Scientific Productivity of INTSORMIL Long-Term Trainees: A Bibliometric Analysis

Timothy J. Dalton
Sandra Contreras
Graciela Cristina Andrango

Department of Agricultural Economics
Kansas State University
Manhattan, KS 66506

tdalton@ksu.edu
sandrac@ksu.edu

DRAFT-Not for CITATION

May 2012

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Research support is gratefully acknowledged from USAID through the Sorghum, Millet and other Grains CRSP. The Sorghum, Millet and Other Grains CRSP is made possible by the United States Agency for International Development and the generous support of the American people through USAID Cooperative Agreement No. EPP-A-00-06-00016-00. All opinions are of the authors and may not reflect those of USAID.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The Sorghum, Millet and Other Grains (SMOG) Collaborative Research Support Program (CRSP), based at the University of Nebraska, Lincoln, is an international agricultural development program funded by the United States Agency for International Development (USAID). This program was established in 1979 and efforts initially focused on sorghum and millet issues under the name of "INTSORMIL," the International Sorghum and Millet CRSP. However, in 2006 it was extended to other grains (finger millet, folio and tef) and renamed the "Sorghum, Millets, and other Grains" or "SMOG" CRSP. We use SMOG and INTSROMIL synonymously. SMOG is a consortium of researchers from several U.S. universities leading

A central component of the program's activity concentrates on developing human capacity through the long- and short-term training of students. This is done in addition to the research, development and outreach activities. Long-term training (i.e. for B,S, M.S, Ph.D. degrees and post-doctoral fellows) has long been viewed as a large cost center for many international development programs. Training students is expensive as tuition, fees and living stipends are invested yet the contribution of a student to future research and development is unknown. Hence the investment can be viewed as risky and it is born by the U.S. taxpayers for a future benefit that is unknown. At the same time, education and human capital development has been a central component of USAID development objectives and it is at the heart of the land grant university mission. The purpose of this report is to evaluate the impact of long-term trainees under the INTSORMIL program through the metric of scientific productivity, as measured though publications.

We first present a descriptive analysis of primary data related to INTSORMIL long-term trainees emphasizing trends and distributions and an overview the data. Then, take this information and quantify the scientific productivity of trainees funded by INTSORMIL through bibliometric indices. This covers individuals trained during 32 years of the program. Lastly, we briefly present trends and characteristics of the financial resources INTSORMIL has received during the years that has been in existence and attempt to relate productivity to the cost of training these individuals.

LITERATURE REVIEW

Production and Measurement of knowledge

Agricultural research is an economic activity that involves the investment of scarce resources in the production of knowledge in order to increase future agricultural productivity (Alston *et al*. 1998), Consequently, research investments can lead to a change in productivity by positive changes in the quality of conventional inputs (or their prices), as well as the implementation of more effective practices of production. In the latter case, more effective practices evolve through scientific input and knew knowledge about sorghum production and usage.

While capacity-building investments are judged as essential to creating the scientific base for knowledge creation, Gordon and Chadwick (1997) assert that there is little hard evidence to demonstrate how important these investments are. They argue that most studies have commonly measured capacity built (such as skills gained) but not output or productivity gained. Lee and Bozeman (2005) propose to assess scientific productivity and research collaboration as as a dynamic knowledge variable. In this matter, they evaluate scientific productivity as a function of the number of publications and collaboration, motivation for collaboration, collaborators, collaboration scale, grants, job satisfaction, discrimination, individual characteristics, field, and publishing productivity.

An alternative model is the one that links education with spending to create knowledge and make changes in the stock of useable knowledge. The accumulation of this knowledge generates changes in productivity, through new technology, and hopefully output. In the present report we consider this later model to measure the creation of tangible knowledge, as measured by scientific publications and their impact on the scientific community through a bibliometric framework.

Bibliometric Indicators

Assessing the investment in production of knowledge is difficult, and there have been some attempts to measure the outputs of this investment. Most recently an academic sub-discipline focusing on data source, data metrics and citation analysis has developed to explore for a good proxy of the outputs coming from the production of knowledge.

In the bibliometric sub-discipline, published findings, particularly in refereed scientific and technical journals, are a highly visible feature of modern science and scientific productivity. Publications serve an easily accessible measure, specific and representative of the most recent knowledge; it keeps record of what is being done in a particular field, and recognize the intellectual property of scientific work in a specific discipline. In the past, publications were more difficult to access but now days with journal being electronic, access, and reachability of this recorded knowledge is much easier and allows for deeper introspection.

Quality is based on citations received by the paper, and this is an extensively used measure of assessing impact of publications and research (Rigby J. 2009). Citations are a variable that controls for quality because it is seen as “a collective verdict of the market and larger number of users decide on the impact of one’s work” (Rigby J. 2009). The opposite happens when the analysis is conducted using the type of journal a scientist has published, since the publication in a top journal can be influenced by a small number of referees that can be bias in the review process.

Undoubtedly, publications and citations are not the only tool to measure productivity Any analysis of publication and citation patterns must be sensitive to the contexts, the areas of knowledge (Social sciences, Natural Sciences), the era in which papers are written, published, and cited. Nonetheless, careful observation of scientific material that is written and how it is used can support in evaluating of an organization's performance such as INTSORMIL whose ultimate goal is the improvement in education and creation of research communities. This is our objective, to utilize the publication record, and citations of these academic documents and articles written by participants of the INTSORMIL trainings, and some bibliometric indices to evaluate different aspects of a scientific research work.

As with any other methodology it is necessary to acknowledge the tendency of improper use of the bibliometric indicators due to relatively easy availability and calculations. Hence, we were selective in choosing some indexes for our analysis as well as other tools such as the Gini coefficients to make cross-sectional comparisons.

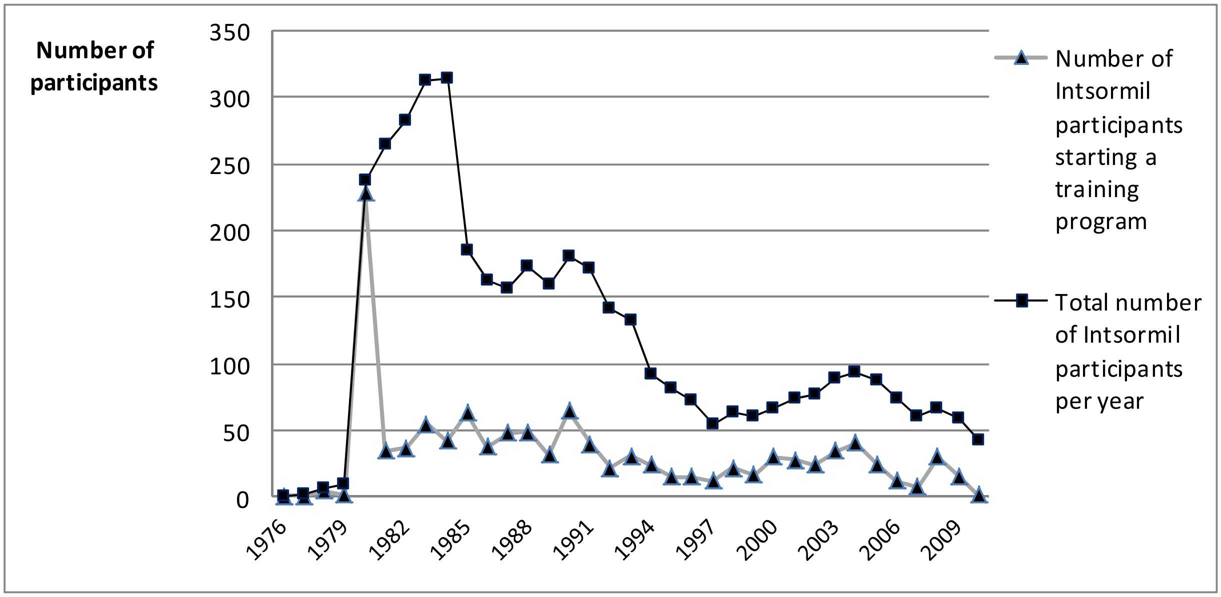
INSTITUTIONAL BACKGROUND

The present section aims to identify the trends of the trainees through time as well as to analyze whether or not there is statistically significant difference between categories (i.e. gender, type of program). Additionally, it will be possible to find out what type of program and category of trainees INTSORMIL has been invested. The data for the present report was obtained from the INTSORMIL database at the University of Nebraska, Lincoln. Such data has been compiled in the database FileMaker and the information contained on it is explained in Appendix 1.

General Facts

INTSORMIL has trained more than 1150 students during the last 32 years. The first years of the program the number of trainees was very low, but this number increased rapidly during the mid-1980 where the highest number of trainees starting a program was reported, and on average 205 people were being trained each year. This trend declines from 1994 to 2010 and the average number of people trained decreased to 72. Figure 1 presents the trends of the number of participants starting a training program each year from 1976 to 2010 and of the total number of participants being trained by INTSORMIL each year in the same period.

Figure 1. Comparison between total number of annual INTSORMIL participants and total number of participants starting training per year from 1976-2010



From 1981 to 2008, INTSORMIL had on average 30 trainees starting a program each year. Furthermore, from 1980 to 1993, on average 205 people were being trained each year while from 1994 to 2010 this number decreased to 72 people.

INTSORMIL has followed a decreasing tendency when allocating resources on education and training. In fact, both, total number of trainings and number of starting trainees per year, show a decreasing trend through 1981 to 2010. In order to determine whether this declining is statistically significant, the number of participants starting a training program each year and the total number of participants per year were regressed on time. Equation (1) shows the quadratic function:

$$Participants = \alpha + \beta_1 time + \beta_2 time^2 \qquad (1)$$

Equation (1) shows the linear regression of participants on time, where the dependent variable “Participants” refers to the starting trainees per year for the first regression, and it refers to the total number of participants within the program per year for the second regression. As clarified before, time is the number of years from 1980 to 2010.

In the following table we show the result for starting participants on time and time square. The declining of the number INTSORMIL participants starting a training program each year is statistically significant at 10% of confidence level on time, and it is not significant on time square.

Table 1.Regression of number of INTSORMIL participants starting a training program each year on time – quadratic relationship

Variable	Coefficient
Constant	50.585 ***
	(7.115)
Time	-0.1533 *
	(1.058)
$Time^2$	0.0115
	0.033

Numbers in parenthesis are standard errors
* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Obs: 30 $R^2 = 0.441$

On the other hand, table 1 shows the results for the quadratic relationship of total number of participants per year. The declining of the total number of INTSORMIL participants per year is statistically significant at 1% of confidence level on time and time square. Therefore, in this case, the quadratic function fits better for explaining such trend. On the other hand, table 2 shows the results for the quadratic relationship of total number of participants per year. The declining of the total number of INTSORMIL participants per year is statistically significant at 1% of confidence level on time and time square. Therefore, in this case, the quadratic function fits better for explaining such trend.

Table 2. Regression of total number of INTSORMIL participants per year on time – quadratic relationship.

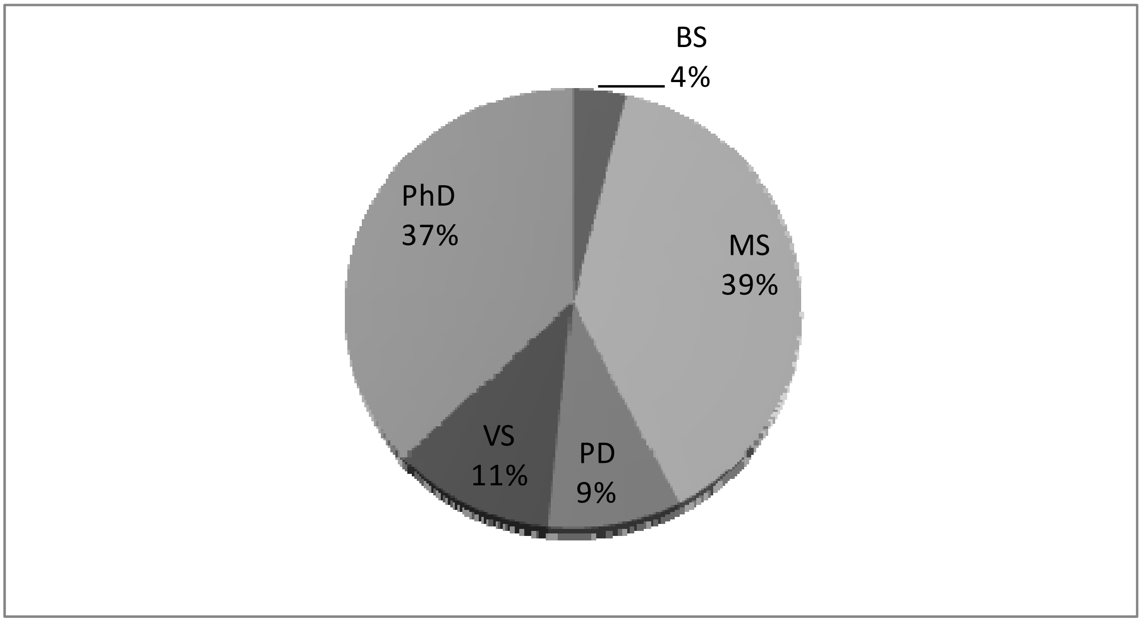
Variable	Coefficient
Constant	317.93 ***
	(17.0567)
Time	-20.586 ***
	(2.5364)
$Time^2$	0.406 ***
	0.07938

Numbers in parenthesis are standard errors
* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Obs: 30 $R^2 = 0.88$

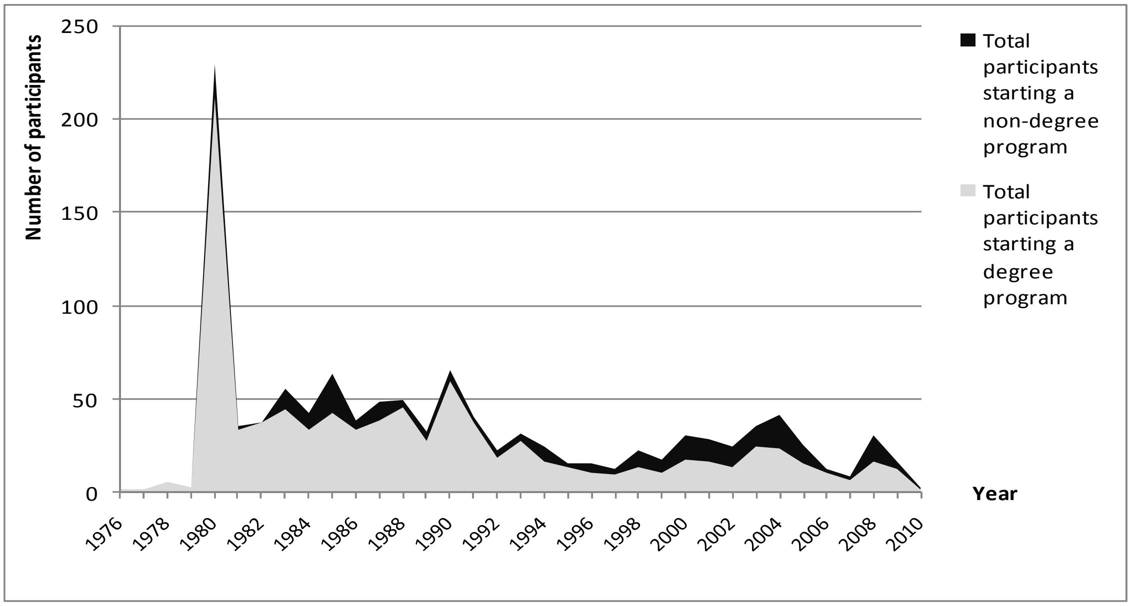
The total number of trainees reported in the INTSORMIL's database by June 14th, 2010 was 1,151 from which 428 trainees are reported under the PhD category, 46 under BS, 443 under MS, 105 under PD (post-doctoral fellow), and 129 were trained as visiting scientists who were non-degree seeking (VS). Figure 2 shows such distribution, INTSORMIL participants have been trained mainly under the degree category: PhD and MS.

Figure 2. Percentage of INTSORMIL participants by type of program from 1976-2010



Furthermore, when discriminating the number of starting participants by type of program, that is those starting a non-degree program from those starting a degree program, the tendency is similar to that for the total number of participants starting a program. Thus, both groups show a decreasing trend from 1981 to 2010 as it is illustrated in figure 3.

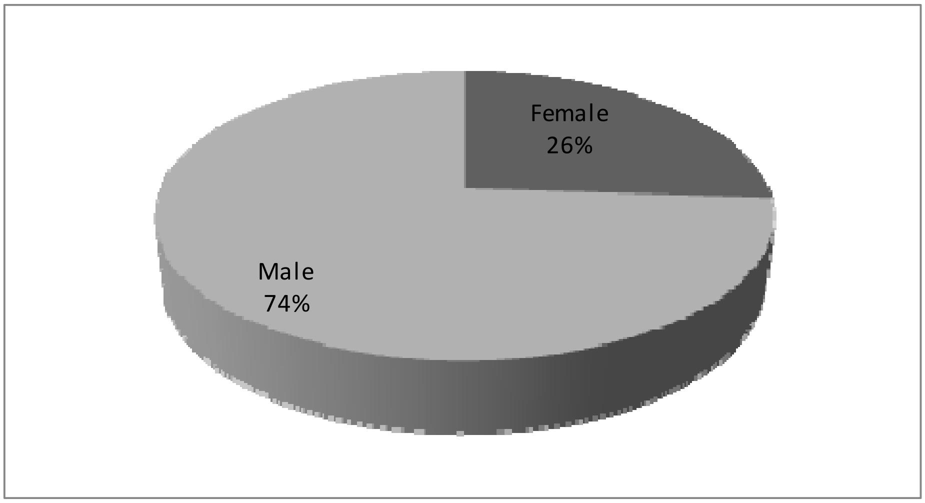
Figure 3. Number of INTSORMIL participants starting a non-degree or degree program from 1976-2010



Gender

In addition to the trend in education and training, it is also important to analyze the presence of women participating within those programs from the time INTSORMIL started until the present year. The following section presents a description about gender and how the presence of women has changed through the time. INTSORMIL has considered women's participation as an important issue throughout the program. In fact, USAID changed its policy in 1998 towards a women empowerment vision. However, as it can be seeing in figure 4, from the total trainees 74% were male, and 26% were female.

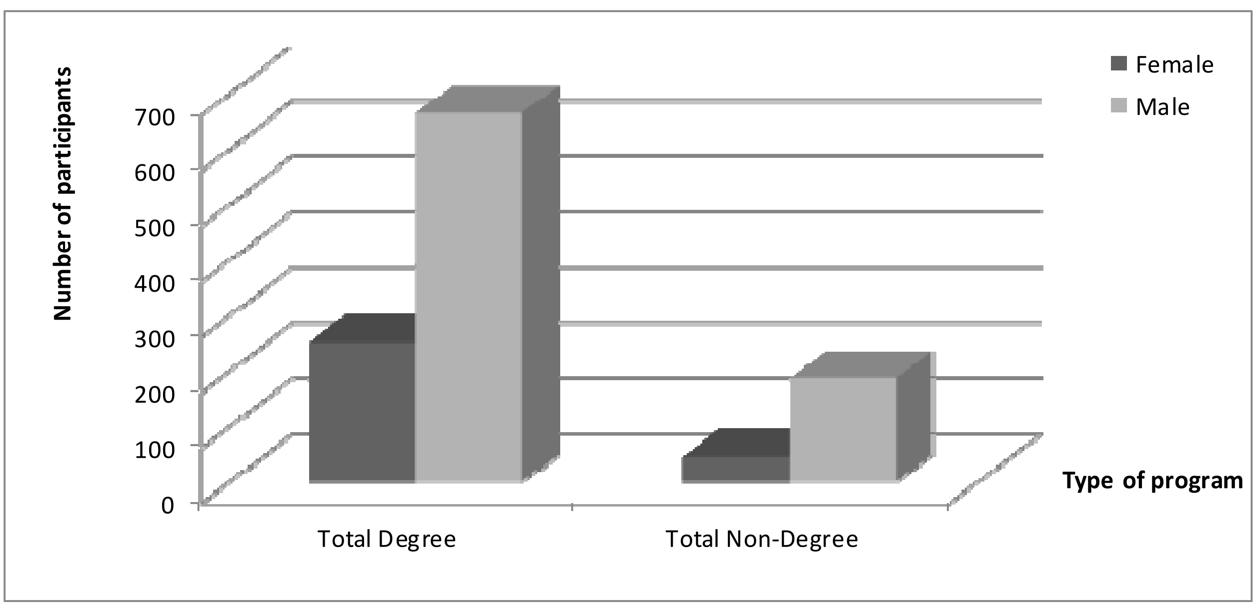
Figure 4. Percentage of INTSORMIL participants by gender from 1976-2010



It can be hypothesized that women have participated more in short programs such as MS, VS and PD since the length of those allow women with family to have the opportunity to be trained and not to be apart from their families.

Overall the proportion between female and male that are being trained in any of the programs is not equitable. In addition, the disparity of this proportion is greater in the degree programs where only 25% of the total participants in this type of program are female (figure 5).

Figure 5. Total number of INTSORMIL participants from 1976-2010 by type of program and gender

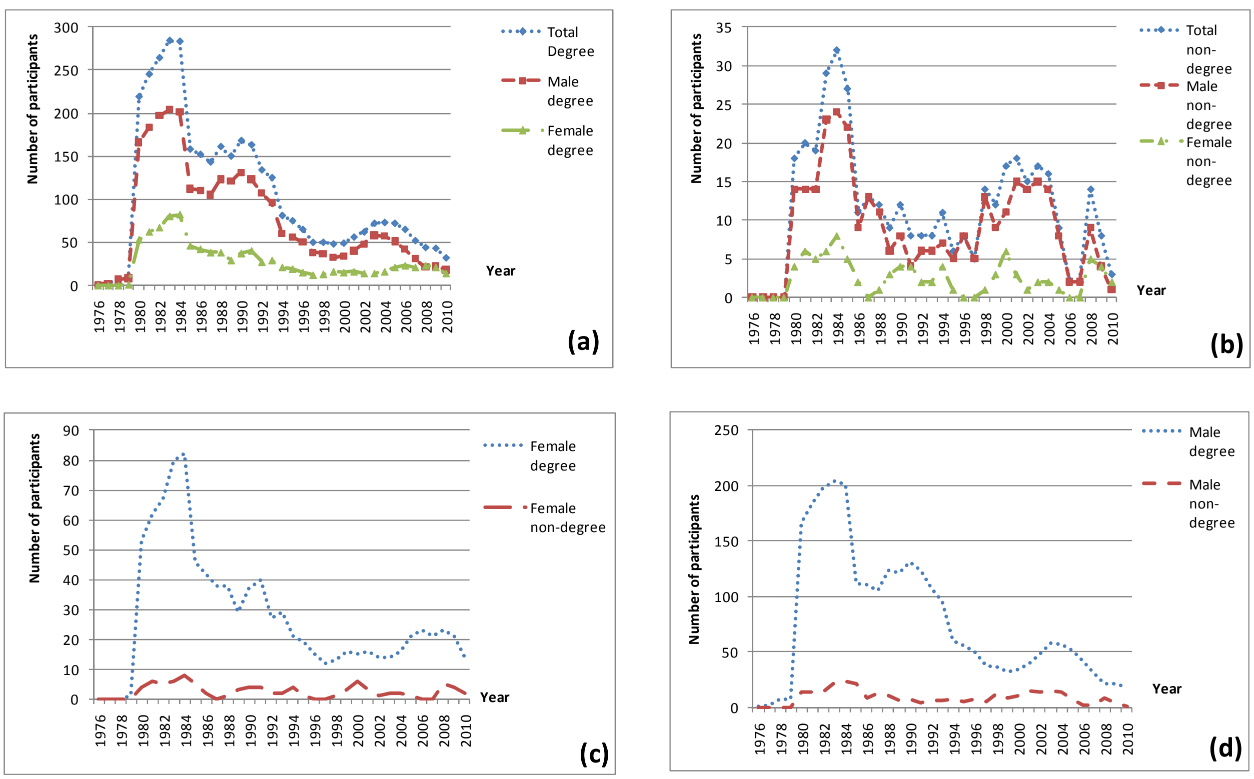


Females are more frequently trained in non-degree programs. The percentage of females train in a non-degree program is a higher percentage (28%), but still there is a disproportion between the number of females and males that have been trained.

We conducted a Chi-square test in order to find out whether or not the proportion of females and males trained in degree and non-degree programs have a statistically significant difference. The result indicates that there are no equal proportions of males and females trained in the degree as well as in the non-degree programs. In other words, the null hypothesis is rejected with a 95% of confidence. Chi-square with one degree of freedom = 6.0996, p = 0.0014. In addition, the tendency of total INTSORMIL participants by year discriminated by gender and by type of program is presented in figure 6.

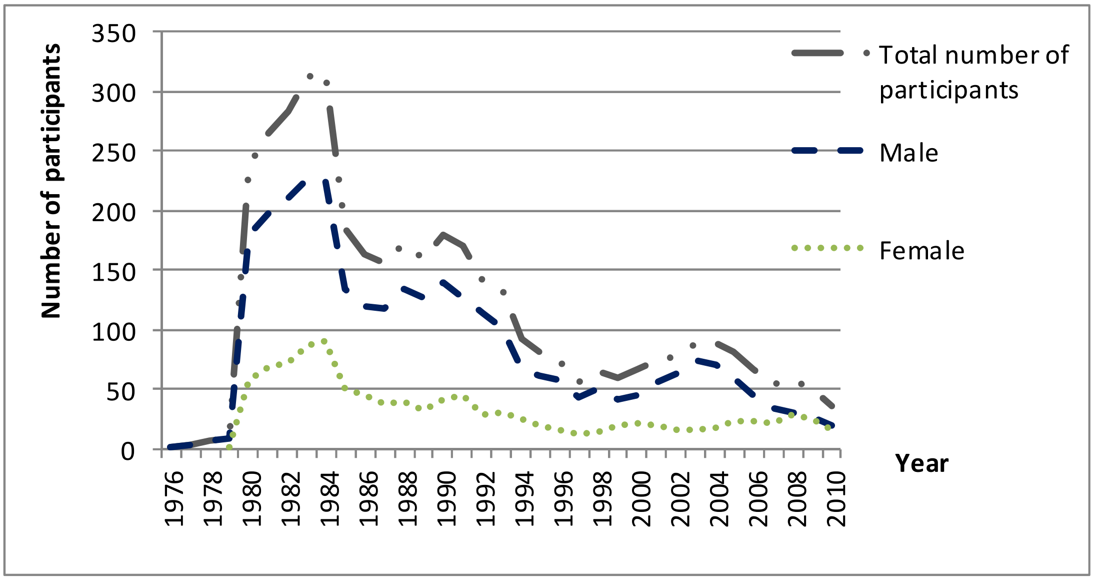
Panel (a) shows that the tendency of the total INTSORMIL participants compared to that of male and female involved within a degree type of program. Such tendency was decreasing from 1981 to 2010. In contrast, panel (b) presents the same groups involved within a non-degree program. In this case, the trend presents cycles, which makes more difficult to analyze this situation.

Figure 6. Trends of cumulative number of INTSORMIL participants by year involved within a degree or non-degree program by gender from 1976-2010



On their part, panel (c) and (d) illustrates the trend of female and male, respectively, involved in a degree program compared to those involved in a non-degree program. The number of women and men who participate within a degree program is higher than those who participated within a non-degree program. In the same way, when referring to the trend of total number of INTSORMIL participants per year compared to that for male and female (figure 7) it can be concluded that the both, total number of participants and males follow a similar decreasing trend. Conversely, the trend of female shows a declining from 1984 to 1988. After that, the tendency had become flat until 1998 when it starts to increase in a small rate. In 2008, it can be seen that both trends, female and male intersect with each other. The change in such trend could be due to the different policies installed in USAID, and consequently in INTSORMIL which envision a greater participation of women.

Figure 7. Total number of INTSORMIL participants per year by gender from 1976-2010



Consequently, it can be hypothesized that the presence of women in the INTSORMIL project has increased through the time due to changes in cultural behavior and policies of the sponsor institutions such as USAID which implemented the gender equity and women empowerment principles in 1998.

In order to find out whether or not the declining of the trend of the number of starting participants by gender is statistically significant, equation (1) was regressed, where participants is the number of female or male trainees starting the program each year on time. Results are shown in tables 7 and 8.

The trend of women from 1980 to 2010 is better explained when using the quadratic equation. The model presents a better fit (R^2 = 0.846) and it can be concluded that the decreasing of the presence of women through time is statistically significant.

Table 3. Regression of number of INTSORMIL female participants per year on time – quadratic form

Variable	Coefficient
Constant	87.499 ***
	(5.047)
Time	-6.642 ***
	(0.751)
$Time^2$	0.1542 ***
	(0.023)

Numbers in parenthesis are standard errors
* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Obs: 30 R^2 = 0.846

In the same way, when seeing at table 8 it can be concluded that the quadratic model explains better the trend of men through the time.

Table 4. Regression of number of INTSORMIL male participants per year on time – quadratic form

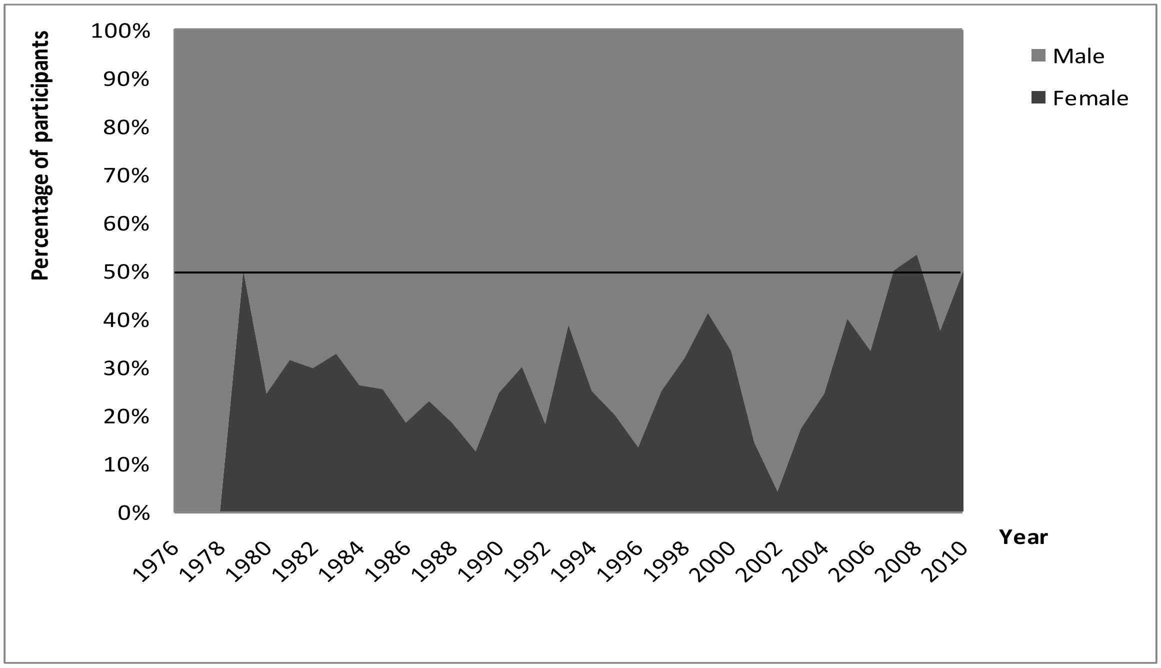
Variable	Coefficient
Constant	230.432 ***
	(12.95)
Time	-13.944 ***
	(0.1.9256)
$Time^2$	-0.2518 ***
	(0.0603)

Numbers in parenthesis are standard errors
* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Obs: 30 R^2 = 0.8762

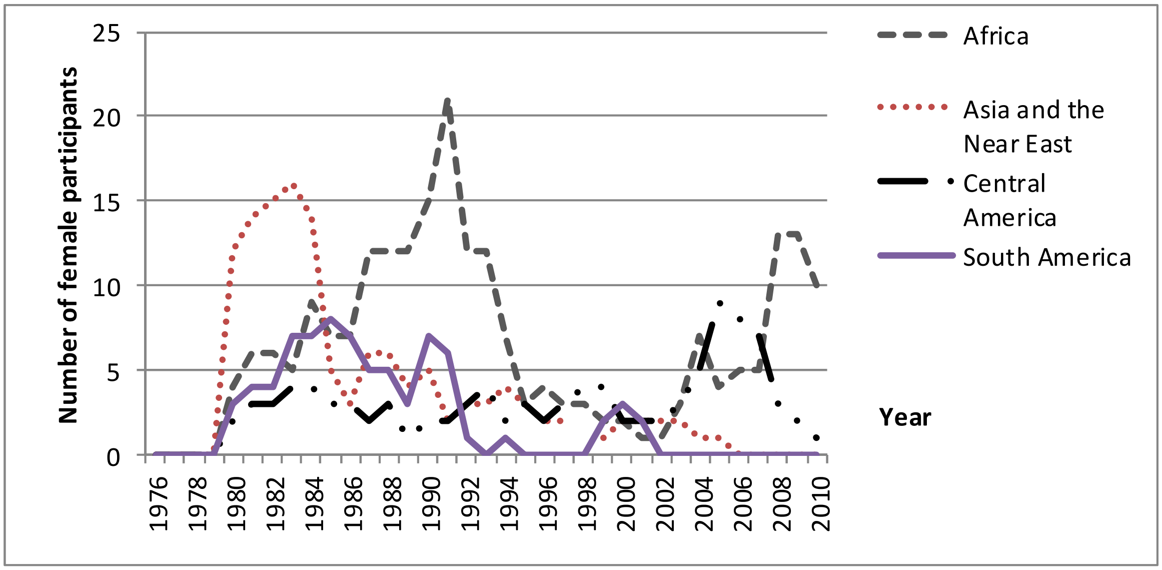
Similarly, figure 8 illustrates the difference on proportion through the time between male and female. Such difference is evident; however, from 2002 there has been a tendency to increase women within the INTSORMIL program.

Figure 8. Percentage of INTSORMIL starting participants by year and gender from 1976-2010



Furthermore, it can be thought that those women who have participated within short term programs are more likely to come from regions outside the USA. Figure 9 illustrates the trend of the total number of women participating within a program according to their regions of origin.

Figure 9. Trend in total number of female INTSORMIL participants by region excluding the USA from 1976-2010



In this particular case, the region of Africa, Asia and the Near East, Central and South America were taken in account. The USA was eliminated since the highest number of women came from that region. The trend is fairly irregular, but overall; most women belong to the Africa region, followed by those who belong to the Asia and Near East. Female participants coming from South America had their highest representation from 1982 to 1990 while those from Central America had it from 2004 to 2008.

The INTSORMIL database provided information about the disciplines within the participants were involved. Looking toward an analysis of productivity measure of the trainees, Section 3 presents a summary about the disciplines and/or sub-disciplines in which INTSORMIL participants had been trained.

Disciplines

In order to overcome the constraints that the growth and production of sorghum and millet present, INTSORMIL has trained more than 1150 participants in different disciplines. The disciplines in which participants have been trained are mostly in the production science. However, INTSORMIL have also trained participants in more than 10 other different fields of knowledge. This section will present descriptive information related to these fields. The disciplines where INTSORMIL has trained participants were classified in five groups:

Plant Protection

- Plant Pathology
- Entomology
- Molecular Biology
- Biometrics
- Bird Control

Animal Science

- Animal Nutrition
- Animal Science
- Forages

Food Science

Social Science

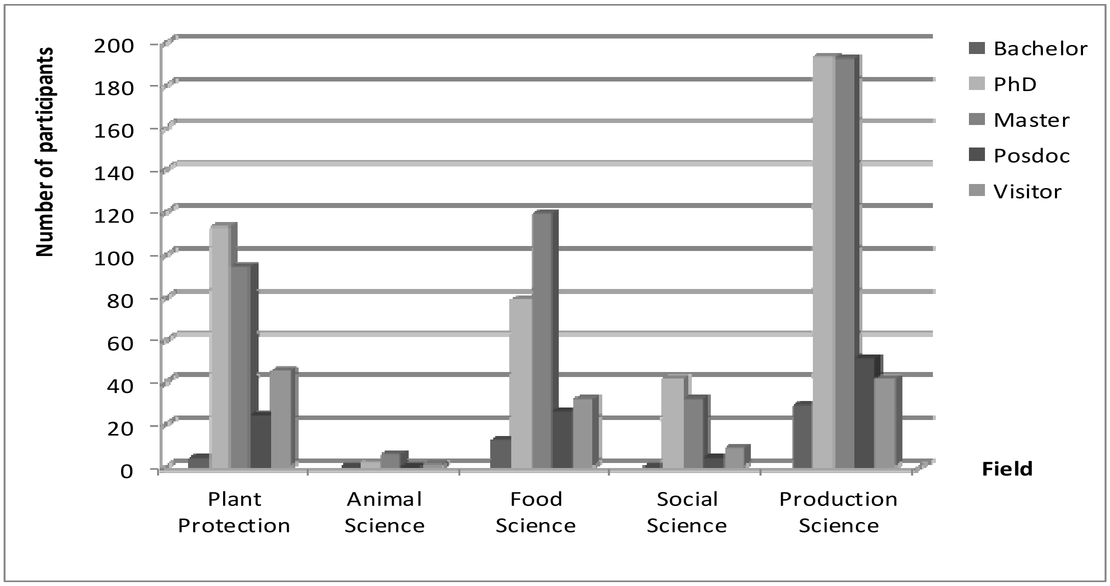
- Economics
- Sociology
- Agricultural Extension

Production Science

- Agronomy
- Breeding
- Soil Sciences
- Physiology
- Agro-climatology

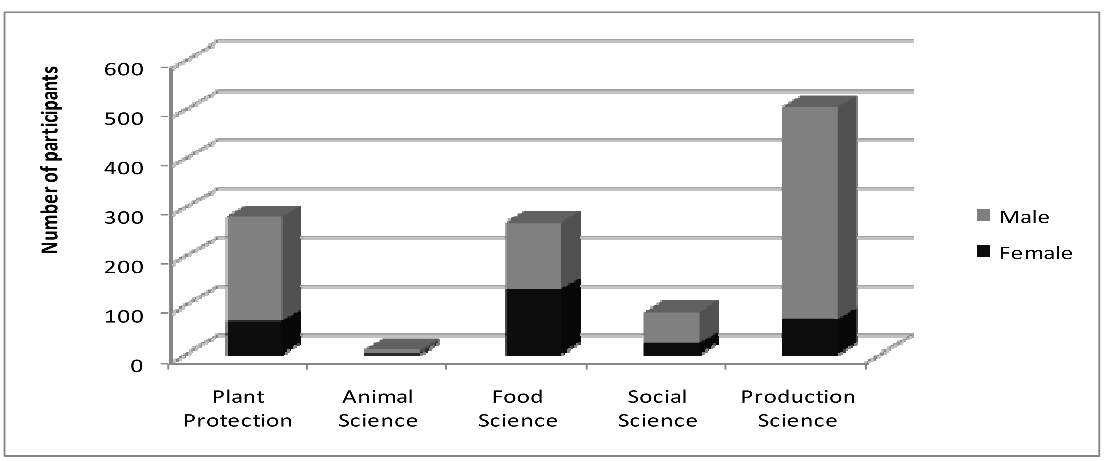
One this classification was done, the number of trainee's participants in the different degrees was located in the different regions. From the figure below we conclude that most of the INTSORMIL participants have been trained at a Master's level followed by PhD's level in all disciplines.

Figure 10. Total number of INTSORMIL participants from 1976-2010 by type of training across disciplines



Considering only the production science disciplines, 76% of participants have received an advance degree education. On the other hand, participants trained in Plant Protection disciplines have the highest number of participants trained as short-term scholars.

Figure 11. Total number of INTSORMIL participants from 1976-2010 by gender and field



Most of the women have been trained in the Agronomy, Breeding, Food Science, Plant Pathology, and Sociology. From those fields, most of the women have come to Food Sciences. In the same way, men have been trained in Agronomy, Breeding, Economics, Entomology, Food Sciences, Physiology, Plant Pathology and Soil Sciences.

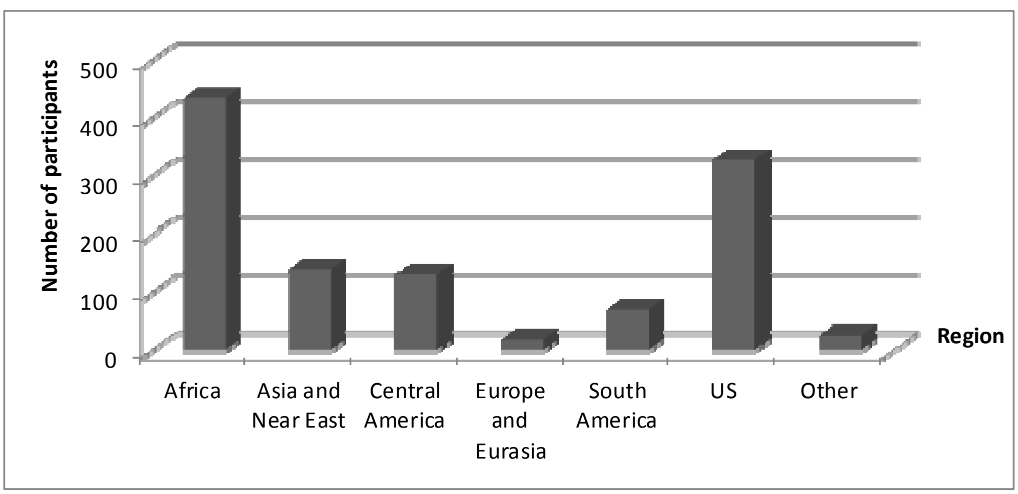
In figure 11, most of the students have been trained in Agronomy, Breeding, Food Sciences, and Plant Pathology. A less number of students have been trained in fields such as Economics, Entomology, Physiology and Sociology.

It was expected to find more participants within the production science field due to the nature of the program. For the same reason, we expect to find higher productivity from individuals trained within the field of Production Science than those individuals within the rest of the fields. Productivity is measured by the number, quality, and impact of publications.

Region

Over the last three decades, INTSORMIL has been supported training participants all over the world. In search of establish collaborative research between millet and sorghum producing countries, and those with higher developed technology, INTSORMIL has sponsored trainees from 85 different countries including the United State. The following section presents trends, and distributions of the places trainees have come from. By identifying these trends, possible inferences can be made about geographical location of higher rates of returns of investment in human capital. Also, the section will indicate if the priorities that INTSORMIL has of training host country scientist whose countries are higher producer of sorghum are being met.

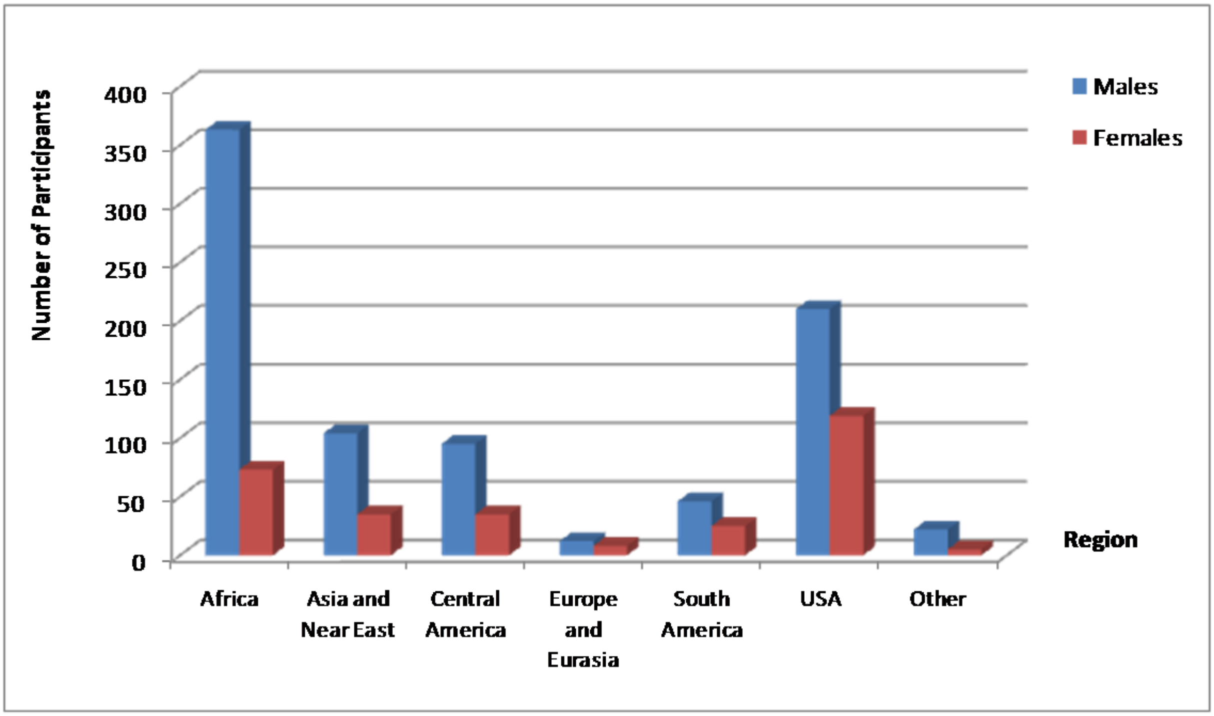
Figure 12. Total number of INTSORMIL participants by region from 1976-2010



The division of the regions in the way they appeared in this section has been set according the original division found in the database. Considering the total number of INTSORMIL participants, data shows that the regions that have had the most number of participants educated through this program are Africa and USA.

The total percentage of participants that come from Africa is 38%, and 29% come from USA (Figure 12). In total, these two regions represent more than 50% of the total number of training participants. Taking into consideration the network effect among regions we can infer that it is expected that African and USA participants have higher rate of returns of investment on training explained by greater level of networking effect, and increasing in knowledge compare to those from the other regions.

Figure 13. Total number of INTSORMIL participants divided by region and gender from 1976-2010



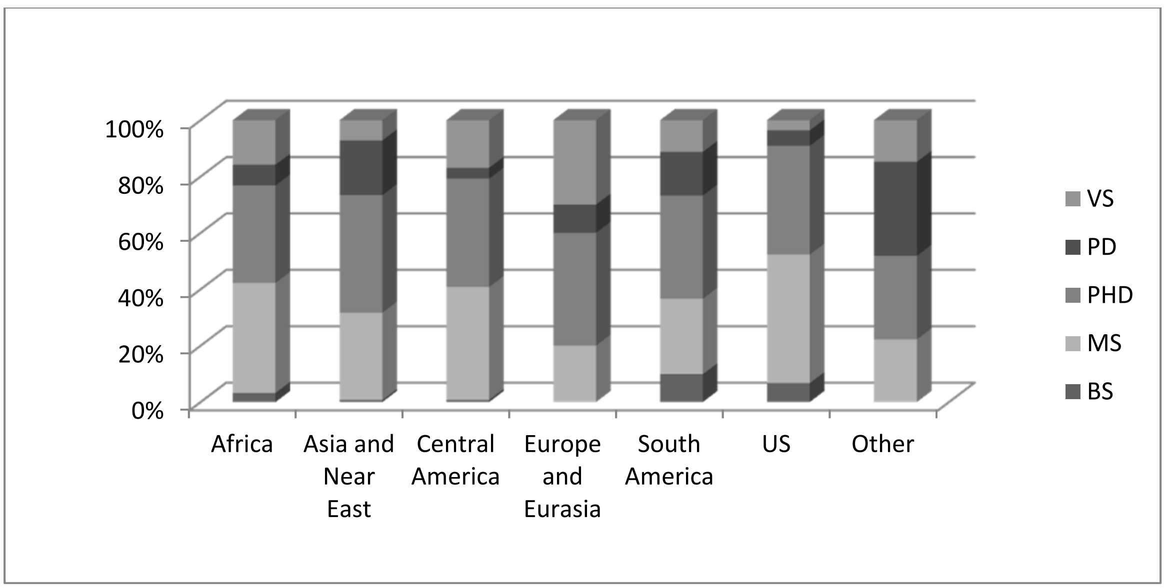
From figure 13 it is evident that among the seven regions where INTSORMIL supports training participants, the majority of participants in all of them have been males. The regions that present less disparity in the distribution are USA and Eurasia. These two regions are the most developed ones also. In contrast the regions with more disparity between the number of females and males been trained are those consider undeveloped regions. The percentages are presented in table 5.

Table 5. Percentage of INTSORMIL participants divided by region and gender from 1976-2010

Regions	Males	Females
	(%)	
Africa	83.3	16.7
Asia and Near East	74.8	25.2
Central America	73.1	26.9
Europe and Eurasia	60.0	40.0
South America	64.8	35.2
USA	63.8	36.2
Other	81.5	18.5

According to figure 13 and table 5 the biggest disparity in relation to gender is Africa, where 83% of the participants are male and only 16% of them are women. Again, it is seen that female African scientist have not received the same opportunity in training than male African scientist and there are likely several reasons for this beyond the control of the ME and participating scientists.

Figure 14. Proportion of INTSORMIL participants by region and degree from 1976-2010



VS=Visitor PD=Post doctorate PhD= Philosophy Doctorate MS=Master BS=Bachelor

There are five different degrees that training participants can pursue. In figure 14, the proportions show that about 85% of the training participants that came from the USA region pursued either a PhD or MS. Also, 30% of participants that came from Europe and Eurasia region were trained as Visitors.

Participants who come from Africa region are mostly involved into the MS or PhD programs, this is 35% and 39% respectively.

To have a more general idea of the geographical distribution of degree and non-degree INTSORMIL trainees the following table 6 was organized. We took into account the type of trained the participants were pursued, and countries where they came from.

Table 6. Percentage of INTSORMIL participants by degree and non-degree programs among regions from 1976-2010

Regions	Degree	Non-degree
	(%)	
Africa	76.83	23.17
Asia and Near East	73.38	26.62
Central America	79.23	20.77
Europe and Eurasia	60.00	40.00
South America	73.24	26.76
USA	90.88	9.12
Other	51.85	48.15

Among regions, the proportion between participants training in degree and non-degree programs for all regions is very similar that 75/25. The only regions that proportions are very apart are USA, and Europe and Eurasia, the most developed ones.

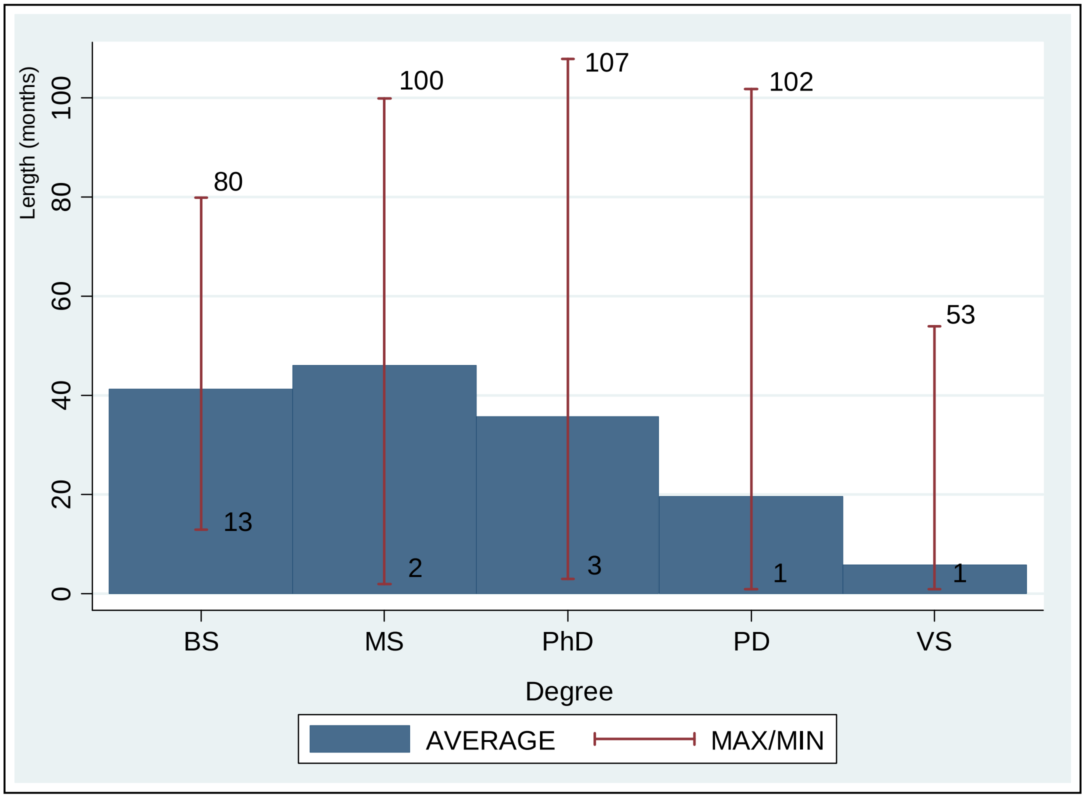
From the proportions found in the data, we hypothesize that there is no difference in productivity among undeveloped regions. Since the percentage of participants trained in a degree program is higher than that from a non-degree program for all less developed regions.

Length of training

Some other variable that we took into consideration for this analysis was the length of a participant in finishing the training. If the majority of trainees finished in the time expected, participants are being productive and any extra cost is being carried by INTSORMIL.

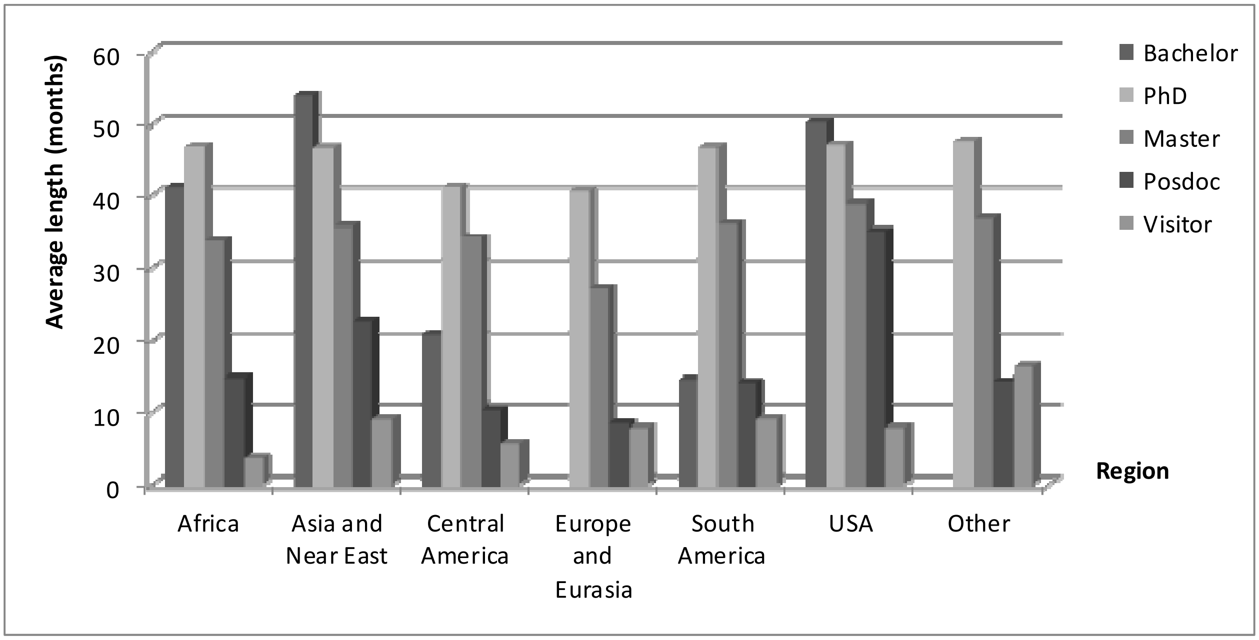
INTSORMIL has partial or totally sponsored participants. They are expected first to finish their trainings, and second to take the reasonable time in doing so. In figure 14 we divided the data by periods of time a participant spent in finishing a particular degree. We looked at a disaggregated data that will show us the average length of each of the programs INTSORMIL participants have attended (PhD, Master, Bachelors ,Visitor Scholar training).

Figure 14. Average, maximum and minimum length of INTSORMIL training in months by degree 1979-2010



The average length of training spent by an INTSORMIL participant fits the standards of a regular university program. Therefore we can infer that participants are finishing their programs in a reasonable period of time but there are some outliers, as in any program. In that matter, INTSORMIL is not assuming higher costs for having students longer periods of time. To consider the success of participants in finishing their programs, we disaggregated the data related to length of training and regions from where participants were from. The following figure contains this information.

Figure 15. Average length in months that an INTSORMIL participant spent within the program by region from 1976-2010

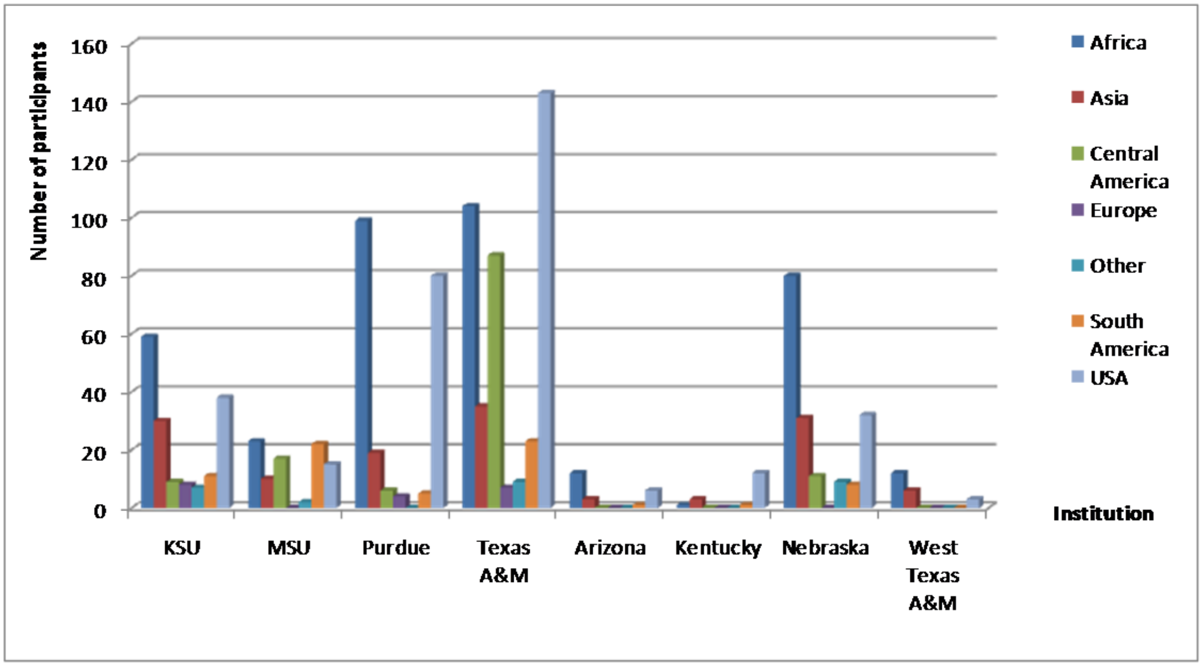


The length of time that participants are spending in pursuing any degree is about 2-4 years in all regions. Taking into consideration that the highest number of participants that INTSORMIL has trained are PhD and Master levels students, participants have spent the adequate time in completing their degrees. Therefore we can expect that participants who belong to a region different from USA in average will spend more time finishing a program than a participant from the USA. This could be explained due to language barriers, cultural and academic adjustment, among others.

When referring to the length in training, it is important to consider also the institutions that trainees have attended to pursue their degrees, and detect if there is a relation between origin of participants, and the institution they attended.

In the following figure the eight institutions that we considered in the report were those with the highest number of INTSORMIL participants.

Figure 16. Total number INTSORMIL participants by institution and region from 1976-2010



Most of the training participants that have come to Texas A&M, Kansas State University, Purdue University, University of Nebraska, and University of Arizona are from Africa and USA. Most of the training participants that came to Mississippi State came from Africa and South America. The majority of the training participants that have come to University of Kentucky came from Asia and USA. Most of the Training Participants that have come to West Texas A&M University came from Africa and Asia. Then we expected that the universities that have the highest numbers of trainees will have a bigger network effect.

One way we propose to measure the network effect is by taking into consideration the number, quality, and frequency cited publications that trainees from these institutions have written. the collaborative research, rate of return in training human capital and network effects can be quantify using these bibliometrics technics. The next section of this report, will present the bibliometric analysis.

BIBLOMETRICS AND PRODUCTIVITY OF TRAINING

We created a bibliometrics dataset from the available INTSORMIL data-base of scientists sponsored by INTSORMIL. The INTSORMIL database contains 16 variables: name of scientist, gender, region of origin, university where the scientist received its degree, etc.(See appendix 1, table B).

The population for the bibliometric study included 766 international and US scientists that were funded to complete their training totally or partially by INTSORMIL between 1976 and 2010. This represents 77% of the total number of trainees reported in the INTSORMIL data base. The total number of people funded partially or totally by INTSORMIL is not included in the new data base because of the following reasons: 1) Trainees dropped the program of studies therefore never completed a degree, 2) trainee's names were incomplete in the INTSORMIL data base, 3) inconsistent use of author initials, 4) author name changes, 5) homographs (ie, scientist with the same names in different disciplines), 6) publications are not register yet in Google Scholars, 7) scientists have not published during their careers.

We used the Harzing's bibliometric analysis software (version 3.2.4150, 2011) *Publish or Perish* (PoP) to analyze the productivity of trainees. PoP is a free software, reviewed by *Nature 2010* and by *Journal of the American Society for Information Science & Technology* 2011, that uses Google Scholar (GS) to retrieve the raw information of academic citations and publications. The software reports the number of papers found in GS for each scientist's name, the number of citations to each paper listed for the author, the name of the journal, and the year of publication, among other information; the information listed before was incorporated in the new data-base.

We identified the scientist (author) articles' information and materials by typing his/her name, and last name in the software; the software retrieves information not only of articles and patents but other important types of published literature such as books, book chapters, conferences, reports, seminars posters, and colloquia. We consider those publications as part of the scientist production outcome, and in particular conference proceedings because as stated by Breschi and Malerba 2011 "the publication of conference proceedings has become (at least in recent years) a suitable alternative either to speed up the dissemination of knowledge before results are published in a peer-reviewed journal or even as the final outlet for research results".

Nevertheless, for all material included in our sample we have extracted a subset of documents corresponding to posters in conference papers, and publications that do not report the year when they were published, also material which name of journal is omitted was not taken into consideration. In addition, the publications we worked with are mainly related to sorghum, millet and/or maize since the main mission of INTSORMIL as well as the training provided by it are related to these crops. Only material elaborated during and after a scientist's training were included in the database. Then, those publications published before the first year of his/her training were not considered. Therefore there is an understimation of the total amount of output from the group of Scientists training and this is discussed in the section on limitation later on.

The data base search found 8,288 documents produced by the subset of INTSORMIL trainees during a period of 32 years (1976-2008). From those documents a total of 68,015 citations were recorded. We later use the number of publications and citation to calculate six different bibliometric indexes described in table 7 and the following paragraphs.

Table 7. Bibliometric Indexes

Index	Description	Formula
H- index (2005) Hirsh	"It aims to measure the cumulative impact of a researcher's output by looking at the amount of citations his/her work has received"	$h = max (j : cit_j \geq j)$ *cit_j is the number of citations of the jth paper*
M Quotient (2005). Hirsch	This is a quotient between h-index, and number of years the academic has been active	$M = \frac{H_{index}}{\text{\# of years of active career}}$
Contemporary H-index (2006). Sidiropoulus, Katsaros, Manolopoulus	This index adds "age-related weighting to each cited article, giving a less weight to older articles"	$S^c(i) = \gamma * (Y(now) - Y(i) + 1)^{-\delta} * \lvert C(i) \rvert$ *$S^c(i)$ for an article i based on citation counting Y(i) is the publication year of an article i and C(i) are the articles citing the article i Gamma is the weight for recent papers and delta for old ones*
H2 index	"A scientist's h(2)-index is defined as the highest natural number such that his *h(2)* most cited papers received each at least $[h(2)]^2$ citations"	*(H-index) ^ 2*
A-index Rousseau R (2006)	The a-index includes in the calculation only papers that are in the Hirsch core. It is defined as the average number of citations of papers in the Hirsch core.	$A = \frac{1}{h} \sum_{j=1}^{h} cit_j$
G-index Egghe (2006)	"Improves the h-index by giving more weight to highly-cited articles". "The highest number g papers that together received g^2 or more citation".	$\sum_{j=1}^{k} cit_j > N^2$ k = 1, 2, 3, . . ,N

The H-index

The H-index is the index of reference in bibliometrics sub-discipline. It quantifies an individual research output, combine with the impact of his/her work, by identifying a set of core high performance journal articles. "A scientist has index *h* if *h* of his or her *Np* papers have at least h citations each and the other (*Np* - *h*) papers have ≤ *h* citations each." (Hirsch JE. 2005).

The M-quotient

In order to allow comparisons between scientists with different length careers, M quotient takes the H-index and divided by the number of years of research activity. We calculated this number of years counting the years since a scientist's first publication. By first publication in our case is the first publication published after the first year of training. In this particular study the m-quotient is of particular interest since the different scientists are in different stages of career.

The Contemporary H-index

In order to account for the "age" of an article, Sidiropoulos A, Katsaros D, Manolopoulos Y (2007), created the contemporary H-index. This index will gradually takes away "value" of the old papers, even if it still gets citations and mainly takes into account the newer articles. It gives greater weight to new articles and less to old ones. The contemporary h-index is expressed as follows: "A researcher has contemporary h-index *hc* if *hc* of its *Np* (number of publications) articles get a score of *Sc(i) = hc* each, and the rest *(Np - hc)* articles get a score of *Sc(i) = hc*." For our results, we used the PoP delta and gamma implementation of 1 and 4 respectively. This parameter can identify who has remained productive and influential. When assigning values of 4 and 1 to the temporal parameters, a weight of 4 is being given to recent articles and the older the article the less weight it receives.

The H(2)-index

The H(2)-index also gives more weight to highly cited articles: "A scientist's h(2)-index is defined as the highest natural number such that his *h(2)* most cited papers received each at least [*h(2)*]2 citations." (Kosmulski 2005). Possibly overly sensitive to few highly cited papers. It is good to measure those top scientists, with high number of citation and high number of publications. This index gives high importance to a constantly productive scientist who publishes papers receiving a significant number of publications and rather than a scientist having few highly-cited papers.

The A-index

The A-index includes in the calculation only papers that are in the Hirsch core. The Hirsch core is a term that refers to a set of highly-cited publications, with respect to the scientist's career. "It is defined as the average number of citations of papers in the Hirsch core. The proposal to use this average number of citations as a variant of the h-index was made by Jin, the main editor of *Science Focus*" (http://sci2s.ugr.es/hindex/#one).Since this index can be very sensitive to highly cited papers, it is used to measure excessive citation or very brilliant scientists. It also evaluates the scientist's production as if it was uniformly distributed.

The G-index

The G index created by Egghe L in 2006 measures the quality of a scientist by giving some more weight to highly cited papers, G-index make some adjustments to the H-index. Egghe defines the G-index "as the highest number g of papers that together received g2 or more citations". The difference from the H-index is that G-index gives more weight to the most cited papers, by doing so, this index will correct the h-index because with the h-index once a paper belongs to the top h papers, its subsequent citations no longer count" (Bornmann L, Mutz R, Daniel HD 2008).

Findings: Scientist Research Productivity

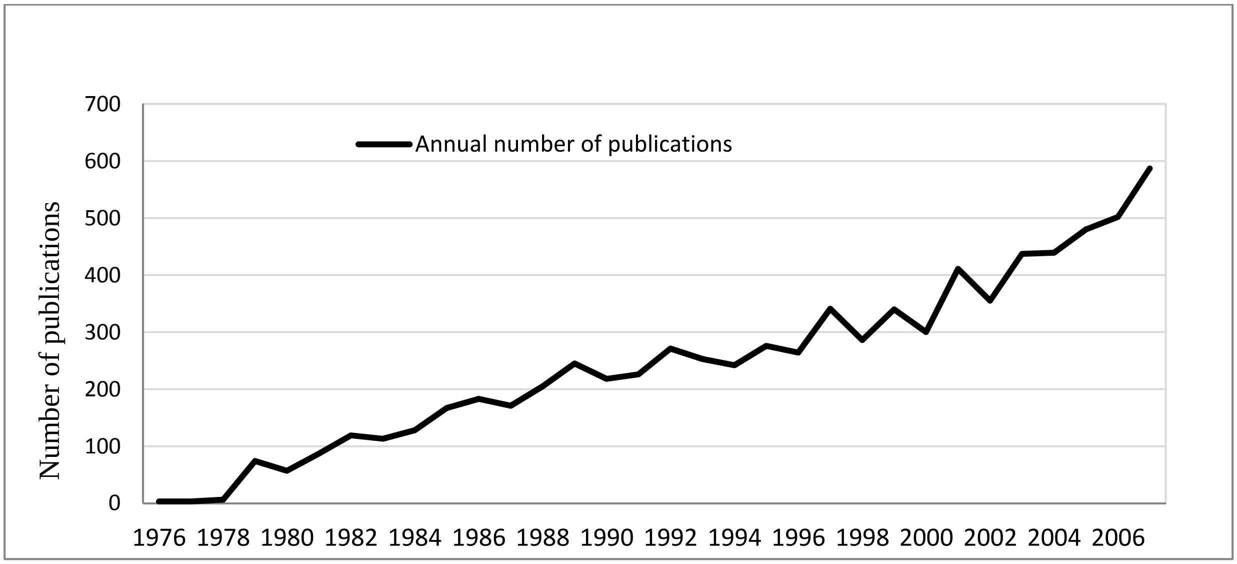
A total of 8,871 articles and academic materials were published by 766 international and US agricultural scientist from the five categories of major disciplines, and five different geographical regions. On average each trainee funded by INTSORMIL has published close to 11 academic documents most of them related to sorghum and millet. These are described in Table 7.

Table 8. Number of author, publications, citation count by region, gender and discipline.

	Number of authors	Number of publications	Citation Count
Region			
Total	747	8,870	68,015
Africa	306	2,922	15,532
Asia and Near East	100	1,447	13,224
US	209	2,700	29,018
Central and South America	116	1,427	6,976
Europe	16	204	2,423
Gender			
Female	183	1,660	16,232
Male	582	7,201	51,783
Disciplines			
Animal Science	7	44	40
Food Science	163	1,729	20,822
Plant Protection	191	1.910	11,161
Production Science	344	3,910	25,573
Social Sciences	53	783	4,575

An average of 259 publications per year have been produced from 1978-2010. As Figure 17 shows during these years, the annual number of publications has an increasing trend. This increasing trend can also be observed when the data is divided by gender, as well as by degree.

Figure 17. Annual number of publications from 1976-2008



During the period of the analysis, the number of people funded partially or totally by INTSORMIL has decreased since 1987, despite this trend the number of publications of scientists funded by INTSORMIL has increased at a very rapid rate, (Figure 18).

Figure 18. Annual number of total trainees and publications by gender from 1976-2008

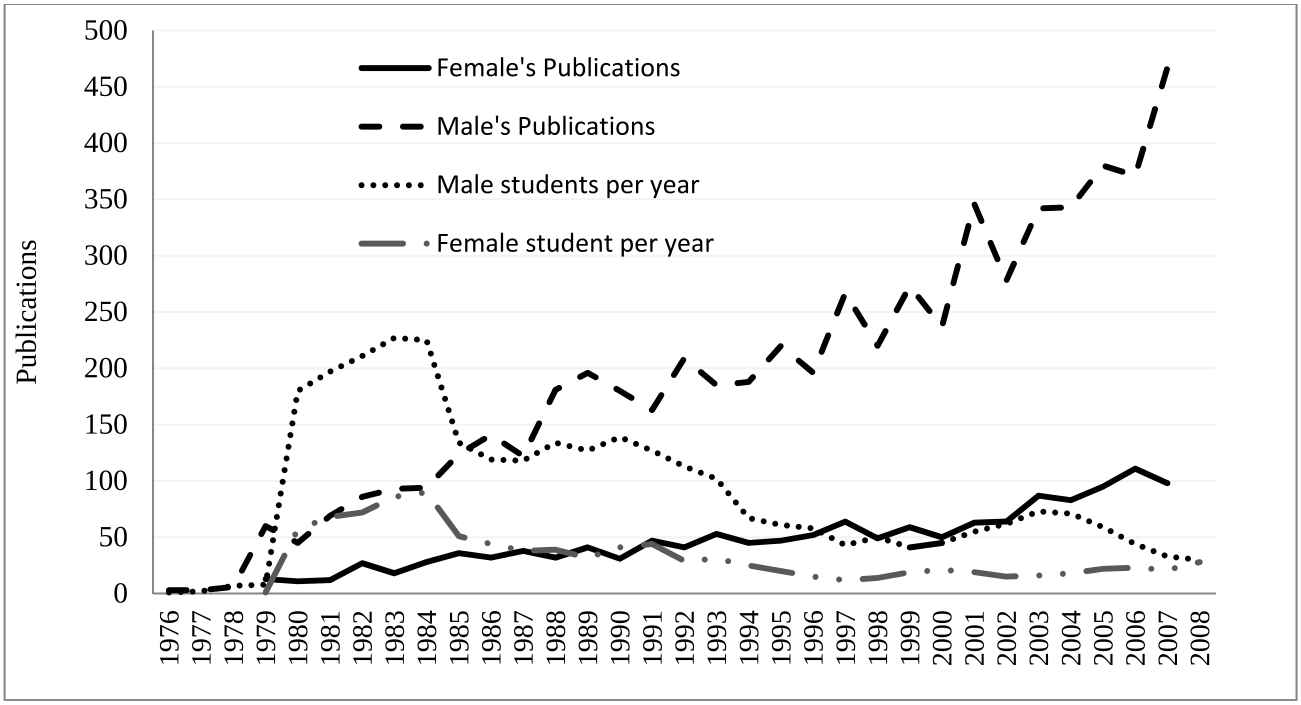
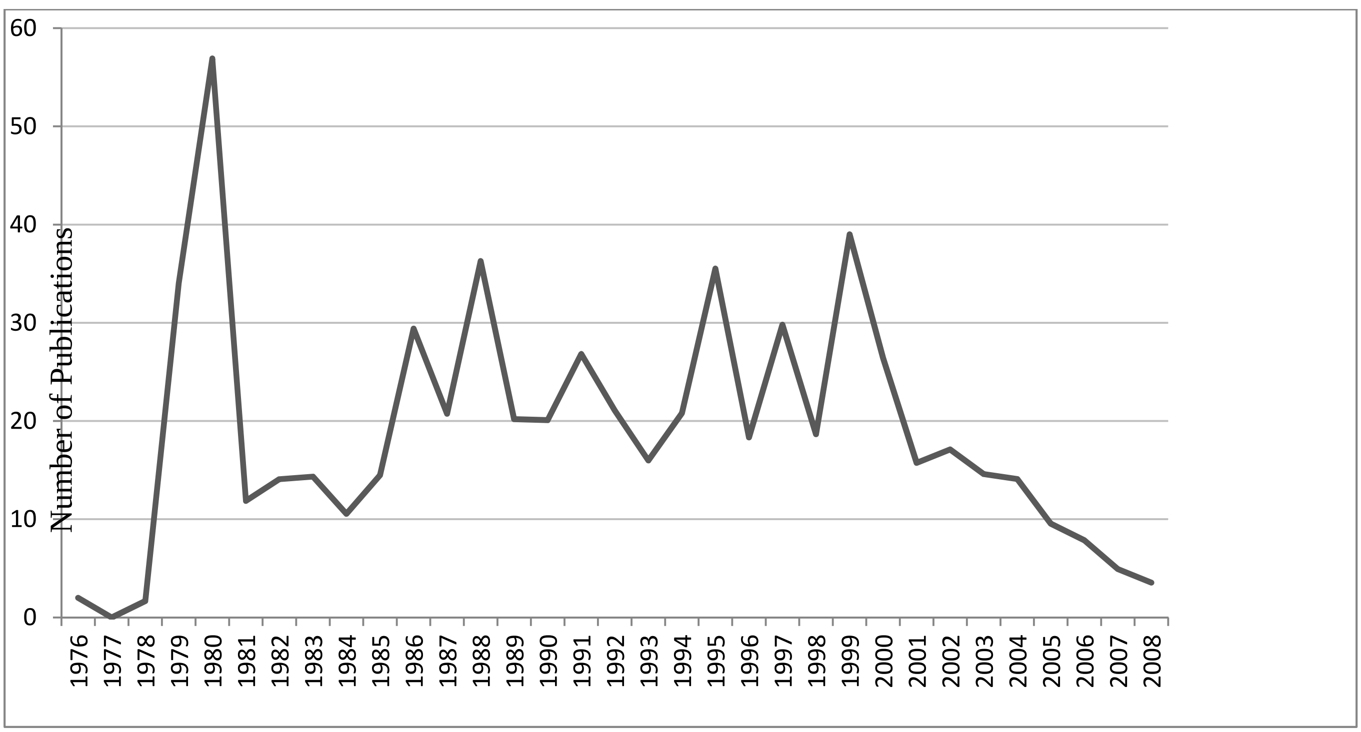


Figure 19. Citation counts of annual publications during 1976-2008

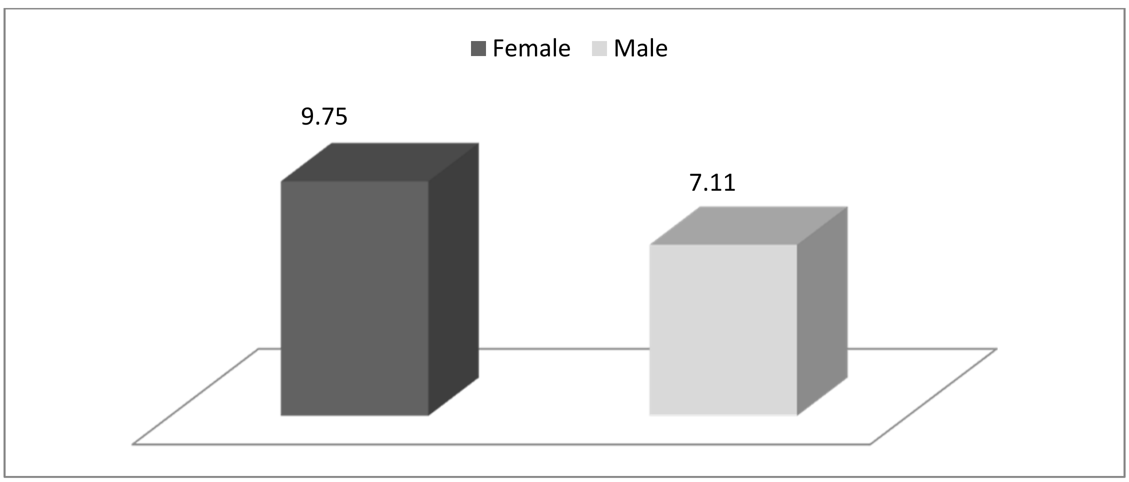


We also calculated the citation counts of annual publications during 1976-2008 and the trend that we found is displayed in Figure 19 indicating that for some years the average number of citations is

considerably higher up to 57 in 1981, then remains in an average of 20 to 40 citations per year, but at the end of the series after 2002 this tends to decline.

When making the analysis dividing the data by gender, the number of per-capita publication produced by male authors is slightly higher (12 per person) than the one produced by female authors (9 per person). Additionally, when searching the information of female authors, some of them were not found by the software, a possible explanation is the fact that some women will change their names when they change their marital status. When considering the average number of citation counts by publication is 9.75 for women and 7.11 for men (Figure 20). Papers published by women on average get cited more frequently than those published by men. If the impact of a publication can be measured by number of citations of that particular paper, then in average publications produced by INTSORMIL female scientist have higher impact in the academic world.

Figure 20. Average citation per publication by gender 1976-2008



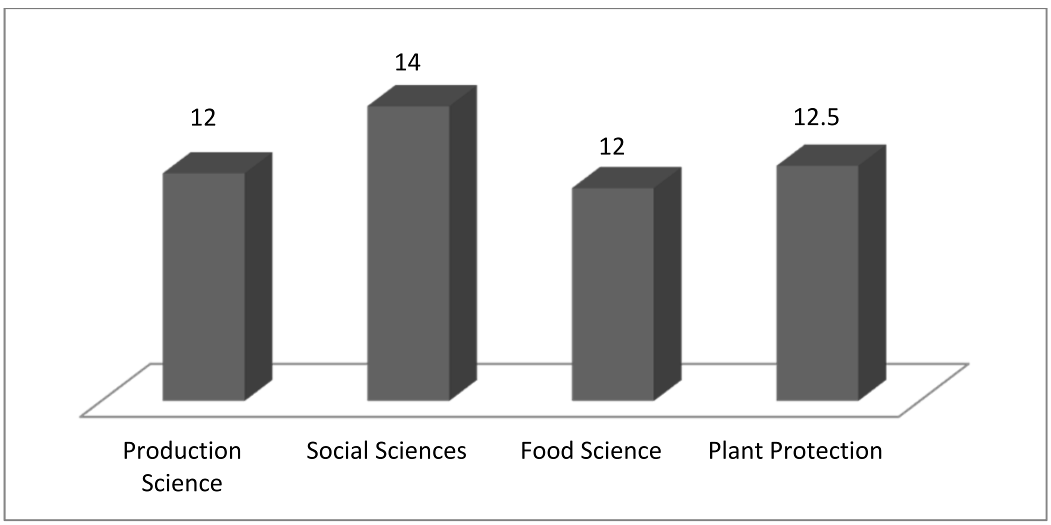
Making the comparison across degrees we found that papers published by Post Doc scientists have been cited more times than the rest of the trainees that obtained other degrees. The second group whose papers were cited the most is the scientist trained as PhDs, with 7 citation by paper on average (see table below).

Table 9. Citation counts and Total number of papers by degree 1978-2008.

Degree	#of authors	Total # of citation	Total # of Papers	# of Citation/ Total # of Papers
VS	61	3,132	678	4.62
PhD	304	31,936	4,607	6.93
PD	70	9,096	898	10.13
BS	17	647	104	6.22
MS	311	9,634	2,576	3.74

As one of the variables that we considered is type of discipline in which the scientist has been trained, we grouped these disciplines in five sets as mentioned before. The findings by discipline show that among disciplines the differences in the per-capita number of publications is fairly small. Social Science scientists have a higher per-capita number of publications, 14 publications per scientist. Production Science, Food Science, and Plant Protection Science have about 12 publication per scientist trained. These four disciplines represent 95% of the scientists in the sample (Figure 5).

Figure 21.Number of Publication per-capita by discipline 1976-2008



When looking at the number of citations per paper in each discipline, the results show that scientists that have been trained in Food Science have been cited almost twice more than the rest of the scientists. We found a lower result for scientist trained in Animal science. (Table 8).

Table 10. Citations and Total number of papers by discipline 1976-2008

Disciplines	Total # of citation	Total # of Papers	Citations per paper
Animal Science	40	44	0.91
Food Science	20,822	1,729	12.04
Plant Protection	11,166	1,910	5.85
Production Science	25,573	3,919	6.53
Social Sciences	4,575	783	5.84

We constructed Gini coefficients of publication and citations rather than the calculation of an arithmetic mean value (as we did with the data when was sorted by gender), since the Gini coefficient is a measure of inequality, defined as the mean of absolute differences between all pairs of individuals for some measure, it is more reliable than a simple arithmetic mean value because in a cross-discipline analysis the arithmetic mean value can lead to wrong finding since each discipline has its own publication patterns and citations habits (Bornmann,L. 2008). Consequently, one measure of concentration should be computed in order to distinguish between research groups with 'collective strength' and groups with 'individual strength' (Daniel & Fisch 1990, Burrell 2006).

The Gini coefficients found for all regions are in average .57 as reported in table 15. No important differences were found among the Gini coefficients of the regions of origin of the scientists. This led us to think that this group of scientist when divided by regions does not show a high concentration in their publications, meaning that the effect of "star" scientist is not significant. This might imply that the number of publications or citations is relying on only few predominant scientists that carry the effort in research productivity. We do not find evidence of a "star scientist". On the contrary, the average Gini coefficient of 0.57 indicates that almost 50% of the publications are being produced by the 50% of the researchers. This is consistent across regions (table 9).

Table 11.Gini Coefficients of Publications by region 1976-2008

Region	Gini Coefficient
Africa	0.54
Asia and Near East	0.55
US	0.58
Central and South America	0.57
Europe	0.57

Bibliometrics and Index Characteristics

A total of 8871 publications produced by 766 international agricultural scientists was found in PoP. As noted in table 16, the mean number of total publications is 11.58 (95% Confidence Interval 10 to 12).

Mean number of total citation was 88 (95% Confidence Interval 71 to 105). Since each index corrects for different patterns of publication, and scientific characteristics; it is always prudent to use several indicators to measure research performance; therefore we calculated six bibliometric indexes as well as the average for the different disciplines scientists were trained in.

Table 12.Research Bibliometric indexes and statistics

Variable	Observations	Mean	Std. Dev	Min	Max
h-index	766	3	3	0	27
g-index	591	6	6	1	55
m-quo	766	0.16	0.17	0	1.42
h-cont	766	2	2	0	18
a-index	752	13	21	0	251
h2-index	766	1	1.4	0	10
Citations	766	88	239	0	3067
Total Publications	766	11	16	1	203

Comparison of indexes for the different regions, gender, or degrees is not generally useful to make (Thompson D, et all. 2009). We worked with the averages of indexes for the entire sample, and average indexes calculated by disciplines. Therefore, the total average of the H-index for the group of scientist observed is 5, which means that at least on average an INTSORMIL scientist has published five papers that have been cited 5 or more times each, and the rest of articles have fewer than 5 citations. By discipline, the highest average h-index reported is from Social Sciences scientists, and Plant Protection Science scientist with 11 and 12 respectively. Social Science scientists and Plant Protection scientists have published on average 11-12 publications that at least have been cited each 11 times or 12 times respectively. This group of scientists represents around 30% of total authors of the sample (n= 215). The rest of the authors in the other two disciplines reported a lower average h-index than the total average (Table 11).

Table 13. Average indexes divided by discipline

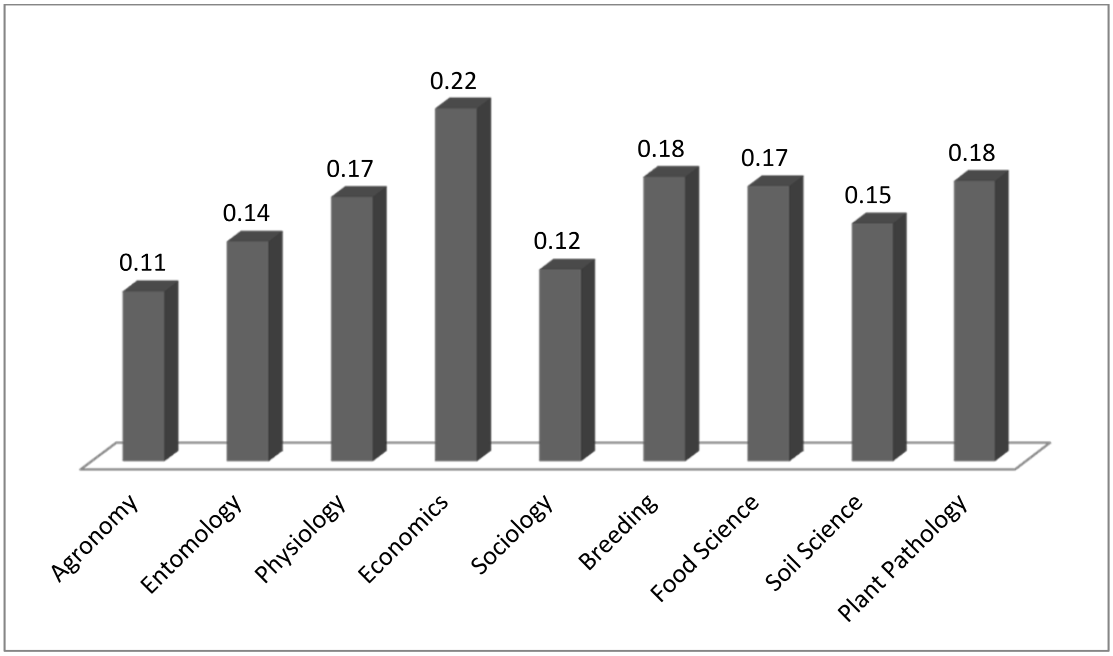
	H-index	G-index	M quo	A-index	H(2)	H-Conte
Production Science	3	4	0.15	9.75	1.52	1.52
Social Sciences	7	11	0.34	22.01	3.72	3.56
Food Science	3	6	0.17	17.99	1.95	2.09
Plant Protection	8	12	0.60	29.45	4.68	5.00
Total average*	5	8	0.31	19.79	2.97	3.04

*The total average is of the 755 scientists, since 5 scientists did have a discipline register as “other”.

For the h-index once a paper belongs to the top h papers, its subsequent citations no longer count, to correct for that, we calculated the g-index. On average the 8 most cited papers of researchers funded by INTSORMIL have received together at least 64 citations. The average g-index by discipline shows that Plant protection and Social science scientists report the highest number with 11-12 respectively. That means that in average 11-12 of the most cited papers of each of these scientist have received together more than 121-144 citations. Production Science has the lowest g-index of 4, a little less than the total average.

To adjust for career length among the scientist, Figure 22 depicts the average of the m-quotient by discipline, with exception of the Social Sciences (Sociology and Economics) and Agronomy, the rest of disciplines have m-quotients very similar from 0.11 to 0.22, meaning that a comparison of the H-index between scientists that differ in seniority is fair since almost all m-quotients are similar.

Figure 22. Average M-quotient of scientist work by discipline 1978-2008



Also, two of the most cited papers of each scientist at least each of them have been cited 4 times in average (h-cont.). To control for scientist that have had few papers but highly cited ones, we used the A-index to count for excessive citation in the work of a scientist, since only uses the Hirsch core (that is the set of highly-cited publications). For the 766 scientist in the sample, an average of the A- index was 19.36. This means that the average number of citations of papers in the Hirsch core of the scientist is about 19 times. That number shows the impact of a researcher's highly cited papers. When looking by disciplines, this A-index is high for the Plant Protection scientists and Social science scientists with 29.4, and 22 respectively. This numbers is higher than the average; on the other hand the lowest number report is 9.75.

For the calculation of scientists that are considered top-scientists the h2-index was considered. According to the h2-index, in the INTSORMIL program the fields with the top scientists are in Plant Protection and Social Sciences with a h2-index of 4.57, 3.72 respectively. These two disciplines are above the total average which is 2.97.

We used the h-contemporary index to see who has been active in publications, active in the sense that have published frequently, and also have been cited frequently. The agricultural scientists who work in disciplines related to Plant Protection are the most influential, follow by the Social Science scientists, a h-contemporary index of 5 and 3.56 were calculated for these two groups respectively. Five of the most cited papers of each scientist at least each of them have been cited 25 times for the Plant Protection disciplines, and 9 for the Social science. The average h-contemporary value is 3 for the 766 scientists. On average 3 of the most cited papers of each scientist at least each has been cited 9 times. If we take the lowest and the highest of theses h-contemporary index there are not considerable differences between groups of disciplines, confirming again that all 766 scientists funded by INTSORMIL have been productive, and there is not a particular group whose performance is noticeably low.

LIMITATIONS

The findings of this study provide some insight on research productivity in a group of trainees supported by the INTSORMIL program. Yet there are some cautions that need to be considered when interpreting the results.

First, productivity of scientists is a multidimensional concept. While publications and citations indexes are an important dimension of productivity, it is only one of several major outputs (e.g., teaching, public service, and outreach, etc.) in most research international institutions, and some of these other outputs might have higher social or political priorities at times, especially at institution in less develop countries. We should mentioned the substantial informal flow of knowledge through more personal (face-to-face) collaborations, that is not officially document in scholarly documents that is part of the production of knowledge. Also, an increase in bureaucratic loads in some research institutions, for example, is likely to lead to reduced research performance due to time constraints or personnel and economic resources. Hence is important to mention that the results presented in this study tend to underestimate the production of knowledge from the researchers as well as their productivity.

Second, even though publications often appears to be an "easy" variable of research's output, there are still some serious difficulties in measuring research performance using only publications and citation counts, but this project is an attempt that tries to have a more comprehensive way to measure this type of output by avoiding any participation of the actual subjects of research.

Third, there are multiple forms of research outputs such as journal articles, books, book chapters, monographs, unpublished conference presentations, and even computer software. This study used the information obtained by PoP journal articles as its main source of research output since data related to other forms were not available. It also must be acknowledged that the current study can have several biases because the use of journal article counts that were obtained came exclusively from Google Scholar as its measure of output and articles. Additionally, the majority of the publications that were taken into consideration are publications related mainly to sorghum, millet, maize. By selecting only these particular key topics we are restricting the vast ability of a scientist to make contributions in other crops or fields that are not necessarily related to these crops.

Four, the control for quality of those publications was done by removing articles and material that were not published in a journal (presentations, brochures, posters) but the quality of the journals was not considered in this study.

Last, the data available for this study limit the development of a more comprehensive research productivity model since there is no control group to compare. We did not have a control group due to the fact that scientists trained by INTSORMIL vary in nationality, discipline, age, gender and degree, as well as time when trainees where in the program. These facts make hard to find names and information of a proper control group.

CONCLUSIONS

The application of citation counts and number of publications is a new powerful set of tools for empirical analysis, and for the evaluation of scientists to measure their contribution to the creation of knowledge, and the advance of a particular discipline.

This manuscrip reports the impact of INTSORMIL funded scientists from 1976 up to 2010, for different agricultural disciplines. We showed that INTSORMIL resources invested in training of agricultural scientists provides continuous and prolific publications and academic materials in international scientific communities, especially in the Social Sciences and Plant Protection fields. This positive impact has being in both quantity (number of papers) and quality (number of times the paper has been cited) of academic research materials. We also found that despite the decreasing number of female scientists they present a slightly higher quality research output measure by citation counts compared to male scientists.

We argue that there is a little concentration in the contribution of publication and citation counts among scientists. On average the productivity in terms of publication is similar among all trainees. Finally trends in publications and citation counts have increased over time except for a decline number in the last three years. This decreasing number coincides in part with a decreasing in the number of students being trained by INTSROMIL. This is also due to the “young” age of many publications recently produced that have not had time to accrue impact.

In addition, in the recent data base we can add supplementary information such as: number of papers for which the scientist is the first author, type of journal in which they published, language in which they published. By incorporating this new information a study could measure in a more precise way the productivity of the authors and could analyze patterns in the coauthors of publications of papers as well as a selection on the importance of the journal. A further study using a control group that can be created by looking at co-authors in each paper or looking at scientists that had the same advisor and graduation time, would give a more plausible way to assess the impact of training of INTSORMIL.

REFERENCES

Adams, J., and Z. Griliches. 1996. Measuring science: An exploration. Discussion Paper No. 1749, February. Cambridge: Harvard Institute of Economic Research, Harvard University.

Alston, Julian M., Norton, George W., and Pardey, Philip G. 1998. Science Under Scarcity: Principles and Practices for Agricultural Research Evaluation and Priority Setting. CABI International in association with International Service for National Agriculture Research (ISNAR).

Bornmann, L. 2009. The state of h index research. is the h index the ideal way to measure research performance? EMBO Reports 10 (1): 2.

———. 2008. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. Journal of the American Society for Information Science and Technology 59 (5): 830.

———. 2008. Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics in Science and Environmental Politics 8 (1): 93.

Breschi, S. 2011. Assessing the scientific and technological output of EU framework programmes: Evidence from the FP6 projects in the ICT field. Scientometrics 88 (1): 239.

Daniel, H. D. 1990. Research performance evaluation in the german university sector. Scientometrics 19 (5-6): 349.

Egghe, L. 2006. Theory and practise of the g-index. Scientometrics 69 (1): 131.

Gordon, Jenny and Chadwick, Kevin. 2007. Impact assessment of capacity building and training: Assessment Framework and Two Case Studies. Australian Centre for International Agricultural Research. Impact Assessment Series Report No. 44.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102 (46): 16569.

Harzing, A.W. (2011) *Publish or Perish*. Version 3.1, available at www. Harzing.com/pop.htm

INTSORMIL Sorghum, Millet and Other Grains CRSP ANNUAL REPORTS (1980 through 2010) INTSORMIL Publication 10-01, University of Nebraska, Lincoln, Nebraska

Lee, Sooho and Bozeman, Barry. 2005. The Impact of Research Collaboration on Scientific Productivity Source. Social Studies of Science 35 (5): 673- 702

Metze, K. 2010. Bureaucrats, researchers, editors, and the impact factor: A vicious circle that is detrimental to science. Clinics 65 : 937.

Reskin, B. F. 1978. Scientific productivity, sex, and location in the institution of science. American Journal of Sociology: 1235.

Rigby, J. 2009. Comparing the scientific quality achieved by funding instruments for single grant holders and for collaborative networks within a research system: Some observations. Scientometrics 78 (1): 145.

Sorghum, Millet, and Other Grains CRSP. Collaborative Research Support Program. From http://intsormil.org/index.htm (accesed July 17, 2010).

Stata FAQ. How can I quickly convert many string variables to numeric variables? UCLA: Academic Technology Services, Statistical Consulting Group. From http://www.ats.ucla.edu/stat/stata/faq/destring.htm (accessed July 17, 2010).

Stata FAQ. How can I turn a string variable containing dates into a date variable Stata can recognize? UCLA: Academic Technology Services, Statistical Consulting Group. From http://www.ats.ucla.edu/stat/stata/faq/string_dates.htm (accessed July 17, 2010).

Stata FAQ. What statistical analysis should I use? Statistical analyses using Stata. UCLA: Academic Technology Services, Statistical Consulting Group. From http://www.ats.ucla.edu/stat/stata/whatstat/whatstat.htm#chisq (accessed August 24, 2010).

Sidiropoulos, A. 2007. Generalized hirsch h-index for disclosing latent facts in citation networks. Scientometrics 72 (2): 253.

Thompson, D. F. 2009. New indices in scholarship assessment. American Journal of Pharmaceutical Education 73, (6).

APPENDICES

Appendix 1.

Table 14. Information available in the FileMaker database

Variable	Description	Type of variable
Name	Complete identification of the trainee (name and last name)	Ordinal
Gender	Male or Female	Binary
Degree*	Five categories: Bachelor (BS), Master's (MS), Doctorate (PhD), Post-doctorate (PD), Visitor Scholar (VS)**	Categorical
Region	Geographic region of origin of trainees - continent/subcontinent	Categorical
Origin	Trainees' country of origin	Ordinal
University	Host institution where the participant is/was trained	Ordinal
Advisor	Major professor	Ordinal
Discipline	Major field of study: Agronomy, Plant Breeding, Plant Pathology, Sociology, and others	Categorical
Subdiscipline	A minor discipline within a major field of study	Categorical
Arrival Date	Starting date on the program (month, year)	Ordinal
Depart Date	Finishing date on the program (month, year)	Ordinal
Funding	Whether trainee is totally or partially funded by INTSORMIL	Binary
Labor information	Information of the trainees' job at their home country	Ordinal
Permanent address	Trainees' address at their home country	Ordinal
E-mail	E-mail account address	Ordinal

* The variable Degree refers to the "type of program" on which the trainees were involved on. Then, type of program can be classified in two categories: Degree and Non-degree. Within the Degree three categories are considered: Bachelor, Master's, and Doctorate. On its part, the Non-degree type of program contains: Post doctorate and Visitor scholar (Diane Sullivan, Personal Communication).

** Visitor scholars can be classified as short term and longer term visitor scholar, depending on the length of the training (Diane Sullivan, Personal Communication).

Table 15. Information added in the new data base

Variable	Description	Type of variable
Year of publication	Year when a particular article was published	Ordinal
Journal	Name of Journal where the article was published	String
Citation	Number of times a paper has been cited	Ordinal
Author	Names of the author or authors of a particular paper	String
Title	Title of the article	String
J	Assigns the ordinal number of papers an author has	Ordinal
Acc j	Cumulative number of times papers of the same author has been cited.	Ordinal
H-index	The calculation of the h-index per author	Ordinal
G-index	The calculation of the g-index per author	Ordinal
M-quo	The calculation of the m-quotient index per author	Ordinal
H-cont	The calculation of the h-contemporary index per author	Ordinal
Sc(i) *	novel score	Ordinal
A-index	The calculation of the A- index per author	Ordinal
H(2)-index	The calculation of the Hsquared- index per author	Ordinal
Total publications	Total number of publications per author	Ordinal
Patents	Total number of patents per author	Ordinal
Top journals	Whether the article was published on a top journal or not.	Binary

* "*Sc(i)* is the number of citations that the article *i* has received, divided by the "age" of the article." (http://sci2s.ugr.es/hindex/ 2010)