Online Appendix Tables 12–15 summarize

---

the plan here because it is mathematically impossible to control for community-pair fixed effects in the disaggregated specification while estimating every single treatment effect. Because of the randomized design, the inability to control for community-pair fixed effects when estimating the HL and V treatment effects relative to control does not bias our estimates, but it does reduce our statistical power.

23. Within each of our outcome families, let $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_m$ be the set of ordered p-values that correspond to the m hypotheses tested. For a given false discovery rate α, let k be the largest value of i such that $p_i \leqslant \frac{i\alpha}{m}$, and reject all hypotheses with rank $i \leqslant k$. The q-value of a hypothesis, an analog to the p-value, is the smallest α for which the hypothesis would be rejected (Anderson 2008). The Benjamini-Hochberg procedure was originally proven to work under the assumption that the test statistics were independent. Subsequent work has shown that the procedure is robust to various dependence structures (Goeman and Solari 2014). Romano, Sheikh, and Wolf (2008) develop a testing procedure that incorporates information about the dependence structure. Benjamini-Hochberg q-values are conservative, and more powerful procedures have been more recently developed (e.g., Storey, Taylor, and Siegmund 2004; Benjamini, Krieger, and Yekutieli 2006). We do not follow these approaches because we wish to stay as close as possible to our preanalysis plan, which specified the more conservative Benjamini-Hochberg procedure.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

the control group's level of religiosity and indicate that many are not maximally religiously fervent. For example, when asked, "To what extent do you consider yourself a religious person?" the average control respondent rates herself at 2.8 on a 4-point scale, where higher numbers indicate greater religiosity. Only 66% say that they have made a personal commitment to Jesus Christ that is still important to them today, and 56% have read or listened to the Bible in the past week.

Tables I–III contain all of our prespecified analyses. Table I, columns (1)–(4) show the treatment effects on the primary religious outcomes. The pooled specification (Panel A) finds that the V curriculum, offered either on its own or in conjunction with the HL curriculum, increases all four measures of religiosity, three of them at $q < .01$.[24] The Any-V effect on the three statistically significant indices ranges from 0.08 to 0.13 standard deviations. The change in the list-randomization outcome—which we have lower statistical power to detect, both because list-randomized questions measure the outcome of interest in only half the sample and because we only have two such questions—is positive, and its 4.8 percentage point magnitude (corresponding to a 0.10 standard deviation movement given the 60.6% control group mean) is economically significant and in line with the magnitudes (in standard deviation space) we get from the three direct elicitation measures. However, the 95% confidence interval for the list-randomization index treatment effect is wide and encompasses zero. Furthermore, recent work has demonstrated a large amount of instability in estimates coming from list randomization. In a developing country context, Chuang et al. (2020) find that within a single survey of about 1,000 respondents, estimates of the prevalence of a given sensitive behavior can vary by as much as 39 percentage points across two list-randomized elicitations, and there is no clear evidence that the list-randomized estimates are systematically less biased than direct responses. Thus, we believe little should be concluded from the treatment effect estimates on the list-randomization outcome. The statistically significant first-stage effect of the treatment on directly elicited religiosity justifies examining differences in downstream nonreligious

24. Although intrinsic and extrinsic religious orientation were originally conceived of as opposing concepts on a unidimensional scale, empirical work has found the two to be orthogonal (Kirkpatrick and Hood 1990).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE I
PRIMARY OUTCOMES, SIX-MONTH SURVEY

	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index (1)	Religion extrinsic index (2)	General religion index (3)	Religion, list-randomized (4)	Monthly consump-tion (PHP) (5)	Food security index (6)	Monthly income (PHP) (7)	Adult weekly labor supply (hours) (8)	Life satis-faction index (9)	Perceived relative econ. status (10)
Panel A: Pooled specification										
Any-V	0.102	0.130	0.077	0.048	–1.1	0.010	386.1	0.9	0.019	–0.113
	(0.024)	(0.024)	(0.023)	(0.037)	(100.4)	(0.023)	(126.8)	(1.1)	(0.022)	(0.047)
Any-HL	0.014	–0.021	0.001	–0.028	–103.0	–0.044	131.2	–1.8	–0.010	–0.040
	(0.024)	(0.024)	(0.023)	(0.038)	(93.3)	(0.023)	(126.3)	(1.1)	(0.022)	(0.047)
FDR q-value, Any-V	.000	.000	.001	.197	.991	.778	.015	.595	.595	.050
FWER p-value, Any-V	.000	.000	.002	.197	1.000	1.000	.015	1.000	1.000	.083
Panel B: Disaggregated specification										
VHL	0.115	0.109	0.077	0.020	–102.2	–0.033	524.4	–0.9	0.009	–0.151
	(0.034)	(0.037)	(0.031)	(0.054)	(159.5)	(0.037)	(175.0)	(1.4)	(0.028)	(0.067)
HL	0.047	0.073	–0.029	–0.002	–314.3	–0.050	287.9	–0.1	–0.031	–0.073
	(0.055)	(0.065)	(0.054)	(0.055)	(203.0)	(0.051)	(278.4)	(2.4)	(0.056)	(0.112)
V	0.123	0.204	0.052	0.070	–167.4	–0.007	574.2	3.0	–0.018	–0.133
	(0.050)	(0.064)	(0.051)	(0.057)	(209.5)	(0.050)	(285.4)	(2.3)	(0.047)	(0.119)
FDR q-value, VHL = HL	.393	.653	.146	.653	—	—	—	—	—	—
FDR q-value, V = C	.058	.013	.416	.393	.637	.885	.271	.529	.850	.529
FWER p-value, VHL = HL	1.000	1.000	.330	1.000	—	—	—	—	—	—
FWER p-value, V = C	.102	.013	1.000	1.000	1.000	1.000	.271	1.000	1.000	1.000

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE I
(CONTINUED)

	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index (1)	Religion extrinsic index (2)	General religion index (3)	Religion, list-randomized (4)	Monthly consumption (PHP) (5)	Food security index (6)	Monthly income (PHP) (7)	Adult weekly labor supply (hours) (8)	Life satisfaction index (9)	Perceived relative econ. status (10)
Panel C: Summary information										
Control mean	0	0	0	0.606	5,001	0	4,213	79.6	0	3.242
Control standard deviation	1	1	1	—	4,720	1	5,567	57.7	1	2.256
# observations in VHL	1,578	1,578	1,578	1,578	1,578	1,526	1,452	1,452	1,578	1,576
# observations in HL	1,549	1,549	1,549	1,549	1,549	1,521	1,440	1,439	1,549	1,548
# observations in V	1,550	1,550	1,550	1,550	1,550	1,517	1,435	1,434	1,550	1,547
# observations in C	1,599	1,599	1,599	1,599	1,599	1,567	1,490	1,490	1,599	1,596

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. See Appendix Table A.1 for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. In Panel B, we do not show VHL = HL q-values and FWER-adjusted p-values for primary economic outcomes because there is no significant first-stage VHL versus HL difference in religiosity. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

outcomes across treatment groups to gain insight into the effects of religiosity.

We also present results for the disaggregated specification in Table I, Panel B, where we estimate the impact of the V curriculum by separately comparing VHL against HL and V against control. Although the point estimates of VHL's effect on religiosity relative to HL are always positive, they are not statistically significant. On the other hand, V significantly increases extrinsic religious orientation (0.20 std. dev., $q = .013$) and marginally statistically significantly increases intrinsic religious orientation (0.12 std. dev., $q = .058$) relative to the control group. Therefore, although we report all treatment effect estimates on downstream outcomes from the disaggregated specification, we only discuss and interpret these outcomes for the V versus control comparisons and only correct for multiple hypothesis tests within the V versus control comparisons.

The primary economic outcome effects are reported in Table I, columns (5)–(10). We find no statistically significant treatment effects on consumption, food security, total adult labor supply, or life satisfaction. We have enough statistical power to reject, at the 95% confidence level, increases in these variables of more than 0.06 standard deviations and decreases of more than 0.04 standard deviations. However, we do find a statistically significant 9.2% increase in income (386 PHP $\approx$ US\$8.6 per month, $q = .015$) in the pooled specification (Panel A).[25] In the disaggregated specification (Panel B), where we have less statistical power (the standard errors are over twice as large as in the pooled specification), the 574 PHP income effect for V compared to C is statistically significant before correcting for multiple hypothesis tests but not after ($p = .045$, $q = .271$). We also find a decrease in perceived relative economic status (-0.11 points on a 10-point scale, which corresponds to -0.05 std. dev., $q = .050$) in the pooled specification. Perceived relative economic status is measured by one question that asks respondents to place themselves on a ladder where the top rung (10) represents the best-off people in their community and the bottom rung (1) the poorest people in their community. We discuss potential interpretations of these results in Section IV.F.

25. The results become more statistically significant when income is winsorized at the 95th or 99th percentile, or when we use the log of income (see Online Appendix Table 42).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Table II reports tests of mechanisms that might generate the primary economic effects. The V curriculum teaches that God's love continues during adversity, which he ultimately uses for good, so participants can find hope in the midst of hardship. Correspondingly, we find in the pooled specification (Panel A) that the V curriculum leads to increases in the sense that God is in control (powerful others index, 0.09 std. dev., $p = .001$)[26] and a marginally statistically significant increase in grit (0.04 std. dev., $p = .065$). However, there is no consistent effect on the three measures of optimism. Perceived self-control falls by a marginally statistically significant extent (-0.03 std. dev., $p = .095$), which could be due to the V curriculum increasing the number of behaviors participants believe to be undesirable temptations rather than an actual reduction in self-control. There is also a marginally statistically significant reduction in perceived locus of control (-0.04 std. dev., $p = .075$), although subcomponent analysis finds that V recipients report that both personal initiative and chance play larger roles in their life (Online Appendix Table 27). Although all three of the treatment arms—VHL, HL, and V—involve group meetings that could increase social capital, we see no consistent or statistically significant effects of any of the treatments on our measures of trust, the presence of a social safety net, or participation in community activities.[27]

Finally, we examine treatment effects on secondary outcomes (Table III). In the pooled specification, we find that the V curriculum leads to statistically significant ($p = .0002$) increases in hygienic behaviors not measured by list randomization (avoiding open defecation and keeping animals in a sanitary way), but no statistically significant increase in the list-randomization response regarding washing hands after using the bathroom and treating water. We note that we find via list randomization an increase in reported domestic violence, although it is only significant at the 10% level. This finding could be interpreted either as an increase in identifying behaviors as abuse or an increase in actual abuse. Although we do not observe a statistically significant

26. Although our preanalysis plan treats the powerful others index as a potential mechanism rather than a primary outcome, the increase in its value could also be seen as evidence that the V curriculum succeeded in increasing religiosity.

27. Online Appendix Table 14 shows that the Any-V effect on religious service attendance frequency is not statistically significant (0.9 times a year increase, standard error = 0.6, with a control mean of 39.5 times a year).

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE II

MECHANISMS, SIX-MONTH SURVEY

	Social capital			Locus of control			Optimism				
	Trust index (1)	Social safety net index (2)	Community activities index (3)	Perceived stress scale index (4)	Powerful others index (5)	Locus of control index (6)	Life orientation index (7)	Expectations index (8)	Optimism index (9)	Grit index (10)	Self-control index (11)
Panel A: Pooled specification											
Any-V	0.004	0.026	0.005	–0.011	0.093	–0.035	–0.050	–0.037	0.053	0.041	–0.034
	(0.022)	(0.024)	(0.025)	(0.020)	(0.027)	(0.020)	(0.027)	(0.025)	(0.024)	(0.022)	(0.021)
Any–HL	–0.023	–0.027	0.041	–0.018	0.044	–0.000	0.016	–0.016	–0.024	0.017	0.006
	(0.022)	(0.024)	(0.025)	(0.021)	(0.027)	(0.020)	(0.027)	(0.025)	(0.024)	(0.022)	(0.020)
p-value, Any-V	.865	.282	.851	.596	.001	.075	.065	.133	.029	.065	.095
Panel B: Disaggregated specification											
VHL	–0.019	0.000	0.045	–0.026	0.135	–0.035	–0.034	–0.055	0.030	0.056	–0.027
	(0.032)	(0.032)	(0.034)	(0.026)	(0.038)	(0.029)	(0.037)	(0.032)	(0.032)	(0.029)	(0.025)
HL	–0.023	–0.076	0.019	–0.009	0.031	–0.064	–0.046	–0.014	–0.007	0.030	0.039
	(0.043)	(0.048)	(0.058)	(0.044)	(0.060)	(0.057)	(0.068)	(0.056)	(0.061)	(0.058)	(0.047)
V	–0.018	–0.023	–0.011	–0.007	0.073	–0.085	–0.103	–0.054	0.069	0.041	–0.001
	(0.046)	(0.048)	(0.059)	(0.043)	(0.059)	(0.050)	(0.069)	(0.057)	(0.066)	(0.058)	(0.050)
p-value, VHL = HL	.927	.140	.655	.684	.085	.605	.862	.468	.541	.671	.155
p-value, V = C	.704	.631	.857	.876	.222	.090	.132	.344	.298	.484	.980

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE II

(CONTINUED)

	Social capital			Locus of control			Optimism				
	Trust index (1)	Social safety net index (2)	Community activities index (3)	Perceived stress scale index (4)	Powerful others index (5)	Locus of control index (6)	Life orientation index (7)	Expectations index (8)	Optimism index (9)	Grit index (10)	Self-control index (11)
Panel C: Summary information											
Control mean	0	0	0	0	0	0	0	0	0	0	0
Control standard deviation	1	1	1	1	1	1	1	1	1	1	1
# observations in VHL	1,578	1,578	1,561	1,577	1,578	1,578	1,578	1,542	1,578	1,578	1,578
# observations in HL	1,549	1,549	1,542	1,549	1,549	1,549	1,549	1,508	1,549	1,549	1,549
# observations in V	1,550	1,550	1,534	1,549	1,550	1,550	1,550	1,518	1,550	1,550	1,550
# observations in C	1,599	1,599	1,592	1,599	1,599	1,599	1,599	1,567	1,599	1,599	1,599

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. Indices have been coded so that more positive numbers are better. See Appendix Table A.1 for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE III

SECONDARY OUTCOMES, SIX-MONTH SURVEY

	Salvation by grace belief index (1)	Assets index (2)	Financial inclusion index (3)	Health index (4)	Hygiene index, non-list-random. (5)	Hygiene, list random. (6)	House index (7)	Migration and re-mittance index (8)	No discord index (9)	No domestic violence, list-rand. (10)	Child labor supply (hours) (11)	# children enrolled in school (12)
Panel A: Pooled specification												
Any-V	–0.036	–0.027	0.020	0.000	0.092	0.043	0.030	0.027	–0.034	–0.072	0.2	–0.02
	(0.020)	(0.021)	(0.024)	(0.020)	(0.024)	(0.033)	(0.025)	(0.019)	(0.024)	(0.040)	(0.2)	(0.02)
Any-HL	–0.005	–0.025	0.157	0.015	0.030	0.066	0.007	–0.015	–0.029	–0.048	0.0	–0.01
	(0.020)	(0.021)	(0.025)	(0.020)	(0.024)	(0.033)	(0.025)	(0.019)	(0.024)	(0.040)	(0.2)	(0.02)
p-value, Any-V	.079	.211	.396	.985	.000	.191	.239	.153	.164	.078	.256	.349
Panel B: Disaggregated specification												
VHL	–0.040	–0.050	0.179	0.015	0.121	0.108	0.036	0.012	–0.063	–0.118	0.3	–0.03
	(0.026)	(0.031)	(0.038)	(0.028)	(0.034)	(0.049)	(0.036)	(0.031)	(0.036)	(0.055)	(0.3)	(0.02)
HL	–0.021	0.014	0.124	–0.027	0.136	0.121	0.045	–0.083	–0.036	–0.081	–0.1	–0.01
	(0.045)	(0.057)	(0.048)	(0.042)	(0.070)	(0.043)	(0.059)	(0.038)	(0.052)	(0.058)	(0.4)	(0.04)
V	–0.061	0.008	–0.010	–0.044	0.208	0.105	0.068	–0.039	–0.049	–0.120	0.1	–0.02
	(0.041)	(0.060)	(0.044)	(0.041)	(0.067)	(0.045)	(0.060)	(0.039)	(0.049)	(0.061)	(0.4)	(0.04)
p-value, VHL = HL	.696	.265	.297	.334	.836	.779	.879	.017	.617	.509	.404	.687
p-value, V = C	.143	.899	.811	.285	.002	.020	.258	.317	.326	.050	.775	.618

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE III
(CONTINUED)

	Salvation by grace belief index (1)	Assets index (2)	Financial inclusion index (3)	Health index (4)	Hygiene index, non-list-random. (5)	Hygiene, list random. (6)	House index (7)	Migration and re-mittance index (8)	No discord index (9)	No domestic violence, list-rand. (10)	Child labor supply (hours) (11)	# children enrolled in school (12)
Panel C: Summary information												
Control mean	0	0	0	0	0	0.558	0	0	0	0.903	1.6	1.67
Control standard deviation	1	1	1	1	1	—	1	1	1	0.037	12.3	1.37
# observations in VHL	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,267	1,579	1,452	1,578
# observations in HL	1,549	1,549	1,549	1,549	1,549	1,549	1,549	1,549	1,297	1,550	1,439	1,549
# observations in V	1,550	1,550	1,550	1,550	1,550	1,550	1,550	1,550	1,263	1,551	1,434	1,550
# observations in C	1,599	1,599	1,599	1,599	1,599	1,599	1,599	1,599	1,331	1,600	1,490	1,599

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. Indices have been coded so that more positive numbers are better. See Appendix for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015 and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

change in the non-list-randomized discord index, we do in post hoc analysis observe a significant increase in one of its components, major arguments regarding interactions with relatives (2.2 percentage points, $p = .009$, Online Appendix Table 39).

The remainder of the secondary outcomes are not statistically significant at the 5% level. We find an unexpected marginally statistically significant decrease in the index for the belief in the doctrine of salvation by grace. This may be because of the counterintuitive nature of the doctrine, which requires one to disagree with two of the three statements in our index: "I follow God's laws so that I can go to heaven" and "If I am good enough, God will cleanse me of my sins." In becoming more religiously fervent, subjects may have felt that they should agree more strongly with these pious-sounding statements despite the efforts of the V curriculum. The V curriculum also increases agreement with the third statement in the index, "I will go to heaven because I have accepted Jesus Christ as my personal savior," even though that statement is consistent with salvation by grace. The pattern of responses is consistent with the V curriculum increasing agreement with all pious-sounding statements.

IV.E. Post Hoc Analyses

In this subsection, we discuss assorted post hoc (nonpreregistered) analyses, many of which address robustness.

1. Controlling the Family-Wise Error Rate. An alternative approach to correcting for multiple hypothesis tests is to control the family-wise error rate (FWER) instead of the FDR. The FWER is the probability of incorrectly rejecting at least one true null hypothesis among all those tested, while the FDR is the expected proportion of rejected null hypotheses that are actually true. The following matrix, taken from Efron (2013), illustrates the difference between these two quantities.

		Decision		
		Null	Nonnull	Total
Actual	Null	$N_0 - x$	x	N_0
	Nonnull	$N_1 - y$	y	N_1
	Total	$N - R$	R	N

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

There are N null hypotheses being tested, of which N_0 are actually true (null) and N_1 are actually false (nonnull). Consider a decision rule that incorrectly decides that x of the true null hypotheses are false, and $N_1 - y$ of the false null hypotheses are true. The FWER is the probability that $x > 0$, R is the number of rejected null hypotheses, and the FDR is the expectation of $\frac{x}{R}$ (defining $\frac{x}{R}$ to be 0 when $R = 0$). Controlling the FWER results in fewer false positives at the cost of lower statistical power relative to controlling the FDR. Controlling the FDR instead of the FWER is appropriate if one judges the cost of false positives to be relatively low compared with the benefit of detecting true positives.

In post hoc analysis, we control the FWER using the procedure of Holm (1979), which has greater power than the Bonferroni correction to detect truly false nulls while preserving the upper bound on the FWER. The FWER-adjusted p-value for a null hypothesis is the FWER tolerance level above which we would reject that null.

Table I shows FWER-adjusted p-values—the only nonprespecified analysis contained in this table. In our setting, both FDR and FWER control lead to similar qualitative inferences, partly because of the relatively modest number of hypotheses tested. In the pooled specification, only the effect on perceived relative economic status crosses a 1% or 5% significance boundary, with an adjusted p-value of .083 versus a q-value of .050. In the disaggregated specification, the V versus control effect on intrinsic religious orientation is no longer significant even at the 10% level (adjusted $p = .102$ versus $q = .058$), but the V versus control effect on extrinsic religion orientation remains significant (adjusted $p = .013$).

2. Naive OLS versus IV Estimates of Religiosity Effect. What would a researcher who naively runs an OLS regression of economic outcomes on religiosity in our control and HL groups find? We construct a composite religiosity index for each respondent by adding her intrinsic, extrinsic, and general religion indices together and normalizing so that its standard deviation in the control group is 1. Online Appendix Table 45 shows that this naive analysis leads to a significant negative coefficient of religiosity on monthly income of –291 PHP and on weekly adult labor supply of –2.3 hours, indicating negative selection into religiosity. This is consistent with a literature that suggests that the club good provision aspects of religion are likely to generate more demand from

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

those with low income (Chen 2010). Also interesting is that despite lower objective economic status among the more religious, the religiosity coefficient on life satisfaction is statistically significantly positive, and the religiosity coefficient on perceived relative economic status is marginally significantly positive ($p = .091$) as well.

In contrast, an IV estimation on our full sample, using receipt of the V curriculum as the instrument, finds that a one standard deviation increase in composite religiosity significantly increases monthly income by 3,073 PHP and decreases perceived relative economic status by 0.9 points on a 10-point scale. These are large estimates, but they should be interpreted with caution because it seems likely that nearly all Transform participants had their religiosity increased by much less than a full standard deviation. If so, the estimated effect of a one standard deviation increase in religiosity achieved through intentional means (as opposed to organic means, as discussed in the introduction, which is probably mostly responsible for the cross-sectional variance in control group religiosity) is a linear extrapolation of an effect that is estimated over a much smaller range. The true effect size curve may be quite concave, so the actual causal effect of increasing religiosity by a full standard deviation through intentional means may be much smaller than our estimate.

Figure I shows suggestive evidence that the V curriculum had an effect on religiosity that is consistently less than one standard deviation, indicating that the IV estimation relies heavily on linear out-of-sample extrapolation to obtain a one standard deviation effect size. The three graphs split the sample by whether the community received the V curriculum (VHL and V groups) or not (HL and control groups), sorts each subsample by one of the directly elicited measures of religiosity, and displays, for each percentile, the difference in the religiosity variable value between the Any-V individual and the No-V individual at this percentile. For example, the leftmost point in the top graph shows the first percentile intrinsic religion index value among the Any-V groups minus the first percentile intrinsic religion index value among the No-V groups. The difference in religiosity never exceeds 0.35 standard deviations for the intrinsic index, exceeds 0.22 standard deviations only once for the extrinsic index, and exceeds 0.18 standard deviations only once for the general index.[28] (The intrinsic

28. Top-coding is significant for the intrinsic and extrinsic indexes; 25% of the sample has the maximum possible intrinsic index value, and 13% of the sample has the maximum possible extrinsic index value.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

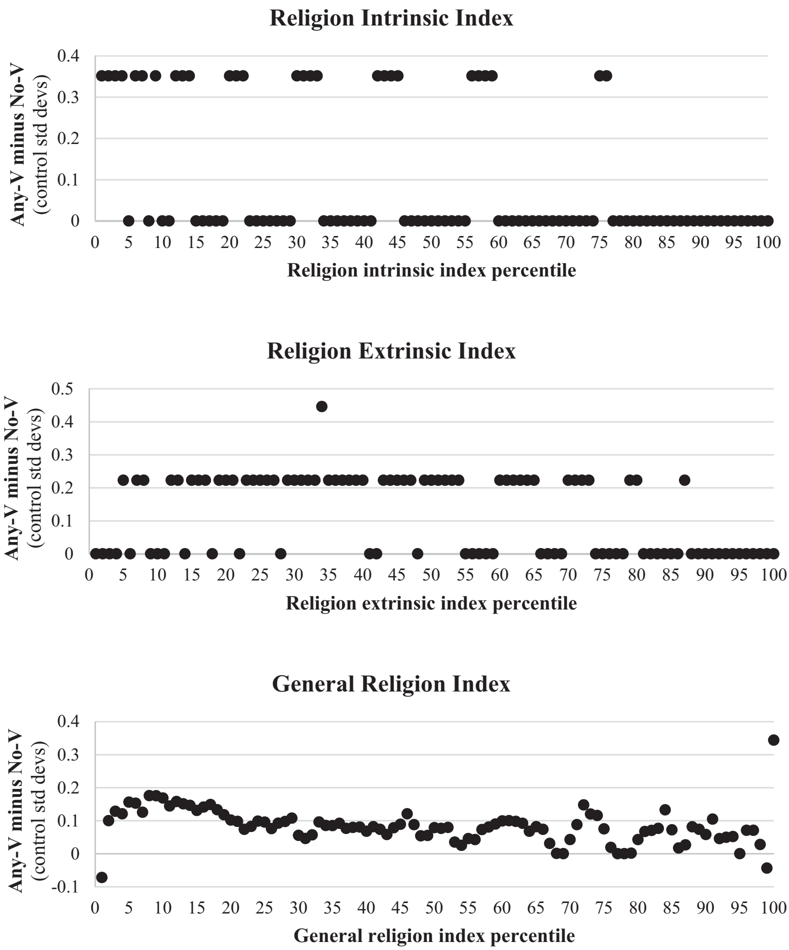

FIGURE I

Religiosity in VHL and V Groups (Any-V) Minus Religiosity in HL and Control Groups (No-V) at each Percentile, Six Months after Treatment

We rank households who were invited to receive the Values curriculum by their religiosity index level at six months. Define V_i to be the index level for the person whose percentile ranking is i. Similarly define NV_i for those not invited to receive the Values curriculum. Each graph plots $V_i - NV_i$ for $i \in \{1, 2, \ldots, 100\}$ for the religiosity index in the graph's title. The y-axis units are multiples of the control group's standard deviation. The intrinsic and extrinsic indices are discrete measures that can take on only a relatively small set of outcomes. A one-point difference in the intrinsic index is 0.35 control standard deviations, and a one-point difference in the extrinsic index is 0.22 control standard deviations. The top two graphs show that there is no percentile at which the difference between V_i and NV_i is greater than one point for the intrinsic index and two points for the extrinsic index.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

and extrinsic indices are measured on a discrete scale, with one point on the scale corresponding to 0.35 standard deviations for the intrinsic index and 0.22 standard deviations for the extrinsic index. The graphs show that at no percentile do Any-V and No-V subjects differ by more than one point on this scale.) Although the V curriculum could cause religiosity ranks to change in a population, these graphs suggest that the V curriculum seldom increases religiosity by anything close to a full standard deviation.

3. Treatment Effect Heterogeneity. The differences plotted in Figure I tend to be smaller at higher percentiles. This suggests that the V curriculum increases our religiosity measures more at lower percentiles of religiosity. If that is true and our religiosity variables map linearly to true religiosity, then the V curriculum has a stronger effect on religiosity for the less religious, in which case downstream economic treatment effects might also be stronger for the less religious.

Our ability to rigorously identify treatment effect heterogeneity is limited because we were unable to collect pretreatment baseline data on most of our sample. What we can do is stratify the sample based on a small number of characteristics collected in the six-month survey that are unlikely to have been affected by the treatment at the time of measurement (respondent age, gender, years of education, literacy, marital status, number of children in the household, and number of adults in the household). Employing the leave-one-out procedure of Abadie, Chingos, and West (2018), we use these variables to predict the composite religiosity index (defined above in Section IV.E.2) at six months in the HL and control groups.[29] We then sort observations into terciles based on their predicted composite religiosity index in the absence of the V curriculum and estimate treatment effects separately within each tercile as before. Throughout this analysis, we restrict the sample to those where the respondent is the targeted or actual Transform invitee.

Using this method, Table IV shows no clear pattern of treatment effect heterogeneity. Although the treatment effect on the composite religiosity index decreases with predicted composite religiosity sans V curriculum (significantly only in the disaggregated specification), the treatment effect on the list-randomized

29. In a multivariate regression that does not leave any observations out, significant positive predictors of religiosity in the HL and control groups are being female, older, literate, less educated, and not divorced.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE IV

TREATMENT EFFECTS BY PREDICTED RELIGIOSITY WITHOUT V CURRICULUM, SIX-MONTH SURVEY

	Predicted religiosity			*p*-value of joint equality across terciles
	Low	Medium	High	
Panel A: Pooled specification				
Composite religiosity index	0.152	0.146	0.074	.101
	(0.040)	(0.039)	(0.035)	
Religion, list-randomized	–0.041	0.054	0.191	.025
	(0.065)	(0.070)	(0.068)	
Monthly income	421.8	318.5	407.3	.912
	(226.3)	(188.4)	(255.9)	
Perceived relative economic status	–0.164	–0.092	0.002	.291
	(0.080)	(0.090)	(0.085)	
Panel B: Disaggregated specification, V versus control				
Composite religiosity index	0.252	0.160	0.027	.021
	(0.078)	(0.070)	(0.071)	
Religion, list-randomized	0.008	0.034	0.212	.200
	(0.092)	(0.098)	(0.090)	
Monthly income	880.6	471.7	408.3	.607
	(420.2)	(298.8)	(518.0)	
Perceived relative economic status	–0.050	–0.242	–0.054	.624
	(0.171)	(0.188)	(0.168)	
Panel C: Summary information				
Mean composite religiosity index value in control and HL groups	–0.123	–0.036	0.227	

Notes. Panel A shows "Any-V" treatment effects on the variable in the left column, separately for each tercile of predicted composite religiosity index value in the absence of the V curriculum. See Appendix Table A.1 for details on variable construction. The composite religiosity index is the normalized sum of the intrinsic, extrinsic, and general religion indices. The predictor variables are respondent age, gender, years of education, literacy, and marital status; number of children in the household; and number of adults in the household. Panel B shows treatment effects estimated by comparing the V group to the control group. All regressions estimating treatment effects control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A control for community-pair fixed effects. The regressions in Panel B control for ICM base fixed effects. The sample is restricted to observations where the survey respondent is the potential or actual Transform invitee. Standard errors clustered by community are in parentheses.

religiosity measure (which was not shown in Figure I) increases with predicted composite religiosity sans V curriculum (significantly only in the pooled specification). There is correspondingly no statistically significant difference across terciles in the treatment effect on monthly income and perceived relative economic

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

status—the two primary economic outcomes for which we found a significant effect over the entire sample.

4. How Much of the Any-V Treatment Effect Operates through the V Curriculum? Those assigned to the V treatment not only received the V curriculum but also socialized with other classmates, spent time away from home to attend class, received medical treatment (with less than 1% probability), and so on. How much of the Any-V treatment effect is due to the V curriculum itself rather than the other accompanying factors?

We can gain some insight into this question by comparing the effect of the HL treatment, which also brought participants together for ICM-sponsored classes, to the effect of the V treatment. Under the assumption that the HL curriculum's treatment effect has the same sign as the V curriculum's treatment effect, the difference between the Any-V and Any-HL effects is a lower bound on the portion of the Any-V effect that comes from the V curriculum.

Comparing magnitudes of the point estimates in Table I, we see that the Any-V treatment effect on income is 386 PHP, whereas the Any-HL treatment effect is only 131 PHP, suggesting that at least 66% of the Any-V treatment effect is due to the V curriculum. Similarly, at least 64% of the decrease in perceived relative economic status caused by the Any-V treatment is due to the V curriculum. An analogous comparison of the V treatment effect to the HL treatment effect in the disaggregated specification suggests a lower bound of 50% for the income effect and 45% for the perceived relative economic status effect due to the V curriculum. However, we note that we cannot statistically reject equality of the Any-V and Any-HL effects on these outcomes in the pooled specification, nor the equality of the corresponding V and HL effects in the disaggregated specification, which means we cannot rule out the possibility that the economic effects we identify are attributable to noncurricular elements that accompany the V curriculum.[30]

30. The p-value of the difference between the Any-V and Any-HL treatment effects is .160 for income and .270 for perceived relative economic status. The p-value of the difference between the V and HL treatment effects is .257 for income and .628 for perceived relative economic status. We can also compare the VHL to HL treatment effects in the disaggregated specification, although this analysis is clouded by the fact that we detected no significant difference in religiosity between these two treatment cells. We find that the incremental addition of the V

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

5. Social Desirability Bias in Survey Responses. Although it is possible that the V curriculum is causing respondents to increase the amount by which they falsely inflate reported income for social desirability reasons, this seems unlikely, because there is no positive V treatment effect on other economic outcomes—in particular, self-reported life satisfaction, a more subjective outcome than income that seems at least as susceptible to social desirability motives.

We can also test for the existence of social desirability bias in some of our survey responses by using the technique of Coffman, Coffman, and Marzilli Ericson (2017). For four of the sensitive statements whose truth we elicited by list randomization, we have direct questions elsewhere in the survey that ask about the same issue. We take respondents whose list randomized question did not include the sensitive statement of interest and compute how many of the list items would have been reported true if their list had included the sensitive statement of interest, using their response to the direct question to impute whether the sensitive statement would have been counted as true in the list-randomized question.[31] Under the null of no social desirability bias (keeping in mind the caveats about the instability of list randomized estimates raised by Chuang et al. 2020), there should be no difference between (i) the number of statements that are indicated to be true by those who did receive the sensitive statement in their

curriculum accounts for 45% of the VHL effect on income and 52% of the VHL effect on perceived relative economic status. The p-value of the difference between the VHL and HL treatment effects is .390 for income and .488 for perceived relative economic status.

31. The directly asked questions are “How much do you agree with this statement: ‘I have made a personal commitment to Jesus Christ that is still important to me today’”; “In the past seven days, how many times did you read or listen to the Bible, the Koran, or other religious literature?”; “Do you wash your hands with ash or soap after using the latrine?”; and “Is the following true or false? Someone in my household is experiencing physical abuse.” We code the “personal commitment to Jesus Christ” statement as true if the respondent slightly agrees, agrees, or strongly agrees; reading or listening to the Bible as true if the respondent did so at least once; and washing hands as true if the respondent answers sometimes or always. The results are directionally identical if we count the “personal commitment to Jesus” statement as true only if the respondent agrees or strongly agrees, and if we count washing hands as true only if the respondent answers “always.” Due to a programming problem in the questionnaire, we only have 1,447 observations for the physical abuse question.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE V

TEST FOR EXISTENCE OF SOCIAL DESIRABILITY BIAS IN RESPONSES, SIX-MONTH SURVEY

	I have made a personal commitment to Jesus Christ that is still important to me	I have read or listened to the Bible in the past week	I wash my hands after going to the bathroom	Someone in my household is experiencing physical abuse (higher = less abuse)
Panel A: Presence of social desirability bias				
Received sensitive statement	–0.262	–0.217	–0.228	–0.093
	(0.023)	(0.024)	(0.021)	(0.045)
Constant	3.609	2.237	2.615	–1.577
	(0.103)	(0.109)	(0.179)	(0.943)
Observations	6,276	6,276	6,262	1,447
Panel B: Interaction of social desirability bias with treatment				
Received sensitive statement	–0.286	–0.197	–0.261	–0.097
	(0.045)	(0.041)	(0.034)	(0.080)
Sensitive statement × Any-V	0.042	0.021	0.031	0.020
	(0.046)	(0.047)	(0.041)	(0.091)
Sensitive statement × Any-HL	0.007	–0.059	0.037	–0.012
	(0.047)	(0.048)	(0.042)	(0.091)
Any-V	0.024	0.074	0.013	–0.101
	(0.029)	(0.031)	(0.025)	(0.061)
Any-HL	–0.001	0.037	0.006	0.032
	(0.029)	(0.030)	(0.025)	(0.060)
Constant	3.608	2.200	2.605	–1.584
	(0.106)	(0.110)	(0.181)	(0.909)
Observations	6,276	6,276	6,262	1,447

Notes. This table shows coefficients for regressions where the dependent variable is the number of statements reported to be true in a list that includes the sensitive statement in the column label. (We use the negative of this number for the physical abuse question.) For respondents who did not actually receive that statement in their list, the dependent variable is the number of statements they reported to be true plus an indicator for whether we impute that the sensitive statement is true for them based on their response to a direct question about it. The key explanatory variables are a dummy for having actually received the sensitive statement in the list, treatment dummies, and interactions between sensitive statement receipt and the treatment dummies. The regressions also control for respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; the number of days between June 1, 2015, and the interview date; and community-pair fixed effects. Standard errors clustered by community are in parentheses.

list randomized question, and (ii) the number of statements that we impute would have been marked as true by those who did not receive the sensitive statement in their list randomized question.

Table V, Panel A shows the results of a regression that tests the null of no social desirability bias, where the dependent variable is the number of statements the respondent said were true

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

(either actual or imputed) and the main explanatory variable is a dummy for having actually received the sensitive statement in the list. We see that the fraction that reports a personal commitment to Jesus, reading or listening to the Bible in the past week, washing their hands after going to the bathroom, or that nobody in their household is experiencing physical abuse is 26, 22, 23, and 9 percentage points lower, respectively, when this is elicited via list randomization rather than directly. This indicates the existence of social desirability bias. However, in Panel B, we see that the size of this bias does not vary significantly with whether the respondent received the V curriculum. Although the standard errors of these interaction coefficients are relatively large, they do suggest that social desirability is not biasing our treatment effect estimates.[32]

It may also be the case that it is more psychologically costly to lie about publicly observable expressions of religiosity, making self-reports about them more truthful. We asked respondents about two religious activities that would have been observed by others: "In the last month, have you tried to convince anyone else to change the way they think about God?" and "How often do you go to religious services?" A binary indicator for the first question and a coding of the second question into the number of attendances per year are positively and significantly ($p < .01$) correlated with the intrinsic, extrinsic, and general religion indices (with the general religion index stripped of these two publicly observable components).

6. Sensitivity of Estimates to Survey Attrition. We noted in Section IV.A that the survey attrition rate did not differ across experimental cells. In this subsection, we examine how our results would be affected if the outcomes of nonresponders systematically differ across experimental cells.

Let j index primary outcomes excluding list-randomized religiosity. For every missing response to outcome j, we impute a value x_j if the household is in the VHL or V group and y_j if the household is in the HL or control group. In the most pessimistic scenario, for all primary outcomes excluding list-randomized

32. An alternative analysis that estimates treatment effects on the responses to the direct questions finds that none of the Any-V treatment effects estimated this way are statistically distinguishable from the Any-V treatment effects estimated using list randomization, although the standard errors of the list randomization estimates are large.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

religiosity, we set x_j equal to the minimum observed value of j in the household's ICM base × treatment arm cell and y_j equal to the maximum observed value of j in the same cell. In the most optimistic scenario, we set x_j equal to the maximum observed value in the household's ICM base × treatment arm cell and y_j equal to the minimum observed value of the outcome in the same cell. We also consider the scenarios $(x_j, y_j) = (\mu_j - Z\sigma_j, \mu_j + Z\sigma_j)$ for $Z = \{-0.25, -0.1, -0.05, 0.05, 0.1, 0.25\}$, where μ_j and σ_j are the mean and standard deviation of observed j within the household's base × treatment cell. For each scenario, we estimate treatment effects for all the primary outcomes, setting missing explanatory variables equal to their observed base × treatment means, and compute q-values.[33]

Online Appendix Table 43 shows that the most pessimistic scenario in which the Any-V treatment effect on religiosity remains statistically significantly positive is if all missing VHL and V observations have religiosity 0.1 standard deviations below their base × treatment means and all missing HL and control observations have religiosity 0.1 standard deviations above their base × treatment means. The most pessimistic scenario in which the Any-V treatment effect on income remains statistically significantly positive is if all missing VHL and V observations have primary economic outcomes 0.05 standard deviations below their base × treatment means and all missing HL and control observations have primary outcomes 0.05 standard deviations above their base × treatment means. Even the smallest optimistic perturbation considered suffices to eliminate the statistical significance of the negative perceived relative economic status effect.

IV.F. Discussion of Six-Month Results

A puzzle regarding the treatment effect on income is that we do not observe movement in other variables that would be expected to rise with income—total labor supply, consumption, food security, and assets—and perceived relative economic status decreases.

For labor supply, although there is no change in total hours, we do see a shift from agriculture to nonagricultural self-employment, livestock tending, fishing, and other employment of

33. In the q-value calculation, we use the p-value from the list-randomized religiosity treatment effect without imputed observations.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

unclear formality (Online Appendix Table 19), which could increase income. Furthermore, we cannot observe labor effort per hour worked, which may increase with grit and which the V curriculum encourages as “a sacred ministry” that “merits heavenly reward.” In post hoc analysis, we examine two subscales within the grit index (Duckworth et al. 2007; Duckworth and Quinn 2009) and find that all of the movement in grit is coming from the “perseverance of effort” subscale ($p = .00003$ for Any-V, $p = .041$ for V = C)—which is the sum of agreement with the statements “I am a very hard worker,” “I finish whatever I begin,” “Setbacks don’t discourage me,” and “I am diligent”—and not the “consistency of interests” subscale ($p = .396$ for Any-V, $p = .655$ for V = C). This is consistent with the doctrine of hard work promoted by the V curriculum.[34]

A simple explanation could in principle account for the lack of observed movement in consumption and assets: all of the additional income was consumed, but we do not have the statistical power to detect this. However, when we test whether the Any-V income and consumption effects are equal to each other, we reject this hypothesis at $p = .003$. This leaves open the possibility that there was an increase in expenditures on the goods, services, and assets that we did not measure.[35]

Of course, it is possible that the income result is a purely random Type I error despite the multiple-testing correction. Further evidence seems inconsistent with this interpretation. Among the 88% of households where the individual identified as a potential Transform invitee was the survey respondent, the Any-V effect on labor income is 236 PHP ($p = .0006$) for the respondent herself and 164 PHP ($p = .151$) summed across all other household members. Hence, the labor income effect is strongly concentrated on the Transform beneficiary.

Another possibility is that control and HL group respondents are understating their income to the surveyor as part of a general practice of understating their resources to avoid having to

34. In Online Appendix Table 30, columns (3), (5), (8), and (9) are the subcomponents that sum up to the perseverance of effort subscale, and columns (2), (4), (6), and (7) are the subcomponents that sum up to the consistency of interests subscale.

35. For example, we did not collect data on tithing. ICM reports that its pastors collect on average 570 PHP a month from their entire congregation, and the average congregation has about 25 adults. Thus, the gap between the income and consumption treatment effects is unlikely to be entirely explained by tithing.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

share them with others, and the V curriculum raises reported income because it causes respondents to be more honest about their income. But this is inconsistent with the lack of a V curriculum effect on the number of meals the household gave to others in the local community in the past 30 days (Online Appendix Table 23), although it is possible that the V treatment increases actual meals given and reduces exaggeration in the number of meals that respondents reported having given by approximately the same amount.

The negative effect on perceived relative economic status could arise from participants realizing that Transform targeted those in extreme poverty. However, the HL treatment used the same targeting process, and we do not observe a significant negative effect on perceived relative economic status for the HL curriculum. Furthermore, Banerjee et al. (2015) find that other programs targeting those in extreme poverty do not generate a negative effect on perceived relative well-being, although their measurements occurred two years after program completion, rather than six months. The V curriculum did move participants into work activities where they earned more per hour (as noted already, income increases but hours of labor supply did not increase) and from agricultural labor to enterprise labor, both of which may have increased their contact with higher-income individuals. Alternatively, the Values curriculum, by attempting to build hope and aspiration, may make salient to attendees that others are living without as much economic hardship.

V. 30-Month Survey

V.A. *Survey Administration*

Thirty months after the end of the Transform program, we started sending IPA surveyors to households again and successfully interviewed 5,878 of them (73%) over a six-month span (November 27, 2017–June 6, 2018). Surveyors attempted to interview the potential/actual Transform invitee, and if he or she was not available, the potential/actual invitee's spouse or partner. In 84% of successfully interviewed households, the respondent was the potential/actual invitee. Insurgent violence prevented surveyors from entering eight communities—the six affected by violence during the six-month survey plus two others. Respondents were compensated with 100 PHP. We again drop from our analysis

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

sample the five community pairs that were not treated in accordance with their treatment assignment. Online Appendix Table 46 shows that the attrition rate does not differ significantly across the four experimental groups and that the groups are balanced on observable characteristics in joint tests of equality. Online Appendix Table 47 shows that among those successfully surveyed at 6 months, attrition at 30 months is statistically significantly higher for younger and male respondents, but is not statistically significantly related to education, household or respondent income, or religiosity measured at 6 months.[36]

V.B. Econometric Strategy and Outcome Variables

We did not separately preregister the analysis for the 30-month survey, but generally follow the preanalysis plan used for the 6-month survey.

Because of the trouble respondents had in the six-month survey with the three reversed questions in the intrinsic religiosity index, we replaced those reversed questions with analogous questions for which stronger agreement indicates greater religiosity.[37] In the analysis that follows, we construct the intrinsic religiosity index excluding these three revised questions, but including them does not qualitatively change our results.

Based on feedback from ICM and surveyors in the field, we modified some of the other questions that comprise our outcome variables. Although we sacrificed comparability across the two surveys to gain precision and surveying efficiency, we do not believe that any of the changes bias the treatment effect estimates

36. In untabulated results, we find that the predictiveness of education and income measured at 6 months for 30-month attrition significantly varies across treatment arms. The significance of the education variation is driven by each year of education being associated with a 0.6 percentage point higher probability of attrition in the HL group relative to the control group. But Online Appendix Table 46 shows that when testing the equality of education levels between the control and HL groups at 30 months, the *p*-value is .880, indicating that this differential attrition created minimal imbalance in practice. The significance of the income variation is driven by 1,000 PHP of extra income being associated with 1.0 percentage point lower probability of attrition in the VHL arm than the control arm. This would bias us toward finding a positive Any-V income effect, but we in fact estimate a null effect.

37. The three revised questions ask about agreement with the statements, "My religious beliefs are important as well as my behavior," "My religion affects my daily life," and "My religion is one of the most important things in my life."

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

by affecting some treatment cells differently in expectation than the others.

We added questions about spending on gambling and gaming, snacks, water, and electricity, which we then include in our consumption variable. Recall periods were changed from one week to 30 days for the following spending categories: phone credit, transportation, clothing and shoes, soaps, and cosmetics and detergents. We stopped asking about spending on gifts because we separately ask about spending on weddings, funerals, festivals, anniversaries, and birthdays, so the response to the gifts question may lead to double-counting of spending. As in the six-month survey, we scale all reported spending to obtain monthly spending rates.

We shifted from measuring household business and nonbusiness income in separate sections to measuring both in the same section in a uniform manner. The recall period for nonbusiness income was changed from 30 days to 7 days, and household business profit was also measured over the past seven days rather than over the most recent month with "normal sales." We scale all income categories up to monthly rates for the purposes of analysis. To reduce the frequency of income sources falling into the "other" category, we changed the set of available categories in the survey's income classification question, and labor supply categories were changed to match the income categories.[38]

38. In the six-month survey, the income categories were agricultural labor for a nonhousehold member, salaried/formal employment outside the household, housework in an outside household, animal tending in an outside household, operating a business that is not the household's, daily labor, and other. In the 30-month survey, the income categories were self-employed/household business/own business, wage labor, casual labor, piece worker, and other. For those who were reported to be in wage or casual labor, we asked whether the work fell into one of 11 subcategories. For those who were self-employed or working in a household business or in their own business, we asked whether the business fell into one of eight subcategories. For those who were doing piece work, we asked whether it involved food products or nonfood products. We added an additional income question asking about any other income received over the last 30 days that had not been mentioned yet, such as money from friends and family, remittances, additional labor income, pensions, and government transfers. These income sources were not measured in the six-month survey, so we exclude it from the main 30-month income variable. Online Appendix Table 60 shows that the 30-month Any-V treatment effect on this other nonlabor income is a 139 PHP increase ($p = .055$), which approximately offsets the –117 PHP Any-V treatment effect on the main income variable shown in Table VI.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Due to budget constraints, we dropped some questions from the 30-month survey, most of which had high overlap with other questions. We dropped three sets of questions from the life satisfaction index—whether taking all things together, the respondent would say they are happy; whether the respondent experienced enjoyment/happiness/worry/sadness during a lot of the day yesterday; and whether the respondent smiled or laughed a lot yesterday. From the community activities index, we dropped a question on attendance at village leaders' meetings. From the three mechanism measures related to locus of control, we dropped the perceived stress scale index. From the three mechanism measures related to optimism, we dropped the life orientation index and optimism index. Among secondary outcomes, we dropped the questions about open defecation from the non-list-randomized hygiene index (leaving only a question about whether animals are kept in a stable separate from the house), the question about whether the primary latrine is in the house from the six-component house index, and the number of days migrators in the household were gone in the past six months from the five-component migration and remittance index.

V.C. Treatment Effects on Primary Outcomes, Mechanisms, and Secondary Outcomes

Table VI shows 30-month treatment effect estimates for the primary outcomes. (Online Appendix Tables 54–81 show the treatment effect estimates on each component of the outcome variables.) There is no statistically significant treatment effect for any of the primary religious outcomes; in fact, three of the four Any-V point estimates are negative, and these are significantly different from their corresponding six-month treatment effects ($p \leqslant .002$). However, we also investigated whether the treatment had an effect on denominational affiliation. These regressions, reported in Table VII, were not included in our six-month preanalysis plan. The results show that there is a shift in religious affiliation at 30 months. Receiving the V curriculum is associated with a 3.6 percentage point decline in the likelihood of the survey respondent identifying as a Catholic ($p = .014$), a 2.3 percentage point increase in the likelihood of identifying as a Protestant ($p = .102$), and a 1.3 percentage point increase in the likelihood of identifying with some other religion ($p = .025$). The increase in "other" affiliation is mostly driven by a

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE VI
PRIMARY OUTCOMES, 30-MONTH SURVEY

	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index (1)	Religion extrinsic index (2)	General religion index (3)	Religion, list randomized (4)	Monthly consumption (PHP) (5)	Food security index (6)	Monthly income (PHP) (7)	Adult weekly labor supply (hours) (8)	Life satisfaction index (9)	Perceived relative econ. status (10)
Panel A: Pooled specification										
Any-V	–0.052	–0.008	–0.023	0.001	131.9	–0.014	–116.9	–0.8	–0.004	0.097
	(0.025)	(0.026)	(0.025)	(0.040)	(88.2)	(0.024)	(189.0)	(1.1)	(0.022)	(0.044)
Any-HL	0.035	0.018	–0.047	0.021	–77.5	–0.050	246.1	0.8	0.036	0.019
	(0.025)	(0.026)	(0.025)	(0.040)	(88.5)	(0.024)	(191.9)	(1.1)	(0.022)	(0.044)
FDR q-value, Any-V	.163	.980	.726	.980	.408	.669	.669	.669	.841	.168
FWER p-value, Any-V	.163	1.000	1.000	1.000	.680	1.000	1.000	1.000	1.000	.168
p-value, Any-V 6 vs. 30 mo.	.000	.000	.002	.370	.235	.309	.012	.227	.755	.000
Panel B: Disaggregated specification										
VHL	–0.013	0.012	–0.069	0.019	56.2	–0.065	134.3	0.1	0.032	0.120
	(0.030)	(0.033)	(0.031)	(0.055)	(115.5)	(0.037)	(287.6)	(1.6)	(0.026)	(0.062)
HL	0.013	0.035	–0.027	0.083	254.6	–0.012	842.1	4.3	0.077	0.234
	(0.067)	(0.074)	(0.063)	(0.055)	(195.5)	(0.052)	(393.9)	(2.1)	(0.052)	(0.109)
V	–0.070	0.007	–0.004	0.052	481.4	0.024	501.2	2.8	0.044	0.340
	(0.066)	(0.071)	(0.065)	(0.054)	(207.2)	(0.050)	(434.7)	(2.2)	(0.054)	(0.114)
FDR q-value, VHL = HL	.954	.954	.954	.914	—	—	—	—	—	—
FDR q-value, V = C	.914	.954	.954	.914	.062	.625	.375	.375	.507	.019
FWER p-value, VHL = HL	1.000	1.000	1.000	1.000	—	—	—	—	—	—
FWER p-value, V = C	1.000	1.000	1.000	1.000	.104	.844	.795	.795	.844	.019
p-value, V 6 vs. 30 months	.017	.032	.470	.827	.005	.601	.864	.958	.313	.002

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE VI
(CONTINUED)

	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index (1)	Religion extrinsic index (2)	General religion index (3)	Religion, list randomized (4)	Monthly consumption (PHP) (5)	Food security index (6)	Monthly income (PHP) (7)	Adult weekly labor supply (hours) (8)	Life satisfaction index (9)	Perceived relative econ. status (10)
Panel C: Summary information										
Control mean	0	0	0	0.347	6,378	0	8,162	67.8	0	3.662
Control standard deviation	1	1	1	—	3,789	1	10,500	52.5	1	2.050
# observations in VHL	1,441	1,441	1,441	1,441	1,441	1,440	1,441	1,441	1,441	1,440
# observations in HL	1,366	1,366	1,366	1,366	1,366	1,365	1,366	1,366	1,366	1,365
# observations in V	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
# observations in C	1,481	1,481	1,481	1,481	1,481	1,479	1,481	1,481	1,481	1,480

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. See Appendix Table A.1 and Section V.B for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. In Panel B, we do not show VHL = HL q-values and FWER-adjusted p-values for primary economic outcomes because there is no significant first-stage VHL versus HL difference in religiosity in the six-month survey. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects. When testing the null that the 6- and 30-month treatment effects are equal, we reestimate the 6-month treatment effects on the life satisfaction index using a variable definition that is harmonized with the 30-month variable definition.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE VII

TREATMENT EFFECTS ON RELIGIOUS AFFILIATION

	6-month survey			30-month survey		
	Catholic	Protestant	Other	Catholic	Protestant	Other
Panel A: Pooled specification						
Any-V	–0.027	0.004	0.023	–0.036	0.023	0.013
	(0.015)	(0.012)	(0.009)	(0.014)	(0.014)	(0.006)
Any-HL	–0.004	–0.004	0.008	0.009	–0.001	–0.007
	(0.015)	(0.013)	(0.009)	(0.015)	(0.014)	(0.006)
p-value, Any-V	.064	.765	.008	.014	.102	.025
Panel B: Disaggregated specification						
VHL	–0.032	0.000	0.031	–0.027	0.022	0.005
	(0.020)	(0.016)	(0.012)	(0.018)	(0.018)	(0.007)
HL	0.007	–0.004	–0.004	0.006	0.013	–0.020
	(0.039)	(0.035)	(0.019)	(0.040)	(0.039)	(0.012)
V	–0.017	0.003	0.014	–0.042	0.040	0.002
	(0.038)	(0.032)	(0.021)	(0.038)	(0.036)	(0.015)
p-value, VHL = HL	.334	.910	.089	.407	.820	.053
p-value, V = C	.654	.920	.517	.273	.264	.911
Panel C: Summary information						
Control mean	0.700	0.209	0.091	0.707	0.241	0.052
# obs. in VHL	1,568	1,568	1,568	1,437	1,437	1,437
# obs. in HL	1,537	1,537	1,537	1,364	1,364	1,364
# obs. in V	1,539	1,539	1,539	1,385	1,385	1,385
# obs. in C	1,585	1,585	1,585	1,477	1,477	1,477

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are dummies for identifying as a member of the denomination indicated in the column title. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects. Standard errors clustered by community are in parentheses.

0.7 percentage point increase in affiliation with Iglesia Filipina Independiente (p = .044), which is in full communion with the Anglican Communion and can thus be thought of as a quasi-Protestant denomination. Thus, even though the V curriculum effect on the intensity of religiosity—which is what our primary religious outcomes mostly measure—dissipates at 30 months, its overall effect on religiosity may not be null. Table VII also shows that this shift in religious affiliation was already under way six months after Transform, although the decrease in Catholic

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

affiliation was only marginally statistically significant at that time ($p = .064$).[39]

Like in the 6-month survey, we analyze the 30-month effects on primary economic outcomes by comparing V against control in the disaggregated specification and estimating the effect of Any-V in the pooled specification. In the main analysis of the disaggregated specification, we do not consider the comparison between VHL and HL, as there was no statistically significant difference in religiosity between these two experimental cells at six months.

Table VI shows that in the disaggregated specification, V households perceive their relative economic status to be 0.34 points higher ($q = .019$) on a 10-point scale than control households (this has the opposite sign from the six-month point estimate, with the p-value of the difference between the effects across surveys being .002). In addition, their monthly consumption is 481 PHP $\approx$ US\$9.6 higher than control households, a 7.5% increase that is marginally statistically significant ($q = .062$, p-value of difference versus six-month effect $= .005$).[40] This higher consumption appears to be supported by monthly income that is 501 PHP $\approx$ US\$10.0, or 6.1%, higher than control households, although this income effect is estimated with a great deal of noise and is not statistically significant ($q = .375$).[41] Online Appendix Table 83 shows that if we estimated the V effect on log income instead, we would get a statistically significant 0.11 log point increase ($q = .027$).

39. Nearly all of the remaining increase in "other" religious affiliation at 30 months is accounted for by a 0.6 percentage point increase in affiliation with Iglesia Ni Cristo ($p = .166$), a nontrinitarian Christian sect that denies the deity of Jesus and the Holy Spirit. At six months, the Any-V treatment effects are a 1.2 percentage point increase for Iglesia Filipina Independiente ($p = .008$) and a 0.5 percentage point increase for Iglesia Ni Cristo ($p = .150$).

40. If we excluded consumption categories that were not measured at 6 months, the 30-month effect of V on consumption would be 378 PHP ($q = .117$).

41. The control group's average monthly income at 30 months is 9,707 PHP, which is much higher than the 4,213 PHP we measured at 6 months. The control group's average monthly consumption level also grew from 5,001 PHP to 6,378 PHP. Although some of this growth may be due to changes in how we measured income and consumption between surveys, at least some of it is likely to reflect real economic improvements. Food security was measured in a consistent way across surveys; the fraction of control households that reported that no household member has gone hungry in the past 6 months rose from 82% at 6 months to 94% at 30 months (Online Appendix Tables 17 and 59). This improvement is probably due largely to regression to the mean, as households were selected for being among the poorest 30 in their community before Transform. In addition, Filipino GDP per capita grew 22% from 2015 to 2018.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

In contrast, in the pooled specification, none of the primary economic outcomes has statistically significant Any-V effects. (The change between the 6- and 30-month effects for income and perceived relative economic status has $p = .012$ and .0004, respectively.) The difference in this pattern of results relative to the disaggregated specification comes from VHL households being generally worse off at 30 months than HL households, even though the point estimates of most of the VHL treatment effects relative to the control group are positive. The HL group, which showed little indication of being better off than the control group at 6 months, is doing substantially better at 30 months. In tests that do not correct for multiple comparisons, the HL group has statistically significantly higher income (842 PHP $\approx$ US\$16.8, 10.3% greater than the control mean, $p = .033$), adult labor supply (4.3 hours a week, 6.3% greater than the control mean, $p = .045$), and perceived relative economic status (0.23 points relative to a control mean of 3.66 on a 10-point scale, $p = .033$). It is not obvious why the V curriculum would have a negative marginal effect when combined with the HL curriculum in the long run but not the short run.

Table VIII presents results for potential mechanisms. We no longer see a positive treatment effect of the V curriculum on grit, although this is not statistically distinguishable from the six-month effect. Unlike at 6 months, there is no statistically significant treatment effect on the “perseverance of effort” subscale of grit (p-value of Any-V effect = .982; p-value of change in Any-V effect from 6 to 30 months = .003), and the “consistency of interests” subscale continues to have no statistically significant treatment effect (p-value of Any-V effect = .724; p-value of change in Any-V effect from 6 to 30 months = .364). There is, however, a statistically significant increase in optimism in both the disaggregated specification (0.12 standard deviations, $p = .029$) and the pooled specification (0.05 standard deviations, $p = .034$), which is driven equally by expectations of greater life satisfaction and expectations of higher relative economic status five years in the future (Online Appendix Table 69).[42] One objective of Transform is to increase hope in participants. Although an increase in optimism can be the result of improved circumstances, many scholars have argued that optimism, at least in moderate quantities, causes better outcomes through a motivational channel (e.g., Scheier and Carver 1985; Puri and Robinson 2007). We note that in the

42. These are unadjusted p-values. As discussed previously, we do not control for multiple testing when exploring possible mechanisms.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE VIII
MECHANISMS, 30-MONTH SURVEY

	Social capital			Locus of control		Optimism		
	Trust index (1)	Social safety net index (2)	Community activities index (3)	Powerful others index (4)	Locus of control index (5)	Expectations index (6)	Grit index (7)	Self-control index (8)
Panel A: Pooled specification								
Any-V	–0.021	0.038	–0.023	–0.047	–0.000	0.047	0.006	–0.014
	(0.023)	(0.019)	(0.023)	(0.024)	(0.021)	(0.022)	(0.020)	(0.018)
Any-HL	–0.027	0.032	–0.012	–0.004	0.007	0.016	–0.037	–0.017
	(0.022)	(0.019)	(0.023)	(0.024)	(0.021)	(0.022)	(0.020)	(0.018)
p-value, Any-V	.354	.046	.324	.047	.989	.034	.761	.458
p-value, Any-V 6 vs. 30 mo.	.376	.655	.093	.000	.205	.010	.215	.437
Panel B: Disaggregated specification								
VHL	–0.047	0.068	–0.035	–0.050	0.008	0.062	–0.031	–0.031
	(0.030)	(0.026)	(0.031)	(0.031)	(0.030)	(0.028)	(0.028)	(0.026)
HL	–0.083	0.035	–0.031	0.027	0.010	0.073	–0.053	–0.107
	(0.057)	(0.047)	(0.053)	(0.069)	(0.058)	(0.049)	(0.059)	(0.063)
V	–0.067	0.054	–0.042	–0.017	0.013	0.116	–0.006	–0.101
	(0.048)	(0.047)	(0.055)	(0.069)	(0.056)	(0.053)	(0.063)	(0.065)
p-value, VHL = HL	.528	.495	.939	.247	.979	.826	.694	.224
p-value, V = C	.167	.247	.437	.800	.810	.029	.921	.119
p-value, V 6 vs. 30 months	.443	.192	.260	.320	.189	.033	.562	.200

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE VIII
(CONTINUED)

	Social capital			Locus of control		Optimism		
	Trust index (1)	Social safety net index (2)	Community activities index (3)	Powerful others index (4)	Locus of control index (5)	Expectations index (6)	Grit index (7)	Self-control index (8)
Panel C: Summary information								
Control mean	0	0	0	0	0	0	0	0
Control standard deviation	1	1	1	1	1	1	1	1
# observations in VHL	1,441	1,441	1,441	1,441	1,441	1,440	1,441	1,441
# observations in HL	1,366	1,366	1,366	1,366	1,366	1,365	1,366	1,366
# observations in V	1,389	1,389	1,389	1,389	1,389	1,388	1,389	1,389
# observations in C	1,481	1,481	1,481	1,481	1,481	1,479	1,481	1,481

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. Indices have been coded so that more positive numbers are better. See Appendix Table A.1 and Section V.B for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015, and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects. When testing the null that the 6- and 30-month treatment effects are equal, we reestimate the 6-month treatment effects on the community activities index using a variable definition that is harmonized with the 30-month variable definition.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

six-month survey, where we had measured three different optimism scales, we estimated one significant positive Any-V effect and one marginally negative Any-V effect on optimism. In the 30-month survey, we have only one optimism scale. Therefore, this positive effect at 30 months should be interpreted with caution.

In addition, in the pooled specification only, we see a positive and statistically significant Any-V effect on social safety net strength (0.038 standard deviations, $p = .046$). This effect comes from an increased belief that the household could access 40 PHP or 1,000 PHP from outside the household for an urgent need (Online Appendix Table 65).

At six months, we had estimated a positive Any-V treatment effect on the powerful others index—the sense that God is in control of one's life. Table VIII shows that at 30 months, this treatment effect has reversed to become negative and statistically significant. This is in accord with the negative (albeit not statistically significant) effects of Any-V on the directly elicited religiosity measures at 30 months, reported in Table VI. Relatedly, among secondary outcomes in Table IX, the strongest Any-V treatment effect is an increase in stated belief in salvation by grace. We saw that at 6 months, increases in religiosity due to the V curriculum are associated with decreases in agreement with this doctrine, so increases in agreement with this doctrine at 30 months could be interpreted as a decrease in religiosity. However, Protestants express significantly more agreement with the doctrine than Catholics at both 6 months (0.13 standard deviations, $p = .000002$) and 30 months (0.21 standard deviations, $p < .000001$). In light of the increased self-identification with Protestantism caused by the V curriculum, it may be better to interpret greater agreement with the doctrine at 30 months as being the result of the V curriculum having its intended effect in the long run, although the contradictory results at 6 months make this interpretation uncertain.

The statistically significant Any-V treatment effects on other secondary outcomes in Table IX are mostly positive. There is a statistically significant positive effect on the non-list-randomized hygiene index (which in the 30-month survey only measured whether animals are kept in a stable separate from the house). There are marginally statistically significant positive effects on financial inclusion and the list-randomized hygiene outcome, which measures hand washing and treatment of drinking water, and a marginally statistically significant negative effect on the number of children

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE IX

SECONDARY OUTCOMES, 30-MONTH SURVEY

	Salvation by grace belief index (1)	Assets index (2)	Financial inclusion index (3)	Health index (4)	Hygiene index, non-list-random. (5)	Hygiene, list-random. (6)	House index (7)	Migration and re-mittance index (8)	No discord index (9)	No domestic violence, list-rand. (10)	Child labor supply (hours) (11)	# children enrolled in school (12)
Panel A: Pooled specification												
Any-V	0.085	0.013	0.039	–0.017	0.050	0.073	0.021	–0.021	–0.029	–0.042	0.0	–0.03
	(0.020)	(0.024)	(0.023)	(0.022)	(0.022)	(0.041)	(0.024)	(0.023)	(0.020)	(0.039)	(0.1)	(0.01)
Any-HL	0.009	0.018	0.057	–0.017	–0.008	0.019	0.037	0.028	–0.004	–0.059	–0.0	–0.02
	(0.020)	(0.024)	(0.023)	(0.022)	(0.022)	(0.041)	(0.024)	(0.023)	(0.020)	(0.039)	(0.1)	(0.01)
p-value, Any-V	.000	.590	.090	.452	.025	.079	.401	.352	.146	.287	.813	.062
p-value, Any-V 6 vs. 30 mo.	.000	.149	.514	.586	.892	.539	.689	.070	.872	.583	.359	.588
Panel B: Disaggregated specification												
VHL	0.093	0.031	0.096	– 0.034	0.042	0.091	0.058	0.006	– 0.034	– 0.100	– 0.0	– 0.04
	(0.028)	(0.039)	(0.033)	(0.032)	(0.033)	(0.058)	(0.034)	(0.031)	(0.027)	(0.053)	(0.2)	(0.02)
HL	– 0.014	0.065	0.075	– 0.022	0.038	0.002	0.086	0.046	– 0.032	– 0.078	0.3	0.02
	(0.054)	(0.062)	(0.059)	(0.046)	(0.063)	(0.055)	(0.058)	(0.043)	(0.046)	(0.056)	(0.3)	(0.03)
V	0.066	0.059	0.059	– 0.021	0.093	0.048	0.084	0.008	– 0.047	– 0.046	0.4	0.01
	(0.053)	(0.063)	(0.054)	(0.050)	(0.058)	(0.063)	(0.050)	(0.053)	(0.043)	(0.055)	(0.3)	(0.03)
p-value, VHL = HL	.047	.594	.749	.793	.948	.103	.644	.352	.965	.687	.317	.053
p-value, V = C	.215	.345	.273	.670	.108	.447	.097	.886	.282	.400	.219	.844
p-value, V 6 vs. 30 mo.	.056	.501	.240	.705	.475	.401	.408	.465	.979	.373	.557	.467

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE IX
(CONTINUED)

	Salvation by grace belief index (1)	Assets index (2)	Financial inclusion index (3)	Health index (4)	Hygiene index, non-list-random. (5)	Hygiene, list-random. (6)	House index (7)	Migration and re-mittance index (8)	No discord index (9)	No domestic violence, list-rand. (10)	Child labor supply (hours) (11)	# children enrolled in school (12)
Panel C: Summary information												
Control mean	0	0	0	0	0	0.405	0	0	0	0.939	1.1	1.68
Control standard deviation	1	1	1	1	1	–	1	1	1	–	8.2	1.33
# observations in VHL	1,441	1,441	1,441	1,441	1,441	1,441	1,441	1,441	1,441	1,441	1,441	1,441
# observations in HL	1,366	1,366	1,366	1,366	1,366	1,366	1,366	1,366	1,366	1,366	1,366	1,366
# observations in V	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,388	1,389	1,389	1,389
# observations in C	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481

Notes. Panels A and B show treatment effect estimates relative to control. The dependent variables are indicated in the column title. Indices have been coded so that more positive numbers are better. See Appendix Table A.1 and Section V.B for details on variable construction. In Panel A, "Any-V" refers to the "Values only" and "Values, Health, and Livelihood" treatment groups, and "Any-HL" refers to the "Health and Livelihood only" and "Values, Health, and Livelihood" treatment groups. Standard errors clustered by community are in parentheses. All regressions control for the respondent's gender, marital status, and education; the number of adults in the household; the number of children in the household; and the number of days between June 1, 2015 and the interview date. The regressions in Panel A and the regressions estimating the VHL effect in Panel B control for community-pair fixed effects. The regressions estimating the HL and V effects in Panel B control for ICM base fixed effects. When testing the null that the 6- and 30-month treatment effects are equal, we reestimate the 6-month treatment effects on the non-list-randomized hygiene index, the house index, and the migration and remittance index using variable definitions that are harmonized with the 30-month variable definitions.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

enrolled in school. In the disaggregated specification, the only marginally statistically significant effect is an improvement in the house quality index, which is driven by an increase in the prevalence of electricity being the primary source for lighting (Online Appendix Table 77).

V.D. Were No-V Communities Evangelized between the 6- and 30-Month Surveys?

One hypothesis for why the V curriculum effects on religious intensity disappeared at 30 months is that the pastors evangelized the No-V communities after the first Transform implementation.[43] To test whether this occurred, in October 2018 (about three years after the program), we surveyed 131 of the 160 pastors involved in the study. Each pastor was presented with 45 people's names sorted alphabetically: 15 Transform invitees from the community in which the pastor had taught the V curriculum, 15 potential Transform invitees from the community in which the pastor had identified potential invitees but which had not been selected to receive the V curriculum, and 15 from a placebo community that is far from where the pastor worked, randomly selected from the communities served by a different ICM base than the one associated with the pastor.

The survey prompt read, "We have a list of people you may have interacted with in a ministry context during and after the ICM Transform Values training that you led from February to June 2015, three years ago. We believe that some of these people participated in your Transform program, and some did not." To test whether there had been any evangelism of the control group, the survey asked whether each person in the list had ever

43. Consistent with this story, the directly measured religiosity variables are higher at 30 months than at 6 months in all the treatment cells. For example, among control group respondents who were interviewed at both times, intrinsic religiosity rises by 0.18 standard deviations, extrinsic religiosity rises by 0.34 standard deviations, and general religiosity rises by 0.20 standard deviations (where standard deviation is measured at six months over control group respondents who appear in both surveys). On the other hand, religiosity measured via list randomization is much lower at 30 months than at 6 months in all treatment cells. The proportion of 30-month respondents for whom list randomization was expected to successfully anonymize their response about the targeted sensitive statement (see note 16) is similar to the 6-month survey's proportions: 76% for commitment to Jesus, 82% for reading the Bible, 85% for water treatment, 84% for hand washing, and 90% for domestic abuse.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

participated in a Transform Values program with the pastor. To test whether this interaction occurred between the 6- and 30-month surveys, the survey asked whether this participation in the Transform program occurred in 2015 or after 2015. The survey also asked whether the pastor had interacted with the listed person in any ministry context (defined as "an occasion where spiritual matters were discussed, or an event sponsored by a religious ministry"), and if yes, whether that interaction happened in 2015, after 2015, or both in and after 2015.

If pastors evangelizing the control group explains the fading of the religiosity treatment effect from 6 months to 30 months, we would expect to see that pastors report more contact with people in their No-V community than in their placebo community and that a significant amount of the reported No-V contact occurred exclusively between the 6- and 30-month surveys.

Pastors report that 79% of actual Transform invitees, 58% of No-V individuals, and 25% of placebo group members participated in the Transform Values program. Similarly, pastors report having interacted in a ministry context with 65% of actual Transform participants, 46% of No-V individuals, and 19% of placebo group members (some pastors did not classify Transform as a "ministry context"). Thus, there is some reason to believe that part of the No-V group may have been treated, which would attenuate our estimated treatment effects. However, conditional on believing that an individual participated in Transform, pastors report that that participation happened in 2015 for 99% of the individuals. Among No-V individuals whom the pastor reports interacting with in a ministry context, only 2% of those interactions happened exclusively after 2015, 75% of them happened exclusively in 2015 or earlier, and the remaining 23% happened in both periods. Because the 6-month survey completed data collection in January 2016, we see little evidence that a significant portion of the No-V group was treated exclusively between the 6- and 30-month surveys.

If the reported evangelism of the control group is real, rather than attributable to recall error, then we would expect to see larger six-month religiosity treatment effects for pastors who recall ministering to relatively few people in their No-V group compared to their Any-V group. In fact, there is a slightly negative and not statistically significant (p = .447, .425, and .864 for intrinsic, extrinsic, and general religiosity, respectively) relationship between pastor-level religiosity treatment effects at six months and the difference between the fraction of Any-V and No-V group

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

members the pastor reports to have participated in Transform.[44] This suggests that the high fraction of No-V individuals reported to have been in Transform is due to recall error rather than noncompliance with the treatment assignment. (Pastors may recall more No-V individuals having been in Transform than placebo members because pastors did interact with No-V individuals when identifying potential Transform invitees.)

Although it is possible in principle that Transform participants' evangelization of No-V communities is responsible for the erosion of the estimated V curriculum effect, we believe that this is unlikely given the geographic distance between the communities and the fact that any evangelization effort in a No-V community would have been dispersed among those who were identified as potential Transform invitees (and hence were in our survey) and those who were not.

V.E. Discussion of 30-Month Results

The 30-month results provide a mixed message. There is reason to believe that the Values curriculum had an ongoing effect on religiosity. The six-month impact on intensity of religiosity dies down, but there is evidence of a shift away from Catholicism toward Protestantism. It is less clear that this change had ongoing economic effects. Although we see some evidence of positive consumption and perceived relative economic status differences, supported by an increase in income, the statistically significant consumption and relative economic status effects only appear in the disaggregated specification when comparing V to control, and the income effect is not statistically significant. There are also positive income, adult labor supply, and perceived relative economic status effects of the HL treatment relative to control. This generates a negative estimated marginal effect of the V curriculum when it is added to the HL curriculum, since there are low levels of well-being in the VHL group relative to the HL group. There is no obvious reason why such a strong negative interaction between the V and HL curricula would exist at 30 months but not 6 months. If one's prior belief put significant weight on the V and HL curricula having additive treatment effects, then the best estimate

44. The pastor-level treatment effect is estimated as the difference in mean religiosity between the pastor's Any-V and No-V community, with no further control variables. This analysis excludes 13 pastors who said that 10 or more of the 15 placebo names participated in Transform.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

of the V curriculum effect at 30 months would come from the pooled specification, which finds no significant economic effects.

VI. Ethical Considerations

As this is a study on the impact of religious outreach, a contentious topic, it seems appropriate to briefly discuss the ethics of the study. We make an important distinction between our study and the program we studied. Our discussion is about the decision to study Transform and the research protocols chosen, and not the ethics of the religious outreach itself.[45]

First, our study had no impact on the number of people who received the V curriculum during the study period, although it did change the identities of those who received the curriculum. In this case as with many others, we believe that random assignment of treatment is not only valuable for estimating causal impacts, but also a fair way of allocating a program that would go ahead even without the study. Additional ethical concerns may have been relevant had the research design required changing the number of people receiving the V curriculum relative to what would have happened otherwise. This was not the case in our design.

Second, there is equipoise: religious programming is pervasive throughout the world, often delivered in coordination with poverty alleviation programs, yet there is no consensus on its impacts. We were agnostic about what the findings would be. Indeed, we believe the enormous scale of religious outreach creates an ethical obligation to study its impacts, as challenging and sensitive as the topic may be.

Third, we did not decide to conduct the study with the goal of generating a positive result for Transform or ICM. We chose to work with ICM because they were willing to randomize the inclusion of religious programming with their health and livelihoods curriculum. We were committed to publishing the results regardless of whether they were positive, null, mixed, or negative.

45. Our role as researchers consisted of suggesting that ICM remove the V curriculum from randomly chosen Transform implementations and deliver that V curriculum elsewhere; working with ICM to construct guidance to pastors on how to choose eligible communities; randomly assigning communities to experimental arms; securing funding for the no-religion HL treatment arm and our direct research expenses such as surveying costs; securing Institutional Review Board approval for human subjects research; designing and conducting surveys of potential participants; analyzing the data; interpreting the results; and writing this paper.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Finally, we alone are responsible for the way the paper is written. We believe we have presented the material in a dispassionate way and drawn appropriate conclusions, but would like to remind readers that using our results to extrapolate to other settings; to answer questions about Filipino history; or to predict impacts at a larger scale of operations, at longer time horizons, or on economic, social, or psychological variables that we did not measure all require great care.

VII. Conclusion

Our work demonstrates that a randomized controlled trial is a viable tool to study the effect of religiosity on social and economic outcomes. As with all program evaluations, our results are, strictly speaking, specific to the program and setting we study. Having said that, Transform's curriculum and dissemination method are similar to efforts by many religious organizations around the world, and evangelization of Catholics by evangelical Protestants is a widespread phenomenon (Pew Research Center 2014).

We find that increasing religiosity via a four-month Protestant pastor-led program increases income while decreasing perceived relative economic status in the short run. The effects on the intensity of religiosity dissipate 30 months after the program ends, but there is a shift in affiliation from Catholicism to Protestantism. There is mixed evidence on whether the positive economic effects of the curriculum persist to 30 months. When comparing those who received only the Protestant Christian theology, values, and character virtues curriculum against the no-treatment control group, we find that religious curriculum recipients have higher consumption and perceived relative economic status. But in a pooled specification that identifies the religious curriculum effect by comparing both the religious curriculum-only group against the no-treatment control and those who received the full religious, health, and livelihood skills curriculum against those who received only the health and livelihood skills curriculum, we find no statistically significant effects on primary economic outcomes. Although church-based programs may represent a method of increasing noncognitive skills and reducing poverty in the short run among adults in developing countries, more work is required to understand whether the effects can persist and if not, why not.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

APPENDIX

TABLE A.1

SIX-MONTH SURVEY OUTCOME VARIABLE CONSTRUCTION

Variable	Components	Subcomponents (if any)	Details	Possible answers
Panel A: Primary religious outcomes				
Religion intrinsic index	I enjoy thinking about my religion.		From Gorsuch and McPherson (1989). Index formed by adding together responses without first normalizing	1 Strongly disagree–5 Strongly agree
	It is important to me to spend time in private thought and prayer.			1 Strongly disagree–5 Strongly agree
	I have often had a strong sense of God's presence.			1 Strongly disagree–5 Strongly agree
	I try hard to live all my life according to my religious beliefs.			1 Strongly disagree–5 Strongly agree
	My whole approach to life is based on religion.			1 Strongly disagree–5 Strongly agree
	Although I am religious, I don't let it affect my daily life.		This question not used in our main analysis	1 Strongly agree–5 Strongly disagree
	It doesn't much matter what I believe so long as I am good.		This question not used in our main analysis	1 Strongly agree–5 Strongly disagree
	Although I believe in my religion, many other things are more important in life.		This question not used in our main analysis	1 Strongly agree–5 Strongly disagree
Religion extrinsic index	I go to religious services because it helps me to make friends.		From Gorsuch and McPherson (1989). Index formed by adding together responses without first normalizing	1 Strongly disagree–5 Strongly agree
	I pray mainly to gain relief and protection.			1 Strongly disagree–5 Strongly agree

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	What religion offers me most is comfort in times of trouble and sorrow.			1 Strongly disagree–5 Strongly agree
	Prayer is for peace and happiness.			1 Strongly disagree–5 Strongly agree
	I go to religious services mostly to spend time with my friends.			1 Strongly disagree–5 Strongly agree
	I go to religious services mainly because I enjoy seeing people there.			1 Strongly disagree–5 Strongly agree
General religion index	To what extent do you consider yourself a religious person?		From the Brief Multidimensional Measure of Religiousness/Spirituality (Fetzer Institute 1999)	1 Not religious at all–4 Very religious
	In the last month, have you tried to convince anyone else to change the way they think about God?		From ICM survey	No = 0, Yes = 1
	How many people [have you tried to convince]?		Adapted from ICM survey	Integer $\geqslant 0$
	How often do you go to religious services?			Daily = 365, More than once a week = 104, Once a week = 52, Once or twice a month = 18, Every month or so = 9, Once or twice a year = 1.5, Never = 0.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	In how many of the past seven days did you pray privately in places other than at a place of worship?			Integer 0–7
	How satisfied are you with your spiritual life right now?		From ICM survey	1 Not at all satisfied–5 Very satisfied
	The Bible is accurate in all that it teaches.		From ICM survey. These three responses are added together before standardizing, and then given triple weight when averaging the components to construct the general religion index. Asked only of Christians.	1 Strongly disagree–5 Strongly agree
	I believe the Bible has decisive authority over what I say and do.			1 Strongly disagree–5 Strongly agree
	I believe the Christian God—Father, Son, and Holy Spirit—is the only true God.			1 Strongly disagree–5 Strongly agree
Religion, list-randomized	I have made a personal commitment to Jesus Christ that is still important to me today.		Adapted from ICM survey. Both questions elicited using list randomization. Outcome variable is average of two responses.	False = 0, True = 1
	I have read or listened to the Bible in the past week.			False = 0, True = 1

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Panel B: Primary nonreligious outcomes				
Monthly consumption	Food consumption in the last week		Total amount spent in the last week on viand, rice/corn/beans/etc., bananas/cassava/potatoes/yams/starches/etc., fruits/vegetables, milk/eggs, nonalcoholic beverages. Multiplied by 30/7.	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Nonfood consumption in the last week		Total amount spent in the last week on alcoholic beverages, cigarettes, phone credit, transportation, clothing/shoes, soaps/cosmetics, gifts. Multiplied by 30/7.	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Average monthly celebration spending in last six months		Total amount spent on weddings, funerals, festivals, anniversaries, and birthdays in the last six months divided by 6	Amount in PHP (US$1 ≈ 45 PHP in 2015)
Food security index	No household member has gone to bed hungry in last six months		Constructed from question, “In the last six months, did you or any other person in this household ever go to bed hungry because there were not enough resources for food?”	No = 1; Yes = 0; Yes, but during lean season only = 0 [Lean season in the Philippines is usually July and August]

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	No household member has gone to bed hungry in last six months outside of lean season		Constructed from question, "In the last six months, did you or any other person in this household ever go to bed hungry because there were not enough resources for food?"	No = 1; Yes = 0; Yes, but during lean season only = 1 [Lean season in the Philippines is usually July and August]
	Number of days where no household member has gone to bed hungry in past seven days		Constructed as 7 minus the number of days a member of the household has gone to bed hungry in past seven days	Integer 0–7
Monthly income	Total household payments received for agricultural labor on behalf of non-household member		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total household payments received for formal employment		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total household payments received for housework		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total household payments received for tending animals in an outside household		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total household payments received for operating business that is not the household's		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	Total household payments for daily labor		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total household payments received for other work outside the household		Payments in the past 30 days	Amount in PHP (US$1 ≈ 45 PHP in 2015)
	Total profit from household businesses		In most recent month with normal sales	Amount in PHP (US$1 ≈ 45 PHP in 2015)
Adult weekly labor supply	Total hours spent in outside agricultural labor for nonhousehold member		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent in formal employment		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent doing housework in an outside household		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent tending animals in an outside household during past 7 days		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent operating business that is not the household's		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent on daily labor		During past 7 days, only household members age $\geqslant 17$	Integer
	Total hours spent on other work outside the household		During past 7 days, only household members age $\geqslant 17$	Integer

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Life satisfaction index	Kessler K6 nonspecific distress scale	About how often during the past 30 days did you feel nervous?	From Kessler et al. (2002). Index formed by adding together responses without first normalizing.	1 All of the time–5 None of the time
		About how often during the past 30 days did you feel hopeless?		1 All of the time–5 None of the time
		About how often during the past 30 days did you feel restless or fidgety?		1 All of the time–5 None of the time
		About how often during the past 30 days did you feel so depressed that nothing could you cheer you up?		1 All of the time–5 None of the time
		About how often during the past 30 days did you feel that everything was difficult?		1 All of the time–5 None of the time
		About how often during the past 30 days did you feel worthless?		1 All of the time–5 None of the time

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	Sum of four Gallup World Poll questions	Did you experience enjoyment during a lot of the day yesterday?		No = 0, Yes = 1
		Did you experience happiness during a lot of the day yesterday?		No = 0, Yes = 1
		Did you experience worry during a lot of the day yesterday?		No = 1, Yes = 0
		Did you experience sadness during a lot of the day yesterday?		No = 1, Yes = 0
	Did you smile or laugh a lot yesterday?		From Gallup World Poll	No = 0, Yes = 1
	How would you describe your satisfaction with life?		Elicited using Cantril's ladder	1 Very dissatisfied–10 Very satisfied
	Taking all things together, would you say you are . . .		From World Values Survey	1 Not at all happy–4 Very happy
Perceived relative economic status	Where would you place your household on the ladder in terms of economic status?		Elicited using Cantril's ladder	1 Poorest individuals of your community–10 Best-off members of your community

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Panel C: Mechanisms				
Trust index	In general, would you say that most people can be trusted or that most people cannot be trusted?			Most people can't be trusted = 0, Most people can be trusted = 1
	Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?		From World Values Survey	Try to take advantage of you = 0, Try to be fair = 1
	Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?		From General Social Survey	Looking out for themselves = 0, Try to be helpful = 1
Social safety net index	In the case where someone in your household did not have 40 PHP available for an urgent need, how likely is it that you could access this 40 PHP from a source outside your household?			1 Very unlikely–5 Very likely
	In the case where someone in your household did not have 1,000 PHP available for an urgent need, how likely is it that you could access this 1,000 PHP from a source outside your household?			1 Very unlikely–5 Very likely
	Do you discuss personal issues with anyone outside your close family?			No = 0, Yes = 1

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	How often do you usually speak to this person?			Daily = 365, A few times a week = 104, Weekly = 52, A few times a month = 24, Monthly = 12, Every month or so = 9, A few times a year = 6, Yearly = 1. If there is no such person, coded as 0.
	Did anyone from the household receive any meals from another household in your local community?			No = 0, Yes = 1
	How many meals [were received]?		Top-coded at 99th percentile	Integer
	Did this household give any meals to anybody from another household in your local community?			No = 0, Yes = 1
	How many meals [were given]?		Top-coded at 99th percentile	Integer
Community activities index	Did you attend any village leaders' meetings in the last six months?			No = 0, Yes = 1
	In the past six months, have you participated in any community activities?			No = 0, Yes = 1

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	How frequently did you participate in community activities?			Daily = 365, A few times a week = 104, Weekly = 52, A few times a month = 24, Monthly = 12, Every month or so = 9, A few times a year = 6, Yearly = 1. If the respondent did not participate, coded as 0.
Perceived stress scale index	How often have you felt that you were unable to control the important things in your life?		From Cohen, Kamarck, and Mermelstein (1983). Index formed by adding together responses without first normalizing.	1 Very Often–5 Never
	How often have you felt confident about your ability to handle your personal problems?			1 Never–5 Very Often
	How often have you felt that things were going your way?			1 Never–5 Very Often
	How often have you felt difficulties were piling up so high that you could not overcome them?			1 Very Often–5 Never

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Powerful others index	I feel like what happens in my life is mostly determined by God.		From Levenson (1981) Powerful Others scale, modified to apply to God's control of one's life. Index formed by adding together responses without first normalizing.	1 Strongly disagree–5 Strongly agree
	Although I might have good ability, I will not be successful without appealing to God.			1 Strongly disagree–5 Strongly agree
	My life is chiefly controlled by God.			1 Strongly disagree–5 Strongly agree
	Getting what I want requires pleasing God.			1 Strongly disagree–5 Strongly agree
	Whether or not I have an accident and hurt myself physically depends mostly on God.			1 Strongly disagree–5 Strongly agree
	In order to have my plans work, I make sure that they fit with God's plan for me.			1 Strongly disagree–5 Strongly agree
Locus of control index	Internality subscale	Whether or not I am successful depends mostly on my ability.	From Levenson (1981). Index formed by adding together responses without first normalizing.	1 Strongly disagree–5 Strongly agree
		Whether or not I have an accident and hurt myself depends mostly on how careful I am on a daily basis.		1 Strongly disagree–5 Strongly agree

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
		When I make plans, I am almost certain to make them work.		1 Strongly disagree–5 Strongly agree
		How many friends I have depends on how nice a person I am.		1 Strongly disagree–5 Strongly agree
		I can pretty much determine what will happen in my life.		1 Strongly disagree–5 Strongly agree
		I am usually able to protect my personal interests.		1 Strongly disagree–5 Strongly agree
		When I get what I want it's usually because I worked hard for it.		1 Strongly disagree–5 Strongly agree
		My life is determined by my own actions.		1 Strongly disagree–5 Strongly agree
	Chance subscale	To a great extent my life is controlled by accidental happenings.	From Levenson (1981). Index formed by adding together responses without first normalizing.	1 Strongly agree–5 Strongly disagree

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
		Often there is no chance of protecting my personal interests from bad luck happening.		1 Strongly agree–5 Strongly disagree
		When I get what I want, it is usually because I am lucky.		1 Strongly agree–5 Strongly disagree
		I have often found that what is going to happen will happen.		1 Strongly agree–5 Strongly disagree
		Whether or not I get into an accident and hurt myself physically is mostly a matter of luck.		1 Strongly agree–5 Strongly disagree
		It is not wise for me to plan too far ahead because many things turn out to be a matter of good or bad fortune.		1 Strongly agree–5 Strongly disagree

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
		Whether or not I am successful depends on whether I am lucky enough to be in the right place at the right time.		1 Strongly agree–5 Strongly disagree
		It is chiefly a matter of fate whether or not I have a few friends or many friends.		1 Strongly agree–5 Strongly disagree
	World Values Survey locus of control	Which comes closest to your view on a scale on which (1) means "everything in life is determined by fate" and (10) means "people shape their fate themselves"?	From World Values Survey	1 fate–10 people
Life orientation index	In uncertain times, I usually expect the best.		From the Life Orientation Test – Revised index by Scheier, Carver, and Bridges (1994). Index formed by adding together responses without first normalizing.	1 I disagree a lot–5 I agree a lot
	If something can go wrong for me, it will.			1 I agree a lot–5 I disagree a lot
	I'm always optimistic about my future.			1 I disagree a lot–5 I agree a lot

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	I hardly ever expect things to go my way.			1 I agree a lot–5 I disagree a lot
	I rarely count on good things happening to me.			1 I agree a lot–5 I disagree a lot
	Overall, I expect more good things to happen to me than bad.			1 I disagree a lot–5 I agree a lot
Expectations index	Which step [of the life satisfaction ladder] do you believe you will be on in five years?		Elicited using Cantril's ladder	1 Very dissatisfied–10 Very satisfied
	Where do you think you will be on this [relative economic status] ladder five years from now?		Elicited using Cantril's ladder	1 Poorest individuals–10 Best-off members
Optimism index	How optimistic are you in general, on a scale of 1 to 7?		From Scale Optimism-Pessimism-2 by Kemper et al. (2017).	1 Not at all optimistic–7 Very optimistic
	How pessimistic are you in general, on a scale of 1 to 7?		Pessimism scale shown to respondents had 1 be "not at all pessimistic" and 7 be "very pessimistic"	1 Very pessimistic–7 Not at all pessimistic

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Grit index	New ideas and projects sometimes distract me from previous ones.		From the Short Grit Scale (Duckworth and Quinn 2009). Index formed by adding together responses without first normalizing.	1 Very much like me–5 Not like me at all
	Setbacks don't discourage me.			1 Not like me at all–5 Very much like me
	I have been obsessed with a certain idea or project for a short time but later lost interest			1 Very much like me–5 Not like me at all
	I am a very hard worker.			1 Not like me at all–5 Very much like me
	I often set a goal but later choose to pursue a different one.			1 Very much like me–5 Not like me at all
	I have difficulty maintaining my focus on projects that take more than a few months.			1 Very much like me–5 Not like me at all
	I finish whatever I begin.			1 Not like me at all–5 Very much like me
	I am diligent.			1 Not like me at all–5 Very much like me

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Self-control index	I have a hard time breaking bad habits.		Subset of the Brief Self-Control Scale by Tangney, Baumeister, and Boone (2004). Index formed by adding together responses without first normalizing.	1 Very much like me–5 Not like me at all
	I get distracted easily.			1 Very much like me–5 Not like me at all
	I say inappropriate things.			1 Very much like me–5 Not like me at all
	I refuse things that are bad for me, even if they are fun.			1 Not like me at all–5 Very much like me
	People would say that I have very strong self-discipline.			1 Not like me at all–5 Very much like me
	Pleasure and fun sometimes keep me from getting work done.			1 Very much like me–5 Not like me at all
	I'm good at resisting temptation.			1 Not like me at all–5 Very much like me
	I do things that feel good in the moment but regret later on.			1 Very much like me–5 Not like me at all
	Sometimes I can't stop myself from doing something, even if I know it's wrong.			1 Very much like me–5 Not like me at all
	I often act without thinking through all the alternatives.			1 Very much like me–5 Not like me at all

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Panel D: Secondary outcomes				
Salvation by grace belief index	If I am good enough, God will cleanse me of my sins.		Question asked only of Christians	1 Strongly agree–5 Strongly disagree
	I follow God's laws so that I can go to heaven.		Question asked only of Christians	1 Strongly agree–5 Strongly disagree
	Which of the following best describes your belief about what happens after death?			There is no life after death = 0; I will go to heaven because I tried my best to be a good person and to live a good life = 0; I will go to heaven because I tried to be involved in my religion, pray, and live the way I think God wants me to = 0; I will go to hell = 0; I'm not sure if I will go to heaven or hell = 0; I will be reincarnated = 0; My belief is not well-described by any of these choices = 0; I will go to heaven because I have accepted Jesus Christ as my personal savior = 1

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Assets index	Chance that you, or someone in your household, would have 40 PHP available for your use in this circumstance of urgent need?			1 Very unlikely–5 Very likely
	Chance that you, or someone in your household, would have 1,000 PHP available for your use in this circumstance of urgent need?			1 Very unlikely–5 Very likely
	Number of productive assets acquired in past six months		Number of the following acquired in the past six months: tractors, sewing machines and farm tools. Top-coded at 99th percentile.	Integer $\geqslant 0$
	Value of the productive assets in the household acquired in the past six months		Sum of the amount paid for the above categories of assets. Top-coded at 99th percentile.	Value of assets in PHP (US\$1 $\approx$ 45 PHP in 2015)
	Number of house assets acquired in past six months		Number of the following acquired in the last six months: TV, VTR/VHS/VCD/DVD player, radio/transistor/stereo, electric fan, refrigerator/freezer, telephone/mobile phone, sala set, bicycle or pedicab, motorcab or motorcycle, boat, washing machine, chair/stool, bed or cot, table, watch or clock, jewelry, gas stove. Top-coded at 99th percentile.	Integer $\geqslant 0$

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	Value of the house assets acquired in the past six months		Sum of the amount paid for the above categories of assets. Top-coded at 99th percentile.	Value of assets in PHP (US$1 $\approx$ 45 PHP in 2015)
	Number of productive assets (level)		Number of tractors, sewing machines, and farm tools owned. Top-coded at 99th percentile.	Integer $\geqslant 0$
	Value of productive assets (level)		Sum of the amount paid for the above assets. Top-coded at 99th percentile.	Value of assets in PHP (US$1 $\approx$ 45 PHP in 2015)
	Number of house assets (level)		Number of the following owned: TV, VTR/VHS/VCD/DVD player, radio/transistor/stereo, electric fan, refrigerator/freezer, telephone/mobile phone, sala set, bicycle or pedicab, motorcab or motorcycle, boat, washing machine, chair/stool, bed or cot, table, watch or clock, jewelry, gas stove. Top-coded at 99th percentile.	Integer $\geqslant 0$
	Value of house assets (level)		Sum of the amount paid for the above assets. Top-coded at 99th percentile.	Value of assets in PHP (US$1 $\approx$ 45 PHP in 2015)

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	How much money do you have set aside in savings?			Amount in PHP (US$1 ≈ 45 PHP in 2015)
Financial inclusion index	Do you or anyone in your household currently have money set aside as savings?			No = 0, Yes = 1
	Do you—by yourself or with other people—currently have an account at a bank?			No = 0, Yes = 1
	Have you made a deposit at a financial institution in the past six months?			No = 0, Yes = 1
Health index	Number of serious health events in the household (past six months)		We top-code at the 99th percentile and multiply by −1	Integer
	Total number of workdays missed by household members due to illness in past 30 days		We top-code each household member at 30 days and multiply by −1	Integer
	Number of household members that have suffered an illness that has kept them from working (last 30 days)		We code this as the negative of the response	Integer
Hygiene index, non-list-randomized	Own or lease animals that are not kept in a separate stable			No = 1, Yes = 0
	At least one household member practices open defecation		Coded yes if primary latrine is forest, bushes, fields, bodies of water, hanging latrine, uncovered pit latrine, open pit	No = 1, Yes = 0

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
Hygiene, list-randomized	I treat my water before drinking it, for example by using solar disinfection, boiling it, or using a water filter		Both questions elicited using list randomization. Outcome variable is average of two components' responses	No = 0, Yes = 1
	I wash my hands after going to the bathroom.			No = 0, Yes = 1
House index	Are all rooms leak-free?			No = 0, Yes = 1
	Are at least some rooms leak-free?			No = 0, Yes = 1
	Are all rooms able to be safely locked?			No = 0, Yes = 1
	Are at least some rooms able to be safely locked?			No = 0, Yes = 1
	Primary source of energy for lighting is electricity.			No = 0, Yes = 1
	Primary latrine is inside the house.			No = 0, Yes = 1
Migration and remittance index	Number of migrators in the household		Number of household members who have slept outside the house for more than two consecutive nights for work in the past six months	Integer
	Number of days migrators in the household were gone in the past six months			Integer
	Number of migrators who sent remittances or brought money home to the household in the past six months			Integer

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(Continued)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	Household had at least one migrator who sent remittances or brought cash home in the past six months			No = 0, Yes = 1
	Amount received in remittances or cash brought home by household migrators in the past six months			Amount in PHP (US$1 ≈ 45 PHP in 2015)
No discord index	During the last one month, did you have any major arguments with your spouse or partner over spending on major household items or assets?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over saving decisions?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over the behavior and disciplining of children?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over interactions with relatives?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over alcohol consumption?			No = 1, Yes = 0

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

TABLE A.1
(CONTINUED)

Variable	Components	Subcomponents (if any)	Details	Possible answers
	During the last one month, did you have any major arguments with your spouse or partner over any other issues?			No = 1, Yes = 0
No domestic violence, list-randomized	Someone in my household is experiencing physical abuse.		Question elicited using list randomization.	No = 1, Yes = 0
Child labor supply	Total hours spent in outside agricultural labor for nonhousehold member		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent in formal employment		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent doing housework in an outside household		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent tending animals in an outside household during past seven days		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent operating business that is not the household's		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent on daily labor		During past 7 days, only household members age $\leqslant 16$	Integer
	Total hours spent on other work outside the household		During past 7 days, only household members age $\leqslant 16$	Integer
No. children enrolled in school			Age $\leqslant 16$	Integer

Notes. Unless indicated otherwise in the table, the variable in the first column is created by summing its components in the second column. Some components are made up of subcomponents, which are shown to the right of the components.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

London School of Economics, Innovations for Poverty Action, MIT Jameel Poverty Action Lab, and Centre for Economic Policy Research
Yale University and National Bureau of Economic Research
Northwestern University, Innovations for Poverty Action, MIT Jameel Poverty Action Lab, National Bureau of Economic Research, and Centre for Economic Policy Research

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

Data Availability

Data and code replicating the tables and figures in this article can be found in Bryan, Choi, and Karlan (2020), in the Harvard Dataverse, doi: 10.7910/DVN/RNGHDV.

References

Abadie, Alberto, Matthew M. Chingos, and Martin R. West, "Endogenous Stratification in Randomized Experiments," *Review of Economics and Statistics*, 100 (2018), 567–580.

Allport, Gordon W., and J. Michael Ross, "Personal Religious Orientation and Prejudice," *Journal of Personality and Social Psychology*, 5 (1967), 432–443.

Anderson, Michael, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103 (2008), 1481–1495.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348 (2015), 1260799.

Barro, Robert J., and Rachel M. McCleary, "Religion and Economic Growth across Countries," *American Sociological Review*, 68 (2003), 760–881.

Basten, Christoph, and Frank Betz, "Beyond Work Ethic: Religion, Individual, and Political Preferences," *American Economic Journal: Economic Policy*, 5 (2013), 67–91.

Becker, Sascha O., and Ludger Woessmann, "Was Weber Wrong? A Human Capital Theory of Protestant Economic History," *Quarterly Journal of Economics*, 124 (2009), 531–596.

Bell, Matthew, "Alpha Rising: The Slickest, Richest, Fastest-Growing Division of the Church of England," Spectator, November 30, 2013. https://life.spectator.co.uk/articles/alpha-rising/.

Benjamin, Daniel J., James J. Choi, and Geoffrey Fisher, "Religious Identity and Economic Behavior," *Review of Economics and Statistics*, 98 (2016), 617–637.

Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, 57 (1995), 289–300.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, “Adaptive Linear Step-up Procedures That Control the False Discovery Rate,” *Biometrika*, 93 (2006), 491–507.

Blattman, Christopher, Julian Jamison, and Margaret Sheridan, “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia,” *American Economic Review*, 107 (2017), 1165–1206.

Bottan, Nicolas L., and Ricardo Perez-Truglia, “Losing My Religion: The Effects of Religious Scandals on Religious Participation and Charitable Giving,” *Journal of Public Economics*, 129 (2015), 106–119.

Bryan, Gharad, James J. Choi, and Dean Karlan, “Replication Data for: ‘Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes’,” (2020), Harvard Dataverse, doi: 10.7910/DVN/RNGHDV.

Cantoni, Davide, “The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands,” *Journal of the European Economic Association*, 13 (2015), 561–598.

Chen, Daniel L., “Club Goods and Group Identity: Evidence from Islamic Resurgence during the Indonesian Financial Crisis,” *Journal of Political Economy*, 118 (2010), 300–354.

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban, “Sex, Lies, and Measurement: Do Indirect Response Survey Methods Work?,” Stanford University Working Paper, 2020.

Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer, “Estimating the Impact of the Hajj: Religion and Tolerance in Islam’s Global Gathering,” *Quarterly Journal of Economics*, 124 (2009), 1133–1170.

Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson, “The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated,” *Management Science*, 63 (2017), 3168–3186.

Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein, “A Global Measure of Perceived Stress,” *Journal of Health and Social Behavior*, 24 (1983), 385–396.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati, “The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application,” in *Measurement Errors in Surveys*, 185–210, Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Seymour Sudman, eds. (New York: Wiley, 1991).

Duckworth, Angela Lee, Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly, “Grit: Perseverance and Passion for Long-Term Goals,” *Journal of Personality and Social Psychology*, 92 (2007), 1087–1101.

Duckworth, Angela Lee, and Patrick D. Quinn, “Development and Validation of the Short Grit Scale (Grit–S),” *Journal of Personality Assessment*, 91 (2009), 166–174.

Efron, Bradley, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Cambridge: Cambridge University Press, 2013).

Ellison, Christopher G., “Religious Involvement and Subjective Well-Being,” *Journal of Health and Social Behavior*, 32 (1991), 80–99.

Fetzer Institute, *Multidimensional Measurement of Religiousness / Spirituality for Use in Health Research* (Kalamazoo, MI: John E. Fetzer Institute, 1999).

Freeman, Richard B., “Who Escapes? The Relation of Churchgoing and Other Background Factors to the Socioeconomic Performance of Black Male Youth from Inner-City Tracts,” in *The Black Youth Employment Crisis*, 353–376, Richard B. Freeman and Harry J. Holzer, eds. (Chicago: University of Chicago Press, 1986).

Goeman, Jelle J., and Aldo Solari, “Multiple Hypothesis Testing in Genomics,” *Statistics in Medicine*, 33 (2014), 1946–1978.

Gorsuch, Richard L., and Susan E. McPherson, “Intrinsic/Extrinsic Measurement: I/E-Revised and Single-Item Scales,” *Journal for the Scientific Study of Religion*, 28 (1989), 348–354.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Gruber, Jonathan, "Religious Market Structure, Religious Participation, and Outcomes: Is Religion Good for You?," *B.E. Journal of Economic Analysis & Policy*, 5 (2005).

Gruber, Jonathan, and Daniel Hungerman, "The Church vs. the Mall: What Happens When Religion Faces Increased Secular Competition?," *Quarterly Journal of Economics*, 123 (2008), 831–862.

Hackett, Conrad, and Brian J. Grim, *Global Christianity: A Report on the Size and Distribution of the World's Christian Population* (Washington, DC: Pew Research Center, 2011).

Hilary, Gilles, and Kai Wai Hui, "Does Religion Matter in Corporate Decision Making in America?," *Journal of Financial Economics*, 93 (2009), 455–473.

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6 (1979), 65–70.

Horton, John J., David G. Rand, and Richard J. Zeckhauser, "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics*, 14 (2011), 399–425.

Iannaccone, Laurence R., "Introduction to the Economics of Religion," *Journal of Economic Literature*, 36 (1998), 1465–1495.

Iyer, Sriya, "The New Economics of Religion," *Journal of Economic Literature*, 54 (2016), 395–441.

Johnson, Byron R., Ralph Brett Tompkins, and Derek Webb, *Objective Hope: Assessing the Effectiveness of Faith-Based Organizations: A Review of the Literature* (Waco, TX: Baylor University, 2008).

Karlan, Dean, and Jonathan Zinman, "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds," *Journal of Development Economics*, 98 (2012), 71–75.

Kautz, Tim, James Heckman, Ron Diris, Bas ter Weel, and Lex Borghans, "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success," NBER Working Paper no. 20749, 2014.

Kemper, Christoph J., Maria Wassermann, Annekatrin Hoppe, Constanze Beierlein, and Beatrice Rammstedt, "Measuring Dispositional Optimism in Large-Scale Studies: Psychometric Evidence for German, Spanish, and Italian Versions of the Scale Optimism-Pessimism-2 (SOP2)," *European Journal of Psychological Assessment*, 33 (2017), 403–408.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. T. Normand, E. E. Walters, and A. M. Zaslavsky, "Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress," *Psychological Medicine*, 32 (2002), 959–976.

Kirkpatrick, Lee A., and Ralph W. Hood, "Intrinsic-Extrinsic Religious Orientation: The Boon or Bane of Contemporary Psychology of Religion?," *Journal for the Scientific Study of Religion*, 29 (1990), 442–462.

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75 (2007), 83–120.

Levenson, Hanna, "Differentiating among Internality, Powerful Others, and Chance," in *Research with the Locus of Control Construct*, 15–63, Hebert M. Lefcourt, ed. (New York: Academic Press, 1981).

Mazar, Nina, On Amir, and Dan Ariely, "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research*, 45 (2008), 633–644.

Pew Research Center, "Religion in Latin America: Widespread Change in a Historically Catholic Region," Pew Research Center, 2014, http://www.pewforum.org/2014/11/13/religion-in-latin-america/.

———, "The Future of World Religions: Population Growth Projections, 2010–2050: Why Muslims Are Rising Fastest and the Unaffiliated Are Shrinking as a Share of the World's Population," Pew Research Center, 2015, https://www.pewforum.org/2015/04/02/religious-projections-2010-2050/.

Puri, Manju, and David Robinson, "Optimism and Economic Choice," *Journal of Financial Economics*, 86 (2007), 71–99.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, “Control of the False Discovery Rate under Dependence Using the Bootstrap and Subsampling,” *TEST*, 17 (2008), 417–442.

Samaritan’s Purse, “Along the Samaritan Road: Samaritan’s Purse 2016 Annual Report,” Samaritan’s Purse, 2017, https://s3.amazonaws.com/static.samaritanspurse.org/pdfs/ANNUAL_REPORT_web_download.pdf.

Scheier, Michael F., and Charles S. Carver, “Optimism, Coping, and Health: Assessment and Implications of Generalized Outcome Expectancies,” *Health Psychology*, 4 (1985), 219–247.

Scheier, Michael F., Charles S. Carver, and Michael W. Bridges, “Distinguishing Optimism from Neuroticism (and Trait Anxiety, Self-Mastery, and Self-Esteem): A Reevaluation of the Life Orientation Test,” *Journal of Personality and Social Psychology*, 67 (1994), 1063–1078.

Shariff, Azim F., and Ara Norenzayan, “God Is Watching You: Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game,” *Psychological Science*, 18 (2007), 803–809.

Shariff, Azim F., Aiyana K. Willard, Teresa Andersen, and Ara Norenzayan, “Religious Priming: A Meta-Analysis with a Focus on Prosociality,” *Personality and Social Psychology Review*, 20 (2016), 27–48.

Spenkuch, Jörg L., “Religion and Work: Micro Evidence from Contemporary Germany,” *Journal of Economic Behavior & Organization*, 135 (2017), 193–214.

Storey, John D., Jonathan E. Taylor, and David Siegmund, “Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66 (2004), 187–205.

Tangney, June P., Roy F. Baumeister, and Angie Luzio Boone, “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success,” *Journal of Personality*, 72 (2004), 271–324.

Weber, Max, *The Protestant Ethic and the Spirit of Capitalism* (New York: Scribner, 1958).

Woodberry, Robert D., “The Missionary Roots of Liberal Democracy,” *American Political Science Review*, 106 (2012), 244–274.

Downloaded from https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjaa023/5860785 by London School of Economics user on 24 August 2020