The Impact of the Provision of Near Vision Glasses by the Vision for a Nation Foundation on Handicraft Weavers' Productivity in Rwanda*

Final Report

Paul Glewwe
University of Minnesota

Julie Schaffner
Tufts University

January 28, 2014

* We are grateful for the opportunity to partner with Vision for a Nation Foundation in carrying out this study and for the Foundation's financial support. We are also grateful to Sjoerd Hannema, Sebastian Ford and other Vision for a Nation staff for many valuable conversations and for support of many types. We would like to thank Rebekah Mierau and Nathaniel Dowling for excellent management of the field work, and the staff of Azizi Life for helping connect the research team with weaving cooperatives. Finally, we thank the 241 women who participated in the study.

Executive Summary

Near vision glasses are low-cost devices with great potential to improve vision and raise earnings for people with poor vision, but very few vision-impaired people living in developing countries have glasses. The aim of this study is to quantify the effects of poor vision, and of the distribution of glasses to people with poor vision, on productivity. More specifically, it examines the impact of presbyopia or farsightedness (the most common near vision problem), and of the distribution of corrective glasses, on the speed and quality of production by farsighted Rwandan women who produce woven handicraft items for sale. Current and former members of handicraft weaving cooperatives were invited to participate in baseline and follow-up study sessions, during which they wove handicraft items (earring sets) for pay under conditions of minimal distraction, while researchers measured the speed and quality of their work. Of the eight groups of 30 weavers each (all women) that participated, four groups were randomly assigned to receive vision testing and, if needed, near vision glasses, immediately after the baseline study session. Using baseline data only, and linear regression estimation, we find strong evidence that productivity is much lower for farsighted weavers than for weavers with good vision. Point estimates (which pass the test of statistical significance) indicate that farsighted women's speed was 13 percent lower. More important, the share of their work rated as medium or high quality was 25 percentage points lower. By a measure that takes into account both speed and quality (the Rwandan francs they would earn per minute if paid at quality-differentiated market prices), their income generating capacity was 26 percent lower. Using both qualitative and quantitative methods, this study attempted to go further and assess the effect on farsighted women's productivity of receiving glasses. According to qualitative data gathered during the follow-up study sessions, of the 47 women who received glasses immediately after the baseline sessions, 46 reported that the glasses allowed them to weave more quickly or more accurately, and the great majority of them (41 of 47) reported their willingness to pay at least 2,000 Rwandan francs (Rwf) to replace their glasses. This is more than the 1,300 Rwf cost of a pair of near vision glasses. Moreover, nearly half of them were willing to pay much more (at least 10,000 Rwf). Quantitative evidence derived from the randomized control trial (RCT) design proved less useful than anticipated because, for reasons examined in the full report, the results suffer from statistical weakness and ambiguity in interpretation. It is likely that a follow-up study employing a larger number of study groups, and a compensation scheme that provides participants with stronger and more balanced motivation to pursue quality as well as quantity, would yield more clear-cut results. Despite the weaknesses of the RCT results, the study provides strong evidence that vision problems reduce the productivity of handicraft weavers in Rwanda, and thus further research is needed on the extent to which provision of glasses increases that productivity.

I. Introduction

In developing countries, most low income families obtain almost all of their income from working, and remain in poverty primarily because they can earn only low wages or low incomes in self-employment. In the long run, higher levels of education, and better health, are arguably the best pathways to reduce poverty for today's children and for future generations. Yet these long-term remedies will have little or no effect on individuals who are already adults. This raises the question of what can be done to increase the labor incomes of working adults.

One relatively simple intervention that has the potential to increase incomes for certain types of workers is the provision of glasses. Many adults in developing countries work in handicraft production, and many of them require adequate eyesight to produce handicraft items in sufficient quantity and quality. As these workers get older, their vision typically deteriorates. One common problem is presbyopia, which is more commonly called farsightedness. Older adults with presbyopia lose their ability to focus their vision on objects that are very near to them. Thus they lose their ability to read books and other documents and, more generally, experience a reduction in their ability to do any work for which they must focus on objects that are less than 1-2 meters from their eyes. Almost all types of presbyopia can be corrected with simple near vision glasses, yet in many developing countries the use of any kind of glasses is rare, especially among the poor.

This paper presents results from a randomized control trial (RCT) that investigates the impact on worker productivity of providing near vision (reading) glasses to adults who have developed presbyopia. More specifically, it focuses on Rwandan women in weaving cooperatives who make baskets, earrings, and other handicraft objects from grasses, reeds and other natural fibers. The study was carried out with the financial and logistical support of Vision for a Nation Foundation, a non-governmental organization that works with the Rwandan government to expand access to vision assessments and glasses.

II. Design and Implementation of the Evaluation

In the late summer of 2013, the research team invited eight groups of women, all current or former members of handicraft weaving cooperatives, to participate in a study of their weaving productivity.[1] Each group included approximately 30 women, and a total of 241 women participated. Women in four of the study groups were members of small weaving cooperatives affiliated with the non-governmental organization Azizi Life. Women in the other four groups were members of a single large weaving cooperative. Two groups from Azizi Life and two groups from the large cooperative were randomly selected to serve as treatment groups, while the remaining four groups (two from Azizi Life and two from the large cooperative) served as control groups.

All women participated in four research sessions, each six to eight hours long. In all four sessions women were paid to weave earring sets, which typically take between 45 and 150 minutes to produce. The first two sessions were "baseline" sessions, the purpose of which was to observe the weaving productivity of all women before any were provided with glasses. These were conducted from September 23 to October 1, 2013. Immediately after the baseline sessions, all women were offered vision assessments and were told that they would be given glasses if the assessments indicated that they needed them. All of the women accepted the offer of vision assessment. They were not told at what point in the study period they would receive the glasses.

[1] We included all interested cooperative members in the study, even though our primary interest was in women with presbyopia, in order to avoid generating jealousy or divisions within the cooperatives.

Women in the treatment groups who were found to have presbyopia were offered near vision glasses after the baseline sessions. All who were offered glasses accepted them.

Two weeks later, all but 2 of the original 241 women returned for another two days of "post-treatment" or "follow-up" observation, allowing for an assessment of whether those who were provided with glasses had become better weavers. After the follow-up sessions, the women in the control groups with presbyopia were offered free near vision glasses.

The research sessions were designed to measure the quality and quantity of the earrings these women produced under conditions of minimum distraction and reasonably strong incentives to work efficiently. This required providing women with motivation to pursue both speed and accuracy in their weaving, because weavers' earnings from the production of woven handicraft items such as the earrings in this study depend on both the quantity and quality of the items they produce. This study, and the markets in which the women participate, distinguish between three levels of quality for the earring sets. The highest quality is "fit for the export market" (henceforth high quality), the intermediate quality is "fit for the domestic market" (henceforth medium quality), and the lowest quality is "not fit for either market" (henceforth low quality). In local markets, women expect to earn 500 Rwandan Francs (Rwf) for medium quality earrings. If they can find buyers, they might expect 200Rwf for low quality earrings and 800Rwf for high quality earrings. Non-governmental organizations that buy woven goods from handicraft cooperatives (and then export them) reject low quality earrings, and pay 500Rwf for both medium and high quality earrings. (Women find it attractive to sell to the non-governmental organizations at this price, because the NGOs arrange bulk orders and provide a way for women to sell the earrings without having to engage in time consuming marketing activities.)

Ideally, study participants would have been paid prices for the earrings they produced that varied with quality in the same way as prices vary in the market. Logistical constraints, however, made it impossible to grade the quality of items on the same day they were produced. Because it was important to pay the women at the end of each day, the pay could not be linked to quality, and the women received a fixed rate of 500 Rwf per pair of earrings, regardless of quality. In order to provide women with an incentive to pursue quality as well as quantity, the women were told that any woman completing at least seven pairs of earrings of export quality over the course of the research would receive a certificate of excellence. While the study participants initially appeared enthusiastic about this possibility, and while working in public with their peers may have provided additional motivation to demonstrate skill, the study probably provided the women with significantly less motivation to pursue quality than they face in the normal course of this work.

In addition to collecting data for a quantitative analysis of productivity impacts, the study also gathered some qualitative information from the women in the treatment group who received glasses. This was done to learn about their use of, and satisfaction with, the glasses they received.

Of the 118 women in the treatment group who attended both baseline and follow-up sessions, eye exams given after the end of the baseline sessions indicated that 47 (about 40%) needed near vision glasses. Because presbyopia is most common among older women, assessments for presbyopia were done only for women 40 years old or older. Accordingly, all of the women diagnosed with presbyopia were at least 40 years old. Of the 61 women age 40 and older, 47 (that is 77%) needed near vision glasses. (All of the 22 women age 49 and older needed them.) Each of the 47 women who needed them were offered near vision glasses, and all of them accepted the offer. Eight of the women in the treatment group who needed near vision glasses already owned glasses prior to the study.

III. Productivity Effects of Poor Vision

This study offers strong evidence that poor vision greatly reduces productivity and potential earnings in handicraft weaving. Table 1 presents weaving productivity statistics in the baseline data. Four productivity measures are defined for each participant

> **Sets produced per day** = Total number of earring sets the participant would produce in an eight hour day if she worked at the average speed she exhibited during the study session. This is equal to the number of earring sets completed during the (baseline) research session, including fractional quantities for the last item, divided by the total number of minutes spent weaving, and then multiplied by 8*60=480.
>
> **Percent high quality** = percent of earring pairs produced during the (baseline) research session judged to be of export quality (including incomplete last items)
>
> **Percent medium or high quality** = percent of earring pairs produced during the (baseline) research session judged to be of domestic market or export quality (including incomplete last items)
>
> **Value per minute** = the number of Rwandan francs the participant would have earned per minute if she had been paid prevailing prices. This measure is equal to the total value of items produced divided by total time spent weaving, where the items were assigned the value of 800 if they were of high quality, 500 if they were of medium quality and 200 if they were of low quality. As with the other measures, it incorporates data on incomplete last items.

Table 1 demonstrates the great importance of good vision for productivity in handicraft weaving. On average, farsighted women produced 25 percent fewer earring sets than women who are not farsighted (4.7 versus 6.3 sets, respectively). The share of earrings that are of high quality is also 11 percentage points lower among farsighted women than among women with good vision (15% vs 26%), and the share of earrings that are of medium or high quality is 22 percentage points lower (40% vs. 62%). The mean value per minute measure, which takes into account both speed and quality, is 40 percent lower for farsighted women (3.7 vs 6.2). The differences are somewhat smaller when women under 40 (none of whom are farsighted) are excluded from the comparison, but they remain striking.[2]

A more accurate estimate of the impact of poor vision on weaving productivity can be obtained by regressing these measures of productivity on variables indicating presbyopia and on other factors that could affect productivity, such as age, using only the baseline data. Such regressions are shown in Table 2. The top half of Table 2 presents regressions of the four measures of productivity on a dummy variable indicating that the woman is farsighted, the woman's age, and a dummy variable indicating that the woman belongs to one of the Azizi Life cooperatives (rather than the large non-Azizi cooperative). The first productivity measure is the (average) number of earring sets that a women could make in an eight hour day, based on the speed of her weaving at baseline. The average woman who is not farsighted could make about 6.3 earring sets per eight hour day at baseline. The results indicate that presbyopia reduces this productivity by approximately 0.8 sets, a reduction of about 13 percent that is statistically significant at the 5 percent level. As one would expect, increased age also reduces this indicator of productivity, and women who belonged to an Azizi Life cooperative were also somewhat less productive. The second column indicates that the percentage of earring sets produced by farsighted women that were high quality sets declined by about six percentage points, but this is not statistically significant. However, when medium and high quality earring sets are combined, as shown in the third column, the decline due to presbyopia in the percentage of earring sets that

[2] The one difference when limiting the sample to woman 40 and older is that the percentage of earring sets that are high quality is slightly higher among women with hyperopia than among women without hyperopia. Yet it is still the case that the percentage of earring sets that are of medium or high quality is lower among women with hyperopia than among women without hyperopia.

are medium or high quality is quite large, about 25 percentage points, and highly statistically significant. Finally, the last column indicates that being farsighted reduces value per minute by 1.6Rwf, an effect that is both highly significant and very large, amounting to a reduction of 26% from the average value of 6.2 for all women in the sample who did not have presbyopia.

The bottom half of Table 2 distinguishes between four different categories of presbyopia. The least serious is the first level and the most serious is the fourth level. Of the farsighted women in our sample, the percentages diagnosed in each of these categories are 37, 33, 21 and 9 percent, respectively. All four levels of vision impairment have significant negative effects on the first measure of productivity (sets completed per eight-hour day), and the largest effects are for those with the highest degree of presbyopia; the impact of level-four presbyopia is twice as large as the average effect shown in the top half of the table.[3] The other three indicators of productivity yield similar results; the more farsighted a woman is the less productive she is in making earring sets and the negative impact on weaving productivity for women with level-four presbyopia is about twice as large as the average impact seen at the top of Table 2.

The effect of farsightedness on productivity and potential earnings is probably even greater than suggested by the results in Table 2. Many women who become farsighted as they age drop out of handicraft weaving entirely; and informal conversations suggest that potential earnings are further reduced for women with presbyopia who continue to weave, because they are less able than other women to weave into the twilight hours of the day.

[3] Probably because sample sizes are small, the differences in effects across levels of vision impairment are not precisely estimated. The null hypothesis that the four coefficients are the same cannot be rejected at the 10 percent level, except when the dependent variable is the percent of earring sets that are of medium or high quality.

IV. Qualitative Assessment of the Value of Distributing Glasses to Farsighted Women

After the baseline productivity session, farsighted women in the treatment group were given near vision glasses. When they returned two weeks later for the follow-up session, the research team administered a short questionnaire to obtain information about their experiences with the glasses.

Of the 47 women who received glasses, 46 reported that the glasses allow them to weave more quickly or more accurately, and also more comfortably. Forty-one reported that they would be willing to pay at least 2,000 Rwandan francs (about 3 US$) to replace them if they were lost, and this is more than the 1,300 Rwf cost of a pair of near vision glasses. Thirty would be willing to pay at least 5000 (about 7.5 US$), and 22 would be willing to pay at least 10,000 (about 15 US$). When asked what they liked best, most reported that the glasses helped them weave better, and many also reported that they liked being able to read better.

V. Quantitative Assessment of the Impact of Glasses Distribution on Weaving Productivity

The study was designed with the aim of obtaining high-quality estimates of the impact of glasses distribution on the productivity of farsighted handicraft weavers using a randomized control and simple estimation methods. In practice, however, complications that arose during the field work rendered the simplest estimation methods inappropriate and led to the use of more complicated methods. Unfortunately, they also rendered the results somewhat equivocal, because the more complicated estimation methods yield valid estimates only under certain assumptions (discussed below) that cannot be tested. The primary complication relates to increases in weaving speed and reductions in weaving quality observed among untreated as well as treated groups, and at different rates for the eight study groups. This pattern of change raises the possibility that participants realized during or after the baseline session that the reward for quality was low relative to the reward for quantity and accordingly increased their speed (at the risk of reducing quality). This shift was greater in some study groups than others, and not identical across treatment and control groups, which complicated the analysis. Many estimates were also statistically insignificant. Thus, while the results do not provide reason to doubt that the provision of glasses improved the productivity of handicraft weavers, neither do they provide strong evidence in favor of large impacts. Table 3 reports productivity statistics for the follow-up productivity sessions, while Table 4 reports the changes in the productivity statistics from the baseline to follow-up sessions. The focus of this section is on Table 4, and the regression estimates in Tables 5 and 6.

A first question one might ask of the change statistics in Table 4 is "What happened to the productivity of the farsighted women in the treatment group, all of whom received near vision glasses between the baseline and follow-up?" The statistics for this group in Table 4 suggest that the women who received glasses managed to weave faster (average sets produced per day rose by 1.9) while increasing quality (the percentage of earring sets that were high quality sets increased by 6.6 percentage points, and the percentage of earring sets that were medium or high quality increased by 7.8 percentage points). The net result is that the mean value per minute rose by 51 percent (from 3.7 to 5.6, an increase of 1.9Rwf per minute).

Caution must be exercised in interpreting these numbers, however, because several patterns in the data suggest that women's productivity changed over time not only because of receiving glasses, but also for other reasons. A first complicating pattern in the data suggests that the before-after change in productivity for the women who received glasses may overstate the impact of glasses on their productivity; productivity measured in terms of value per minute improved by a very similar amount (1.8 Rwf per minute) among farsighted women in the control group, who did not receive glasses. In fact, the numbers of items completed (in an eight-hour day) appeared to rise markedly between baseline and follow-up for all groups of women. It seems unlikely that this increase in speed was the result of learning, because most of the women in the study have been weaving earrings for many years. The more likely explanation is that women in all groups experienced a change in motivation that caused them to pursue greater speed during the second session than they pursued during the first session. While the increase in speed meant risking lower quality, the women may have (correctly) perceived little penalty for letting quality decline, because maintaining quality would yield at most only a certificate, not an increase in payment. They also received no feedback on the quality of the earring sets they produced during the baseline session prior to the follow-up session. The participants' greater comfort with the research setting may also have contributed to an increase in speed. In all cases the increases in speed were sufficiently large relative to the reductions in quality that value per minute measures increased.

A second complicating pattern in the data probably counteracts the first to some extent. Among women without presbyopia (none of whom experienced any vision change between baseline and follow-up), value per minute rose less in the treatment group (1.0) than in the control group (2.0). For some reason not yet fully understood, the shift toward greater speed and lower quality seems to have been stronger in the treatment groups than in the control groups, as evidenced by productivity changes for women who are not farsighted. Group psychology may have played a role in determining how strongly motivation shifted between baseline and follow-up, leading to productivity changes of different degrees in different study groups. If the increased pursuit of speed (at cost of lower quality) that was relatively strong among women with good vision in the treatment group (compared to women with good vision in the control group) also extended to the farsighted women in the treatment group, then the simple comparisons of productivity changes between farsighted women in the treatment group and farsighted women in the control group are likely to be misleading. Such comparisons probably overstate the impact on speed, and probably understate the impact on quality, of providing presbyopic women in the treatment groups with glasses. Yet what matters most is the impact of providing glasses on women's incomes, as measured by the mean value per minute indicator. Simple comparisons of presbyopic women in the treatment and control groups suggest almost no benefit: value per minute increased by 1.8Rwf in the treatment group and 1.9Rwf in the control group. But notice in the last column of Table 4 that for the women without presbyopia, for some reason that is not clear, women in the control group had a much larger increase in value per minute (2.0) than women in the treatment group (1.0). If that "unknown effect" had the same differential impact on women with presbyopia in treatment and control groups, one would expect that the change in average value per minute would also be 1.0 higher for the control group relative to the treatment group among women with presbyopia, but instead it is lower by 0.1. This suggests that providing glasses increased value per minute by 1.1 for women in the treatment group with presbyopia. The remainder of this section develops an approach for estimating the impact of glasses distribution on the productivity of farsighted women that takes these complications into account.

It is useful to introduce notation for describing the various comparisons of interest in the impact assessment. The simplest, but not necessarily the most accurate, way to estimate the impact on weaving productivity of providing near vision glasses to those who needed them (i.e. those suffering from presbyopia), which we presented above, is to calculate how much productivity increased from baseline to follow-up among women with presbyopia who received glasses. This can be expressed as:

$$\text{Prod}_{\text{treat,after}} - \text{Prod}_{\text{treat,before}} \qquad (1)$$

where $\text{Prod}_{\text{treat,after}}$ is the productivity indicator (i.e., sets per day, percent high quality, or value per minute) of presbyopic women in the treatment group in the follow-up sessions and $\text{Prod}_{\text{treat,before}}$ is the productivity indicator of presbyopic women in the treatment group in the baseline sessions. (For simplicity, there is no subscript to indicate presbyopic women since only those women are compared; below another subscript will be introduced to distinguish between women with and without presbyopia.) Tables 1 and 3 indicate that, for the measure of sets completed, this before-after estimate of the impact of providing glasses would be 6.3 – 4.4 = 1.9. This is also shown as 1.9 in Table 4.

This simple comparison may be misleading because, as seen in Tables 1 and 3, women with presbyopia in the control group also experienced an increase in total items completed. Specifically, the number of items they completed increased from 5.0 to 7.1, an increase of 2.1 earring pairs, which is also seen in Table 4. This may reflect a shift in motivation toward the pursuit of greater speed, and perhaps also increased comfort with the study setting, as suggested above. Thus it is possible that much or all of the apparent increase in productivity based on equation (1) is not due to the provision of glasses but to something else.

If this were the only complication in the data, it would suggest that a better way to estimate the productivity impact of receiving glasses is to compare the change in productivity over time for presbyopic women who received glasses to the same change for presbyopic women who did not receive glasses (i.e. presbyopic women in the control group). This "double difference" comparison can be expressed as:

$$(\text{Prod}_{\text{treat,after}} - \text{Prod}_{\text{treat,before}}) - (\text{Prod}_{\text{control,after}} - \text{Prod}_{\text{control,before}}) \qquad (2)$$

The change in productivity over time for the treatment group (i.e. the difference in the first set of parentheses) includes changes driven by both the receipt of glasses and other factors, such as changing motivation. The change in productivity over time for the control group (i.e. the difference in the second set of parentheses) includes only the change over time driven by the other factors. If the changes in productivity over time driven by these other factors were the same in both groups, then the calculation expressed in (2) yields an estimate of the impact of receiving glasses.[4] For the measure of sets completed, this would yield an estimate of 1.9 – 2.1 = -0.2; this negative but very small estimate is clearly much different from the estimate produced in equation (1).

A visual depiction of this approach to estimating the program's impact is shown in Figure 1. The vertical distance marked BA (for before-after) indicates the full before-after increase in productivity for the treatment group, which is also the simplest possible estimate as given in equation (1). The solid line with positive slope indicates the observed increase in productivity from baseline to follow-up experienced by presbyopic women in the control group. The treatment group started with lower productivity than the control group at baseline; that initial

[4] If we let PI denote the productivity impact of receiving eyeglasses and M indicate the change over time arising out of other factors such as motivational changes, then the before-after change for the treatment group equals PI+M, while the before-after change for the control group equal M (assuming that the effect on productivity of motivational changes and other factors was the same size in both groups). Then the difference between the before-after change for the treatment group and the before-after change for the control group equals (PI+M)-M=PI, and is, therefore, an estimate of the productivity impact of receiving eyeglasses.

productivity level is given by the dot at the left end of the dashed line. The dashed line indicates what would have happened to the productivity of the treatment group had it experienced only the same change in productivity induced by other factors (such as changing motivation) that was experienced by the control group. Under the assumption that the change in productivity from these other factors was the same for both groups, the gap labeled PI (for productivity impact, or program impact) represents an additional improvement in productivity, over and above that driven by changing motivation and other factors, and thus represents the effect of receiving glasses.

While the double difference estimate in (2) accounts for one problem that might bias the estimate of equation (1), it still rests on an assumption one might question: that changes in other factors such as motivation caused treatment and control group women's productivities to rise by the same amount. (Visually, this is the assumption that the solid and dashed lines in Figure 1 have the same slope.) The typical change in motivation between baseline and follow-up sessions may have differed between treatment and control groups, however, because random differences in circumstances or the nature of conversations within each of the eight groups could have led to different changes in group psychology. Had we been working with a much larger number of study groups, the average of these differences might have been largely identical in the sets of treatment and control study groups, but with only four treatment groups and four control groups, the average changes could have been different by chance. It is also possible that the difference in motivation change between treatment and control groups arose because the very distribution of glasses in the treatment groups altered the group psychology in a way that the control groups did not experience.

In many analyses there is little one can do to avoid this potential problem, but in this case it is possible to check the reasonableness of the assumption underlying the estimation approach of equation (2) by comparing changes over time among women in the treatment and control groups who do not suffer from presbyopia. Such a comparison can be quite revealing. Of particular interest is that the average value per minute rose by only 1.0 Rwf among non-presbyopic women in the treatment group but rose by 2.0 Rwf among non-presbyopic women in the control group. This and other patterns of change among women without presbyopia suggest that the motivation-based productivity changes between baseline and follow-up were indeed different for treatment and control study groups. If women with and without presbyopia within the same study group experienced the same motivation-based tendencies for change in speed and quality, and more specifically if changes over time in average value per minute were higher in the control group than in the treatment group among women with presbyopia, then the difference in baseline-to-follow-up productivity changes between treatment and control group women with presbyopia (i.e. the double difference estimates) probably understates the effect of treatment.

This motivates interest in a third way of estimating the impact of receipt of glasses, called "triple difference" estimation. This method yields an unbiased and more precise estimate of program impact if, within each study group, women with and without presbyopia experienced the same motivation-based changes in speed and quality. It is calculated by calculating the double difference in equation (2) for women without presbyopia, and subtracting that from the double difference in equation (2) for women with presbyopia. This "triple difference" estimate of the impact of providing glasses can be expressed as:

$$(Prod_{tr,aft,BV} - Prod_{tr,bef,BV}) - (Prod_{con,aft,BV} - Prod_{con,bef,BV}) \quad (3)$$
$$- [(Prod_{tr,aft,GV} - Prod_{tr,bef,GV}) - (Prod_{con,aft,GV} - Prod_{con,bef,GV})]$$

where BV ("bad vision") refers to women with presbyopia and GV ("good vision") refers to women without presbyopia.

Figure 2 shows a visual depiction of the triple difference estimator in equation (3). The lowest point on the left side of the diagram indicates the baseline productivity level for presbyopic women in the treatment group. As before, the vertical distance BA indicates the overall before-to-after change in productivity for this group. The lowest solid line links the productivity levels in the two time periods for presbyopic women in the control group, and its slope thus indicates the change in productivity for women in that group. The upper of the two lowest dashed lines indicates what would have happened to productivity for presbyopic women in the treatment group had their productivity risen at the same rate as that of presbyopic women in the control group, so that the gap marked PI_{DD} represents the double difference estimate of impact (the same estimate depicted in Figure 1). The two solid lines at the top of the figure indicate productivity levels and changes for women with good vision (i.e. non-presbyopic women) in the control and treatment groups. The dashed line between them indicates what would have happened to productivity for non-presbyopic women in the treatment group if it had grown at the same rate as for non-presbyopic women in the control group. The gap A at the right end of the figure indicates how much less productivity rose for women in the treatment group than for women in the control group among non-presbyopic women. This represents a best guess regarding the extent to which the motivation-based changes between baseline and follow-up were different for the treatment groups relative to the control groups. If presbyopic women experienced the same motivation-based productivity changes between baseline and follow-up as the non-presbyopic women in their study groups, then productivity for presbyopic women in the treatment group would have grown along the bottom-most dashed line in the figure. That is, rather than growing at the rate associated with the upper dashed line (and the double-difference impact estimate), their productivity would have grown more slowly than that by the amount A. The difference between the follow-up productivity associated with the lower dashed line and the actual follow-up productivity for presbyopic women in the treatment group is the triple difference estimate of impact, which is denoted by PI_{TD}.

After calculating estimates of the impact of glasses based on either equation (2) or (3), one must determine whether those estimates are statistically significant. This can be easily done using regression estimation. Another advantage of using the regression method is that it allows the inclusion of controls for additional factors that might influence productivity and might differ across women or study groups. For the double difference estimate the regression equation (without additional regressors) is:

$$Prod_{P,T} = \alpha_{dd} + \beta_{dd}T + \gamma_{dd}F + \delta_{dd}T\times F + u_{dd} \qquad (4)$$

where T indicates random assignment to the treatment group (T = 0 for control group and = 1 for the treatment group), F is a time period variable that equals zero for baseline measurements and equals one for follow-up measurements, and the "dd" subscript indicates double difference estimates. While randomization and large samples would imply that $\beta_{dd} = 0$, this need not hold exactly in practice, and does not seem to hold in our case. The γ_{dd} coefficient measures an average change in productivity over time shared by both groups, δ_{dd} measures the impact of the program, and standard calculations allow assessment of the statistical significance of the estimate.

An analogous equation can be used to determine whether the triple difference estimate of the impact of glasses is statistically significant. The regression equation is:

$$Prod_{P,T,H} = \alpha_{td} + \beta_{td}T + \gamma_{td}F + \delta_{td}P + \lambda_{td}T\times F + \theta_{td}T\times P + \tau_{td}F\times P + \pi_{td}T\times F\times P + u_{td} \quad (5)$$

where P is a dummy variable that equals 1 if a women has presbyopia and 0 if she does not, and the "td" subscript is for triple difference. The program impact is measured by π_{td}.

Table 5 presents estimates of the double difference specification in equation (2) for the same four indicators of weaving productivity shown in Tables 1 - 4. Note that the standard errors are adjusted for clustering at the group level (eight groups). This approach is important because the randomization of treatment was performed at the study group level; more generally, standard errors that are not clustered at the group level implicitly assume that the correlation in unobserved differences (i.e. the error terms) of two women in the same group is zero, i.e. that two such women have no unobserved characteristics in common, which is quite doubtful. Note that clustering tends to increase the size of standard errors and thus renders it more difficult to find statistically significant results, because unobserved similarities between women in the same group reduce the effective sample size. (In the extreme case, if a woman added to the sample is identical to one already in the sample, that woman provides no useful additional information for the purposes of estimation).

The first column in Table 5 focuses on the number of sets completed. The significantly positive coefficient on Follow-up indicates that, for presbyopic women in the control group, the average speed rose by 2.1 items per eight-hour day. The coefficient on Treatment indicates that, at baseline, the average speed was 0.6 baskets lower in the treatment group than in the control group, but this difference is not statistically significant (which one would expect given that the treatment was randomly assigned). Finally, the Follow-up × Treatment interaction term, is negative but small (-0.3) and statistically insignificant. If taken at face value, this would indicate that the provision of glasses did not increase the speed at which women made baskets.

The next two columns of Table 5 show the double difference estimates of the program's impact on the percentage of high quality baskets produced (column 2) and the percentage of medium or high quality baskets produced (column 3). Focusing on the Treatment × Follow-up interaction term (PT, the evidence indicates that providing glasses led to an 8 percentage point increase in the proportion of baskets that were of high quality, but this estimate is not statistically significant. The analogous estimate in column (3) shows a slightly negative but very small change in the percentage of baskets that are either medium or high quality.

The last column of Table 4 is the most important since it combines any changes in speed and quality into a single variable, the value per minute of basket production. If taken at face value, the interaction term would indicate that there is no evidence the program had any impact: the point estimate is positive but very small (0.04) and not even close to statistical significance. While these results reveal no evidence of impact, a prudent interpretation notes that the standard error is quite large (1.1) and so the 95% confidence interval includes a wide range of possible values for the impact, from -2.1 to 2.2 (not shown in Table 5). Given that the average value per minute for women with presbyopia in the treatment group was about 4.0, this is a wide confidence interval that does not rule out, for example, a large increase of 2.0 (a 50% increase).. Quite simply, the double difference estimate is too imprecise to detect large increases (or large decreases) in handicraft weaver productivity.

For reasons argued above, the triple difference estimates in Table 6 should do a better job of accounting for the changes in motivation that differed across study groups, if the presbyopic women in any one study group experienced the same changes in motivation as the nonpresbyopic women in the same study group. Note that the estimate of the impact is in the Follow-up×Treatment×Presbyopic row of the table.

According to the first column of Table 6, the provision of glasses does not seem to have increased the speed at which farsighted women produce earring sets; in fact, their speed is lower (by 1.1 items in an eight-hour day). In contrast, the point estimates do suggest that the provision of glasses leads to more high quality baskets being produced, the percent of high quality increasing by 12.6 percentage points, though this is also not statistically significant. Similarly, the third column indicates a 13.7 percentage point increase in baskets that are either of high or medium quality. This impact is also not statistically significant. Finally, the last column indicates that value per minute increased by 1.1, which is a large increase; while it is not statistically significant at conventional levels, it is significant at the 20% level (p-value of 0.181).[5]

VI. Analysis of Study Design

The study did not deliver the clean impact estimates we had hoped for when we designed it. It is, therefore, useful to analyze the choices we made regarding study design in light of what we now know, with an eye to suggesting cost-conscious modifications to the original study

[5] When the triple difference estimation is performed separately for Azizi Life cooperative members, on the one hand, and the members of the other cooperative, on the other, the patterns of productivity change and impact estimates differ markedly across the two subsamples. More specifically, impact estimates appeared much stronger in the non-Azizi cooperative than in the Azizi cooperatives, though small sample sizes render it difficult to draw strong conclusions about these differences. This provides an additional reason to conclude that while the results of this study are consistent with important impacts of glasses distribution on productivity, further research with an improved study design a and larger sample will be required for deriving definitive results.

design that would be most likely to produce more conclusive estimates. Two dimensions of the study design particularly merit attention. The first is the sample design, and especially the choices regarding the number of study groups to include and the choice to include all interested cooperative members, regardless of age. The second pertains to the design of study sessions, and more specifically to choices regarding participant payment schemes, choice of handicraft item, and oversight over study sites.

Consider first the sample design choices. The large sizes of many of the above standard errors suggest that the sample size and structure were not, in fact, adequate for producing sufficiently precise results. It is, therefore, useful to revisit the assumptions we made for the calculations that supported our choice of using eight groups of 30 participants each. We had to make assumptions about three critical parameters: the standard deviation of our productivity measures (across farsighted women before glasses distribution); the extent to which the values of a given productivity measure are correlated across women within study groups; and the percentage of study group attendees who are farsighted. For the first two parameter values we had to make educated guesses based only on conversations with the field staff of Azizi Life, because, to our knowledge, this is the first study to attempt productivity measurement for handicraft weavers and we could not draw on any pre-existing data on our productivity measures. We were able to draw on a better empirical base for choosing the third parameter value (regarding the fraction of participants who are farsighted), because presbyopia rates among women in various age groups are better understood.

Our educated guess regarding the standard deviation of our productivity measures across women turned out to be far too small. Even the measure with the smallest standard deviation (expressed as a percentage of the measure's mean), the number of items completed, had a standard deviation that was twice the maximum rate we used in our sample size calculations. The other measures had standard deviations from 4 to 9 times larger than the value we used in our calculations. Had we used a standard deviation value more in line with these values, we would have determined that we needed a much larger number of study groups (at least 10 times as many, as explained below). As discussed below, the measures of variability found in our study may be misleadingly large, however. It seems possible that these surprisingly large rates of variability were inflated by the motivation problems mentioned above; if some women were working to achieve a balance of quality and speed, while other women were focused more exclusively on high speed, this might have increased the degree of variation in speed across all women. An improved study design that induces all women to pursue a balance of speed and quality might reduce the relevant standard deviations.

Our choice of value for the parameter that quantifies the extent to which overall variation in productivity measures is driven by factors that are common to women within a study group but that differ across study groups (the intra-cluster correlation coefficient or ICC) was also too small, but adjusting this parameter value does not have as large an impact on required sample size as the adjustment to the standard deviation (as seen below).

In contrast, the choices of value for the percentage of study group participants who are farsighted was reasonably accurate.

When we re-calculate the adequate sample size taking into account the fact that the standard deviations in the distributions of outcome variables are higher than the values we had expected, and that the ICC was also higher than expected, the calculations indicate that the required number of study groups should have been increased by a factor of 10.[6] That is, the

[6] The formula used contains the ratio of the standard deviation divided by the square root of the number of groups. To attain a given level of precision, holding other factors constant, this ratio should not change. Thus if the standard

study would have required 80 groups rather than 8. This may be too pessimistic, however. As mentioned above, it is possible that adjustments to the details shaping what happens during study sessions might reduce the standard deviations of the productivity measures, rendering it possible to obtain statistically significant results with sample sizes smaller than a 10-fold increase (though still much more than only eight groups).

It is also possible that the cost of expanding the number of study groups could be mitigated somewhat by reducing the number of participants per study group. Our decision to use study groups of approximately 30 was influenced by one choice and one assumption. The choice was to invite as study participants any interested current and former cooperative members, regardless of their age or vision status. Our concern was that providing the potentially lucrative opportunity to participate in the study sessions to some members but not others might create divisions or jealousies within cooperatives. The implication was that we could only expect 30 or 40 percent of participants to be farsighted. We believe this choice is worth revisiting. In particular, we would suggest exploring the possibility of inviting only current and former cooperative members who are at least 35 years old to participate. (We also suggest that all of them be assessed for presbyopia, rather than assessing only those over 40 years old.) This would increase the fraction of women who are presbyopic to perhaps 60 or 70 percent of participants, reducing the size of study group required to obtain an adequate number of presbyopic women, while still allowing us the opportunity to compare productivities and productivity changes for prebyopic and non-presbyopic women. If this were done, the same number of presbyopic women could be obtained with groups of about 20 women, instead of 30 women. We suggest that qualitative research be undertaken to establish whether cooperative members would accept

deviation doubles the number of groups must increase by four, and if the standard deviation triples the number of groups must increase by nine.

without complaint or jealousy the limitation of study participation to women at least 35 years old.

The second set of important design choices involves: what exactly to ask the women to do; how to motivate them; and how to control the study environment. We believe that better choices in these areas could improve study results in two ways. First, if the women perceived stronger and more consistent incentives to strive for quality as well as quantity, this might have prevented the puzzling and diverse motivational changes – in particular the increase in speed at which the items were produced and the accompanying reduction in the quality of the items – that we believe introduced confounding changes in productivity measures over time, even for untreated women. Second, stronger incentives might also have reduced the variation in productivity measures across women (thereby reducing the sample size required for obtaining statistically significant results). We suggest that the following measures be undertaken to improve participants' incentives and reduce the variability of their productivity measures. First, and most obvious, we suggest re-working the logistics of the field work to make it possible to evaluate the quality of participants' work and pay them at quality-dependent rates at the end of each day. Second, we recommend a small pilot study employing one or more improved incentive schemes, to determine whether the incentives are effective for giving women interest in pursuing quality as well as quantity. Third, it may be worthwhile to ask women to produce more substantial handicraft items, such as baskets that take on the order of 4 hours to make, rather than the earring sets. It is possible that speed and quality would be less variable (relative to mean productivity) for such items; such a possibility could be tested during a small additional pilot measurement session. Fourth, we suggest that one of the two researchers be present during the pilot session and initiation of data collection. It is possible that we would have noticed details of context and operation that would have alerted us to problems and that we could have introduced modifications to the design that would have improved the outcomes.

In this section we have set out suggestions for next steps to take should Vision for a Nation wish to pursue an improved assessment of the productivity impact of glasses distribution for handicraft weavers. Before taking such steps it would be important to evaluate whether it would be feasible to find large enough numbers of handicrafts weavers who are willing to participate and have not already participated in the present study.

VII. Discussion and Conclusion

This study provides clear evidence that handicraft weaving productivity is much lower for presbyopic women without glasses than for non-presbyopic women. This suggests that providing presbyopic women with glasses has great potential to raise their productivity and income.

Because the receipt of glasses may not fully raise presbyopic women's productivity to the level of non-presbyopic women's productivity, we hoped also to provide direct estimates of the impact on presbyopic women's productivity of providing them with near vision glasses. Unfortunately, our estimates of such impacts are somewhat equivocal as the result of unforeseen challenges arising during the implementation of the evaluation. To be clear, our results should not be taken as evidence against strong impacts. The results are consistent with strong productivity impacts (especially in weaving quality), but the data allow only imprecise estimates and rest on assumptions that cannot be fully verified.

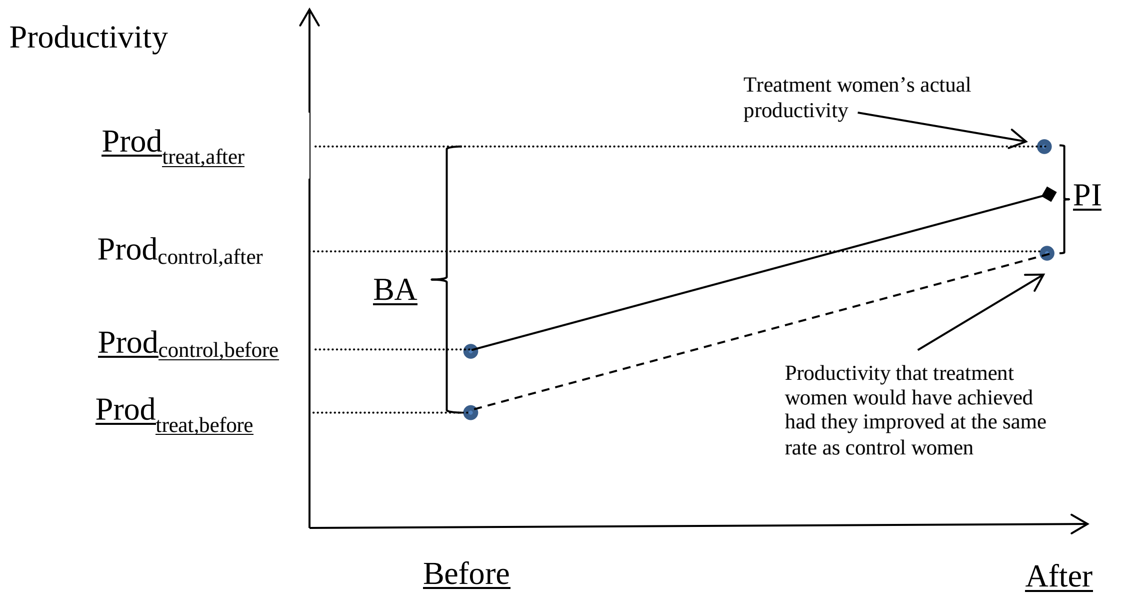

Figure 1: Double Difference Comparison of Presbyopic Women over 2 Time Periods

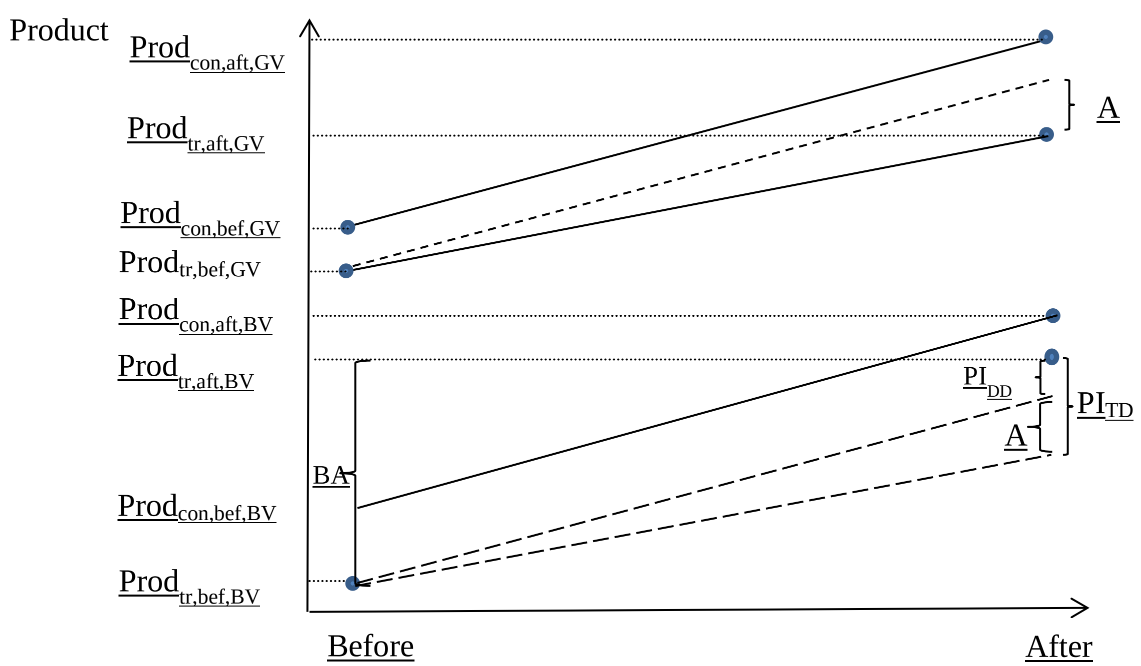

Figure 2: Triple Difference Comparison of Both Types of Women over 2 Time Periods

Table 1. Baseline Productivity Statistics

	Women with Presbyopia (N=89)				Women without Presbyopia (N=150)			
Sample	Sets Completed in 8 hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute	Sets Completed in 8 hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute
All (N=239)	4.7	14.7	40.1	3.7	6.3	25.5	61.6	6.2
Women at least 40 years old (N=114)	4.7	14.7	40.1	3.7	5.6	12.8	51.0	4.7
Women in treatment group (N=119)	4.4	15.6	44.5	3.7	6.2	19.0	57.5	5.7
Women in control group (N=120)	5.0	13.7	35.2	3.8	6.4	31.4	65.4	6.6

Table 2. Impact of Poor Vision on Weaving Productivity at Baseline

A. Farsighted vs. Not Farsighted

VARIABLES	(1) Sets Complete in 8 Hours	(2) Percent High	(3) Percent Medium	(4) Average Value per Minute
Presbyopic	-0.805**	-5.867	-24.74***	-1.610***
	(0.276)	(3.241)	(6.093)	(0.391)
Age	-0.0479**	-0.289	0.317	-0.0453
	(0.0193)	(0.291)	(0.503)	(0.0357)
Azizi	-0.681*	-9.539	-23.78	-1.710*
	(0.345)	(9.382)	(13.36)	(0.891)
Constant	8.217***	39.61**	62.30**	8.478***
	(0.714)	(14.26)	(21.54)	(1.560)
Observations	238	238	238	238
R-squared	0.311	0.062	0.148	0.237

B. Impact by Severity of Farsightedness

VARIABLES	(1) Sets Complete in 8 Hours	(2) Percent High	(3) Percent Medium	(4) Average Value per Minute
Level 1 Presbyopia	-0.787*	-6.584	-28.42***	-1.769**
	(0.395)	(4.460)	(8.029)	(0.610)
Level 2 Presbyopia	-0.732*	-3.439	-16.75	-1.140
	(0.333)	(6.943)	(14.05)	(0.712)
Level 3 Presbyopia	-1.387**	-11.91	-36.73***	-2.631**
	(0.405)	(8.327)	(5.707)	(0.801)
Level 4 Presbyopia	-1.590**	-14.85	-56.99***	-3.285**
	(0.522)	(10.14)	(10.70)	(0.961)
Age	-0.0371	-0.186	0.593	-0.0273
	(0.0197)	(0.421)	(0.567)	(0.0448)
Azizi	-0.716*	-9.987	-24.92	-1.790*
	(0.374)	(9.133)	(13.26)	(0.877)
Constant	7.874***	36.38*	53.67*	7.916***
	(0.734)	(18.23)	(23.60)	(1.845)
Observations	238	238	238	238
R-squared	0.323	0.067	0.176	0.254

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3. Follow-up Productivity Statistics

	Women with Presbyopia (N=88)				Women without Presbyopia (N=148)			
Sample	Sets Completed in 8 Hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute	Sets Completed in 8 Hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute
All (N=236)	6.7	18.1	48.3	5.6	8.8	23.9	53.5	7.7
Women at least 40 years old (N=111)	6.7	18.1	48.3	5.6	7.3	18.3	46.7	5.9
Women in treatment group (N=118)	6.3	22.2	52.4	5.6	9.1	15.0	41.9	6.6
Women in control group (N=118)	7.1	13.5	43.7	5.6	8.5	32.1	64.2	8.6

Table 4. Baseline to Follow-up Changes in Productivity Statistics

	Women with Presbyopia				Women without Presbyopia			
Sample	Sets Completed in 8 Hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute	Sets Completed in 8 Hours	Percent High Quality	Percent Medium or High Quality	Mean Value per Minute
All	2.0*	3.5	8.2	1.9*	2.5*	-1.6	-8.1*	1.5*
Women at least 40 years old	2.0*	3.5	8.2	1.9*	1.7*	5.5	-4.3	1.3
Women in treatment group	1.9*	6.6	7.8	1.9*	2.9*	-4.0	-15.5*	1.0*
Women in control group	2.1*	-0.2	8.5	1.8*	2.0*	0.7	-1.3	2.0*

* Change statistically significant at the 10 percent level.

Table 5.
Double Difference Estimates of Provision of Glasses on Weaving Productivity

VARIABLES	(1) Sets Complete in 8 Hours	(2) Percent High	(3) Percent Medium	(4) Average Value per Minute
Follow-up	2.091**	-0.235	8.284	1.792*
	(0.674)	(4.345)	(4.655)	(0.759)
Treatment	-0.593	1.526	10.72	-0.0721
	(0.568)	(10.87)	(15.58)	(1.065)
Follow-up x Treatment	-0.289	7.671	-1.013	0.0353
	(0.866)	(4.414)	(8.485)	(1.055)
Age	-0.0601**	-0.142	0.0261	-0.0541
	(0.0218)	(0.310)	(0.762)	(0.0449)
Azizi	-0.987**	-5.673	-17.18	-1.522
	(0.348)	(9.178)	(12.59)	(0.810)
Constant	8.491***	23.57	42.53	7.230**
	(1.442)	(19.22)	(36.38)	(2.528)
Observations	175	175	175	175
R-squared	0.349	0.027	0.065	0.175

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6.
Triple Difference Estimates of Provision of Glasses on Weaving Productivity

VARIABLES	(1) Sets Complete in 8 Hours	(2) Percent High	(3) Percent Medium	(4) Average Value per Minute
Follow-up	2.053**	0.742	-1.059	1.986**
	(0.860)	(4.753)	(2.618)	(0.809)
Treatment	-0.272	-13.00	-9.767	-1.078
	(0.500)	(10.10)	(13.99)	(0.977)
Presbyopic	-0.513	-18.80*	-40.91**	-2.573**
	(0.667)	(8.981)	(13.72)	(1.021)
Follow-up x Treatment	0.823	-4.815	-14.57	-1.034
	(1.108)	(8.198)	(8.919)	(0.952)
Follow-up x Presbyopic	0.0275	-1.057	9.162	-0.215*
	(0.219)	(8.350)	(6.210)	(0.104)
Treatment x Presbyopic	-0.235	15.22	22.02*	1.183
	(0.709)	(8.608)	(11.35)	(1.119)
Follow-up x Treatment x Presbyopic	-1.103***	12.57	13.74	1.090
	(0.275)	(10.29)	(10.86)	(0.733)
Age	-0.0527*	0.0595	0.628	-0.0160
	(0.0251)	(0.256)	(0.514)	(0.0370)
Azizi	-1.662***	-8.675	-21.36*	-2.528**
	(0.451)	(7.783)	(10.72)	(0.819)
Constant	8.973***	33.83***	55.49**	8.408***
	(0.780)	(8.847)	(17.15)	(1.243)
Observations	473	473	473	473
R-squared	0.468	0.076	0.138	0.303

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1