**Project Concept Note**

**Impact Evaluation of the Adolescent Girls Employment Initiative (AGEI)- NEPAL**

**20 January 2010**

1. **Motivation**

The Adolescent Girls Employment Initiative (AGEI) in Nepal is part of a Bank-led Adolescent Girls Initiative (AGI) to promote the economic empowerment of adolescent girls and young women in five low-income and post-conflict countries (Afghanistan, Liberia, Nepal, Rwanda, and South Sudan). All projects in the AGI offer training and mentorship to facilitate young women’s transition to work. They share a number of common characteristics, including modest budgets, designs to sustainably increase women’s economic participation, and rigorous impact evaluations. These programs are seen as pilots in which to promote new and innovative approaches, measure the results, and provide evidence for possible scaling-up of similar interventions in the future.

Because the evidence on what works in facilitating the transition of adolescent girls and young women to productive work is limited, impact evaluations are an essential part of this initiative. The impact evaluations of all AGI projects will rely on rigorous methods to measure the effects of the training and life skills programs on the socio-economic well-being of young women and their households. At the end of the three year initiative, a core set of indicators as well as lessons learned will have been developed across countries, providing valuable evidence as to what works, and for whom, for promoting the economic empowerment of young women.

1. **Background on the AGEI**

The AGEI in Nepal will build upon and strengthen an existing model for training and job placement currently being implemented by the Employment Fund (EF). The EF operates through private training providers that are considered “franchisees” of the EF. The EF screens proposals from private training providers on a rolling basis, approving and funding training courses as needed. Training providers must demonstrate that their proposed trades are in demand in particular markets through the use of rapid market assessments; hence, the EF does not prescribe the trades, the geographical locations, or the size of training courses. Training providers have strong incentives to identify employment opportunities for their trainees, as the EF makes full payment to these training providers only after 80% of the training participants have been employed for a minimum of six months. The choice of the EF as the implementing agency for the AGEI was based on its strong track record of implementing this unique model of job training and placement. The EF currently serves youth of both sexes aged 18-35.

The AGEI will adapt the current EF model to specifically target disadvantaged young women aged 16-24 who are constrained by social and economic barriers to enter the labor market. To facilitate their participation in the program and subsequent entry into the labor market, the project will develop and test a special outreach-and-communication strategy. Furthermore, the project will include an enhanced life skills curriculum to address the particular learning needs of young women in addition to providing the basic technical skill-oriented training that the EF already offers to all its trainees. This life skills curriculum will assist young women to build self-confidence, communication skills, and social networks of other young women who can provide support as they embark on their careers. The program will also provide instruction on reproductive health issues and STI prevention to help the trainees avoid risky behaviors.

By combining technical skills training, a life skills curriculum, and job placement focusing on disadvantaged women, and using a demand-driven process for identifying the trades, this program goes far beyond more traditional youth employment programs which focus narrowly on skills training in predetermined areas and tend to attract male participants. Given the unique and innovative nature of this model, along with the increased call for measuring results among the Bank and other donors, it will be essential to rigorously assess the AGEI program’s impacts in order to provide justification for scale-up and draw lessons learned for this and other contexts.

As with the other AGI projects, the impact evaluation of the AGEI will be designed and implemented through Bank-executed resources independently of the rest of the project. Local and international consultants will be contracted to implement the data collection effort. This will ensure that the evaluation is independent and also that the evaluation does not draw resources away from project implementation.

1. **Evaluation Questions and Indicators**

The goal of the evaluation is to identify the impact of the AGEI program on the subjective and objective well-being of program participants. The well-being of program participants will be measured through specific indicators of interest and statistically compared to a control group of young women with similar characteristics who have not participated in the program. In addition, a sample of students from the larger EF program, including males and students aged 25-35, will also be included in the impact evaluation, allowing for the examination of heterogeneous program impacts on different demographic groups.

The ***key research questions*** for the impact evaluation are:

What is the impact of a comprehensive employment program on the socioeconomic well-being of male and female participants aged 18-35?

What is the impact of a comprehensive employment program, combined with targeted life skills training, on the socioeconomic outcomes of young women aged 16-24?

What is the impact a comprehensive employment program on a wide range of social behaviors and outcomes, such as marriage and fertility, time use, control over household resources, aspirations for the future, and communication and negotiation skills?

What are the characteristics of people who succeed in this type of employment program, and what lessons can be learned for improving such programs in the future?

Data on the indicators of interest will be collected both before and after the program through individual interviews using structured questionnaires. Training providers will also be responsible for collecting data on some of these indicators (particularly those related to attendance and course completion). The key indicators of interest for the AGEI program include the following:

* At the level of the applicant / participant: Basic demographic and social characteristics, including age, living arrangements, and childhood circumstances; Employment and/or Entrepreneurship status *after* the 6 month period tracked in the project’s monitoring system; Job sector, wages, income; Knowledge and access to microcredit; Assets, Savings, Loans; Social assets (friends, social networks); Empowerment measures such as physical mobility, control over spending
* At the household level: Financial and physical assets of the household; Income-generating activities of all household members; Household-level self-assessed food security; Time allocation by household members to domestic tasks; Household expenditures on health and education; Educational enrollment of all household members, particularly school-age children
1. **Evaluation Methodology**

The proposed design for the impact evaluation is a **Regression Discontinuity (RD)**[[1]](#footnote-1)comparison, in which successful applicants just above a threshold entry score are compared with unsuccessful applicants just below the threshold. In this evaluation design, Training providers will identify and interview a set of applicants for each training course, assigning each applicant a numerical score. Because applicants with higher scores are different in observable and unobservable ways from those with lower scores, the entire pool of successful applicants cannot be used as the "treatment" group for this impact evaluation. Instead, successful applicants with scores just above the threshold will comprise the treatment group and will be compared to those with scores just below the threshold. This type of statistical comparison is frequently used to evaluate programs in which quantitative scoring criteria are used to determine eligibility.

This RD design will be integrated into the existing EF framework, allowing a flexible, decentralized, and rolling recruitment process. For each training course, two rounds of surveys (baseline and follow-up) will be conducted to collect data on both the treatment and control groups. The surveys will be conducted on a rolling basis as new training courses are scheduled. Baseline surveys will occur after beneficiary selection and before the start of training for each training course. Follow-up surveys will occur approximately 9 months after the completion of training. Specific aspects of the impact evaluation are discussed below.

1. ***Identification and selection of survey respondents***

In accordance with the current practices of the EF, training providers will have the freedom to select young people who apply and are accepted into the program by a selection committee. A standard procedure of recruitment and selection employed across the training providers will be important for establishing statistically valid comparison groups for the impact evaluation. This procedure will require each training provider contracted by the EF to screen and identify at least 50% more applicants than there are spaces, according to a standard set of eligibility criteria that will be developed in conjunction with the EF. Selection committees will score applicants along various dimensions, including one subjective sub-score for motivation level, or ‘commitment’. These scores will be then be aggregated into a single numerical score for each applicant, and applicants with the highest scores will be offered a space in the training course. For each training course, a *threshold score*, or minimum required score to gain entrance to the program, will be recorded. Individuals with scores just above and below this threshold will be asked to participate in the impact evaluation study. Their participation will be entirely voluntary and will not affect their eligibility for the training course. Note that applicants in the AGEI target population (women 16-24) will be considered a separate pool from other EF applicants, thus ensuring that AGEI participants will be adequately represented in the study population.

1. ***Sample size***

The sampling for this impact evaluation will be stratified into two groups: AGEI applicants and non-AGEI applicants. For each group, the training provider will provide a list of all individuals interviewed for the course along with each applicant's numerical score. The training provider will also indicate the minimum score that was required to gain entry to the course. From these lists, the evaluation will enroll applicants with scores just above the threshold entry score, who receive the training, and applicants just below the threshold score, who do not receive the training. Because it is not feasible for the EF to commit to screening at least two applicants for each training slot, the control group will be smaller than the treatment group. The EF anticipates that the applicant pool will be 150% of the size of the training course; that is, 2 out of every 3 applicants will gain admission on average.

It is impossible to know *a priori* how the scores of applicants will be distributed around the threshold entry score. Individuals with scores far away from the threshold, or cutoff score, will be given less weight in the regression analysis than those with scores very close to the threshold.[[2]](#footnote-2) Given that individuals with scores very far from the threshold will be given a weight of almost zero, the evaluation team proposes to collect information only on those individuals whose scores fall within a certain bandwidth around the cutoff score. The current budget of the impact evaluation allows for a sample size of 4000 individuals, which will be divided evenly into 4 groups: AGEI Treatment (i.e., AGEI women with scores above the cutoff), AGEI Control, Non-AGEI Treatment, and Non-AGEI Control. Since the pool of successful applicants is larger than the pool of rejected applicants, this implies that the evaluation will include approximately 25% of the successful AGEI applicants, and 50% of the rejected AGEI applicants. The pool of non-AGEI applicants is much larger, and hence the bandwidth around the threshold entry score for non-AGEI applicants will be much tighter. By creating treatment and control groups of equal size (rather than enrolling a set percentage of individuals from either side of the cutoff), this division of the sample into four equal groups will maximize the statistical power of the evaluation.

1. ***Statistical power***

Statistical power calculations have been performed using the Optimal Design software package using the sample sizes described above. Since outcomes will be correlated among trainees in the same courses, the power calculations must account for intra-cluster correlation. Since the specific number and size of clusters (i.e., training courses) will be determined on a rolling basis after the project begins, the calculations were done for a variety of cluster sizes, ranging from 200 courses of 20 respondents each to 80 courses of 50 respondents each. The results below depict the case in which the 4000 respondents are divided into 100 courses (20 Treatment and 20 Control). These 40 respondents are assumed to be split evenly into AGEI and non-AGEI samples.

Figure 1 represents the power of the proposed statistical test to detect changes in the primary outcome of interest: employment level. The graph shows that given the overall sample of 4000 respondents (2000 Treatment and 2000 Control), the evaluation is well-powered (at 90%) to detect an increase in the employment rate of at least 20%. Given that most applicants to the EF's training programs are currently unemployed (or engaged in agricultural self-employment), a change in employment rate of at least 20% for the treatment group (relative to control) is reasonable to expect and well below the EF's targets.

**Figure 1 Statistical Power for Impact Evaluation**



Figure 2 shows a comparable analysis for the power of the evaluation to detect impacts of the AGEI program specifically. Given the AGEI sample of 1000 Treatment and 1000 Control, split evenly across approximately 100 training courses, the analysis will be adequately powered (at 80%) to detect an increase in the employment rate of approximately 20%. Although less powered than the overall evaluation on all EF students, the very low baseline rate of non-agricultural employment for women in the target population (14% according to the 2006 Nepal Demographic and Health Survey) means that a 20% increase is still well below the project's targets.

Figure 2 Statistical Power for Impact Evaluation on AGEI sub-sample



1. ***Data Collection and Analysis***

This evaluation will consist of at least two rounds of surveys: baseline and follow-up. The survey firm contracted for the data collection will have the primary responsibility for the design and implementation of the survey, in close collaboration with the EF and the World Bank team. In addition, the evaluation will include some key-informant-type interviews with local community and business leaders in the selected sites in order to collect background data on the opportunities available in these communities, and the constraints to entering the labor market or establishing a business in the area.

* + **Baseline Questionnaire.** The instrument for the baseline survey will collect information on the young woman and the head of the household in which she lives. The survey instrument will include all indicators described in Section 3, including employment, education, assets, as well as subjective measures of attitudes, aspirations, and self-confidence. These latter questions will most likely be self-reported beliefs, but the survey may include modules to measure discount rates and other characteristics more quantitatively. It is essential that these baseline interviews *be conducted before the training program begins*. After the baseline survey has been completed for both the treatment and control groups of a particular training course, training can begin.
	+ **Follow-Up Questionnaire.** The follow-up survey instrument will resemble the baseline questionnaire as closely as possible. This will enable the comparison of changes in attitudes and behavior over time. In addition, the follow-up survey will include questions on the business and employment experiences of the young women since the completion of their training.

In addition to these more conventional surveys, the project may undertake to elicit the impact of the training on the participants’ ability to negotiate problems and work in groups, and on their levels of trust and leadership. There exist standard batteries of tests that are currently being used to evaluate the impact of similar programs on young women in other contexts.

Also in addition to the baseline and follow-up surveys, a longer period of follow-up study may be conducted for the study respondents from the first year of project implementation, subject to budget availability. The instruments for these additional follow-up surveys will resemble those from the baseline as much as possible. These additional surveys will allow the evaluation to provide some evidence of longer-term outcomes up to two years after the end of training. By following up multiple times with a smaller sample of respondents, the study can examine indicators that may take longer to evolve, such as changes in self-confidence and accrual of savings, and assess whether the effects of training attenuate or increase with time.

1. ***Timeline***

The AGEI project will be conducted over approximately 3 years, beginning in early 2010. In January 2010, the survey firm will design and test the survey questionnaires. The EF is scheduled to begin requesting proposals from training providers in early 2010, with recruitment of participants at each training site proceeding immediately after. The baseline survey at each training site, for both treatment and control groups, will be conducted *after* beneficiary selection ends and *before* training begins. Baseline surveys will continue over the 3 year project period as new training events are scheduled. At the end of 2010, a baseline report will summarize the characteristics of study respondents surveyed in the first year of project effectiveness.

Because the timing of training will vary across the T&E providers, and individual courses will be offered at different times, the baseline and follow-up surveys will be conducted at different times for each training event. The survey firm will bear primary responsibility for coordinating with the EF and the training providers to ensure that baseline surveys are conducted for each training event.

The follow-up survey will be scheduled nine months after the end of the training period. This is partly to give the intervention some time to work, and partly to gauge the status of the young women *after* the six months of employment mandated by the T&E provider’s contract. This longer period will also distinguish the indicators in the impact evaluation from those in the project’s monitoring framework, which end with the verification of employment and income at the six-month mark. The reason for not waiting longer to conduct the follow-up (for example, 12 months after training) is to increase the probability that the survey team will be able to find and follow-up with the respondents, and to ensure that follow-up occurs for as many respondents as possible before the project ends.

1. **Expected Outputs**

The Bank’s impact evaluation team will prepare at least four types of outputs. First, there will be reports which feed into the monitoring system of the project, to ensure alignment with the project’s results framework. Examples of this type of output include a baseline data report to examine the characteristics of participants in the first year of project implementation. Second, the evaluation will produce a set of analytical reports, to be prepared after follow-up data collection. These analytical reports will be shared with the EF, the T&E providers, and the Bank team to provide direct feedback into their program and operations. Third, the results of this evaluation along with the evaluations of the other 4 AGI projects in Liberia, Rwanda, Afghanistan, and will be consolidated to provide comparative, cross-country evidence on the successes and lessons learned of the overall Initiative. Fourth and finally, policy notes and academic articles will be prepared for dissemination to a wide audience, including policy-makers in Nepal, in order to share the findings with key stakeholders within and external to the Bank. These findings will also be presented at conferences and workshops, with the goal of influencing the design and implementation of future economic empowerment programs.

1. **Risks and Limitations**

*External Validity*: Given the AGEI’s size and national scope, the results from this evaluation will potentially validate an effective and replicable model of an innovative approach to women’s empowerment. Furthermore, by including individuals outside of the AGEI target population (girls aged 16-24), this evaluation can examine the relative effectiveness of the program on different demographic groups. Also, by combining the results from this evaluation with those from the other AGI countries, this effort is uniquely positioned to produce a set of policy recommendations to inform existing and future programs. However, while the evaluation can provide insight into what components of the program are most successful, it will not explicitly test the EF model against other training models, nor will it test the individual components (training, placement, etc.) separately.

*Risks*: The proposed design of this evaluation will need to be flexible enough to adapt to the realities of project implementation. Once the program begins, the evaluation team will pay close attention to the numbers of trainees enrolled into the program, refusal rates among targeted evaluation participants (especially from the control group), the adherence among the training providers to the standardized selection criteria, and other implementation details. Sufficient numbers of individuals from the target population need to apply for the program to ensure construction of a sizeable control group. If sufficient numbers of women are not enrolled in the evaluation study, especially in the control group, various measures can be taken: advertising and recruitment efforts can be bolstered, and the training providers can receive further training on how to screen applicants according to the standard criteria. If the RD design proves unworkable, because of lack of adherence to the strict scoring criteria, the evaluation can instead employ propensity score matching techniques, which will be made easier with the detailed scoring information collected from training providers and the data gathered at baseline.

*Ethical Considerations*: The evaluation study will be conducted in line with international standards for ethical research. All potential respondents will be asked to provide voluntary informed consent before enrolling in the study. Her choice to enroll in the study or not will not affect her status in the AGEI. Further, any respondent can terminate her participation in the research study at any time and will be free to refuse to answer any question at any time. The respondents’ answers to survey questions will be kept confidential at all times and will only be reported in aggregate terms. Finally, survey enumerators will be trained thoroughly on the proper methods of engaging with respondents, respecting confidentiality, and eliciting truthful information, particularly on sensitive topics.

1. **Estimated Budget**

|  |  |  |
| --- | --- | --- |
| Activity | Timing | Estimated amount (USD) |
| First baseline survey | Feb-May 2010 |  |
| Second baseline survey | Feb -May 2011 |  |
| First follow-up survey | Feb-May 2011 |  |
| Third baseline survey | Feb- May 2012 |  |
| Second follow-up survey (of all 2011 respondents plus half of 2010 respondents) | Feb- May 2012 |  |
| Third follow-up survey (including all 2012 respondents, plus half of 2010 respondents) | Feb-May 2013 |  |
|  | TOTAL |  |

1. **Impact Evaluation Team**

The Impact Evaluation (IE) Team will consist of two Economists from the World Bank, a field coordinator (STC), and a survey firm (to be contracted). The IE team will work closely with the Bank supervision team and the EF so that the evaluation aligns with all aspects of the project implementation.

1. **References**

### J Hahn, P Todd, W Van der Klaauw. [Identification and estimation of treatment effects with a regression-discontinuity design](http://www.jstor.org/stable/2692190). Econometrica, Vol. 69, No. 1, 2001.

Imbens, G. and Lemieux, T. Regression discontinuity designs: A guide to practice. Journal of Econometrics. [Volume 142, Issue 2](http://www.sciencedirect.com/science?_ob=PublicationURL&_tockey=%23TOC%235940%232008%23998579997%23676734%23FLA%23&_cdi=5940&_pubType=J&view=c&_auth=y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=766df653a420e5fd657b84e3141dccb1), February 2008.

1. Background information and more details on regression discontinuity designs can be found in Hahn, Todd, Van der Klauw (2001) and Imbens and Lemieux (2008). [↑](#footnote-ref-1)
2. The precise weighting methodology will be developed during the data analysis phase of the evaluation. [↑](#footnote-ref-2)