Measuring the Impact of the International Fellowships Program (IFP) in the Area of Education: *Development of a Valid and Reliable Measurement Instrument*

ASSANE KONE

Towards fulfillment of the Master degree in Educational
Science and Technology (EST)
Track: Educational Management Evaluation and Assessment
(EMEA)

Twente University – Enschede the Netherlands
Faculty of Behavioural Sciences

Mentors:
Drs. M.A. Hendriks
Dr. M.C.M. Ehren

June, 2013

Acknowledgements

Dedication

This work is dedicated to my mother, my brothers and sisters, and to my children (Sidian and Yanisa). To all of you I say sorry – sorry - sorry.

Mom I could not make it to see you before the almighty God decided to call you and to show you my children. Sorry! Things do not work out the way we hoped since we last met in the village. I still remember your words. I will do my best for the family. Rest in peace.

To my brothers and sisters I say sorry for the long time I spent in the Netherlands and therefore not being actively involved in family affairs. Thanks for your comprehension.

To Sidian and Yanisa I say sorry. Sorry for not being there when you needed me. Sorry for not being the father I should be. I ask for your forgiveness and hope to get a second chance.

Abstract

This present study was conducted on behalf of the Center for Higher Education Policy Studies (CHEPS) as a pilot study in order to develop a valid and reliable instrument that could measure the impact of the International Fellowships Program (IFP) in the area of education of countries selected in the program. IFP is an educational program that provides fellowships for graduates who lack access to post-graduate education and who are resident of developing countries or territories.

The instrument (IFP Ed-instrument) was developed through four steps. In the first step, 11 indicators and 23 variables were pre-selected from literature and interviews conducted with academics and practitioners specialized in the field of education as well as experts in the field of impact assessment. A two rounds Delphi method was used in the second step to evaluate the relevance and measurability of the pre-selected indicators and variables. Evaluations were conducted with experts through mail questionnaires in the first round and through structured interviews in the second round. Content validity indexes (I-CVI) and content validity ratio's (CVR) were then computed for each of these individual indicators and variable. The content validity of the overall IFP Ed-instrument was estimated in the third step by experts through mail questionnaires. The Fleiss multiple rater kappa was then computed to estimate the degree of agreement between experts during the third step. Following refinement the IFP Ed-instrument was further field tested through mail questionnaires in Senegal in the fourth step.

Results of this sudy show that seven indicators and 21 variables are content valid measures of outcomes and impacts of IFP with I-CVI values higher than 0,78 and CVR values higher than 0,75. The developed IFP Ed-instrument was found content valid by consulted expert with a multiple rater Fleiss kappa value of 0,36.

It is recommended to further field test the IFP Ed-instrument in triangulation studies in order to establish its construct validity and its reliability.

Table of contents

List of figures and tables

Appendix G

Appendix K

Abbreviations

AFGRAD	African Graduate Fellowship program
ATLAS	Advanced Training for Leadership Program
CFSP	Canadian Francophone Scholarship Program
CHEPS	Center for Higher Education Policy Studies
CIDA	Canadian International Development Agency
CIPP	Context Input Process Product
CREAM	Clear Relevant Economic Adequate and Monitorable
CVR	Content Validity Ratio
DANIDA	Danish International Development Agency
EUROSTAT	Statistical Office of the European Union
I-CVI	Content Validity Index of individual variables
IFP	International Fellowships Program
INES	OECD Indicators of Education Systems programme
ISCED	International Standard Classification of Education
JJ/WBGSP	Joint Japan/World Bank Graduate Scholarship Program
JSP	Japan Scholarship Program
NGO	Non-Governmental Organization
NORAD	Norwegian Agency for Development Cooperation
NUFFIC	Dutch Higher Education Organization
OECD	Organization for Economic Co-operation and Development
PISA	Program for International Student Assessment
S-CVI	Content Validity Index of the overall instrument
SIDA	Swedish International Development Cooperation Agency
SMART	Specific Measurable Attainable Relevant and Time-bound
UIS	UNESCO Institute for Statistics
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
USAID	United States Agency for International Development
UOE	UNESCO-UIS/OECD/ EUROSTAT
WEI	World Education Indicators

Introduction

Fellowships are programs that provide training to individuals in specific fields of studies that are relevant to them and that match with the training needs of their countries (Aguirre International, 2004; Boeren, 2005; CIDA, 2005; Rotem, Zinovieff & Goubarev, 2010; USAID, 1995). The general objective of fellowship programs is to help recipient countries to address their economic and social development issues by increasing their pool of qualified people (Strömbom, 1989). Donors invest high costs in fellowships and expect alumni to strengthen organizations' and institutions' performances beyond individual benefits from training (Boeren, Bakhuisen, Christian-Mak, Musch & Pettersen, 2008). Despite high investment, evaluation of fellowship is still the "Achilles heel" of many organizations (Rotem et al., 2010; Krasulin, Ouedraogo & Quijano, 1998). Few evaluations of scholarship and fellowship schemes have been conducted by donor agencies around the world (Boeren, 2005; Norad, 2005; Rotem et al., 2010). External reviews or evaluations are seldom undertaken and only tracer studies that focus, for example, on alumni degree attainment and alumni return in own country are occasionally conducted (Boeren, 2005). Little is therefore known about evaluation of the impact of fellowship programs (Boeren, 2005; Norad, 2005; Lamont, 2002; Searle, Hatem, Perkowski & Wilkerson, 2006). However, in recent years evaluation of fellowship programs received more attention by agencies because of accountability demands of donors and in order to improve the effectiveness and efficiency of the programs (Boeren et al., 2008; Zinovieff, 2008). This present study builds on evaluation practices of fellowship donor agencies in order to develop a valid and reliable instrument that could assess the impact of the International Fellowships Program (IFP) in the area of education.

The International Fellowships Program (IFP)

The International Fellowships Program (IFP) is an educational program that provides fellowships for graduates who lack access to post-graduate education and who are resident of one of the 22 developing countries or territories selected by the program. IFP was launched in 2001 with a grant of the Ford Foundation (IFP, n.d.a). From 2001 to the time this research was conducted, IFP has given fellowships to more than 4300 academics (IFP, n.d.d). The core objective of the International Fellowship Program (IFP) is to "provide opportunities for postgraduate study to outstanding individuals from social groups and communities that lack systematic access to higher education in IFP countries and to enable them to become social change leaders in their fields" (Enders, Theisens & Westerheijden, 2005, p.7). IFP expects to address social injustice, to promote community development, and to promote access to higher education in its selected countries (IFP, n.d.b). IFP – through organizations/institutions in selected countries - makes an effort to reach out and to select potential future leaders and role models, male and female, from social groups that otherwise lack opportunities to accede higher education (IFP, n.d.c). In doing so, the program aims to diversify the "leadership pool with individuals from marginalized and disadvantaged communities" (Enders & de Boer, 2002, p.33) with the following goals in mind:

At individual level
- Provide opportunities for advanced study to excluded individuals who will use this education to become leaders in their respective fields in their home countries

At institutional level
- Strengthen the development of organizational networks for educational service provision
- Have an impact on other fellowships programs

At system level
- Further development in the fellows' home countries
- Enhance economic and social justice worldwide
- Stimulate the public debate on access to higher education in the fellows' home countries
- Contribute to research and public policy on social returns to higher education
- Diversify the leadership pool with individuals from marginalized and disadvantaged communities
- Strengthen the link between education and development
- Promote global citizenship.

To meet such goals, the program targets graduates from disadvantaged communities who commit themselves to address social issues in their communities and who have some leadership capacity. Therefore, selected fellows are individuals with some experience in community service or development-related activities; that possess some leadership potential; and that clearly state to serve their communities and countries of origin after the end of the fellowship (IFP, n.d.b). Further, IFP enables these fellows to get access to post-graduate education in higher education institutions of their own choice worldwide including higher education institutions in their country of residence. After their study, IFP stimulates and assists alumni to return to their communities or countries of origin (IFP, n.d.c). Alumni who return are expected to "use their leadership skills and knowledge to work toward positive change in their home communities and countries" (IFP, n.d.b). IFP enables alumni to become people that foster social changes in their own communities or countries because "trained people who are also socially committed and morally responsible leaders can identify and address urgent needs in their communities" (IFP, n.d.b). IFP provides opportunities for academic trainings to individuals and expects that they will use the acquired knowledge to address issues targeted by the program in their home countries.

Implementation of IFP in the selected countries is evaluated by the Center for Higher Education Policy Studies (CHEPS) (Enders & de Boer, 2002). CHEPS is a research institute located at the School of Management and Governance of the University of Twente that conducts research, and provides education and training on higher education.

Practical inducement for the study

Since the launch of the IFP program in 2001, different formative evaluations have been conducted by CHEPS in order to measure whether or not the program has been successfully implemented. Evaluations pointed out that implementation of IFP is a success. IFP is able to target and select fellows from marginalized and disadvantaged groups; IFP is able to place fellows in universities of their own choices; the majority of alumni return to their country of origin; and fellows and alumni are satisfied with the program (Enders, Kottmann & Leisyte, 2007). IFP has thus been able to empower people needed to foster social changes in countries selected in the program. However, the impacts of IFP (defined in this study as changes in IFP countries caused by activities that IFP alumni conduct) in the selected countries are not yet investigated. After a high investment (IFP, n.d.d) and a successful implementation, IFP is willing to investigate whether alumni are bringing changes in their own countries, that is, whether such an investment was worthwhile. This present study was conducted on behalf of the Center for Higher Education Policy Studies (CHEPS) as a pilot study in order to develop a valid and reliable instrument that could measure the impact of IFP in the area of education in its selected countries.

The central research question

With the background information mentioned above, the central research question was formulated as follows: *How to develop a valid and reliable instrument that can measure the impact of the International Fellowships Program (IFP) in the area of education in its selected countries?*

This study was performed in an exploratory way as few is known about the assessment of the impact of fellowship programs (Boeren, 2005; Norad, 2005; Searle et al., 2006) and a generally accepted instrument for such assessments is not available in the literature. Exploratory studies are useful to explore new phenomena as well as to develop and test instruments (Fraenkel, Wallen & Hyun, 2012).

To answer this central question, three additional questions were elaborated:

1- *What frameworks are available in the literature for the assessment of the impact of educational programs and to which extent are these frameworks appropriate for the assessment of the impact of IFP?*
2- *What kinds of outcome and impact indicators are necessary to assess the impact of IFP in the area of education?*
3- *How could the quality in terms of validity and reliability of the developed instrument be tested?*

Within this study, outcomes refer to activities that alumni conduct and impacts refer to changes in IFP countries caused by the conducted activities in the area of education. The term indicator refers to any quantitative and qualitative variables appropriate to measure the outcomes and impacts of IFP in its selected countries. Instrument validity and reliability are technical properties that indicate the quality and usefulness of instruments (Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000; Scheerens, Glas & Thomas, 2003). Reliability is similar to reproducibility or consistency of results provided by the measurement instrument and validity is the extent to which an instrument is able to measure well what it is intended to measure (Litwin, 1995; Fraenkel et al., 2012).

How the central question was answered

Four steps were used to answer the research questions (see figure 1.1). The first step was conducted to answer the first research question. The second step was conducted to answer the second research question. The third and fourth steps were conducted to answer the third research question.

In the first step, explorative orientation interviews followed by a literature study were conducted. The aim of the orientation interviews was to understand how impact of fellowships programs could be assessed. Therefore, semi-structured interviews were conducted with academics and practitioners specialized in the field of education as well as experts in the field of impact assessment. Based on the findings of these interviews, a literature study related to evaluation of training in the context of fellowship programs was conducted in order to find out a framework that could be used to assess the outcomes and impacts of IFP and to identify and pre-select indicators and variables that could be used to assess the outcomes and impacts of IFP. During this literature review the focus was on indicators and variables included in available instruments that have been validated in the context of fellowship programs but also on indicators and variables, wherever possible, with evidence of validity and reliability as recommended by literature (Cook & Beckman, 2006; Moore & Benbasat, 1991; Straub, Boudreau & Gefen, 2004). During the second step, final indicators and variables were selected from the pool of pre-selected indicators and variables. Therefore, experts in the area of education and fellowship programs were asked through a two rounds Delphi study to evaluate the relevance and measurability of the pre-selected indicators and variables and to indicate whether the indicators and variables were either essential or not

essential to be included in the draft of the instrument. This evaluation aimed to establish content validity of individual indicators and variables. The received feedback from these two steps led to the development of the draft of the instrument. In the third step, the draft of the instrument was validated by experts. The aim of this step was to establish content validity of the overall instrument. Therefore, the overall instrument was evaluated for clarity and completeness, that is, whether items should be deleted or added. Once validated, the instrument was field tested during the fourth step in Senegal. The aim of this field testing was to check the clarity and the length of the instrument. For that purpose IFP alumni from Senegal graduated in the area of education were asked to evaluate the instrument.
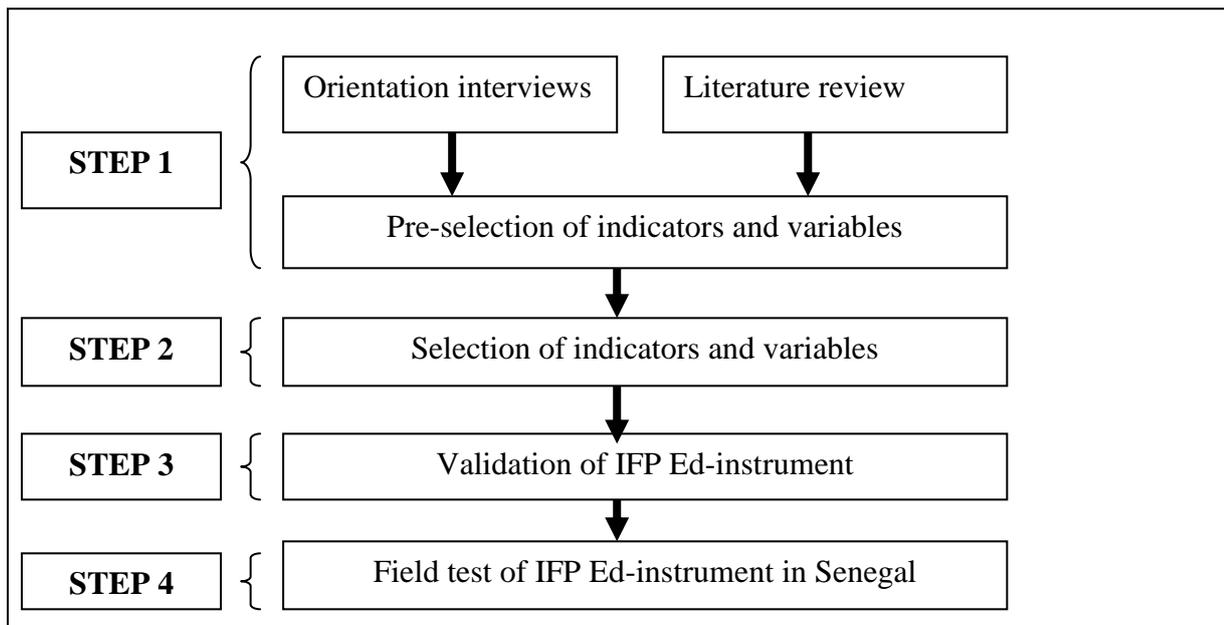
Figure 1.1. Flowchart development IFP Ed-instrument

Purpose and significance

The purpose of this study was to develop a valid and reliable instrument that could measure the impacts of IFP in the area of education in its selected countries. The final product of this study is therefore a list of indicators and variables and an instrument (hereinafter referred to as IFP Ed-instrument) that could be used to assess the impact of IFP in the area of education. This study is the initial stage of the development process of the IFP Ed-instrument. For further improvement of the developed instrument, it is suggested to use the results of this study in triangulation studies in order to establish consistency of the instrument during field tests.

The scientific significance of this study is that it could provide more insights about the assessment of fellowship programs as little is known about the assessment of the impact of such programs (Boeren, 2005; Norad, 2005; Searle et al., 2006). Additionally, this study might constructively contribute to the improvement of impact assessment practices of fellowship programs. It is acknowledged that the developed IFP Ed-instrument may not be applicable to all fellowship programs since the objectives of the programs may differ from each other. Nevertheless, this IFP Ed-instrument can be used as a model for further studies that aim to develop instruments to assess impact of fellowship programs particularly in the area of education.

Overview of the study

This study includes six chapters. The first chapter describes the context within which this study was conducted. It briefly outlines the results of different formative evaluations

conducted by the Center for Higher Education Policy Studies (CHEPS). Literature is reviewed in the second chapter in order to find out frameworks appropriate to assess the impact of educational programs. The conceptual framework within which indicators and variables were pre-selected is presented in the third chapter. The first step of the development process of the IFP Ed-instrument is described in the third chapter as well. Chapter three outlines the criteria used during pre-selection and describes the pre-selected indicators and variables. The research methodology is described in chapter four. The chapter includes the procedures, the instruments and the sample used during each step in order to develop the instrument. The findings of this study are presented in chapter five, followed by the conclusions and recommendations of the study in chapter six. This last chapter six also describes the limitations and areas for further research.

1. Context of the study

Since the start of the IFP program in 2001 many fellows have been selected and placed in different host institutions around the world. The overall mid-term achievements of IFP at fellows and alumni level are evaluated and reported frequently through formative evaluations conducted by the Center for Higher Education Policy Studies (CHEPS) of the University of Twente.

1.1 The evaluation framework of IFP

The evaluation framework of IFP was developed in 2003 by CHEPS in close cooperation with the IFP secretariat and other key stakeholders such as the IFP country partners (Enders & de Boer, 2002). The evaluation framework is an "input-throughput-output" model (figure 1.2) which attempts to relate the activities conducted by IFP to the outputs, outcomes and impacts of the program in its selected countries (Enders & de Boer, 2002, p.9).
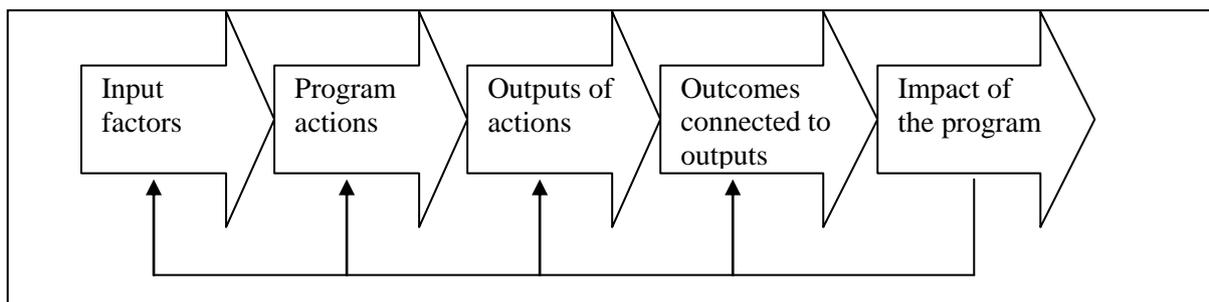
Figure 1.2. The evaluation-chain: basic blocks of the evaluation framework (source: Enders & de Boer, 2002, p.9)

In this framework, input factors refer to the resources generated by IFP to undertake the program's actions (e.g. financial, fellows etc.). Program actions refer to all conducted activities (e.g. recruitment, selection, placement etc.). Outputs of actions refer to "the direct products of the IFP-actions" (Enders & de Boer, 2002, p.13). Examples of outputs include among others the characteristics of selected fellows; placement of fellows; graduation rate of fellows (degree and field of study); the return rate of alumni in their countries etc. (Enders & de Boer, 2002). Outcomes refer to "the changes in the attitude, behaviour, functioning and performance of the service-takers" (Enders & de Boer, 2002, p.13). Three types of outcomes are distinguished in the evaluation framework of IFP: short-term (1-3 years), medium-term (4-6 years), and long-term (more than 7 years) (Enders & de Boer, 2002, p.13). Finally, impact of the program refers to the long-term outcomes, that is, "the fundamental change in organizations or communities as a consequence of the outcomes" (Enders & de Boer, 2002, p.14). Examples of impacts include increased social justice, more debate on access to higher education etc.

Within this framework, several formative evaluations were conducted by CHEPS in order to measure the mid-term outputs of IFP. These outputs are prerequisites needed for IFP in order to determine if an impact occurred in its selected countries.

1.2 Mid-term outputs of IFP

IFP is able to target and select graduates from disadvantaged groups that lack opportunities to accede higher education (Enders et al. 2007; IFP, n.d.e). According to IFP (n.d.e) 3,4% of 1482 fellows are people with disabilities; 80% of over 1000 Latin America fellows are indigenous or afro-descendant; about two-third of the fellows from Africa and the Middle

East come from rural areas or small cities and towns; and three-fourth of fellows from Russia/Asia come from rural areas or small cities and towns.

Mid-term outputs of IFP were evaluated by Kottmann & Enders (2011). For this purpose, Kottmann & Enders (2011) analysed 1457 alumni responses from three cohorts of IFP alumni: fellowship ended before 2005, fellowship ended before 2008, and fellowship ended before/in 2010. They found that 91% of alumni attained a degree and only 1% had discontinued their studies; 95% of alumni who pursued a Master's degree obtained a degree at the time of the study against 68% of alumni who pursued a doctoral degree. Female and male did not differ with regards to the completion rate (92% female and 90% male). IFP seemed therefore successful in selecting and matching fellows and study programs (Kottmann & Enders, 2011). Further analyses show that the majority of alumni return to their home countries/home communities after the end of the fellowship; 82% of alumni are living in their home country/home community (47% in home community and 35% in home country) and 18% are living in another country. According to Kottmann & Enders (2011) 69% of alumni were employed at the time of the study. Others were pursuing another advanced academic study or were combining an advanced study and employment.

Formative evaluations also pointed out that fellows and alumni are satisfied with the program and that they have the feelings to have been empowered by the program. According to Enders et al., (2007) "most of alumni agree that the IFP has contributed to strengthen their commitment to social justice, to build intercultural competencies, to develop leadership skills, and to enhance their knowledge about what is needed in their home communities/regions" (p.3). For example an Asian alumnus said: "IFP allowed me to do what I wanted to do. It gave me a lot of faith and satisfaction and I am really happy and proud about myself. … In general I am really happy and IFP really suits me. I improved my personal skills, my professional skills" (Enders et al., 2007, pp.51-52). Similar to this alumnus, an African alumnus mentioned that "… I have been improved much by the program and for me that's just the greatest benefit. In terms of my own self-esteem, my own self-drive, my own self-vision for what I could really do. And my own self-believe in what I could do. That has really been enhanced by the program" (Enders et al., 2007, p.52). Furthermore, formative evaluations pointed out that IFP alumni are conducting different activities in their communities/countries (Enders et al., 2007; Kottmann & Enders, 2009).

1.3 IFP alumni as social change agents

IFP alumni have deliberately committed themselves to serve their communities after the end of the fellowships (IFP, n.d.b). In their area of involvement alumni are acting as so called social change agents. Social change agents are people who make things happen by conducting actions (Bandura, 2001; Havelock & Zlotolow, 1995). According to Havelock & Zlotolow (1995, p.21), a change agent is "someone who tries deliberately to bring about a change or innovation in a social organization".

The social commitment of IFP alumni was evaluated by Kottmann & Enders (2009) and Kottmann & Enders (2011). For this purpose Kottmann & Enders (2009) analysed 856 alumni responses from four different cohorts of IFP alumni: fellowship ended before 2005, fellowship ended in 2005, fellowship ended in 2006, and fellowship ended in 2007/2008. They found that "the vast majority of alumni (88%) employment was related to social commitment" (Kottmann & Enders, 2009, p.10). This percentage was even higher for the recent alumni (fellowship ended in 2007/2008) cohort where 91% were involved in social commitment work. The main sectors of employment are: education (the most important), community development, and environmental issues. Furthermore, IFP alumni (63%) are also engaged in voluntary social commitment (Kottmann & Enders, 2009, p.13). The most important areas of voluntary work are community development, education, and environmental

issues (Kottmann & Enders, 2009, p.13). Important voluntary activities that are conducted include: training (the most important), networking, and the provision of technical assistance (Kottmann & Enders, 2009, pp.13-14). Similar findings about IFP alumni' social commitment are reported by Kottmann & Enders (2011) after analysis of 1457 responses of alumni from three cohorts: fellowship ended before 2005, fellowship ended before 2008, and fellowship ended before/in 2010. Kottmann & Enders (2011) report for example that the majority of alumni are employed in the public sector (54%) and 26% in the non-profit sector and that education (universities or higher education institutions) is the most important sector of employment. 33% of male alumni work in the education sector against 32% of the female alumni (Kottmann & Enders, 2011).

In sum, IFP selects fellows from social groups and communities that lack systematic access to higher education. The majority of IFP alumni attain a higher educational degree and return to their own countries and communities. In countries and communities respectively IFP alumni conduct social commitment activities particularly in the areas of community development, education, and environment. The majority of alumni are employed in paid as well as in volunteer work related to their stated social commitment. Education is the most important sector of employment.

2. Literature review

For this research literature is reviewed in order to find frameworks appropriate for the assessment of the impact of educational programs. Through this review the first part of the research question could be approached: *What frameworks are available in the literature for the assessment of the impact of educational programs and to which extent are these frameworks appropriate for the assessment of the impact of IFP?*

To answer this question, several scientific sources about evaluation of educational programs or fellowship programs were explored to find out how educational programs are evaluated and to find frameworks that are used to assess the impact of educational programs and the impact of fellowships programs in particular. The main sources used in order to find applicable scientific literature were the Twente University library, Internet search engine (e.g. Scopus, PiCarta, and Google scholar), and databases of international development agencies active in the area of education.

In this chapter the results of the literature review are presented. The chapter is divided into three sections. In the first section the most common methods used to evaluate educational programs are explained. Then, in the second section the findings in literature about the availability of frameworks used to assess impact of educational programs are presented and the quality framework developed by Scheerens (2004); the OECD framework; and the four-level framework developed by Kirkpatrick are described. Finally, in the last section of this chapter the answer of the first part of the research question is given.

2.1 Evaluation of educational programs

Educational evaluation literature is explored in this section in order to understand how educational programs are evaluated. This study adopts the definition of educational programs provided by the International Standard Classification of Education (ISCED-97) and used by OECD (2004). According to OECD (2004) an educational program can be defined as "a collection of educational activities which are organized to accomplish a pre-determined objective or the completion of a specified set of educational tasks" (p.80). Educational activities could refer to courses organised into programmes as well as freestanding courses, interludes of work experience in enterprises, research projects, and preparation of dissertations (OECD, 2004, p.80). This section provides first the types and approaches to educational evaluation and further describes the components included in frameworks used to evaluate educational programs.

2.1.1 Educational program evaluation: types and approaches

Evaluation of educational programs is well described in the literature. Depending on the evaluation objectives, educational evaluation can be formative or summative as well as internal or external (Eseryel 2002; Nevo, 1995; Scheerens et al., 2003; Worthen & Sanders, 1987; Zinovieff, 2008). Formative evaluations (e.g. process or implementation evaluation) are conducted during implementation of programs. They are improvement driven in that they provide ongoing feedbacks that are intended to correct the implementation process of the program. Summative evaluations (e.g. outcome/impact evaluation) are conducted after implementation of programs and are intended to check whether the objectives of a program were reached. Both types of evaluation could be conducted by an internal evaluator (internal evaluation) or by an external evaluator (external evaluation) (Nevo, 1995). Next to these types of educational evaluation, different approaches are mentioned in educational evaluation literature. Examples of approaches to educational evaluation are goal/objective-based evaluation, goal-free evaluations, management oriented evaluations, and system evaluation (for more details see Eseryel, 2002; Madaus et al., 1983; Nevo, 1995; Owen 2007; Scheerens et al., 2003; Worthen & Sanders, 1987). Depending on the approach used, scholars provide a

variety of models/frameworks (see Eseryel, 2002; Madaus et al., 1983; Worthen & Sanders, 1987; Zinovieff, 2008). The models/frameworks include different indicators components used to evaluate educational programs.

2.1.2 Indicator components included in evaluation frameworks
Various indicator components are included in frameworks or models described in educational evaluation literature. In literature it is suggested to categorize indicator components according to logical classification (e.g. see European Commission, 2002; Scheerens et al., 2003; Stufflebeam in Madaus et al., 1983) and to thematic classifications respectively (e.g. French Ministry of Education, 2010; OECD, 2011).

Logical framework classifications
Logical frameworks classification attempts to establish relationships between investments, program activities/processes, and outcomes/results (W.K. Kellogg Foundation [Kellogg Foundation], 2004). Indicators components mostly included in logical frameworks are context, input, process, output, and outcome/impact (Bakewell & Garbutt, 2005; Kellogg Foundation, 2004; Mikkelsen, 2005; Norad, 1999; Örtengren, 2004). Context refers to the external factors that influence the program; inputs are the resources devoted to the program; processes are activities conducted during implementation; outputs refer to short-term results; and outcomes/impacts refer to longer-term results. According to Scheerens et al. (2003) there is no consensus about the number of indicators components and about terminologies used in logical frameworks. Some logical frameworks make no distinction for example between output, outcome and impact components (World Bank, 1996). Examples of logical frameworks in the field of educational evaluation include: the Context Input Process Product (CIPP) model developed by Stufflebeam (Madaus et al., 1983) and the Context-Input-Process-Output model developed by Scheerens (Scheerens et al., 2003).

The Context Input Process Product (CIPP) model was developed by Stufflebeam in the 60's (Madaus et al., 1983; Nevo, 1995; Worthen & Sanders, 1987). The CIPP model is based on the management-oriented approach to education in which decisions are to be made about inputs, processes, and outputs (Worthen & Sanders, 1987) and where evaluation is seen as an integral part of an institution and not as an isolated specialized activity (Stufflebeam in Madaus et al., 1983; Worthen & Sanders, 1987). The CIPP model suggests focusing evaluation of educational programs on four components: context evaluation, input evaluation, process evaluation, and product evaluation (Stufflebeam in Madaus et al., 1983). Context evaluation attempts to find out the needs that are to be addressed in an educational program; the problems to overcome; and the assets (e.g. services already available) that could be available in the environment (Stufflebeam in Maudaus et al., 1983; Stufflebeam, 2000). Input evaluation is intended to check the program strategies, service strategies, work plans and budget invested in the program (Stufflebeam, 2000, p.291). It also checks the approach used and relevant alternatives. Process evaluation identifies the degree to which plans are implemented and identifies problems related to implementation. It provides feedback to the staff and managers about the extent to which planned activities are carried out on schedule and whether they are carried out efficiently (Stufflebeam in Maudaus et al., 1983; Stufflebeam, 2000). Product evaluation is intended to measure the program achievements. According to Stufflebeam (in Madaus et al., 1983, p.135) a product evaluation should be extended to positive and negative results as well as to intended and unintended outcomes. Further, it should be extended to long term outcomes (Stufflebeam in Maudaus et al., 1983; Stufflebeam, 2000).

Similar to the CIPP model, the model (figure 2.1) developed by Scheerens et al. (2003) includes four components: context − input - process − and output. According to Scheerens et

al. (2003), one way to evaluating educational programs at system (e.g. national education system) as well as at lower (e.g. classroom) level is to use a Context-Input-Process/throughput-Output model. The model is developed by using a system approach to educational evaluation in which education is viewed as a production function where educational inputs are transformed to educational outputs in a contextual environment.
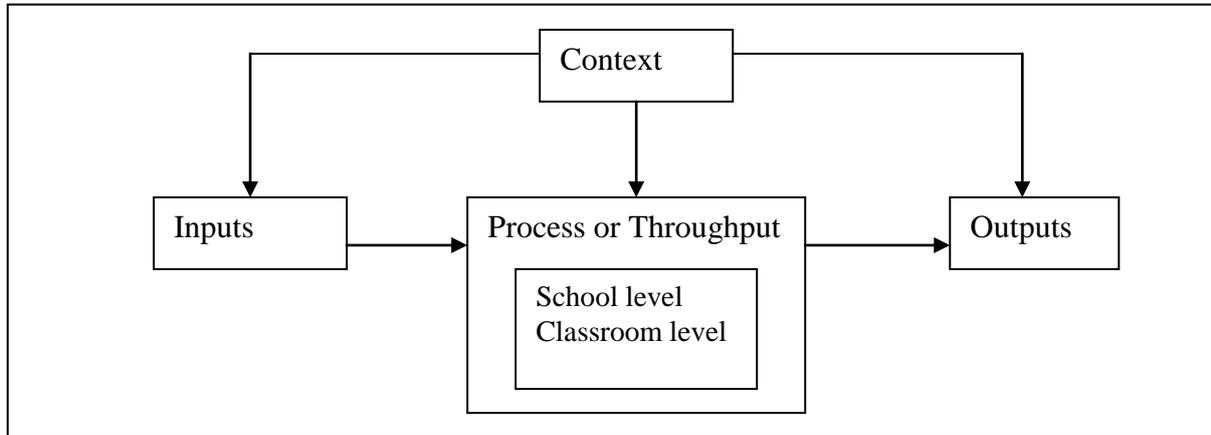
Figure 2.1. Basic systems model (Scheerens et al., 2003, p.18)

Similar to Stufflebeam, Scheerens mentions that each component of the basic conceptual framework should be separately evaluated. In this basic model context refers to the social, economic, and policy environment of the educational object within which the program takes place (Scheerens et al., 2003, p.57); inputs refer to material and financial resources provided; process refer to organizational and instructional structures; outputs could refer to attainment targets such as achievement scores (Scheerens et al., 2003, p18). According to Scheerens et al. (2003, p18-19) context evaluation attempt to understand whether the environment is favourable or not to the functioning of the educational object; input evaluation should describe and judge the material and financial resources of educational object; process indicators should be compared to accepted educational good practices; output evaluation should measure attainments compared to pre-established standards (attainment levels). According to Scheerens et al. (2003), the basic framework can be extended to outcome and impact components in order to classify educational indicators at system level. Outcomes could therefore refer to "statistics on access and participation, attainment statistics and aggregated data on educational achievement" (Scheerens et al., 2003, p.217). Impact or long-term outcome could refer to "changes in other sectors of the society that can be seen as the effects of education" (Scheerens et al., 2003, p.217).

Thematic classifications
Next to logical framework classification, indicators are thematically classified during evaluation of educational programs. Examples of such classification are provided by the Organization for Economic Cooperation and Development (OECD) and the French Ministry of Education (see State of Education in France).
Since the 80's the Organization for Economic Co-operation and Development (OECD) develops internationally comparable educational indicators through the International Indicators of Education Systems (INES) project. Four key themes of indicators are mostly used by OECD (OECD, 2004; OECD, 2012).
- The output of educational institutions and the impact of learning on economic and social outcomes

- The financial and human resources invested in education
- Access to education, participation and progression
- The learning environment and organisation of schools.

The aim of these indicators is to compare educational systems in order to support educational policy makers and practitioners in reforming their policies (OECD, 2011).

Another thematic classification of indicators is provided by the French Ministry of Education since 1991 in order to depict the state of education in France (French Ministry of Education, 2010; Sauvageot, 1997). The State of Education includes educational indicators organized into three themes: costs/expenditures, activities/context, and results (Sedel in Scheerens & Hendriks, 2004; Sauvageot, 1997; French Ministry of Education, 2010). Costs refer to educational expenditures; activities and mode of operations are related to processes that take place in the educational system; results refer to outcomes of education such as knowledge and skills attainments (French Ministry of Education, 2010). Sauvageot (1997) argues that the three themes are interrelated and are influenced by the socio-demographic environment. According to Sauvageot (1997) similarities exist between the classification used by the State of Education and the classification used by OECD.

Summary
Educational evaluation is conducted in order to improve the implementation process (formative) or to find out whether an educational program has met its objectives (summative). Logical or thematic classifications of educational indicators are used in available models/frameworks in order to evaluate educational programs. According to logical classification, processes in the environment of the program can transform inputs into targeted outputs and desired outcomes and impacts. Thematic classification use themes mostly related to expenditure in education; to conducted activities; and to results (in terms of outputs, outcomes, and impact).

2.2 Existing evaluation frameworks of educational programs
Many logical and thematic frameworks are available in literature in order to evaluate the impact of educational programs. In spite of the number of available frameworks, this section describes three frameworks in order to find out whether they could be used as conceptual framework in this study. The described frameworks are the framework developed by Scheerens (2004), the OECD framework, and the four-level framework developed by Kirkpatrick (1996). These frameworks are described in this study for the following reasons: The framework developed by Scheerens (2004) (hereinafter referred to as Scheerens' quality framework) is a logical framework that builds on empirical findings of school effectiveness research in developed and developing countries. The framework includes education indicators and variables that are empirically proven to be relevant for the evaluation of the quality of education at system, program, school, and classroom level. The OECD framework is jointly developed by international organizations such as UNESCO, OECD, and EUROSTAT (UOE). The framework includes internationally comparable education indicators that could be used to evaluate the outcomes of education systems. It was also used to develop the World Education Indicators framework (WEI) in order to assess the outcomes of educational systems in developing countries. The Kirkpatrick four-level framework was developed to assess the outcomes of trainings in corporate organizations. The Kirkpatrick framework explains how to proceed in evaluating outcomes of trainings. Evidence of its use in the context of organizations that provide fellowships is mentioned in the literature. IFP is a fellowship program and fellowship programs are defined as programs that provide training to fellows in specific fields of studies that are relevant for them and that match with the training needs of their countries (Aguirre International, 2004; Boeren, 2005; CIDA, 2005; Rotem et al., 2010;

USAID, 1995). Therefore, it was necessary to understand how the outcomes of trainings could be evaluated. The following sections describe successively the Scheerens' quality framework, the OECD framework and the Kirkpatrick framework. In each framework, focus is on components used by the framework and on indicators included in these components.

2.2.1 Scheerens' quality framework

The Scheerens' quality framework refers to the input-process-outcome-context framework (see figure 2.2 below) introduced by Scheerens (2004) as "a basis for defining quality and to categorize different measures of quality of education" (p.49). The quality framework is based on findings of school effectiveness research in developed and developing countries. The quality framework was developed for the UNESCO Education for All Global Monitoring Report 2004. It builds on the comprehensive model of the school effectiveness framework provided by Scheerens (1991). School effectiveness research attempts to find out malleable inputs, process and context factors that are positively associated with outputs/outcomes of schooling (Scheerens 1991; Scheerens 2000; Scheerens 2001; Scheerens, 2004). According to Scheerens (2004) school effectiveness research provides "malleable conditions" that "matter" at system, school, and classroom level and on which international consensus exist (p.39). Malleable conditions refer to factors that can be manipulated by actors active in the area of education such as "policy planners, local constituencies, school managers, and teachers" in order to increase the quality of education (Scheerens, 2004, p.54). The quality framework includes malleable input and process factors as well as factors that cannot be manipulated ("given factors") by actors in the area of education (Scheerens, 2004, p.54). According to Scheerens (2004) the factors included in the quality framework are positively associated with educational achievement and are able to measure the quality of education in terms of productivity, effectiveness, efficiency, equity, and responsiveness (Scheerens, 2004, p.68). The factors are based on the findings provided by School Effectiveness Research (SER) movement in developed and developing countries.

INPUTS	PROCESS	OUTCOMES

System level financial and material and human resources indicators	System level process indicators	**Output indicators** - Subject matter based - Literacy (reading, mathematical, scientific) - Competencies (learning to learn)
School level financial and material and human resources indicators	School level process indicators	**Outcome/attainment indicators** - Graduation rates - Proportion of students graduated without delay - Drop-out rates - Class repetition rates
Students background characteristics	Effective teaching variables	**Impact indicators** - (for each attainment level) % of employed at a certain job level - % of unemployed - (for lower school levels) % of enrolled in follow-up education - Degree of social participation (social capital) - Adults literacy rates - Average income, for each attainment level

Areas for describing responsiveness

CONTEXTUAL FACTORS

The social context
- Demographic information
- Aspects of the culture and cultural tradition of a country
- Economic aspects
- The institutional infrastructure
- The general health situation in a country
- Disasters of nature and war

Antecedent conditions at national level

- Demographics
- Cultural aspects
- Institutional infrastructure

Context indicators of the local community

- The organizational infrastructure
- Local cultural conditions

Figure 2.2. Synthetic overview of educational input, process, outcome and context indicators (Scheerens, 2004, p. 102)

According to Scheerens (2004, p.60) inputs and process factors are the "means" that bring about outcomes. The following provides successively more details on inputs, process, outcomes and context components by describing the elements included in each component.

Input component of the quality framework
In the quality framework inputs are the pre-conditions for transformation processes (Scheerens, 2004, p.89). The quality framework makes a distinction between three kinds of malleable input factors: Malleable financial and material resources at system and school level

(e.g. spending on salaries, classroom equipment), human resources at system and school level (e.g. teacher training), and malleable background conditions of the students (e.g. language of instruction) (Scheerens, 2004).

Malleable financial and material resources
Financial and material resources are defined at system as well as school level (Scheerens, 2004, p.89). According to Fuller & Clarke (1994) in Yu (2007) the majority of school effectiveness studies in developing countries have found a significant positive association between material resources and educational achievement. Next to material resources, provision of financial resources is also found positively associated with achievement (Aoki et al., 2001; Boissiere, 2004; Scheerens, 2001). Examples of financial and material resources indicators included in the quality framework are: proportion of Gross Domestic Product spent on education (system level) – Proportion of the school budget that is acquired through other than public funding (school level) (Scheerens, 2004, p.90). Table A1 (appendix A) provides an overview of financial and material resources indicators.

Malleable human resources
Malleable human resources refer to variables related to teachers. These variables influence the quality of education (Scheerens, 2004). There is empirical evidence that teacher factors such as teacher subject knowledge, teacher training, and teacher experience are effective and have impact on student achievement in developing countries (Boissiere, 2004; Fuller & Clarke, 1994; Heneveld & Craig, 1996; Pennycuick 1993). Five categories of teachers' indicators and related variables are included in the quality framework (Scheerens, 2004, pp.91-92). Examples of such categories include: teacher background characteristics and teacher professional knowledge and skills (Scheerens, 2004, pp.91-92). Table A2 and table A3 (appendix A) provide an overview of categories of teachers' indicators at school level as well as details about indicators related to these categories.

Malleable background students' characteristics
Student background characteristics refer to the home situation of students that could interact with the conditions for learning at school (Scheerens, 2004, p.92). Empirical studies point out that such conditions have effect on learning achievement (Boissiere, 2004; Fuller & Clarke 1994; Pennycuick, 1993; UNESCO, 2000; UNESCO, 2002; Yu, 2007). Two categories of student background characteristics are included in the quality framework: general background characteristics and specific situational constraints (particularly in developing countries). General background characteristics (e.g. Mother's level of educational attainment or social economic status) are associated with educational performance and cannot be manipulated (Scheerens, 2004, p.93). The specific situational constraints on students (e.g. distance a student has to walk to school) can be manipulated (Scheerens 2004, p. 93). For example building more schools and classrooms could reduce the distance that a student has to walk (Boissiere, 2004). Table A4 (appendix A) provides an overview of Student background characteristics.

Process component of the quality framework
Process indicators are central in measuring quality of education in that they are considered as malleable conditions that influence outcomes of schooling (Scheerens, 2004, p.81). According to Scheerens (2004) "process indicators are interesting from the point of view of policy and management since they refer to conditions that are malleable and thus the subject of active policies to improve education" (p.85). Process factors are educational characteristics that explain the differences between schools (Scheerens, 1991, p.374; Scheerens, 2004, p.21).

According to Scheerens (2004, p.85) these malleable conditions can improve the quality of education because they are found to have a high positive association with educational outputs and outcomes by empirical school effectiveness research. Compared to developed countries few empirical studies related to process factors have been conducted in developing countries (Boissiere, 2004; Heneveld & Craig, 1996; Scheerens, 2000; Scheerens, 2004; Yu, 2007). However, research in developing countries (Boissiere, 2004; Fuller & Clarke, 1994; Heneveld & Craig, 1996; Pennycuick, 1993; Scheerens, 2001; Yu, 2007) has focused on malleable process factors such as instructional time, community and parent participation, new teaching methods, and educational leadership (leadership of headteacher/principals).

In the quality framework, Scheerens (2004) distinguishes three types of malleable process indicators: process indicators at the level of national education systems (e.g. evaluation capacity of the system); process indicators of school functioning (e.g. instruction time, parent's involvement); and effective teaching and learning variables (e.g. opportunity to learn, adaptive teaching).

Process indicators at system level
Process indicators at system level refer to policy measures and structures defined at the level of national education systems (Scheerens, 2004, p.83). System level indicators could be considered as context indicators at school level. Examples of system level process indicators include teaching time per subject and total hours of instruction per year (Scheerens, 2004, p.83). Table A5 (appendix A) provides the overview of process indicators at system level included in the quality framework.

Process indicators at school level
At school level ten process indicators are included in the quality framework (Scheerens, 2004, p.85). Examples of such indicators include community involvement and school financial and human resources (Scheerens, 2004, p.87). Table A6 and table A7 (appendix A) provide an overview of process indicators at school level as well as details about variables related to these indicators.

Teaching and learning variables
Next to process indicators at school level, the quality framework provides effective teaching and learning variables. Examples of such variables include: opportunity to learn and climate aspects (Scheerens, 2004, p.88). Table A8 (appendix A) provides an overview of these variables (Scheerens, 2004, p.88).

Outcome component of the quality framework
In the quality framework, distinction is made between output, outcome and impact indicators. Outputs are defined as the direct outcomes of schooling (e.g. results on a standardized achievement test) (Scheerens, 2004, p.80). Outputs refer to achievement indicators. Outcomes refer to attainment indicators (e.g. graduation rates) (Scheerens, 2004, p.80). According to Scheerens outputs (achievement) and outcomes (attainment) are important criteria used to judge educational quality. Impact indicators refer to "social status of students that have reached certain levels of educational attainment" (Scheerens, 2004, p.69). An example of impact indicator is the percentage of employed at a certain job level (Scheerens, 2004, p.80). Table A9 (appendix A) summarizes the indicators included in the outcome component of the quality framework.

Context component of the quality framework
In the quality framework context could refer to given (not malleable) as well as malleable conditions that influence schooling (Scheerens, 2004, p.95). However, Scheerens acknowledges that it is difficult to separate malleable conditions to conditions that are considered as given (e.g. conditions at community level). Three categories of context indicators are included in the framework: conditions at the level of society at large, national level conditions, and conditions of the local community.

Context indicators related to society include among other demographic and cultural aspects of the society (Scheerens, 2004, p.95). These conditions can also be translated to conditions at national level within the education sector. Conditions at national level (related to education sector) are related to demographics, cultural aspects, and institutional infrastructure (Scheerens, 2004, p.97). At community level, Scheerens (2004, p.98) provides two categories of context factors: the organizational infrastructure and the local cultural conditions. Cultural aspects at local community are given contextual conditions that constraint functioning of schools (Scheerens, 2004, p.99). An example is the parents' values concerning school participation of their children (Scheerens, 2004, p.99). These aspects can also be observed at regional or national level. Table A10 to table A15 (appendix A) provide an overview of the context factors/variables included in the quality framework.

Responsiveness component of the quality framework
Responsiveness mechanisms refer to infrastructures and mechanisms adopted by educational systems to deal with "societal demands to education" (Scheerens, 2004, p.100). Factors included in the responsiveness component are intended to understand whether schools or educational systems have mechanisms and infrastructures to address responsiveness questions (Scheerens, 2004, p.100). For example, whether an institutional infrastructure is available for curriculum development (Scheerens, 2004, p.101). Table A16 (appendix A) summarizes areas of responsiveness to context at system and school level.

2.2.2 OECD education indicators framework
Since the 80's the Organization for Economic Co-operation and Development (OECD) develops internationally comparable educational indicators through the International Indicators of Education Systems (INES) project. The aim of the project was to assess and monitor the effectiveness of education systems by comparing their performance with each other (OECD, 2012). For that purpose, OECD has developed a framework that includes internationally comparable sets of indicators (OECD, 2012, p.3). This was done in collaboration with other international institutions such as the United Nations Educational, Scientific and Cultural Organization Institute of Statistics (UNESCO-UIS), the Statistical Office of the European Union (EUROSTAT) and EdStats (World Bank). The findings of the INES project are annually published – by OECD - in 'Education at a Glance'. The following describes the components of the framework within which OECD indicators are organized.

Components of the OECD framework
In the OECD framework indicators are organized thematically according to three dimensions: actors in education systems, groups of indicators, and policy issues (OECD, 2004, p.21).

Actors in education systems (first dimension) refer to individual learners, instructional settings and learning environments within institutions, educational service providers (educational institutions), and the education system as a whole. Data on indicators included in the OECD framework are collected from these levels (OECD, 2004, p.22).

Three groups of indicators are used in organizing the framework (second dimension): learning outcomes for individuals and countries, policy levers, and antecedents (OECD, 2004, p.21).

According to OECD (2004, p.22) output and outcomes of education include indicators on observed outputs of education systems and indicators related to the impact of knowledge and skills for individuals, societies and economies. Policy levers refer to activities "seeking information on the policy levers or circumstances which shape the outputs and outcomes". Antecedents refer to "factors that define or constrain policy" (OECD, 2004, p.22).

Policy issues (third dimension) make a distinction between three issues: quality of educational outcomes and educational provision, issues of equity in educational outcomes and educational opportunities, and the adequacy and effectiveness of resource management (OECD, 2004, p.21).

The first two dimensions are used to develop the OECD framework (see figure 2.3 below). Each cell of the framework can be used to address issues related to the third dimension.

	1. Educational and learning outputs and outcomes	2. Policy levers and contexts shaping educational outcomes	3. Antecedents or constraints that contextualise policy
I. Individual participants in education and learning	1.I The quality and distribution of individual educational outcomes	2.I Individual attitudes, engagement and behaviour	3.I Background characteristics of the individual learners
II. Instructional settings	1.II The quality of instructional delivery	2.II Pedagogy and learning practices and classroom climate	3.II Student learning conditions and teacher working conditions
III. Providers of educational services	1.III The output of educational institutions and institutional performance	2.III School environment and organization	3.III Characteristics of the service providers and their communities
IV. The education systems as a whole	1.IV The overall performance of the education system	2.IV System-wide institutional settings, resource allocations and policies	3.IV The national educational, social, economic and demographic contexts

Figure 2.3. Matrix describing the first two dimensions of OECD framework (OECD, 2004, p.21)

Indicator themes used by OECD

Different themes of indicators are used by OECD throughout the years. The themes are described in 'Education at a Glance' (e.g. see OECD, 2012). However, the number of themes, the number of indicators, and the semantic of themes could vary per publication (Hendriks, Barzanò, Brumana & Cremonesi in Scheerens & Hendriks, 2004). For example the 1998 publication includes six themes (Hendriks et al. in Scheerens & Hendriks, 2004) while the 2012 publication includes four themes (OECD, 2012). Further, the 2000 and 2001 editions include 31 indicators while the 1998 edition includes 34 indicators (Hendriks et al. in Scheerens & Hendriks, 2004).

Since 2002 the following four key themes of indicators are used by OECD (OECD, 2004):
- The output of educational institutions and the impact of learning (referred to as chapter A)
- The financial and human resources invested in education (referred to as chapter B)
- Access to education, participation and progression (referred to as chapter C)
- The learning environment and organisation of schools (referred to as chapter D)

The indicators included in these themes "provide information on the human and financial resources invested in education, on access to education, progression, completion and education-work transitions, on the learning environment and the organization of schools, on

the quality of learning outcomes, and on the economic and social returns to learning" (OECD, 2004, p.21).

Chapter A indicators: the output of educational institutions and impact of learning
Chapter A includes indicators that measure educational attainment and the output of the education system. These indicators help shaping policies (OECD, 2012). Examples of such indicators are for example graduation rates and education attainment of the population/labour force (OECD, 2004, p.134). In the 2012 publication of 'Education at a Glance', eleven indicators were included in chapter A (OECD, 2012, p.25). Examples of such indicators are the level of study of adult (indicator A1) and the number of students who complete tertiary education (indicator A3). Table A17 (appendix A) provides all indicators included in chapter A of the 2012 publication.

Chapter B indicators: Financial and human resources invested in education
Chapter B provides input indicators that are policy levers or antecedents to policy (OECD, 2012, p.19). Educational financial indicators are for example expenditure on educational institutions per student and cumulative expenditure on educational institutions per student over the average duration of tertiary studies (OECD, 2004, p.149). In the 2012 publication of 'Education at a Glance', seven indicators were included in chapter B (OECD, 2012, p.213). Examples of such indicators are expenditure per student (indicator B1) and proportion of national wealth spent on education (indicator B2). Table A18 (appendix A) provides all indicators included in chapter B of the 2012 publication.

Chapter C indicators: Access to education, participation and progression
Chapter C includes indicators that are "a mixture of outcome indicators, policy levers and context indicators" (OECD, 2012, p.19). Indicators on access, participation and progression are for example enrolment rates and school expectancy including expected years in tertiary education (OECD, 2004, p.141). In the 2012 publication of 'Education at a Glance', six indicators were used for chapter C (OECD, 2012, p.317). For example, indicator C1 (who participate in education) includes variables related to participation in education. Table A19 (appendix A) provides all indicators included in chapter C of the 2012 publication.

Chapter D indicators: Learning environment and organization of schools
Indicators related to learning environment and organization of schools include for example average class size and ratio of students to teaching staff (OECD, 2004, pp.146-147). In the 2012 publication of 'Education at a Glance', seven indicators were used for chapter D (OECD, 2012, p.423). Examples of such indicators are time students spend in classroom (indicator D1) and student- teacher ratio (indicator D2). Table A20 (appendix A) provides all indicators included in chapter D of the 2012 publication.

2.2.3 Kirkpatrick four-level evaluation framework
The four-level evaluation framework (figure 2.4. below) developed by Donald Kirkpatrick in 1959 is the most influential evaluation framework for evaluating outcomes of academic trainings (Bates, 2004; Eseryel, 2002; Ford, 2004; Holton, 1996; Rajeev, Madan & Jayarajan, 2009). The framework is popular because of its simplicity, practicality and because it addresses training evaluation in a systematic way (Bates, 2004; Ford, 2004; Kirkpatrick, 1996; Rajeev et al., 2009).

```
┌─────────────────────────────────────────────────────────────────────────┐
│  ┌─────────────────────┐                                                  │
│  │ Training environment │                                                  │
│  └─────────────────────┘                                                  │
│  ┌──────────────┐      ┌──────────────────┐    ┌──────────────────┐      │
│  │ Learning     │ ───▶ │ Level 1: Reactions│──▶ │ Level 2: Learning │      │
│  │ event        │      │ - Learner         │    │ - Learner         │      │
│  │              │      │ - Client          │    │ - Organization    │      │
│  └──────────────┘      └──────────────────┘    └──────────────────┘      │
└─────────────────────────────────────────────────────────────────────────┘
```
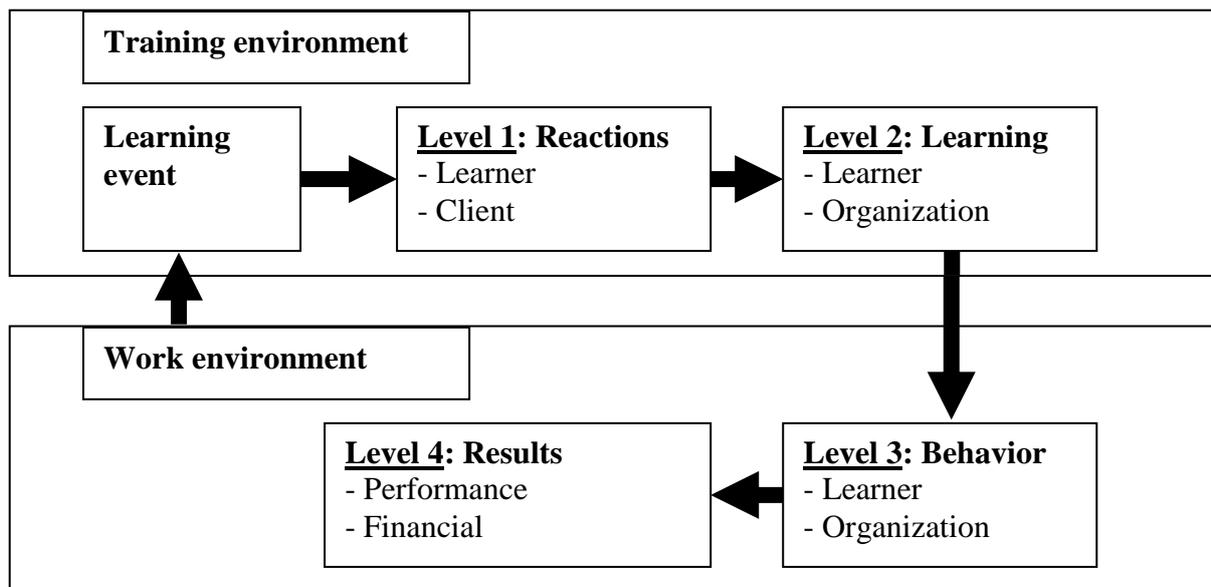
Figure 2.4. Kirkpatrick's Evaluation Model (Ford, 2004, p.37)

The following describes the Kirkpatrick framework in more details by providing its different components and by explaining how the framework should be used to measure outcomes of trainings.

Components of Kirkpatrick four-level evaluation framework
The framework (see figure 2.4) includes four components (levels or criteria by some authors) of evaluation known as reaction, learning, behavior, and results. The components are defined by Kirkpatrick (1996, pp.55-59) as follows:

Level 1 (**Reaction criteria**): This level refers to participants' feelings about aspects of a training program such as topic, speaker, schedule etc. Reactions measure the overall satisfaction of participants and help to understand whether they are motivated and interested in learning. According to Kirkpatrick (1996) "if they don't like a program, there is little chance that they'll put forth an effort to learn" (p.56).

Level 2 (**Learning criteria**): This level measures the extent to which participants' knowledge, skills and attitude changed due to the training received. At this level one can ask whether participants have acquired new knowledge during the training; whether their skills have improved due to participation in training; and whether their attitudes have changed due to participation in training. Evaluators are asked to evaluate, if possible, knowledge, skills and attitudes before and after the training. The use of before-after-training approach helps to attribute any change in knowledge, skills and attitudes to the training program. The use of control groups is recommended if feasible.

Level 3 (**Behavior criteria**): The behavior level measures the extent to which participants "change their on-the-job behavior because of training" (Kirkpatrick, 1996, p.56). Evaluators are therefore asked to evaluate the extent to which participants can apply (or transfer) their acquired knowledge, skills, and attitudes while on the job. In order to attribute change to the training received evaluators are asked to use control groups if feasible. Evaluators are also advised to survey or interview trainees, trainees' bosses, trainees subordinate and others "who often observe trainees' behavior on the job" (Kirkpatrick, 1996, p.57). However, Kirkpatrick

(1996) acknowledges that such an attribution will be difficult to realize because of different external factors that can influence participants.

Level 4 (**Results criteria**): At this level evaluators are asked to measure changes brought to organizations or institutions because of participants' participation in training. For that purpose, evaluators are asked to use control groups if feasible or to use before-and-after training approaches in order to attribute changes in organizations to the training received by participants. Data collection at level 4 can be difficult, costly and time consuming compared to other levels.

Table 2.1. Summary Kirkpatrick's four-level model

Levels	Level of measurement	Definition (what should be measured)
1.Reaction	Individual	Participants' reactions to training received.
2.Learning	Individual	Knowledge, Skills, Attitudes participants acquired during training.
3.Behavior	Institutional/organizational	Transfer of training to the job (The extent to which participants are able to implement or to transfer what they learned to the job).
4.Results	Institutional/organizational	Results in institution and organization due to training received (changes brought to institutions/organizations due to participants' participation in training).

These components are used to measure the outcomes of trainings in business and organizational contexts (Kirkpatrick, 1996). In addition to this traditional use, the components are also used by non-profit organizations that provide fellowships in developing countries. The Kirkpatrick framework is used for example by organizations such as the United States Agency for International Development (USAID), the Canadian International Development Agency (CIDA), the United Nations (UN), and the Danish International Development Agency (Danida) (Aguirre International, 2004; CIDA, 2005; Danida Fellowship Centre [DFC], n.d.; Rotem et al., 2010). Except Danida that includes only the first three components in their monitoring and evaluation system (DFC, n.d.), the other organizations use all four levels as recommended by Kirkpatrick (see table 2.2 below). Further, in the context of fellowship programs, results are sometimes measured beyond organizational and institutional levels. According to Rotem et al. (2010), fellows are selected and trained in order to influence not only institutions but also to influence communities and beyond. Therefore, organizations use an additional component in order to measure results at institutional levels as well as at community, country, regional or international level. This component is referred to as the higher level component by Aguirre International (2004) or as the mega impact component by Rotem et al. (2010). Table 2.2 presents a comparison of levels used by fellowships programs in evaluating their programs with the original Kirkpatrick framework.

Table 2.2. Kirkpatrick original levels compared to levels used by fellowship programs

Kirkpatrick framework (Kirkpatrick, 1996)	USAID fellowship (Aguirre International, 2004)	UN fellowship (Rotem et al., 2010)	Danida fellowship (DFC, n.d.)
1. Reaction	**1.** Reaction	**1.** Reaction	**1.** Reaction or response
2. Learning	**2.** Learning	**2.** Learning	**2.** Learning
3. Behavior	**3.** Behavior	**3.** Behavior	**3.** Behavior
4. Results	**4.** Results	**4.** Results	
	5. Higher level impact	**5.** Mega-impact (long term)	

In sum, Kirkpatrick (1996) recommends to using four levels in evaluating any training program if money and expertise are available. Evaluation of trainings should therefore not be limited to the first two criteria as this is the case by many evaluations (Aguirre International, 2004; Holton, 1996; Kirkpatrick, 1996; Rajeev et al., 2009; Rotem et al., 2010). However, the quality of the Kirkpatrick framework is questioned in literature about evaluation of training programs. The following provides the main critics and describes briefly new frameworks developed in order to address them.

Quality issues of the Kirkpatrick framework
Critics (e.g. Alliger & Janak, 1989; Bates, 2004; Holton, 1996) argue that the framework is not complete as many intervening variables remain unmeasured. For critics the four-level framework is oversimplified; fails to address the important issue of transferring knowledge into the workplace (Holton, 1996); and fails to consider individual and organizational contextual factors that could influence training (Bates, 2004; Holton, 1996). Holton argues for example that reactions to training should be considered as a learning environment variable that influence learning behavior rather than a primary outcome of training as recommended by Kirkpatrick. Further, critics mention that no linear causal relationships exist between the different levels as implicitly assumed by Kirkpatrick (Alliger & Janak, 1989; Bates; 2004; Holton, 1996). In reaction to such critics, Kirkpatrick (1998) argues that the framework refers to different perspectives and that the levels are not intended to be hierarchical.

Following critics, new frameworks were proposed in order to measure outcomes of trainings. Examples of such frameworks include: The HRD Evaluation Research and Measurement Model developed by Holton (Holton, 1996); Phillips' return on investment framework (ROI) (Phillips, 1996); the augmented framework proposed by Alliger, Tannenbaum, Bennett, Traver, & Shotland (1997). However, most of these frameworks were modified versions of Kirkpatrick four-level framework (Bates, 2004; Ford, 2004; Rajeev et al., 2009; Rotem et al., 2010).

Holton (1996) proposes a framework that excludes the reaction level. He recommends (p.9) considering three outcomes of training: Learning (learning outcomes), individual performance (change in individual performance as a result of the learning being applied on the job), and organizational performance (results at the organizational level as a consequence of the change in individual performance). The terminologies individual performance and organizational performance were introduced instead of the terms behavior and results from Kirkpatrick framework. The framework further introduces concepts supposed to influence outcomes such as ability (to learn), motivation (to learn and to transfer), and environmental influences (p.10) (for more details see Holton, 1996). Phillip's (1996) return on investment (ROI) framework adds a fifth level to the four levels of Kirkpatrick framework in order to calculate the return

on investment of training (monetary value of the impact compared to the costs of the training) (for more details see Ford, 2004; Phillips, 1996). Alliger et al. (1997) proposes to augment the Kirkpatrick framework by refining terminologies and criteria. The proposed augmented framework uses the term transfer criteria instead of behavior criteria and divides the criteria reaction into affective reactions and utility reactions (for more details see Alliger et al., 1997). Table 2.3 compares these new frameworks (HRD; ROI; augmented framework) to the original Kirkpatrick framework.

Table 2.3. Comparison between training evaluation frameworks

Kirkpatrick framework (Kirkpatrick, 1996)	HRD framework (Holton, 1996)	ROI framework (Phillips, 1996)	Augmented framework (Alliger et al., 1997)
1. Reaction		**1.** Reaction	**1.** Reaction
2. Learning	**1.** Learning	**2.** Learning	**2.** Learning
3. Behavior	**2.** Individual performance	**3.** Behavior	**3.** Transfer
4. Results	**3.** Organizational performance	**4.** Results	**4.** Results
		5. Return on investment	

The previous sub-sections have described the components of the Kirkpatrick framework. The following explores literature on fellowship programs in order to find out outcomes and impact indicators that are included in the components used by the Kirkpatrick framework.

Kinds of indicators included in the components of the Kirkpatrick framework
This sub-section is influenced by four evaluations of fellowships programs conducted by Aguirre International (2004) in order to evaluate the African Graduate Fellowship (AFGRAD) and Advanced Training for Leadership Program (ATLAS) (in the following referred to as ATLAS/AFGRAD program); by CIDA (2005) in order to evaluate the Canadian Francophone Scholarship Program (CFSP); by the World Bank Institute (2010) in order to evaluate the Joint Japan/World Bank Graduate Scholarship Program (JJ/WBGSP); and by the Asian Development Bank [ADB] (2007) in order to evaluate the Japan Scholarship Program (JSP).
These programs were selected in this study because components of indicators used during their evaluations were similar to the components recommended by Kirkpatrick (1996). Further, the programs were selected for this study because of their similarities with the IFP program in terms of core objectives, field of study and level of study.
The following sub-sections focus on indicators used during evaluation of these four fellowship programs. Each sub-section starts with a brief description of the program. Further, the sub-section outlines the indicators used during evaluation of the programs.

Evaluation of the ATLAS/AFGRAD program
The USAID funded fellowships program AFGRAD Program (African Graduate Fellowship Program) launched in 1963 was provided to recipients countries till 1990. Its successor the ATLAS Program (Advanced Training for Leadership and Skills) started in 1991 and ended in 2003. According to Aguirre International (2004, p.xv) the core goal of the program was to assist "young African nations with a supply of trained mid- and upper-level 'manpower' in key sectors needed for development". For that purpose, the AFGRAD program provided "high-caliber university education to Africans who will assume critical positions in

universities, governmental and parastatal institutions, and the private sector, through which they can contribute to the development of their countries" (USAID, 1995, p.vi). ATLAS/AFGRAD fellows were selected in 52 African countries such as Benin, Ghana, Mali, Madagascar, Namibia, Uganda, etc. From 1963 till 2003, ATLAS/AFGRAD fellowships have trained 3263 African professionals. The ATLAS/AFGRAD fellowship provided PhD and master trainings in U.S Universities. Fellows may study in any development related fields of study important to their country's growth such as agriculture and animal science, business administration and economics, engineering, education and health (USAID, 1995, p.5).

Impact evaluation of ATLAS/AFGRAD program was undertaken from 2003 till 2004 by Aguirre International with the assistance of USAID (Aguirre International, 2004). The aim of the impact evaluation was to "assess whether any impact occurred from long-term U.S. academic training provided to participants in the ATLAS/AFGRAD projects" (Aguirre International, 2004, p.46). For that purpose the evaluation team employed a modified Kirkpatrick evaluation framework (see table 2.4 below). To the original four levels of Kirkpatrick model, the team added a fifth level (higher level impact) in order to measure the program impact at sectoral, national, regional or international levels. The evaluation team argues that alumni contributions can be measured beyond institutional levels since national leaders and high ranking officials are among alumni (Aguirre International, 2004).

Table 2.4. The modified Kirkpatrick framework used for the ATLAS/AFGRAD program (Aguirre International, 2004, p.xvi)

Level one (reaction): the trainee's impression of the program; the level of satisfaction with the course, trainer, pace of instruction, content and materials)
Level two (Learning): the acquisition of skills and knowledge from the training.
Level three (Application): the performance of the trainee on the job following training.
Level four (Institutional Results): changes that the trainee's performance brought to the organization in efficiency, productivity, or profitability.
Level five (Higher-Level Results): changes that the trainee's performance brought directly to a nation, region or beyond to an international sector or institution.

According to Aguirre International (2004) changes attributed to training at levels 2,3,4,5 can be viewed as impact. According to the evaluation team, levels 2 and 1 are less indicative of the impact that USAID is seeking. Level 1 and level 2 indicators produce little information about impact (Aguirre International, 2004). However, the team acknowledges that without the acquisition of Knowledge, Skills and Attitudes (KSA) by alumni (level 2) no impact linked to ATLAS/AFGRAD could occur. Therefore, three types of indicators were used by the evaluation team: before training indicators, about training indicators, and after training indicators (Aguirre International, 2004).

Before training indicators refer to characteristics of fellows before selection in the ATLAS/AFGRAD program such as employment; organizational type and sector of employment. Examples of before training indicators are number of fellows who were employed or not before selection in program and fellows' job title before selection. Before training indicators were collected in order to facilitate the use of a before-and- after approach during the evaluation as recommended by Kirkpatrick (Aguirre International, 2004). About training indicators refer to characteristics of fellows gained during the program such as degree obtained and follow-up activities done during training. About training indicators refer to level 2 indicators. Examples of about training indicators are degree obtained during ATLAS/AFGRAD training and ATLAS/AFGRAD contribution to alumni knowledge, skills and attitudes. After training indicators refer to information about alumni after participation in

the ATLAS/AFGRAD program such as employment, positions held, and participant's impact in institutions and beyond institutions. After training indicators include level 3, level 4, and level 5 indicators. Level 3 indicators CFSP measure the extent to which alumni are able to apply the knowledge, skills and attitudes acquired during participation in the ATLAS/AFGRAD program. An example of level 3 indicators is the applicability of acquired knowledge, skills and attitudes at institutions. Levels 4 and 5 indicators measure the results of ATLAS/AFGRAD program at individual and institutional level as well as at sectoral, national, regional or international levels. Examples of level 4 and 5 indicators are alumni career progression and ATLAS/AFGRAD contribution to institutional changes. Table B1 to B4 (appendix B) provide all kinds of indicators used during evaluation of the ATLAS/AFGRAD program with their operational definitions.

In sum, the framework used to assess the impact of ATLAS/AFGRAD program is a modified Kirkpatrick framework that focus on learning, behavior, results, and higher impacts indicators. No reaction indicators were used during evaluation of ATLAS/AFGRAD program.

Evaluation Canadian Francophone Scholarship Program (CFSP)

The Canadian Francophone Scholarship Program (CFSP) was launched in 1987. From its inception the program is managed by the Canadian International Development Agency (CIDA). According to CIDA (2005) the program aims to increase opportunities for higher education and to establish closer ties between recipient countries and Canada. Further, the program aims to contribute in the development of recipient countries by "giving priority to technical and vocational trainer training, enhanced teaching and research skills in universities, specialist and manager training to build the capacities of private entrepreneurs, appropriate vocational training to build the capacities of public-service managers, and a greater number of women in all sectors" (CIDA, 2005, p.8). The CFSP provides scholarship to graduates from 37 Francophone countries eligible for Canada's Official development assistance (ODA) (CIDA, 2005, p.1). Fellows are selected in Africa, Asia, Indian Ocean and Caribbean countries. From 1987 till 2004, CFSP has provided 1461 scholarships particularly to fellows from African countries. Fellows may follow education at college, bachelor, master and doctoral levels in Canadian universities. They may study in any development related fields of study such as management, health, education, engineering, and science.

A modified Kirkpatrick framework was used during evaluation of the CFSP program. According to CIDA (2005, p.2) CFSP assessment framework (see CIDA, 2005, p.9) is based on the modified Kirkpatrick four-level evaluation framework developed by Aguirre International (2004) for USAID. The framework was designed in order to measure the operational objectives (outputs), the outcomes and impacts of the program. It includes different outcome and impact indicators.

Two kinds of outcome indicators were used in the framework: immediate outcomes and medium-term outcomes. Immediate outcomes refer to reaction and learning indicators. An example of an immediate outcome indicator is the benefits of academic training to alumni. Medium-term outcomes refer behavior as well as results (e.g. individual benefits from training/results) indicators. Examples of medium-term outcome indicators are alumni employment status after graduation and the use of alumni acquired knowledge, skills and attitudes in home institutions or organizations.

Further, expected and unexpected impact indicators were used by CIDA in order to assess the impact of CFSP in recipient countries. An example of expected impact indicators is CFSP contribution to sustainable development in recipient countries. Unexpected impact indicators are related to the impact of alumni who do not return to their countries but are still able to contribute to the development of their countries. According to CIDA (2005, p.20) their

contributions can be in terms of remittances or informal benefits such as working for an organization that contributes to the development of their own country. An example of unexpected impact is the diaspora's (alumni who do not return to home country) contribution to the development of CFSP recipient countries. Table B5 to B 7 (appendix B) provide all kinds of indicators used during evaluation of the CFSP program with their operational definitions.

In sum, the framework used to assess the impact of CFSP program is a modified Kirkpatrick framework that includes reaction, learning, behavior, and results indicators components.

Evaluation Joint Japan/World Bank Graduate Scholarship Program (JJ/WBGSP)

The Joint Japan/World Bank Graduate Scholarship Program (JJ/WBGSP) was established in 1987 in order to "encourage and strengthen human resource development in developing countries" (World Bank Institute, 2010, p.i). The JJ/WBGSP provides scholarship to mid-career professionals from developing countries in order to complete a master or doctorate degree in a development related field of study. In doing so the program aims to expose these professionals to "the latest techniques and knowledge available through development related graduate studies" (World Bank Institute, 2010, p.2). Examples of study areas include public policy, international development and sustainable development, economics, environment and natural resources, and public health, education, and agriculture. Fellows are selected in Africa, Asia and the Pacific, Latin America and Caribbean, the Middle East, not industrialized European countries, and the so called part 1 countries such as Japan. From 1987 till 2007 the program has provided 3733 scholarships mostly to professionals from Africa, Asia, and the Pacific. Fellows may study in 150 universities in 32 World Bank member countries. After completion of their study alumni are expected to return to their home countries in order to apply and disseminate the newly acquired knowledge and skills to enhance the socioeconomic development of their own and other developing countries (World Bank Institute, 2010).

Since 1994 the Word Bank and the government of Japan have conducted a series of tracer studies in order to evaluate the JJ/WBGSP program. The aims of the studies were to trace the performance of the program by investigating whether fellows (World Bank Institute, 2010, p.1): attained their degrees successfully and benefited from their academic programs; returned to their home or other developing countries; achieved recognition for their enhanced skills, progression and mobility, higher income, better grades and promotion in their jobs; are engaged in senior professional and managerial positions that provided them with the opportunity to disseminate newly acquired skills and knowledge; and contributed to the overall socioeconomic development of their home countries or other developing countries.

For that purpose, three types of indicators are included in the framework of JJ/WBGSP: input indicators, output indicators, and impact or outcome indicators. However, only output and impact/outcome indicators were included in the tracer study questionnaire. Self assessment qualitative indicators (considered output as well as impact indicators) are used in order to measure the perceptions of alumni of the benefits and usefulness of the JJ/WBGSP program (World Bank Institute, 2010, p.5).

Outputs indicators included in the JJ/WBGSP framework are related to learning indicators. An example of such indicators is the degree attainment. Impact or outcome indicators (level 4 indicators) are intended to measure the impact of JJ/WBGSP on the socio economic development of alumni countries as well as alumni individual benefits from the program. The impact of JJ/WBGSP is measured through three main indicators: residence status, employment status, and impact on the development environment (World Band Institute, 2010). Table B8 and table B9 (appendix B) provide all kinds of indicators used during evaluation of the JJ/WBGSP program with their operational definitions.

In sum, the framework used to assess the impact of JJ/WBGSP program includes learning, and results indicators components. No reaction and behavior indicators were used during evaluation of JJ/WBGSP.

Evaluation Japan Scholarship Program (JSP)
The Japan Scholarship Program (JSP) supported by the Japan government in the Asian Development Bank (ADB) was launched in 1988. The aim of the program is to "provide qualified citizens with opportunities to pursue further studies at designated national and international institutions recognized for their programs in economics, business and management, science and technology, or any other development-related field" (ADB, 2007, p.iii). Fellows are selected in ADB's developing member countries (DMC) such as Philippines, Thailand, Afghanistan, India, Indonesia, China, Bangladesh, Vietnam, Cambodia, Pakistan, etc. From 1988 till 2006 JSP has provided 2104 scholarships. Fellows may study at master and doctorate levels in 20 designated institutions in 10 countries. They may study in any development related fields of study such as education, agriculture, health, economics, etc. After their study, alumni are expected to return to home countries in order to contribute to the economical and social development by applying their knowledge, skills and attitudes acquired during the study (ADB, 2007).

Evaluation of JSP was undertaken since1999 by the Operations Evaluation Department (OED) of the ADB at request of and with funding from the government of Japan. The aim of the evaluations was to found out whether JSP has contributed to capacity building of scholarship recipients and whether JSP alumni have contributed to the socioeconomic development of their home countries (ADB, 2007, p.1). In order to achieve these goals, ADB conducted in 2007, next to desk review and fieldwork, a tracer study of JSP scholarship recipients. The study questionnaire was designed in order to examine whether the JSP has contributed to the enhancement of the alumni knowledge and skills, whether the alumni have completed their studies, and whether they are in positions that would contribute to the socioeconomic development of their home countries (ADB, 2007, p.1). Two types of indicators are included in the questionnaire: impact indicators related to JSP contribution to development of alumni and impact indicators related to JSP contribution to socioeconomic development of alumni countries (ADB, 2007). Impact indicators related to JSP contribution to development of alumni measure whether the program had contributed to enhancing alumni knowledge, skills and attitudes or whether the program had contributed in enhancing employment of alumni. Examples of such indicators are degree attainment and employment of alumni. Qualitative impact indicators related to JSP contribution to socioeconomic development of alumni countries were used to measure the contribution of JSP in enhancing socioeconomic development in alumni countries. An example of such indicators is JSP contribution to socioeconomic development. Table B10 and table B11 (appendix B) provide all kinds of indicators used during evaluation of the JSP program with their operational definitions.

In sum, the framework used to assess the impact of JSP program includes reaction, learning and results indicators components. No behavior indicators were used during evaluation of JSP.

2.3 Summary literature review
Various frameworks are used to assess impact of educational programs. In these frameworks indicators components are categorized according to logical classifications or thematically. According to logical classification, processes in the environment of the program can

transform inputs into targeted outputs and desired outcomes and impacts. Thematic classification use themes of indicators mostly related to expenditure in education; to conducted activities; and to results (in terms of outputs, outcomes, and impact). Examples of frameworks used in the area of education programs evaluation are the framework developed by Scheerens (2004), the OECD framework, and the four-level framework developed by Kirkpatrick (1996).

The framework developed by Scheerens (2004) is an example of framework within which indicators components are categorized according to logical classification. The Scheerens' framework is an input-process-outcome-context framework. Scheerens' framework builds on empirical findings of school effectiveness research in developed and developing countries and includes education indicators and variables that are empirically proven to be relevant for the evaluation of the quality of education at system, program, school, and classroom level.

The OECD framework is an example of framework within which indicators components are categorized thematically. The OECD framework includes internationally comparable education indicators that could be used to evaluate the outcomes of education systems.

The Kirkpatrick four-level framework is an example of framework used for the assessment of the outcomes of trainings in corporate organizations as well as in the area of fellowship program. The framework include four indicator components related to participants' reaction to training, learning of participants during training, behaviour of participants after training, and results in institutions or organizations due to training received.

These three frameworks and related indicator components are analysed in the third chapter for their appropriateness for the assessment of the impact of IFP.

3. Development of IFP instrument: Pre-selection of indicators and variables

In previous chapters the research questions were explained (see introduction and chapter 1) and the results of the literature review were presented (see chapter 2). Three frameworks were examined in the literature review. These frameworks are analysed in this chapter. Analyses of these frameworks are conducted for answering the second part of the first research question (*to which extent are these frameworks appropriate for the assessment of the impact of IFP?*) and for answering the second research question which is: *What kinds of outcome and impact indicators are necessary to assess the impact of IFP in the area of education?* In chapter 3 the pre-selection process of indicators and variables is explained. Therefore, the three frameworks were first analysed for their applicability as conceptual framework in this study. Then, the criteria used to pre-select indicators and variables were set. Further, indicator included in these frameworks were analysed in order to find out whether they could be pre-selected for the assessment of the impact of IFP.

This chapter includes four sections. The working definition of impact and the conceptual framework within which indicators are pre-selected in this study are given in the first section. The criteria needed for defining applicable indicators and variables in the context of IFP are discussed in the second section. With these criteria the indicators are evaluated in the third section within the context of the reviewed frameworks. In the fourth and final section the indicators and variables pre-selected for the IFP Ed-instrument according to the three components of the conceptual framework are presented.

3.1 Conceptual framework used in this study

Following the goals of IFP (Enders & de Boer, 2002, p.33), impacts of IFP should be measured at individual, institutional and system (country) level. IFP is a fellowship program and fellowship programs are defined as programs that provide training to fellows in specific fields of studies (Aguirre International, 2004; Boeren, 2005; CIDA, 2005; Rotem et al., 2010; USAID, 1995). Therefore, as IFP provides training activities, the three frameworks reviewed in chapter two were analysed according to their relevance to assess training programs and whether they include components appropriate to measure impacts at individual, institutional and system level. The following sub-sections provide first of all the working definition of impact used in this study and then provide analyses of the Scheerens quality framework followed by the OECD framework and the Kirkpatrick framework. Based on these analyses the conceptual framework used to assess the impact of IFP is provided.

3.1.1 Working definition of impact

Definitions of impact abound in educational evaluation literature. Scheerens et al. (2003) define impact as "changes in other sectors of the society that can be seen as the effects of education" (p.217) and relate impact to long term outcomes. Enders & de Boer (2002, p.13) use a similar definition to define impact in the conceptual framework used for formative evaluations of IFP. According to Enders & de Boer, impact refers to the long-term outcomes, that is, "the fundamental change in organizations or communities as a consequence of the outcomes" (Enders & de Boer, 2002, p.14). Enders & de Boer (2002) argue that impact could refer to unintended effect. In the same line, Stufflebeam (in Maudaus et al., 1983) relates impact to long term outcomes and recommends that evaluation of impact should be extended to positive and negative results as well as intended and unintended outcomes. This recommendation is included in the definitions provided by the OECD and by Roche (1999). OECD defines impact as "positive and negative, primarily and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (OECD, 2002, p.24). Effects refer to "intended or unintended change due directly or indirectly to an intervention" (OECD, 2002, p.20). According to Roche (1999) impacts are "lasting or

significant changes – positive or negative, intended or not – in people's lives, brought about by a given action or series of action" (p.302). Lasting changes are sustainable changes that are expected to be seen in the long term and significant changes are temporary changes that are not sustained throughout the time but that are important for the local beneficiaries' lives (Roche, 1999).

Although some slight differences occur, impact is defined by scholars as a long term change that results from activities conducted by an educational program. In the case of fellowships programs, the activities are directly conducted by alumni or network of alumni (Boeren, 2005; Norad, 2005; Rotem et al., 2010; Searle et al., 2006; Strömbom, 1989; Teichler, 1991). Therefore, following CHEPS (Enders & de Boer, 2002, p.13) this study adopts the following definition of impact: *changes in IFP countries as a result of activities conducted by an individual alumnus (alumna) or by a network of alumni*. Two kinds of changes are considered: first order changes and second order changes. The first order changes refer to activities that alumni conduct in order to bring about attitudinal and behavioural changes as well as changes in the functioning and performances of institutions and organizations where alumni are active. These first order changes are considered as the outcomes of IFP. The second order changes are the changes in IFP countries/communities caused by the activities conducted by alumni. Similar to scholars (e.g. Enders & de Boer, 2002; Mikkelsen, 2005; Stufflebeam in Maudaus et al., 1983; Roche, 1999) these changes may be intended or unintended, positive or negative.

3.1.2 Analysis Scheerens' quality framework
The Scheerens' framework is based on findings provided by the school effectiveness research movement in developed and developing countries and includes indicators and variables that are positively associated with educational achievement. These indicators and variables are empirically found by the school effectiveness research to be relevant to measure the quality of education in terms of educational achievement and attainment. The framework is designed to evaluate the quality of education at system level (national education system), at school level, and at classroom level. However, this framework was not developed specifically to assess the outcomes and impact of trainings or fellowship programs. No mention is made by Scheerens that the framework is intended for such a purpose. Evidence of the use of the quality framework by organizations active in the field of fellowship programs was not found in the literature reviewed. No evidence could be found in literature reviewed that indicators included in the quality framework were validated in a population similar to alumni. Further, the framework includes no component appropriate to assess outcomes and impacts at individual level.

3.1.3 Analysis OECD framework
The OECD framework is designed to compare educational performances of countries at system level. The framework includes international indicators designed by groups of statisticians, policy–makers and academics from different institutions. Indicators and definitions used in the framework reflect a consensus among a wide range of professionals including experts from UNESCO-UIS, OECD, and EUROSTAT (UOE). However, the OECD framework is not designed to assess the outcomes and impact of trainings or fellowship programs. Evidence about the use of the OECD framework by organizations active in the field of fellowship programs was not found in the literature explored. No evidence was also found in literature reviewed that indicators included in the OECD framework were validated in a population similar to alumni. Further, the framework includes no component appropriate to evaluate outcomes and impacts at individual and institutional level as all indicators are related to the system level.

3.1.4 Analysis Kirkpatrick framework

The four-level Kirkpatrick framework is specifically designed to measure the outcomes of trainings programs. Each level of the framework measures different aspects of training. According to critics (e.g. Alliger & Janak, 1989; Bates, 2004; Holton, 1996) the reaction component (considered as happy sheets) should not be considered as measures of outcomes of trainings because reaction indicators are limited to opinions of participants to training and do not measure knowledge, skills and attitudes. Critics argue that reaction should be used as variable that influence training behaviour. In line with critics, evaluators in the area of fellowship programs argue that the first level (reaction level) produces little information about impact of fellowship programs (Aguirre International, 2004; Rotem et al., 2010; Zinovieff, 2008). However, except the first level of the Kirkpatrick framework, the other three levels are used (even by critics) to develop new frameworks or to evaluate outcomes of trainings. This indicates implicitly that unanimity exist about the learning, behaviour and results components. Although designed for training programs in corporate contexts, evidence exists about the use of the Kirkpatrick framework by organizations that provide fellowships programs. Further, the Kirkpatrick framework includes components appropriate to evaluate outcomes and impacts at individual, institutional and system level.

3.1.5 Conclusion of the analysis

Scheerens' quality framework and OECD framework appear to be relevant frameworks that could respectively evaluate the quality of education and the educational performances of countries at system level. However, both frameworks are not appropriate − on their own - to evaluate outcomes and impacts of IFP at individual, institutional and system level. This study adopts therefore the Kirkpatrick framework as a guide in pre-selecting indicators that could be used to evaluate the IFP program because the four-level framework is appropriate for the levels (individual, institutional, and system) recommended for the assessment of the impact of the IFP program. However, the reaction level of the four-level framework was not used in this study as no consensus exist between scholars about its use in the area of evaluation of training and fellowship programs and because it is less indicative of impact (Aguirre International, 2004; Rotem et al., 2010; Zinovieff, 2008). Further, this level was not used in this study as reaction indicators should be measured immediately at the end of the training (DFC, n.d.; Kirkpatrick, 1996; Rotem et al., 2010) and the reaction level has no relationship with the objectives of IFP.
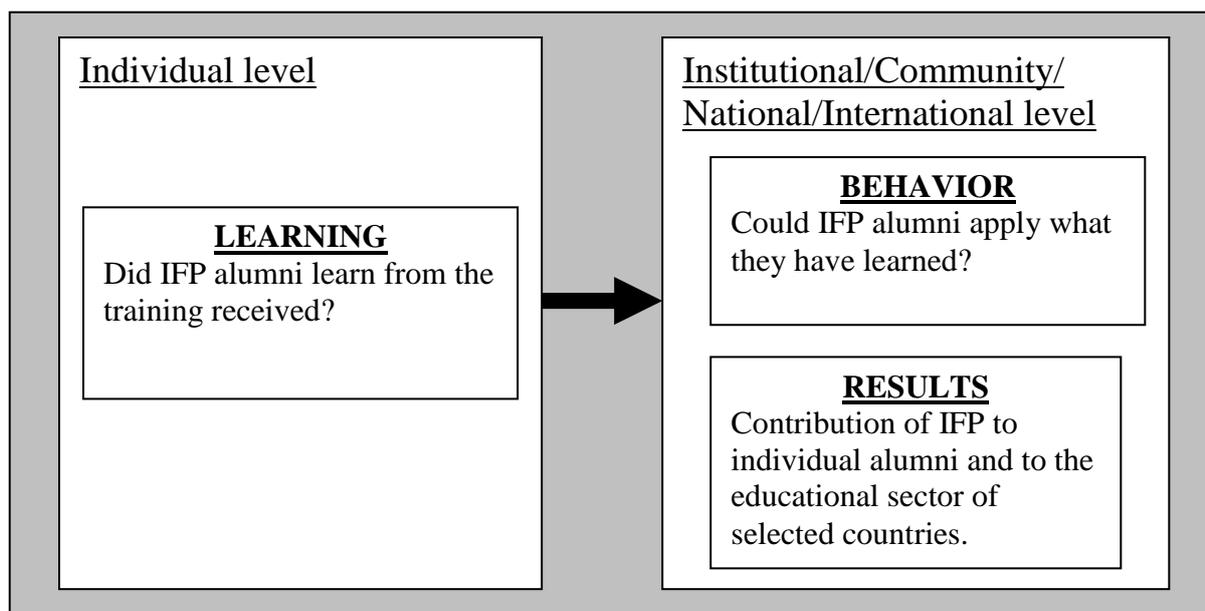
Figure 3.1. Conceptual framework for pre-selection of indicators

The conceptual framework adopted in this study derives from the Kirkpatrick four-level evaluation framework.

Learning refers to whether IFP alumni have gained new knowledge, skills, and attitudes from participating in the training program. Therefore, the learning component should include indicators and variables appropriate to measure the new knowledge, skills and attitudes acquired by IFP alumni.

Behavior refers to the extent to which IFP alumni are able to apply (or to transfer) the acquired knowledge, skills, and attitudes in institutions/organizations and beyond institutions where they are active. Therefore, the behavior component should include indicators and variables appropriate to measure whether IFP alumni are able to apply the acquired knowledge, skills, and attitudes in institutions and organizations and beyond.

Results refer to contribution of the IFP program (through training received by IFP alumni) to the educational sector of recipient countries as well as to individual alumni. Therefore, the result component should include indicators and variables appropriate to measure changes caused by IFP alumni in their countries and communities respectively.

In order to pre-select indicators that could be included in the components of the conceptual framework, the following section provides the criteria needed to define applicable indicators in the context of IFP.

3.2 Selection criteria used in the pre-selection process
This section provides the criteria which were used in this study in order to pre-select indicators from literature. As impact is defined in this study as changes caused by IFP alumni, this study adopts the definition of indicators provided by OECD (2002). According to OECD (2002, p.25) indicators could be defined as "quantitative or qualitative factors or variables that provide a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor" (OECD, 2002, p.25). Therefore, indicators refer in this study to any quantitative and qualitative variables that are appropriate to measure the outcomes and impacts of IFP in the area of education in its selected countries. This section describes first the characteristics of applicable indicators and provides the criteria used in this study in order to pre-select indicators and variables.

3.2.1 Characteristics of applicable indicators

Different approaches of how to choose applicable indicators are described in literature (CIDA, 1997; Kusek & Rist, 2004; Prennushi et al., 2001; Ramos et al. 2004; Roche, 1999). Examples of such approaches are the SMART-model and the CREAM-model.

In order to select applicable indicators, Roche (1999) suggests using the SMART-model. The abbreviation SMART stands for specific, measurable, attainable, relevant, and time-bounded. According to Roche (1999, p.48), indicators should therefore reflect things that the project intends to change (specific), should be defined precisely and should provide objective data that allow comparability and aggregation of changes (measurable), should be achievable and sensitive to changes (attainable), should be easy to collect within reasonable time and cost (relevant), and should indicate when change is expected (time-bound). Next to the SMART-model, Kusek & Rist (2004) recommend to use the CREAM-model as criteria to select applicable indicators. The abbreviation CREAM stands for clear, relevant, economic, adequate, and monitorable. According to Kusek & Rist (2004, p.68) applicable indicators should therefore be precise and unambiguous; should be appropriate to the subject at hand; should be available at a reasonable cost; should provide a sufficient basis to assess performance and should not be too indirect, too much of a proxy, and abstract; should be possible to be validated and verified independently.

Slight differences exist between the SMART and CREAM models. For example the criteria adequate and monitorable of the CREAM-model are not included in the SMART-model. In addition to the SMART-model and the CREAM-model, other characteristics of applicable indicators are mentioned in literature (for more details see CIDA, 1997; OECD, 2004; Prennushi et al., 2001; Ramos et al. 2004; UNESCO, n.d.; World Bank, 2004). Characteristics for applicable indicators mentioned by these scholars slightly differ from each other and slightly differ from the characteristics used in the SMART and CREAM models. For example UNESCO (n.d.) mentions that applicable indicators should be reliable and valid. Validity and reliability are for example not included in the SMART model.

As it can be seen, characteristics for applicable indicators abound in literature and differences in terms of number and wording of characteristics exist. The following attempts therefore to find characteristics which could be used as criteria in this study for the pre-selection of indicators and variables.

3.2.2 Criteria used to pre-select indicators and variables

In order to provide the selection criteria used within this study, selection criteria provided by three organizations active in the area of education and/or fellowship programs were explored and compared with each other. The organizations include the United Nations Educational, Scientific and Cultural Organization (UNESCO), the World Bank, and the Organization for Economic Co-operation and Development (OECD).

Selection criteria of indicators used by the three organizations

According to UNESCO (n.d., p.3) applicable indicators should be:

- *Policy-relevant*, by being capable of providing clear and unambiguous responses to key policy issues and concerns;
- *User friendly*, i.e., comprehensible, timely and few in number;
- *Derived from a framework*, which allows the interpretation of one figure (say enrolment) in the context of other basic variables (say demography and investment in education) of a particular country;
- *Technically sound*, i.e. valid, reliable and comparable;

- *Feasible to measure at reasonable cost*, in that the basic statistics required for deriving them can be either readily available or comparatively easy to collect within a well-defined timeframe.

In the OECD framework data on indicators should meet eight quality dimensions (OECD, 2004, pp.123-124): relevance of the value of data; accuracy of data; credibility of data; timeliness of data, that is, being timely available in order to permit the information to be of value; data should be punctual that is being released in accordance with a publication schedule; data should be easy to locate and to access (accessibility); data should be easy to understand, to use, and analyse by users (interpretability); data should be coherent, that is, being logically connected and mutually consistent. According to the World Bank (World Bank, 2004, pp. 3-4) indicators should meet ten criteria (see table 3.1 below) in order to be considered as applicable. Table 3.1 provides an overview of criteria used by the three organizations.

Table 3.1. Overview of selection criteria (OECD, 2004; UNESCO, n.d.; World Bank, 2004)

UNESCO (n.d., p.3)	OECD (2004, pp.123-124)	World Bank (2004, pp. 3-4)
1- Policy-relevant	1- Relevance	1- Direct, unambiguous measure of progress
2- User friendly	2- Accuracy	2- Vary across group, areas, and over time
3- Derived from a framework	3- Credibility	3- Have a direct link with interventions
4- Technically sound	4- Timeliness	4- Be relevant for policy making
5- Feasible to measure at reasonable cost	5- Punctuality	5- Be consistent with the decision-making cycle
	6- Accessibility	6- Not be easily manipulated
	7- Interpretability	7- Be easy to measure and not too costly to measure
	8- Coherence	8- Be easy to understand
		9- Be reliable
		10- Consistent with data availability and data collection capacity

As it can be seen from table 3.1, criteria for applicable indicators used by these organizations differ from each other. The differences are mostly related to the number and wording of criterion used by each organization. For example: User friendly criterion mentioned by UNESCO means that indicators should be few in number as well as being comprehensible and timeliness. However, timeliness is mentioned as isolated criterion by OECD and the World Bank refers to timeliness as being consistent with the decision-making cycle. In spite of such differences, there is consensus about three criteria: the relevance of the indicators, the scientific quality of the indicators, and the measurability of the indicators. These three criteria are explained hereafter.

Policy relevance of indicators: Indicators should have "a direct link with the intervention" (World Bank, 2004, p.3). In other words, indicators should have a relationship with the direct objectives of an intervention and therefore should be appropriate to measure the changes caused by the intervention. According to UNESCO (n.d., p.3) policy-relevance means that indicators should be "capable of providing clear and unambiguous responses to key policy

issues and concerns" UNESCO (n.d., p.3). Information provided by indicators should therefore be useful to key users (program managers and other stakeholders) and should help guide decisions that these users will need to make.

Scientific quality of indicators: Indicators should be scientifically sound, that is, they should be valid and reliable. An indicator should measure what it is intended to measure (validity) and should provide stable and accurate results when repeated in the same populations and circumstances (reliability) (Kusek & Rist, 2004). According to UNESCO indicators should derive wherever possible from a known framework (UNESCO, n.d.). Further, indicators should be simple, well defined and should provide objective data that allow wherever possible international comparability and disaggregation of changes (OECD, 2004; UNESCO, n.d.; World Bank, 2004). Indicators should not be sensitive to manipulation (World Bank, 2004).

Measurability of indicators: Indicators should be measurable. Measurability is related to the availability of data on indicators in the concerned country or to the possibility to collect objective data on the indicators at reasonable cost (UNESCO, n.d.). Therefore, indicators should be consistent with data availability and data collection capacity of countries (OECD, 2004; UNESCO, n.d.; World Bank, 2004).

Selection criteria used in this study
Policy relevance, scientific quality, and measurability were used in this study as criteria during pre-selection and selection of indicators because consensus exists about their use as criteria for applicable indicators. Therefore, during the pre-selection process, all identified indicators and variables were analysed according to whether they are policy relevant for IFP, that is, whether they have a relationship with IFP intended objectives. According to CHEPS (see Enders & de Boer, 2002, p.33), the intended objectives of IFP should be measured at individual, institutional, and system level. Pre-selected indicators should therefore have a relationship with the objectives of IFP at individual, institutional or system level. Data on pre-selected indicators should be useful for decisions that IFP policy makers will need to take. According to experts interviewed during orientation interviews, data about indicators are useful for IFP policy makers if IFP alumni could influence these indicators. This stresses the importance to pre-select indicators that could be influenced by IFP alumni. Further, all identified indicators and variables were analysed according to whether they have evidence of validity and reliability in literature. Scholars recommend using, wherever possible, indicators with evidence of validity and reliability (Cook & Beckman, 2006; Moore & Benbasat, 1991; Straub et al., 2004). If not available, validity and reliability of indicators could still be evaluated after field testing. Pre-selected indicators and variables should therefore be validated/tested in the context of fellowship programs. Furthermore, all identified indicators and variables were analyzed according to whether they are measurable. Pre-selected indicators should be measurable in the context of IFP, that is, whether data on indicators are available in an up-dated database or are easy to collect.

In the following all identified indicators included in frameworks reviewed in chapter two are analyzed according to the criteria set in this study in order to pre-select indicators and related variables that could be included in the components of the conceptual framework used in this study.

3.3 Analysis of indicators
Analysis of indicators focused on evidence of validity and reliability in literature and on policy relevance and measurability of indicators in the context of IFP. Although the OECD

and the quality frameworks were not found appropriate on their own to assess the impact of IFP, indicators included in these frameworks were analyzed in this section in order to find out whether they could be included in the components of the conceptual framework used in this study. Further, indicators that could be included in the components of the Kirkpatrick framework during evaluation of fellowship programs were analyzed. The following provides successively the analysis of indicators included in the components of Scheerens' framework, in the components of OECD framework, and in the components of the Kirkpatrick framework during evaluation of fellowship programs.

3.3.1 Analysis of OECD indicators

Indicators included in the OECD framework are designed to measure the outcomes of education at system level (see indicators in chapter 2). According to OECD (2012) the aim of indicators included in the OECD framework is to compare educational systems in order to support educational policy makers and practitioners in reforming their policies.

Evidence of validity and reliability

Estimates of validity and reliability of OECD indicators are not available in literature reviewed. However, OECD indicators are the product of ongoing review processes conducted by different groups of international experts (OECD, 2004). Quality of data is for example controlled by the technical group where OECD works jointly with UNESCO-UIS and EUROSTAT. Indicators used in the OECD framework are designed by groups of statisticians, policy–makers and academics. Indicators and definitions used in the framework reflect a consensus among these professionals. Feedback from different group of experts ensures content validity of instruments (Fraenkel et al., 2012; Haynes, Richard & Kubany, 1995; Litwin, 1995; Straub, 1989; Straub et al., 2004; Wynd, Schmidt & Schaefer, 2003). Although estimates of reliability are not available in literature reviewed, methodological and technical issues (definitions – measurement – data sources) are well described in the OECD handbook for internationally comparative education statistics (OECD, 2004). OECD indicators are field tested before being used in survey such as the Programme for International Student Assessment (PISA). OECD has standardized data collection procedures with the use of the International Standard Classification of Education (ISCED) framework for definitions and by providing data collection guidelines to member countries (OECD, 2004). Standardization of procedures and methods is a way of producing consistent findings over time.

Policy relevance and measurability

Indicators included in the OECD framework are designed to measure the outcomes of education at system level. Data on these indicators could be useful for IFP policy makers to evaluate the objectives of IFP at system level. However, literature in the field of fellowship evaluation argues that it is difficult or even impossible to establish influence of alumni at system level because of context factors (Krasulin et al., 1998; Norad, 2005; Rotem et al., 2010; USAID, 1991). At national level, contributions of individual alumni may be small and therefore difficult to identify and to measure (USAID, 1991). Similarly, all experts interviewed during orientation interviews argue that it will be difficult to measure the impact of IFP at system/country level because of context factors that can influence the measurement. Examples of comments made include:
"It is too ambitious to develop an instrument to measure the impact of IFP at country level" (respondent R2).
"Outcomes are measurable but forget the impacts at system level" (respondent R6).
"It will be quite difficult to get the impact clear as politics play a considerable role in these countries" (respondent R2).

"Fellowships are about individuals and their impacts in a country will not be visible and also difficult to measure" (respondent R4).

According to respondent (R2) it will be difficult to establish a link between activities conducted by alumni and changes in the educational system because many decisions could be taken by other stakeholders in the area of education. Similarly, respondent (R6) argued that too many factors will influence the measurement of impacts at system level. Therefore, it will be difficult to attribute changes to alumni. Further, respondents recommend not to include system level quantitative indicators such as gross enrolment rate (GER), net enrolment rate (NER), drop out rate etc. According to respondent (R9) "it would be unrealistic to assume that alumni can influence country level indicators such as enrolment rate, dropout rate etc.". According to this respondent alumni alone are not able to influence such indicators. The statements of these consulted experts stress how difficult it is to attribute changes at system level to activities that IFP alumni will conduct. Because of these attribution issues, OECD indicators were found not measurable in the context of IFP.

In sum, indicators included in the OECD framework are good indicators that can measure the performance of education at system level. However, OECD indicators cannot be influenced by IFP alumni on their own. Data provided by OECD indicators could therefore be difficult to be attributed to IFP alumni because of measurement issues. These data are therefore not useful for IFP policy makers. For these reasons, OECD indicators were not pre-selected for the IFP instrument.

3.3.2 Analysis of indicators included in the quality framework

Indicators included in Scheerens' quality framework are designed to evaluate the quality of education at system level (national education system), at school level, and at classroom level (see indicators in chapter 2).

Evidence of validity and reliability

Evidence of validity of indicators included in the quality framework is not available in literature on school effectiveness research reviewed. However, the quality framework includes indicators and variables that are positively associated with educational achievement. Further, consensus exists in the school effectiveness research community about the variables included in the quality framework. Therefore, it could be assumed that these indicators are content valid measures of the quality of education in terms of student's achievement and attainment. Evidence of reliability of the indicators is also not supported by literature. Measurement issues related to operationalization of variables/factors are mentioned in Scheerens et al. (2003). According to Scheerens et al. (2003), operational definitions vary across school effectiveness studies and "there is little agreement, at the operational level, on the substance of the key factors that are supposed to determine school effectiveness" (p.298). Standardization of research instruments do not exist (Scheerens et al., 2003). Without standardization, consistency of findings is difficult to establish between studies and over time.

Policy relevance and measurability

The framework includes indicators and variables that can be manipulated by actors in the area of education such as "policy planners, local constituencies, school managers, and teachers" in order to increase the quality of education (Scheerens, 2004, p.54). They can help policy makers to make decisions to address educational quality issues. Indicators included in the quality framework are designed to measure the quality of education at system, as well as at school and classroom level. For the same reasons as mentioned in the section above, system level inputs, process and outcome/impact indicators included in the quality framework could

not be influenced by alumni on their own and attribution to alumni will be difficult to establish. Further, such indicators are difficult to be measured in the context of fellowship programs because of the influence of external factors. However, the quality framework includes input and process factors that can be manipulated at school and classroom level. There is evidence that alumni are active at school and classroom levels (Kottmann & Enders, 2009). In spite of such evidence, it is for example difficult to establish that the degree of parents' involvement in schools is only attributable to activities that alumni conduct. Therefore, because of issues of attribution, input and process factors at school or classroom level were not used as impact indicators at institutional level but rather as outcomes variables (activities that alumni could conduct). Data on these indicators could be policy relevant at institutional level as IFP aims to investigate whether alumni conduct activities that foster quality of education.

In sum, indicators included in the Scheerens' quality framework are appropriate to measure the quality of education at system, as well as at school and classroom level. However, system level indicators were not pre-selected in this study as they cannot be influenced by IFP alumni alone. Further, because of attribution and measurement issues data on such indicators are not useful for IFP policy makers. Variables at school and classroom level were pre-selected as outcomes variables in order to understand whether IFP alumni conduct such activities.

3.3.3 Analysis of indicators used during evaluation of ATLAS/AFGRAD
Indicators included in the framework used to assess the impact of ATLAS/AFGRAD program focus on learning, behavior, results, and higher impacts levels of assessment (see chapter 2).

Evidence of validity and reliability
Estimates of validity and reliability of ATLAS/AFGRAD indicators are not available in the literature reviewed. Data collection methodology and procedures and sampling techniques are well described by the evaluation team of ATLAS/AFGRAD program. The evaluation team only reports that different types of statistical reliability were accessed in order to increase the credibility of findings. But no details about these reliability estimates were mentioned. ATLAS/AFGRAD indicators were validated in a population similar to IFP alumni. Different pretesting rounds with expert groups were also reported by the evaluation team. According to Aguirre International (2004), the draft of the questionnaire instrument was first evaluated by evaluation and survey specialists. After evaluation, the draft was revised and tested with ATLAS/AFGRAD participants in countries selected for site visits, then revised and finalized. Pretesting with experts groups is a way to ensure content validity of instruments (Straub, et al., 2004). Further, according to Aguirre International (2004) data collected from alumni are about perceived impact. Therefore, the evaluation team used different methods such as participant surveys, country site visits, internet search, and interviews with alumni supervisors, USAID/Washington education specialists, and USAID mission staff. The use of different methods (triangulation) is a way to ensure consistency and credibility of findings in qualitative research.

Policy relevance and measurability
ATLAS/AFGRAD indicators are appropriate to evaluate impact at individual level (e.g. education degree obtained – employment of alumni), at institutional level (e.g. Applicability of acquired knowledge, skills and attitudes at institutions or at other areas) and at system level (e.g. contribution of alumni to sectoral, national, regional or international levels impacts). Data provided by such indicators are therefore useful for IFP policy-makers as IFP aims to measure the impact at these three levels. According to Aguirre International (2004) measurement issues

could be related to poor contact data. ATLAS/AFGRAD quantitative indicators are measurable if an alumni data base is available with contact data on alumni. Lack of accurate participant contact information is an issue of measurability (Aguirre International, 2004). The qualitative indicators (alumni perceptions) are measurable if alumni accept to participate in the survey and fill in the questionnaire.

In sum
Although the lack of validity and reliability estimates, quality of ATLAS/AFGRAD indicators was addressed through triangulation. The indicators are policy relevant for IFP as they are appropriate to measure the objectives of IFP at individual, institutional and system level. Further, the indicators are measurable since CHEPS has conducted evaluation studies that resulted in a database with contact information on alumni in terms of phone numbers, emails addresses, name employers, country of residence etc.. Therefore, ATLAS/AFGRAD indicators were pre-selected for the IFP instrument as they are policy relevant and measurable.

3.3.4 Analysis of indicators used during evaluation of CFSP
Indicators included in the framework used to assess the impact of CFSP program focus on reaction, learning, behavior, and results levels of assessment (see chapter 2).

Evidence of validity and reliability
Validity and reliability estimates of CFSP indicators are not available in literature provided by CIDA. However, the indicators were validated in a population similar to IFP alumni. The evaluation team of CFSP also used different methods such as online survey, interviews with alumni, and interviews with alumni employers in country site visits in order to ensure consistency and credibility of findings.

Policy relevance and measurability
CFSP indicators are appropriate to evaluate impact at individual level (e.g. benefits academic training), at institutional level (e.g. fellows effects on their organizations' performance) and at system level (e.g. CFSP contribution to sustainable development in recipient countries). Data provided by such indicators are therefore useful for IFP policy-makers. CFSP quantitative indicators are measurable if an alumni data base is available. The qualitative indicators (alumni perceptions) are measurable if alumni accept to participate in the survey and fill in the questionnaire.

In sum
CFSP indicators are policy relevant for IFP as they are appropriate to measure the objectives of IFP at individual, institutional and system level. Further, CFSP indicators are measurable in the context of IFP as CHEPS has a database with contact information on alumni. Although the lack of validity and reliability estimates in literature, quality of CFSP indicators was addressed through triangulation. Therefore, CFSP indicators were pre-selected for the IFP instrument.

3.3.5 Analysis of indicators used during evaluation of JJ/WBGSP
Indicators included in the framework used to assess the impact of JJ/WBGSP program focus on learning, and results levels of assessment (see chapter 2).

Evidence of validity and reliability
Evidence of validity and reliability of JJ/WBGSP indicators is not supported by literature. Data collection methodologies and instruments are reported but details about the quality of indicators are not available in literature provided by the World Bank Institute. However, the indicators were validated in a population similar to IFP alumni. The same questionnaire, with slight modifications, was used during the tracer studies. Different methods were used in order to ensure consistency and credibility of findings.

Policy relevance and measurability
JJ/WBGSP indicators are appropriate to evaluate impact at individual (e.g. benefits academic training) and at system level (alumni contribution to the socio-economic development of home country). Data provided by such indicators are therefore useful for IFP policy-makers. JJ/WBGSP quantitative indicators are measurable if an alumni data base is available. The qualitative indicators (alumni perceptions) are measurable if alumni accept to participate in the survey and fill in the questionnaire.

In sum
JJ/WBGSP indicators are policy relevant for IFP as they are appropriate to measure the objectives of IFP at individual and system level. Further, JJ/WBGSP indicators are measurable in the context of IFP as CHEPS has a database with contact information on alumni. Quality of JJ/WBGSP indicators was addressed through triangulation of methods. Therefore, JJ/WBGSP indicators were pre-selected for the IFP instrument.

3.3.6 Analysis of indicators used during evaluation of JSP
Indicators included in the framework used to assess the impact of JSP program focus on reaction, learning, and results levels of assessment (see chapter 2).

Evidence of validity and reliability
Evidence of validity and reliability of JSP indicators is not supported by literature. Data is gathered through longitudinal studies and comparisons of cohorts (ADB, 2007). However, the indicators were validated in a population similar to IFP alumni. Different methods such as survey questionnaire, interviews with JSP stakeholders, and fieldwork in institutions were used in order to ensure consistency of credibility of findings.

Policy relevance and measurability
JSP indicators are appropriate to evaluate impact at individual level (e.g. employment status and career progression) and at system level (JSP contribution to socioeconomic development). Data provided by such indicators are therefore useful for IFP policy-makers. JSP quantitative indicators are measurable if an alumni database is available. The qualitative indicators (alumni perceptions) are measurable if alumni accept to participate in the survey and fill in the questionnaire.

In sum
JSP indicators are policy relevant for IFP as they are appropriate to measure the objectives of IFP at individual and system level. Further, JSP indicators are measurable in the context of IFP as CHEPS has a database with contact information on alumni. Quality of JSP indicators was addressed through triangulation of methods. Therefore, JSP indicators were pre-selected for the IFP instrument.

As it can be seen from the above analyses, indicators included in the OECD framework and Scheerens quality framework are good indicators that are appropriate to measure the performances of educational systems and the quality of education but are not relevant to evaluate the impact of IFP. OECD indicators were not pre-selected for the IFP Ed-instrument. Variables at school and classroom level used in the Scheerens' quality framework were pre-selected as outcomes variables in order to understand whether IFP alumni conduct such activities. Indicators and variables used during evaluation of the four fellowship programs reviewed in chapter two were pre-selected for the IFP Ed-instrument.

3.4 Overview of pre-selected indicators and variables

All indicators and variables found appropriate for the assessment of IFP were pre-selected and adapted to the context of IFP (see appendix C for more details). However, indicators that were common in frameworks reviewed are only once mentioned. For example the indicator education degree attainment used by all fellowship programs reviewed as indicator of learning is only once mentioned in this study. In line with the conceptual framework used in this study, the following provides respectively learning indicators, behavior indicators, and results indicators pre-selected for the IFP Ed-instrument.

Learning indicators

Learning indicators measure whether IFP alumni have gained new knowledge, skills, and attitudes from participating in the training program. They refer to IFP alumni improvement in knowledge, increase in skills and attitudinal changes caused by participation in the training program. Table 3.2 depicts learning indicators and variables pre-selected for the IFP Ed-instrument.

Table 3.2. Learning indicators and variables pre-selected for IFP Ed-instrument

Type of indicators	Variables
Education attainment	Degree obtained (level of study) – Field/kind of study – place of study
Effectiveness and relevance of training	Perception of effectiveness and relevance of training received during fellowship
Contribution of training to acquisition specific knowledge, skills, and attitudes	Perception of contribution of training received to acquisition of specific knowledge, skills, and attitudes

Learning indicators are related to the objectives of IFP at individual level (see chapter 1). Measurement of learning is important for attribution (Aguirre International, 2004). According to Aguirre International (2004), attribution of alumni' impacts to fellowship program is impossible without learning acquired during training. Further, learning variables (e.g. education degree, year of education) are considered as control variables that influence career progress/success (Heslin, 2005; Lortie-Lussier & Rinfret, 2005; Melamed, 1995; Polk & Armstrong, 2001; Turban & Dougherty, 1994; van der Sluis & Poell 2003; Wayne, Liden, Kraimer, & Graf, 1999; Whitely, Dougherty, & Dreher, 1991).

Behavior indicators

Behavior indicators measure the extent to which IFP alumni are able to apply (or to transfer) the acquired knowledge, skills, and attitudes in institutions/organizations and beyond

institutions where they are active. They refer to IFP alumni job performance that could be attributed to knowledge, skills and attitudes gained through participation in training program. Table 3.3 depicts behavior indicator and variable pre-selected for the IFP Ed-instrument.

Table 3.3. Behavior indicators pre-selected for the IFP Ed-instrument

Type of indicator	Variables
Applicability of acquired knowledge, skills, and attitudes	Applicability of acquired knowledge, skills and attitudes at institutions/organizations or at other areas such as in community where alumni are active

Behavior indicators are related to the objectives of IFP at institutional level. The possibility to apply acquired knowledge is important to be measured (Krasulin, et al., 1998). According to Krasulin, et al. (1998), fellowships will contribute to capacity building in recipient countries only if alumni get the opportunity to apply their newly acquired knowledge or skills. This is supported by Rotem et al. (2010), Norad (2005) and Boeren (2005). According to Rotem et al. (2010 p1) "it is necessary to ascertain that fellows are using what they have learned". Similarly, Boeren (2005) and Norad (2005) argue that fellows may have impact when they can fully apply what they have learned.

Results indicators
Results indicators refer to outcome and impact indicators that measure whether the IFP program (through training received by alumni) has contributed to bring about educational changes in institution/organizations as well as in communities, countries. These indicators also measure the benefits of IFP for individual alumni. Quantitative indicators (objective: e.g. employment status after training) and qualitative indicators (subjective perceptions: e.g. alumni perceptions of program contribution in own country) were included in the framework. Table 3.4 depicts result indicators and variables pre-selected for the IFP Ed-instrument.

Table 3.4. Results indicators pre-selected for the IFP Ed-instrument

Type of indicators	Variables
Employment status	Place of employment – Type of job – Sector of employment – Level of employment
Career advancement/progress	Job position – Number of promotions since graduation – Responsibility in current work – Locus of decision – Satisfaction with career
Income	Income of alumni after graduation
Conducted activities	Activities that alumni conduct after graduation: activities related to financial and resources issues, activities related to human resources issues - activities related to students' background conditions - activities related to community/parents' involvement - activities related to instructional issues
Contribution of fellowship to personal alumni	Personal benefits received by alumni from participation in fellowship program
Contribution of fellowship to alumni home country	Contribution of training received to alumni home country

Results indicators are related to the objectives of IFP at individual (e.g. alumni employment status after training) as well as institutional and system level. Following the goals of IFP (Enders & de Boer, 2002, p.33), outcomes and impacts of IFP should be measured at individual, institutional and system level. System refers to IFP countries as well as communities where alumni are active (Enders & de Boer, 2002, p.33). Similar results indicators are included in the framework of other organizations that provide fellowships such as Norad (Norad, 2005), Commonwealth Scholarship Commission (Commonwealth Scholarship Commission [CSC], 2009), and the UN fellowship programmes (Rotem et al., 2010).

Based on the conceptual framework and the selection criteria used in this study, ten core indicators and 22 related variables were pre-selected in order to develop the IFP Ed-instrument. In addition to these ten indicators, one indicator (socio-biographical backgrounds of alumni) and three related variables (gender – age – marital status) were pre-selected as recommended by respondents during orientation interviews in order to allow international comparability and disaggregation into gender, age etc. A total of 11 indicators and 25 variables were therefore pre-selected for the IFP Ed-instrument. The quality (in terms of relevance and measurability) of these pre-selected indicators and variables was again evaluated by experts active in the field of fellowship programs as well as in the field of education. For that purpose, all indicators/variables were first operationalized before evaluation by experts as recommended by DeVellis (2012). Appendix C includes more details about operationalization of pre-selected indicators and variables. Since each fellowship program has its own objectives, the operationalization was based on specific objectives of IFP. Procedures and methods used for selection of final indicators and variables as well as procedures used for the development of IFP Ed-instrument are explained in the following chapter.

4. Research methodology

The previous chapter has described the first step of the development process of the IFP Ed-instrument. Following the first step, three other steps were used to develop the IFP Ed-instrument. These steps are described in this chapter four. As mentioned in the first chapter these steps are conducted for answering the third and last research question: *How could the quality in terms of validity and reliability of the developed instrument be tested?*

'Quality of instrument' refers in this study to the validity and reliability of the developed IFP Ed-instrument. In this chapter the methods, the procedures, and the samples used in each step of the development process of the IFP Ed-instrument are depicted.

This chapter is divided into three sections. As the objective of this study was to develop a valid and reliable instrument, in the first section the types of instrument validity and reliability used in literature are illustrated and then the types of instrument validity and reliability used to assess the quality of the IFP Ed-instrument are provided. The steps used in this study in order to develop the IFP Ed-instrument are described in the second section, followed by sampling procedures, data collection and data analysis procedures in the third and last section.

4.1 Instrument validity and reliability

Definitions of instrument validity and reliability abound in literature (DeVellis, 2012; Dooley, 2001; Fraenkel et al., 2012; Kusek & Rist, 2004; Litwin, 1995; Neuman, 2000). Instrument validity for example is defined by Litwin (1995, p.33) as a measure of how well an instrument measures what it is intended to measure. According to Fraenkel et al. (2012, p.147) reliability refers to "the consistency of scores or answers from one administration of an instrument to another, and from one set of items to another". However, definitions and procedures to estimate validity and reliability depend on the nature (quantitative or qualitative) of data collected (Fraenkel et al., 2012; Golafshani, 2003; Merriam, 1995; Neuman, 2000). Variables pre-selected for the IFP Ed-instrument can produce quantitative (e.g. number of IFP alumni graduated) or qualitative data (e.g. alumni perceptions of personal benefits received from IFP). Therefore, the following describes first the different kinds of instrument validity and reliability mentioned in literature and explains how they could be estimated in the case of quantitative or qualitative data. Based on these findings this section further describes the types of instrument validity and reliability used to assess quality of the IFP Ed-instrument.

4.1.1 Validity of quantitative data

Quantitative data are data that are reduced to numerical scores (Fraenkel et al., 2012). An example of quantitative data in this study is data that could be generated with the pre-selected variable number of IFP alumni graduated in the area of education. Types of instrument validity used in quantitative research include: face validity, content validity, construct validity and criterion validity (Cook & Beckman, 2006; Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000).

Face validity

Face validity refers to the degree to which items of an assessment instrument appear appropriate to measure what they are supposed to measure (Cook & Beckman, 2006; Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000). Face validity is sometimes associated to content validity in the literature (DeVellis, 2012; Dooley, 2001; Litwin, 1995; Neuman, 2000) or believe to be a component of content validity (Haynes et al., 1995). However, DeVellis (2012, p.71) argues that face validity is an informal and a less structured assessment compared to content validity. Face validity is perceived by some scholars as the least scientific measure of validity while others argue that it could be used as an indirect

approach to estimating content validity (Cook & Beckman, 2006; Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000).

Content validity
Content validity could be defined as "a matter of determining if the content that the instrument contains is an adequate sample of the domains of content it is supposed to represent" (Fraenkel et al., 2012, p.150). Scholars agree that content validity could be established through literature reviews (to identify items) followed by evaluations of these items by experts judges with some expertise of the subject under study (Lawshe, 1975; Litwin, 1995; Fraenkel et al., 2012; Haynes et al., 1995; Straub, 1989; Straub et al., 2004; Wynd, Schmidt & Schaefer, 2003). Content validity can be quantified. One way of quantifying content validity of instruments is to compute Lawshe' (1975) content validity ratio (CVR) or Lynn' (1986) content validity index (CVI) (DeVon et al., 2007; Polit & Beck, 2006; Polit, Beck & Owen, 2007; Straub et al., 2004; Wynd et al., 2003;). As content validity is a matter of judgment between experts, kappa coefficient of agreement could be used as estimate of the degree of agreement between experts (Cohen, 1960; Fleiss, 1971; Polit et al., 2007; Wynd et al., 2003).

Construct validity
Construct validity is "the degree to which an assessment instrument measures the targeted constructs" (Haynes et al., 1995, p.3). Wherever possible scholars recommend to establish construct validity of instruments (Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Straub et al., 2004). However, scholars agree that construct validity of instrument is difficult to establish as many types of evidence are needed (Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Straub et al., 2004). Evidence needed include (among other) evidence of content validity and evidence of criterion validity (DeVon et al., 2007; Fraenkel et al., 2012; Haynes et al., 1995). Construct validity includes two forms of instrument validity: convergent validity and divergent (discriminant) validity (Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Straub et al., 2004). Convergent validity "implies that several different methods for obtaining the same information about a given trait or concept produce similar results" (Litwin, 1995, p.43). For an instrument to have divergent validity "it must be shown not to correlate too closely with similar but distinct concepts or traits" (Litwin, 1995, p.44). Construct validity could be estimated by using factor analysis or multitrait-multimethod matrix (MTMM) methods (for more details about these techniques see Dooley, 2001; Fraenkel et al., 2012; Straub, 1989; Straub et al., 2004). These methods are used to analyze scores from data collected after administration of the instrument.

Criterion validity
According to Neuman (2000, p.168) criterion validity uses a standard or criterion to indicate accuracy of a construct. To estimate criterion validity of an instrument researchers compare performance on the instrument with performance on a second test that measures the same variable (Fraenkel et al. (2012, p.152). Criterion validity includes two kinds of instrument validity: concurrent validity and predictive validity (Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000; Straub et al., 2004). Concurrent validity measures "how well the item or scale correlates with 'gold standard' measures of the same variables" (Litwin, 1995, p.45). Concurrent validity statistic is calculated as a correlation coefficient with that gold standard (Litwin, 1995, p.45). The gold standard is a generally accepted standard instrument with evidence of validity and reliability (Litwin, 1995). Predictive validity refers to how well an instrument predicts future scores of respondents (Fraenkel et al., 2012; Litwin, 1995). In order to estimate predictive validity, instrument scores and criterion scores are

gathered at different period of times (Fraenkel et al., 2012). Predictive validity is calculated as correlation coefficient between the initial test and the secondary outcome (Litwin, 1995).

4.1.2 Reliability of quantitative data

Reliability is an evaluation of instrument accuracy (Fraenkel et al., 2012; Litwin, 1995; Straub, 1989). Reliability refers to consistency of measurement scores from one assessment to another (Cook & Beckman, 2006; Dooley, 2001; Fraenkel et al., 2012; Litwin, 1995). Three main types of reliability are mentioned in literature: test-retest, alternate-form, and internal consistency (Fraenkel et al., 2012; Litwin, 1995; Neuman, 2000; Straub et al., 2004).

Test-retest is a measure of how reproducible a set of results is (Litwin, 1995, p.8). It involves the administration of the same test twice to the same group of respondents after a certain time interval (Fraenkel et al., 2012; Litwin, 1995; Straub et al., 2004). After completion of both tests, test-retest reliability is estimated by calculating the correlation coefficients in order to compare the two sets of responses (Fraenkel et al., 2012; Litwin, 1995). Alternate-form reliability involves the administration of two different but equivalent forms of a test to the same group of respondents (Fraenkel et al., 2012; Litwin, 1995). After administration, a reliability coefficient is then calculated between the two sets of scores obtained (Fraenkel et al., 2012; Litwin, 1995). Internal consistency method is commonly used to assess instrument reliability (Fraenkel et al., 2012; Litwin, 1995, p.21; Straub et al., 2004). Internal consistency indicates "how well the different items measure the same issue" (Litwin, 1995, p.21). It requires only a single administration of an instrument to a group of respondents (Fraenkel et al., 2012; Litwin, 1995; Straub et al., 2004). Internal consistency is estimated by calculating Cronbach's coefficient alpha or by using other procedures such as split-half procedure (Spearman-Brown formula) and Kuder-Richardson approaches (KR20 and KR21) (Fraenkel et al. ,2012; Litwin, 1995; Straub et al., 2004). Cronbach's alpha coefficient is the most frequently used statistic to show internal consistency reliability.

4.1.3 Validity and reliability in qualitative research

Qualitative research focuses more on narrative description than numerical scores (Fraenkel et al., 2012, p.426). An example of qualitative data in this study is data that could be generated with the pre-selected variable alumni perceptions of personal benefits received from IFP. According to Neuman (2000, p.417) qualitative data represent or describe "people, actions, and events in social life". Qualitative data are collected in the form of written words, of impressions, sentences, photos, symbols, images from documents, observations, transcripts or pictures rather than numbers (Dooley, 2001; Fraenkel et al., 2012; Neuman, 2000). Qualitative data are usually obtained from less structured, open-ended data collection procedures such as interviews, observations and focus groups (Fraenkel et al., 2012; Neuman, 2000). The terms reliability and validity (as defined in quantitative research) are not used in qualitative research because of the nature of data (Fraenkel et al., 2012; Golafshani, 2003; Merriam, 1995). According to Lincoln & Guba (1988) in Merriam (1995, p.56) qualitative researchers do not attempt to replicate study results but to find out "whether the results of a study are consistent with the data collected". Terminologies used in qualitative research to refer to validity and reliability include (among others) credibility, consistency or dependability, applicability or transferability, and trustworthiness (Fraenkel et al., 2012; Golafshani, 2003; Merriam, 1995).

A number of procedures are recommended by scholars (Fraenkel et al., 2012; Golafshani, 2003; Merriam, 1995) in order to ensure credibility or consistency of data collected and findings. Examples of such procedures include: the use of multiple methods of data collection and data analysis in order to cross-check data and findings, asking someone else to check whether findings are consistent with the data collected and to check the accuracy of the report

(peer examination), asking an individual outside of the study to review and evaluate the report and the findings (external audit).

4.1.4 Summary validity and reliability
To ensure quality of instruments, validity and reliability should be established. Four types of instrument validity could be used to estimate validity of quantitative data. They include face validity, content validity, construct (convergent and divergent) validity, and criterion (concurrent and predictive) validity. Depending on the objectives of a study, scholars recommend to establish one or more of these types of validity in order to ensure quality of instruments. Three main types of reliability can be used to estimate instrument reliability: test-retest, alternate-form, and internal consistency. Scholars agree that reliability of new instrument is mostly estimated through internal consistency methods. In qualitative research (qualitative data) the terms credibility, consistency or trustworthiness are used rather than validity and reliability. Procedures used to ensure credibility or consistency of qualitative data include triangulation of methods by researcher, peer examination, and external audit.

4.1.5 Types of validity and reliability used for the IFP Ed-instrument
As summarized above, different kinds of instrument validity and reliability tests could be used to ensure the quality of instruments. Depending on the study objectives, scholars recommend to test instruments against one of more of these types of instrument validity and reliability (DeVellis, 2012; Moore & Benbasat, 1991; Radhakrishna, 2007; Sendjaya, 2003; Straub et al., 2004). The following provides therefore the types of instrument validity and reliability used within this study in order to assess the quality of IFP Ed-instrument:

Face validity was not used in this study as a type of instrument validity because scholars do not agree about its use as a type of validity in literature.

Content validity of the IFP Ed-instrument was established in this study to ensure that pre-selected indicators and variables are appropriate to measure the outcomes and impacts of IFP. Content validation during development of instruments is necessary to identify and delete variables not relevant for the objectives of the instruments (DeVellis, 2012; Haynes et al., 1995; Moore & Benbasat, 1991; Sendjaya, 2003; Straub et al., 2004)

Construct validity of the IFP Ed-instrument was not established in this study. As pre-selected indicators and variables are quantitative as well as qualitative, one way of establishing construct validity of the IFP Ed-instrument was to use triangulation of methods such as country site visits in IFP countries and interviews with IFP alumni employers (Fraenkel et al., 2012; Litwin, 1995, Merriam, 1995). However, country site visits were outside the scope of this study as this study is the initial stage of the development process of the final instrument. In addition, evidence of construct validity could not be found in a single study but only after years of applying the instrument (Lawshe, 1975; Litwin, 1995, Fraenkel et al., 2012).

Criterion validity of the IFP Ed-instrument was also not established in this study because the IFP Ed-instrument should, for that purpose, be compared to a gold standard instrument in the area of fellowship evaluation. According to Litwin (1995) a gold standard instrument is a generally accepted standard instrument with evidence of validity and reliability (Litwin, 1995). However, literature explored in chapter 2 provides no instrument that is generally used during evaluation of fellowship programs. Impact assessment of fellowship programs is in its infancy (Boeren, 2005; Lamont, 2002; Norad, 2005; Norad, 2009; Searle et al., 2006; Rotem et al., 2010) and generally accepted instruments with evidence of validity and reliability are not available.

Establishing reliability of the IFP Ed-instrument as defined by quantitative research (reproducibility/replication of scores) could be difficult to achieve in this study as indicators and variables pre-selected were quantitative as well as qualitative. With quantitative and

qualitative indicators/variables, one way of establishing reliability of the IFP Ed-instrument could be to cross-check the responses of respondents through country site visits (Fraenkel et al., 2012; Litwin, 1995, Merriam, 1995). However, country site visits were outside the scope of this study.

In sum, within this study content validity of individual indicators and variables pre-selected and content validity of the whole IFP Ed-instrument were established in order to ensure the quality of the IFP Ed-instrument. The following section provides the steps used to develop a content valid instrument based on recommendations made in literature.

4.2 Steps used to develop the IFP Ed-instrument
In the development process of valid and reliable instruments, different numbers of steps are used without any consensus (DeVellis, 2012; Haynes et al., 1995; Moore & Benbasat, 1991; Radhakrishna, 2007; Sendjaya, 2003). DeVellis (2012, pp.73-114) recommends using eight steps when developing a measurement instrument: determination of what need to be measured in the first step, generation of item pool in the second step, determination of the format of measurement in the third step, review of item pool by experts in the fourth step, inclusion of validation of items in the fifth step, administration of items to a development sample in the sixth step, evaluation of items in the seventh step, and optimization of scale length in the eighth step. Radhakrishna (2007, pp.1-4) suggests five steps in order to develop a valid and reliable questionnaire instrument: literature search to understand the problem in the first step; generation of statements/questions and definition of major variables that will be included in the questionnaire in the second step; development of instrument in the third step; validity test in the fourth step; reliability test of the instrument by conducting a field test with subjects not included in the sample in the fifth step. Three steps were used by Moore & Benbasat (1991) and Sendjaya (2003) when developing measurement instruments. Moore & Benbasat (1991, p.198) created items in the first step (item creation) by identifying items from existing scales and by creating additional items that appeared to fit the construct definitions. During the second step (scale development), Moore & Benbasat (1991) used a panel of judges to sort the items from the first step into separate categories and examined items according to inappropriate wording, ambiguous items etc.. The categories were then combined into an overall instrument for the instrument testing stage. Moore & Benbasat (1991) tested the instrument in the third step (instrument testing stage) in order to establish its reliability. Similar to Moore & Benbasat (1991), Sendjaya (2003, p.2) identified and generated items in the first step. According to Sendjaya (2003) the first step is important to ensure content validity. In the second step, Sendjaya (2003, p.4) submitted items to experts in order to estimate content validity of the instrument. During the second step items not relevant to the content domain were identified and deleted. In the third step the developed instrument was pilot tested in order to establish its reliability in terms of internal consistency and its construct validity (Sendjaya, 2003).

In sum, as it can be seen different steps are used to develop valid and reliable instruments. Consensus does not exist about the number and wording of steps used. Commonly used steps include: items generation, development of instrument, content validity estimation of the instrument, field testing of instrument to estimate its construct validity and reliability.
As instrument validation refers in this study to establishing content validity of the IFP Ed-instrument, the following four steps were used to develop a content valid IFP Ed-instrument:
Step1: Pre-selection of indicators and variables
This step refers to the step of items generation mentioned in literature. The aim of step 1 in this study was to identify and pre-select indicators and variables that could be used to develop

the IFP Ed-instrument. Therefore, orientation interviews with experts and review of literature were conducted.

Step 2: Selection of indicators and variables and development of the IFP Ed-instrument

During the second step all pre-selected indicators and variables were submitted to experts for evaluation. The aim of this evaluation was to estimate content validity of individual indicators and variables. Selected indicators and variables were used to develop the overall IFP Ed-instrument. Determination of content validity of the instrument started in this step by computing content validity indexes and content validity ratios of individual variables.

Step 3: Validation of overall IFP Ed-instrument

The third step was used to further estimate the content validity of the overall instrument. For that purpose, the draft of the developed IFP Ed-instrument was submitted to experts for evaluation. The aim was to identify and delete incoherent items. After evaluation by experts the IFP Ed-instrument was refined.

Step 4: Field test of the IFP Ed-instrument

The refined instrument was field tested in Senegal. For that purpose the instrument was submitted to Senegalese alumni graduated in the area of education for evaluation. The aim of the field test was to check the length and clarity of the instrument and not to establish construct validity and reliability of the developed instrument. As mentioned earlier, procedures needed to establish construct validity and reliability of the IFP Ed-instrument were outside the scope of this study because of the nature of data that should be collected.

The following of this chapter provides more details about procedures and methods used during the above mentioned four steps.

4.3 Data collection procedures and methods used in each step

This section details successively the procedures and methods used in each step described above. Further, it also provides the sample and instruments used to collect data and data collection procedure used each step.

4.3.1 Data collection and data analysis procedures in step 1

Following Sendjaya (2003) and Grant & Davis (1997), two rounds were used in step 1: orientation interviews with experts and review of literature. According to Sendjaya (2003, p.2) and Grant & Davis (1997, p.272) semi-structured interviews are used to generate themes and obtained more insights about the topic under investigation. Therefore, the aim of orientation interviews in this study was to understand how impact of fellowship programs could be assessed. For the orientation interviews an interview guide was developed in order to conduct semi-structured interviews with academics and practitioners specialized in the field of education as well as in the field of impact assessment. The interview guide (see appendix D) included open-ended questions. Respondents were first joined on telephone personally to ask them whether they would like to participate in this research and for their willingness to be interviewed. Respondents were further sent a first mail in order to explain the purpose of the study; to provide more details on IFP; and to emphasize the importance of their contribution. Anonymity and confidentiality of the responses were guaranteed. Ten semi-interviews (6 face to face and 4 via telephone) were conducted. During the interviews respondents were asked to participate in other successive steps. Care was taken to conduct interviews in a standardized way (asking the same questions to each respondent) with the same interview guide and to ask neutral and clear questions to all respondents in order to minimize interviewer's bias and to facilitate data analysis (Patton, 1990). A recorder was used during the interviews in order to increase the accuracy of data collection and to allow the interviewer to be more attentive to the interviewee (Patton, 1990). Based on recommendations of experts during interviews a

literature review (see chapter 2 for more details) was conducted. Following Sendjaya (2003), the aim of literature review was to check the recommendations of experts and to identify and pre-select indicators and variables that could be used to develop the IFP Ed-instrument. Procedures and criteria used to pre-select indicators and variables are detailed in chapter 3. All pre-selected indicators and variables were operationalized before submission to expert's evaluation (see appendix C).

Sample and respondents in step 1
Scholars (e.g. Campbell & Cantrill, 2001; Hsu & Sandford, 2007; Powell, 2003) recommend choosing respondents that have knowledge about the topic under investigation in order to address validity and reliability issues. According to the Standards for Educational and Psychological Testing in Grant & Davis (1997, p.270), experts involved in content review processes should have relevant training, experience, and qualification. Therefore, the target population of this study were academics and practitioners specialized in the area of education (with knowledge about the educational context of developing countries) or specialized in the area of impact assessment and program evaluation (particularly of fellowship programs). As it is difficult to find experts meeting all criteria (Grant & Davis, 1997, p.270), respondents were selected according to one (or more) of the following three criteria:
  - They should be specialized in the area of education (particularly of developing countries in general or of IFP countries in particular).
  - They should be specialized in impact assessment or program evaluation (if possible of educational/fellowship programs).
  - They should have conducted impact assessment activities or program evaluation activities if possible in the area of education in developing countries.

In order to avoid distance concerns (particularly for face to face interviews) the population was restricted to academics, specialists or practitioners active in the Netherlands. Because it was difficult to locate and reach the target population a purposive sampling followed by a snowball sampling (Dooley, 2001; Kumar, 2006; Neuman, 2000) were used in order to select the study sample that was representative of the population of interest. According to Kumar (2006) snowball sampling is a multistage technique appropriate to identify specialized populations that are difficult to locate.

Purposive sampling was first used in order to identify academics and practitioners that met the selection criteria. The purposive sampling helped to identify four academics and practitioners. These four persons were first approached by telephone in order to ask them to participate in the study and to explain the objectives of the study. Because of time constraints only two accepted to participate in the study. The four identified persons were also asked to suggest names of other academics or practitioners that could meet the selection criteria. In the second stage of the sampling technique, a snowball sampling method (based on inputs of the four persons) was used to identify 27 academics and practitioners. These identified respondents were approached by mail and telephone in order to explain the objectives of the study and to ask them to participate in the study. After reminder e-mails only eight expressed their willingness to participate in the study. Reasons for not participation were related to time constraints.

Altogether, based on these two sampling methods, 10 respondents were selected for the orientation interviews. Further, as it can be seen in the table 4.1 below, the educational context of the majority of IFP countries is familiar to selected respondents.

Table 4.1. Profile of respondents

Name	Organization	Position	IFP country of expertise
Respondent R1	- Utrecht University - Private educational consultancy business	- Lecturer International Pedagogical issues. - International consultant on education	Senegal – Uganda - Kenya
Respondent R2	Edburgh consultants	International education consultant	Indonesia – China – India – Egypt – Palestine – Ghana – South Africa – Tanzania – Kenya - Uganda
Respondent R3	Windesheim – University of Wageningen	Consultant internationalization	South Africa
Respondent R4	NUFFIC (Netherlands Organization for international cooperation in higher education)	Senior Policy Officer	Philippines – Thailand – Kenya - Tanzania
Respondent R5	African Studies Centre in Leiden	Main researcher	Nigeria – Ghana - Mozambique
Respondent R6	Private educational consultancy business	Senior education consultant	Thailand – Philippines – Kenya
Respondent R7	International Institute of Social Studies	Lecturer in development economics. Expert in program evaluation	Indonesia
Respondent R8	Stoas university of Applied Sciences and Teacher Education	Project manager International Liaison	Kenya – Ghana - Nigeria
Respondent R9	Amsterdam Institute for International Development (AIID)	Professor in project and program evaluation for international development	Bolivia – Brazil - Indonesia
Respondent R10	Private educational consultancy business	International education consultant	South Africa – Tanzania - Senegal

Instrument used during orientation interviews (step 1)
Following Sendjaya (2003, p.2), data were collected by using semi-structured interviews (face-to-face and telephone). The main purpose of semi-structured interviews (with open-ended questions) was to collect information on how impact of a fellowship program active in developing countries could be measured and to understand what kind of outcomes and impacts indicators could be used for such purpose. Semi-structured interviews were chosen in this study because they are flexible and are appropriate for all kind of questions (Dooley, 2001; Neuman, 2000; Patton, 1990; Sendjaya, 2003). In order to standardize the procedures and to avoid interviewer bias (Neuman, 2000; Patton, 1990), care was taken to elaborate an interview guide that list questions and issues that were explored during the interviews. Most

questions in the guide (see appendix D) were open-ended in order to enable respondents to provide any kind of answer and additional details.

The guide included questions related to personal experiences of respondents in conducting impact assessment or in evaluating programs particularly in developing countries. Most questions were related to the following topics: evaluation of impact of fellowship programs active in developing countries, kinds of outcome and impact indicators, and development of IFP instrument. The interview guide was pilot tested before the start of interviews with two subjects with similar backgrounds and experience to those who participated in this study (one academic specialized in impact assessment and one practitioner who conducted impact assessment activities for a Dutch-Indonesian scholarship program in Indonesia). The aim of the pilot test was to understand whether the questions represent the whole domain of possible questions; are clear and understandable; not time consuming; and are put in a suitable way (Kumar, 2006). Some weaknesses were observed during the pilot test. The interview guide included questions that were too broad, not clear to respondents and not necessary for this study. Further, the test pointed out that the interview guide was time consuming and overlap between questions was also observed. Based on notes made during the pilot studies, the instruments were refined. The refined instruments were sent again to the same subjects for feedback. Once agreement was reached with the two subjects the guide were used for data collection.

Data analysis procedures in step 1 or during orientation interviews
As the interviews were explorative, answers/comments provided by respondents (data gathered during orientation interviews) were sorted in terms of similarities according to categories of questions. Similar recommendations provided by different respondents were taken into consideration in this study as the aim of the interviews was to understand how to measure the outcomes and impacts of fellowship programs active in developing countries.

4.3.2 Data collection and data analysis procedures in step 2
A Delphi method was used in step 2 as method in order to measure the degree of agreement/consensus between experts about the relevance and measurability of pre-selected indicators and variables and to finally select indicators and variables that could be used to develop the IFP Ed-instrument. Delphi methods are iterative methods appropriate to reach consensus among a group of respondents (mostly experts in the field of investigation) about issues where little is known about (for more details about the method see Hasson, & McKenna, 2006; Hasson, Keeney & McKenna, 2000; Keeney, Hsu & Sandford, 2007; Mullen, 2003; Okoli & Pawlowski, 2004; Rowe & Wright, 1999; Skulmoski, Hartman & Krahn, 2007). Therefore, Delphi methods were found appropriate for this study because little is known about how to assess the impact of fellowship programs as agencies that provide fellowship programs have done few attempts to measure the impacts of such programs (Boeren, 2005; Enders & de Boer, 2002; Krasulin et al., 1998; Lamont, 2002; Norad, 2005; Norad, 2009; Rotem et al., 2010; Searle et al., 2006). The choice of Delphi methods and not other judgment methods was due to the characteristics of the method which ensure anonymity/confidentiality; avoid social pressure (respondents do not know each other); allow all respondents to provide input (Hasson et al., 2000; Hsu & Sandford, 2007; Keeney et al., 2006; Mullen, 2003; Okoli & Pawlowski, 2004; Rowe & Wright, 1999 Skulmoski et al., 2007). Further, the choice was made for pragmatic reasons. Because of the agenda of respondents it was impossible to organize meetings where they could meet and share their ideas. In order to avoid a time consuming process that can cause attrition and low response rate between rounds (Mullen, 2003; Skulmoski et al., 2007) this study adopted a two rounds Delphi method as described in Mullen (2003) and by others (see Hsu & Sandford, 2007;

Powell, 2003; Skulmoski et al., 2007). In order to avoid pressure to conformity that influence reliability and validity (Woudenberg, 1991) consensus between experts was not the intended goal of the Delphi technique used in this study. In order to avoid that respondents influence each other, respondents do not meet physically and do not know who made which response. The procedures used in the two rounds are explained in the following.

First round procedures
Data were collected in the first round by using the questionnaire (questionnaire 1) developed after pre-selection of indicators and variables. After pilot testing, questionnaire 1 was administered through electronic mail together with operational definitions of all pre-selected indicators and variables as recommended by DeVellis (2012). The main purpose of this mail questionnaire was to assess the relevance and the measurability of indicators and variables pre-selected from literature. A mail questionnaire was chosen in this study because it has the advantage of reaching a wide group of respondents simultaneously; it offers anonymity; it avoids interviewer bias; and respondents complete the questionnaire when it is convenient (Baker, 2000; Neuman, 2000). In questionnaire 1, respondents were asked to indicate whether indicators are appropriate to measure the outcomes and impacts of IFP. Respondents were asked to assess the relevance and measurability of the related variables according to a four-point Likert-scale ranging from 1 (not at all relevant or measurable) to 4 (very relevant or measurable). A four point Likert scale with no middle neutral value was used in order to force the respondents in their choice and to avoid social desirable answers that arise from Likert scales with middle neutral values (Garland, 1991). Respondents who could not rate the variables according to these four scales were offered an option 'don't know' (scale 5). Further, respondents were also asked to indicate whether the variables were either essential or not essential to be included in the IFP Ed-instrument. Furthermore, respondents were offered the opportunity to explain their scores or to comment on the definition of the indicators and to suggest new relevant and measurable indicators thought to be appropriate for the IFP Ed-instrument. The questionnaire was administered to all respondents (n=10) who agreed to participate in the Delphi method. To address the issue of low return rate of questionnaires, respondents were first joined on telephone personally to ask for their willingness to complete the questionnaire. Respondents were asked to complete the questionnaire individually and to return it – if possible - within two weeks time. After two weeks, reminders mails were sent out to respondents who did not complete the questionnaire in order to avoid attrition. The total number of completed questionnaire received was eight (80% response rate). For professional reasons, two respondents could not complete the questionnaire. After completion of the questionnaires data were analysed according to criteria set for this study. The responses of the first round served as inputs for second round.

Second round procedures
In the second round, data were collected by administrating a questionnaire (questionnaire 2) through telephone interviews. The purpose of this second round was to evaluate indicators and variables from round one which needed revisions according to respondents and to evaluate new variables suggested by respondents. For that purpose, overall group response (how respondents agree with another) and respondent' own responses from round one were computed in a summarized statistics form and fed back only to respondents who asked for it in order to avoid pressure to conformity that could influence reliability and validity (Woudenberg, 1991). It was made clear to respondents that the aim was not to reach per se a consensus. Respondents were free (and not asked) to reconsider their initial response in the light of overall group response from round one. In the second round, it was decided to conduct telephone interviews with respondents (and not a mail questionnaire) for

standardization purposes. Some respondents were conducting – at that period of time – evaluation activities in developing countries where access to and quality of internet were not appropriate for administration of a mail questionnaire. For standardization purposes, it was decided to use the same data collection procedure with all respondents. The aim was to avoid instrument bias (questionnaire to a group of respondents and interviews with another group). Similar data collection procedures (use of telephone interviews rather than a mail questionnaire) are mentioned in literature (see Hill & Fowles, 1975; Aguirre International, 2004). During the interviews, respondents were again asked to evaluate indicators and variables according to the same criteria (relevance, measurability, and selection or not for IFP Ed-instrument) and according to the same Likert-scale used in the first round. Care was taken not to influence respondents by asking neutral questions, that is, reading questions without any comment or assistance as recommended by Aguirre International (2004). The same questions were asked to all respondents.

Sample and respondents in step 2
During the second step all respondents (n=10) who participated in the first step were asked to participate in the Delphi rounds. All ten respondents (n=10) who participated in the orientations interviews (see table 4.1 above) accepted to participate in the first round of the Delphi method. Two weeks after completion of the first questionnaire all respondents who participated in round 1 were again asked to participate in the second round. After reminder e-mails seven respondents (n=7) expressed their willingness to participate in the second round. Reason of not participating in this round was related to the professional agenda of respondents. Unsuccessful attempts were also made to find other experts who meet the experts selection criteria set for this study in order to increase the number of respondents in the second step. However, sample sizes of ten or seven respondents are not uncommon in Delphi studies (see Campbell & Cantrill, 2001; Hasson et al., 2000; Hsu & Sandford, 2007; Keeney et al., 2006; Mullen, 2003; Okoli & Pawlowski, 2004). Further, these sample sizes were found appropriate as data analysis methods used in this second step require a minimum of five experts (Lawshe, 1975; Lynn, 1986 in Polit & Beck, 2006).

Table 4.2. Overview of respondents during Delphi rounds

Name	Round 1	Round 2
Respondent R1	X	X
Respondent R2		
Respondent R3	X	X
Respondent R4	X	X
Respondent R5	X	X
Respondent R6	X	X
Respondent R7	X	X
Respondent R8		
Respondent R9	X	
Respondent R10	X	X

Note. X: indicates that the respondent participate in a Delphi round. R2 and R3 could not participate in Delphi rounds for professional reasons.

Instruments used for data collection in step 2

During the second step, data were collected by using two questionnaires: questionnaire 1 (first Delphi round) and questionnaire 2 (second Delphi round). Questionnaire 1 was developed after the first step. Care was taken to develop a questionnaire that is unambiguous and easy to understand in order to increase the response rate as recommended by Neuman (2000). The initial version of questionnaire 1 included all indicators and variables pre-selected for the IFP instrument. Further, the initial version of questionnaire 1 included socio-biographical indicators and variables as recommended by respondents during orientation interviews. Questionnaire 1 was pilot tested with the same two subjects who participated in the pilot test from step 1. During the pilot test the first version of the questionnaire appeared to be time consuming and many questions were related to alumni perceptions. Because of overlaps between questions, it was suggested to reorganize the questionnaire for practical reasons. For example, it was recommended to use only one indicator in order to refer to alumni contribution to the education sector. It was also advised to use income as a variable of career progress rather than as a different indicator and to use applicability of knowledge as variable of employment. Based on notes made during the pilot test, the questionnaire was refined. The refined questionnaire 1 was sent again to the same respondents for feedback. Once agreement was reached with the two respondents the questionnaire was used for data collection. The final version of questionnaire 1 (see appendix E) was divided into six core indicators and 23 variables: (1) socio-biographical backgrounds of alumni (three variables), (2) educational backgrounds of alumni (three variables), (3) employment (five Variables), (4) career progress (six variables), (5) type of activities in the area of education (five variables), and (6) changes caused by conducted activities (one variable). The content of questionnaire 2 (see appendix F) was based on feedback provided by respondents in the first Delphi round 1. The questionnaire includes indicators and variables that need revision according to respondents and new indicators and variables suggested by respondents in round 1. Based on recommendations made during pilot tests the second questionnaire used the same format as the first one.

Data analysis in step 2

Descriptive statistics were first used to analyse data obtained with questionnaires. The frequency of distribution of the responses was computed for each variable. This statistic is appropriate when analysing data based on Likert scale (Allen & Seaman, 2007; Jamieson, 2004; Mullen, 2003). According to Mullen (2003), the frequency distribution has the advantage of pinpointing bimodal distribution and extreme outliers. As recommended by literature, Lynn' (1986) content validity index (CVI) and Lawshe' (1975) content validity ratio (CVR) were further used as methods to measure content validity of individual variables in round 1 and in round 2 (DeVon et al., 2007; Polit & Beck, 2006; Polit et al., 2007; Straub et al., 2004; Wynd et al., 2003). Both methods require using at least five judges in order to provide a sufficient level of control for chance agreement. The number of respondents who completed the questionnaire from round1 (n=8) and participated in interviews from round 2 (n=7) is above this requirement.

In order to estimate content validity of individual variables, Lynn' (1986) procedures were used. These procedures are appropriate in this study because of the response format of the questionnaire. Following Lynn (1986) procedures (see Polit & Beck, 2006; Wynd et al., 2003), the four categories ranging from 1 (not at all relevant) to 4 (very relevant) were collapsed into two categories of responses (content valid and content invalid) after respondents' ratings. Rating 1 and 2 were considered content invalid. Rating 3 and 4 were considered content valid. Content validity index of individual items (I-CVI) was then computed as the number of experts giving a rating of either 3 or 4 divided by the total number of experts. The content validity of the overall instrument is the proportion of items that

received a rating of 3 or 4 by the experts. It could be computed as the average I-CVI for all items in the instrument or to divide the total number of items rated relevant by all experts combined by the total number of rated items (Polit & Beck, 2006; Wynd et al., 2003). This study adopted I-CVI to refer to content validity index of individual variables as suggested by Polit & Beck (2006) in order to avoid confusion with the content validity index of the overall instrument (S-CVI). According to Lynn, an instrument has good content validity if it includes items that have I-CVIs of 0,78 or higher and S-CVI of 0,80 or higher (Gibson et al., 2006; Polit & Beck, 2006; Polit et al., 2007).

Lawshe (1975) content validity ratio (CVR) was also computed in order to select variables essential for the IFP instrument as respondents were asked to indicate whether variables are essential or not to be included in the instrument. CVR for individual items were calculated by using the following formula (Lawshe, 1975, p.567):

$$CVR_i = [n_e - N/2] / [N/2]$$

CVR_i: content validity ratio for ith item (CVR_i values vary from -1,00 to +1,00).
N: total number of panelists
n_e: number of panelists indicating that ith item is "essential".
Because eight respondents completed the questionnaire, the minimum CVR value (in order to satisfy the 5% significance level) required by Lawshe for an item to be included in the instrument was 0, 75 (Lawshe, 1975, p.568). According to Lawshe (1975, p.567), an item has some degree of content validity is more than half of respondents indicate that the item is essential. For this present study, at least 4 respondents (ICVI = 0,50 and CVR = 0,00) have to indicate that the variable is essential.

Therefore, following recommendations of Lynn (1986) and Lawshe (1975) for eight raters, variables used to develop the IFP Ed-instrument should have the following values: I-CVI $\geq$ 0,78 for relevance and measurability and CVR $\geq$ 0,75 for selection in IFP instrument. Variables with lower values of I-CVI and CVR (I-CVI $\leq$ 0,50 and CVR $\leq$ 0,00) were deleted. Variables with I-CVI values between 0,50 and 0,78 and CVR values between 0,00 and 0,75 were revised (when recommended by respondents) and submitted again in the second round or deleted.

As seven respondents were used in the second round, the following criteria were set based on recommendations of Lynn (1986) and Lawshe (1975) for seven raters: variables used to develop the IFP Ed-instrument should have the following minimum of values: I-CVI $\geq$ 0,78 for relevance and measurability and CVR $\geq$ 0,99 for selection in IFP instrument. All variables not meeting these criteria were deleted.

4.3.3 Data collection and data analysis procedures in step 3

The overall IFP Ed-instrument was first developed before data collection in this third step. Based on literature and recommendations during orientation interviews, different items related to the selected variables (see appendix G) were identified and used to develop the initial overall IFP Ed-instrument. The initial purpose in this third step was to ask respondents to evaluate all individual items included in the overall instrument in order to compute I-CVI of all items and the content validity of the overall instrument. However, this way of proceeding was found time consuming rating process by respondents. Respondents expressed their willingness (through phone calls) to evaluate the whole IFP Ed-instrument but not individual items included in the instrument. Therefore, data were collected by submitting the overall initial instrument together with a mail questionnaire 3 (see appendix H) to respondents. Respondents were first asked to read the instrument and then to answer the questions included in questionnaire 3 according to a five-point Likert scale ranging from 1 (not at all) to 5 (to a

very high extent). Following Grant & Davis (1997), the aim of questionnaire 3 was to evaluate the overall instrument for clarity of items and completeness (identification of items that need to be deleted or to be added). Based on remarks and recommendations of respondents, the IFP Ed-instrument was first refined (see appendix I). Since no major revisions were recommended by respondents, a second content analysis of the overall IFP Ed-instrument was not conducted after refinement.

Sample and respondents in step 3

All experts who participated in the first step were again asked to participate in this third step because it was difficult to find a new panel of experts willing to participate in this study. Selecting respondents from the same pool of experts used in previous steps is supported by Lynn (1986) in Polit et al. (2007). After a number of reminder e-mails six respondents (n=6) completed the questionnaire. According to Polit et al. (2007, p.466) this number of experts is appropriate to evaluate the relevance of items included instrument and therefore to evaluate content validity of the overall instrument.

Table 4.3. Profile of respondents in step 3

Name	Organization	Position
Respondent R1	- Utrecht University - Private educational consultancy business	- Lecturer International Pedagogical issues. - International consultant on education
Respondent R3	Windesheim – University of Wageningen	Consultant internationalization
Respondent R4	NUFFIC (Netherlands Organization for international cooperation in higher education)	Senior Policy Officer
Respondent R6	Private educational consultancy business	Senior education consultant
Respondent R7	International Institute of Social Studies	Lecturer in development economics. Expert in program evaluation
Respondent R10	Private educational consultancy business	International education consultant

Instruments used for data collection in step 3

Data were collected by using a mail questionnaire (questionnaire 3 in appendix H) together with the initial IFP Ed-instrument. Following Radhakrishna (2007) and Grant & Davis (1997), the following questions were included in questionnaire 3: what items should be excluded from the IFP Ed-instrument? – What items are missing in the IFP Ed-instrument? further respondents were asked to indicate the extent to which the IFP Ed-instrument contributes in measuring the outcomes and impacts of IFP and the extent to which the IFP Ed-instrument is content valid (that is, could the instrument measure what it is intended to measure?) according to a five-point Likert scale ranging from 1 (not at all) to 5 (to a very high extent). Before data collection, the overall IFP Ed-instrument and questionnaire 3 were first pilot tested with the same subjects who participated in the first and second steps.

Data analysis in step 3

Lawshe (1975) procedures were used for selection or deletion of items. For n = 6, CVR value is 0,99 (Lawshe, 1975). That means that all respondents should agree in order to include a new item in the instrument (Lawshe, 1975). Further, the frequency of distribution of the responses was computed for each question included in questionnaire 3. The aim was to

estimate the degree of agreement between respondents. Therefore, Fleiss (1971) multiple rater kappa was used as statistic in order to estimate the degree of agreement between respondents. This statistic was appropriate because more than two raters were used in this study. According to Fleiss (1971, pp.378-379) multiple rater kappa could be computed by using the following formula:

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$\bar{P} - \bar{P}_e$ degree of agreement in excess of chance
$1 - \bar{P}_e$ measures the degree of agreement attainable over and above what would be predicted by chance.
$\bar{p}$: The overall proportion of agreement (average of the P_is).
Pi the proportion of agreement among the n raters for the ith item.

$$\bar{p} = \frac{1}{N} \sum_{i=1}^{N} P_i \text{ where } P_i = \frac{1}{n(n-1)} \left(\sum_{j=1}^{k} n_{ij}^2 - n \right)$$

$\bar{P}_e$: Proportion of agreement expected to occur by chance

$$\bar{P}_e = \sum_{j=1}^{k} P_j^2 \text{ where } P_j = \frac{1}{Nn} \sum_{i=1}^{N} n_{ij}$$

P_j: proportion of all assignments which were to the jth category ($\sum_j P_j = 1$)
n_{ij}: number of raters who assigned the ith item to the jth category ($\sum_j n_{ij} = n$)
In these different formula's: N refers to the total number of items; n refers to the number of ratings per item; k refers to the number of categories; the subscript i refers to items (i = 1, …, N); the subscript j refers to categories (j = 1, …, k).
According to Cohen (1960, p.40), the coefficient kappa is "the proportion of agreement after chance agreement is removed from consideration". Kappa values range from +1,00 to -1,00 (Cohen, 1960). The value +1,00 indicates complete agreement beyond chance. A value of 0,00 indicates that obtained agreement could also be predicted/expected by chance. A value of -1,00 indicates total disagreement.

4.3.4 Data collection and data analysis procedures in step 4
During the field test, the IFP Ed-instrument refined in the third step was submitted to alumni from Senegal graduated in the area of education. A mail questionnaire (questionnaire 4 in appendix J) was also submitted in order to evaluate the instrument. Data was therefore collected by using questionnaire 4. The aim was to test the length of the instrument as well as its clarity. Respondents were first asked to read the IFP Ed-instrument and then to answer the questions included in questionnaire 4 according to a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Senegal was selected in this study because Senegalese alumni reacted to mails sent (in Ghana and Senegal) and expressed their willingness to participate in this study.

Sample and respondents in step 4
The sample identified through purposive and snowball sampling includes Senegalese IFP alumni graduated in the area of education. The purposive sampling helped to identify the director of the IFP alumni association in Senegal who was graduated in education. The director was first contacted through e-mail and then by telephone in order to ask him to participate in the study and to explain the objectives of the study. He expresses his willingness to participate in this study. The director helped to identify three other alumni graduated in the

area of education. These three alumni were approached by the director and they accepted to participate in the evaluation. As an alumni meeting was planned the director expressed his willingness to submit the questionnaire to these alumni and to discuss the instrument with them. However, after reminder e- mails and phone calls, only one respondent (the director) completed the questionnaire.

Table 4.4. Profile of respondents in step 4

Name	Organization	Position
Respondent S1	Ministry of education in Senegal	Manager planning, monitoring, evaluation in secondary education and director IFP alumni association in Senegal
Respondent S2	Ministry of education in Senegal	Department of educational planning and reform (DPRE) in charge of project monitoring/evaluation office
Respondent S3	Private business	Consultant as expert in special education
Respondent S4	Ministry of education in Senegal	Head of statistics bureau at the directorate of literacy and member of monitoring and evaluation team

Instruments used for data collection in step 4
Data were collected by using a mail questionnaire (questionnaire 4). Following Kumar (2006) and Roche (1999) questionnaire 4 includes questions related to length and clarity of questions. Respondents were first asked to indicate the extent to which questions included in the IFP Ed-instrument are easy to follow, easy to understand, and easy to answer according to a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Further, they were asked to indicate the time needed for completion and to indicate the questions that were difficult to answer. Questionnaire 4 also offers the opportunity to provide personal comments.

Data analysis in step 4
The aim of data analysis was to compute Fleiss kappa statistics in order to measure the strength of agreement between respondents and as needed to refine the instrument. However, due to low response rate (only one respondent out of four completed the questionnaire); there was too little data to assure a significant analysis. Fleiss kappa statistics were therefore not computed.

5. Results

The purpose of this study was to develop a valid and reliable instrument that could measure the impact of IFP in the area of education. Therefore, four steps were used to develop the IFP Ed-instrument. In the first step (as described in chapter 3) indicators and variables were pre-selected from literature by using a modified Kirkpatrick framework. The relevance and measurability of these indicators and variables were evaluated by experts in the second step. Indicators and variables found relevant and measurable by experts were then used to develop the draft of the IFP Ed- instrument. This draft was evaluated in the third step and then refined. The refined instrument was field tested in Senegal in the fourth step. The results of these evaluations are presented in this chapter. Results of the second step are first described in the first section, followed by results from the third (section 2) and fourth steps (section 3). The major findings of this study are summarized in the fourth and final section. Results of orientation interviews conducted in the first step are not described separately in this chapter but presented where applicable.

5.1 Results from step 2 evaluations: Evaluation of pre-selected variables

This section provides the results about content validity index (I-CVI) and content validity ratios (CVR) of the individual six indicators and 23 variables evaluated in the second step. The 23 variables evaluated were related to the following six indicators: (1) socio-biographical backgrounds of alumni, (2) educational backgrounds of alumni, (3) employment, (4) career progress, (5) type of activities in the area of education, and (6) changes caused by conducted activities. Results on content validity indexes (I-CVI) and content validity ratios of individual variables (CVR_i) were computed by following Lynn (1986) and Lawshe (1975) procedures explained in more details in chapter 4:

- I-CVI = number of experts giving a rating of either 3 or 4 divided by the total number of experts (Lynn, 1986 in Polit & Beck, 2006, p.491).
- $CVR_i = [n_e – N/2] / [N/2]$. Where n_e refers to the number of raters that indicate that the variable is essential to be included in the instrument and N is the total number of raters (Lawshe, 1975, p.567).

As a two rounds Delphi procedures (see chapter 4) were used to select indicators and variables respectively, this section first presents the results of the first round followed by the results of the second round.

5.1.1 Results first Delphi-round

Determination of content validity of the instrument started in this first round by computing content validity indexes and content validity ratios of individual indicators and variables. In the first round, the relevance and measurability of the pre-selected indicators and variables were rated by eight respondents (n=8) on a 4- point Likert-scale ranging from 1 (not at all relevant or measurable) to 4 (very relevant or measurable). Respondents who could not rate the variables according to these four scales were offered an option 'don't know' (scale 5). Further, respondents were also asked to indicate whether the variables are essential to be included in the final instrument. As eight respondents participated in this round, the selection criteria set were as followed:

- Variables selected in the IFP instrument should have the following values: I-CVI ≥ 0,78 for relevance and measurability and CVR ≥ 0,75 for selection in IFP instrument.
- Variables with lower values of I-CVI and CVR (I-CVI ≤ 0,50 and CVR ≤ 0,00) were deleted.
- Variables with values of I-CVI between 0,50 and 0,78 and values of CVR between 0,00 – 0,75 were revised when recommended by respondents and submitted again in the second round or deleted.

The findings of the first Delphi-round for each indicator as well as the suggestions or recommendations provided by respondents are presented below. Details about the frequency distribution of responses for each indicator and variable are presented in appendix K. The appendix depicts also I-CVI and CVR values of each indicator or variable.

Indicator 1: Socio-biographical backgrounds alumni

When asked about relevance of indicator 1, all respondents indicated that socio-biographical backgrounds of alumni are relevant (I-CVI = 1,00) to measure outcomes and impacts of IFP (table K1). This value of content validity index of indicator 1 meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

Three variables related to indicator 1 were further evaluated by respondents: gender of alumnus, age of alumnus, and marital status of alumnus. All respondents found the variables 'gender' and 'age' relevant to very relevant (table K2 in appendix K). Further both variables were found very measurable by seven respondents in contrast to one respondent that found both variables not at all measurable. Finally all respondents agree that both variables should be included in the final instrument. Three respondents found the variable 'marital status' relevant; five found the variable measurable to very measurable; and four found the variable essential to be included in the IFP Ed-instrument (table K2).

Analyses of the responses (appendix K) show that the variables gender and age have acceptable levels of I-CVI (I-CVI ≥ 0,78) and acceptable level of CVR (CVR ≥ 0,75) (see table K3). Therefore, both variables were selected for the IFP instrument. The content validity indexes of the variable marital status (0,38 for relevance and 0,63 for measurement) are below the level of 0,78 set for this study (table K3). Content validity ratio of this variable (0,00) was also below the level of content validity set in this study (0,75). Therefore, the variable marital status was excluded.

Indicator 2: Educational backgrounds alumni

When asked about relevance of indicator 2, all respondents indicated that educational backgrounds of alumni are very relevant (I-CVI = 1,00) to measure outcomes and impacts of IFP (table K1). This value of content validity index of indicator 2 meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

Three variables related to indicator 2 were further evaluated by respondents: education degree obtained through IFP, place/country of study, and kind/content of study. All respondents rated the variables 'education degree obtained through IFP' and 'place/country of study' very relevant (table K4 in appendix K). One respondent rated the variable 'kind/content of study' relevant while seven rated the same variable very relevant. Further, all three variables were rated measurable to very measurable by all respondents. Finally, all respondents found the three variables essential to be included in the final instrument.

Analyses of the responses (appendix K) indicate that content validity and content validity ratios of all variables related to indicator 2 are above the values of 0,78 and 0,75 set for this study (table K5). Therefore, these variables were selected for the IFP instrument.

Indicator 3: Employment

When asked about relevance of indicator 3, all respondents indicated that employment of alumni is very relevant (I-CVI = 1,00) to measure outcomes and impacts of IFP (table K1). This value of content validity index of indicator 3 meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

Five variables related to indicator 3 were further evaluated by respondents: type of job, place of employment, kind of employment sector, level of employment, and applicability knowledge/skills. Except the variable 'type of job', all respondents rated all other variables

relevant to very relevant (table K6). Four respondents rated the variable 'type of job' relevant and very relevant. Concerning measurability, all respondents found the variables 'place of employment' and 'level of employment' very measurable; seven respondents found the variable 'kind of employment sector' very relevant; five respondents rated the variable 'type of job' measurable to very measurable. Four respondents rated the variable very measurable and four respondents rated the variable somewhat measurable. Except the variable 'type of job', all respondents found that the variables essential to be included in the final instrument. Only four respondents recommended including the variable 'type of job' in the final instrument. Details about the frequency distribution of responses for each of these variables are presented in appendix K.

Content validity index and ratio analyses indicate that the variables 'place of employment', 'kind of employment sector', and 'level of employment' have acceptable levels of I-CVI (I-CVI ≥ 0,78) and acceptable level of CVR (CVR ≥ 0,75) (table K7). Therefore, these variables were selected for the IFP instrument. However, in order to avoid confusions, respondents recommended using the term 'kind of employment institution' rather than 'level of employment' (variable 3.4). The variable 'applicability of knowledge' has an acceptable level of I-CVI relevance (1,00) but has somewhat lower level of I-CVI measurement (0,50) (table K7). This variable was again submitted to respondents in round two in order to understand why all respondents agree to include the variable in the instrument. I-CVI (relevance) values and CVR values for the variable 'type of job' are very low (I-CVI ≤ 0,50 and CVR ≤ 0,00) (see table K7). Half of respondents found the variable not relevant and indicate that it should not be included in the instrument. Therefore, the variable 'type of job' was deleted from instrument.

Indicator 4: Career progress
When asked about relevance of indicator 4, all respondents indicated that career progress of alumni is relevant (I-CVI = 1,00) to measure outcomes and impacts of IFP (table K1). This value of content validity index of indicator 4 meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

Six variables related to indicator 4 were further evaluated by respondents: income/salary, number of promotions, job position/title, responsibility in work, locus of decision, and alumni' career satisfaction. Seven respondents rated the variables 'income/salary' and 'number of promotions' relevant to very relevant (table K8). The variables 'job position/title' and 'responsibility' were rated very relevant by all respondents. The variable 'locus of decision' was rated relevant to very relevant by all respondents. The variable 'alumni' career satisfaction' was rated very relevant by six respondents. Concerning measurability, except the variables 'number of promotion' and 'alumni career satisfaction' all other variables were found measurable to very measurable by at least six respondents (table K8). The variable 'number of promotions' was rated very measurable by five respondents and rated somewhat measurable by three respondents. Also, five respondents rated the variable 'alumni career satisfaction' measurable (one respondent) to very measurable (four respondents). Finally six respondents found that the variables 'income/salary' and 'alumni career satisfaction' should be included in the final instrument; five found the variable 'number of promotion' essential to be included in final instrument; all respondents found that the variables 'job position' and 'responsibility' essential to be included in the final instrument; and seven respondents recommended to include the variable 'locus of decision' in the final instrument. Details about the frequency distribution of responses for each of these variables are presented in appendix K.

Content validity index and ratio analyses indicate that the variables 'job position', 'responsibility in work', and 'locus of decision' have acceptable levels of I-CVI (I-CVI ≥

0,78) and acceptable level of CVR (CVR ≥ 0,75) (table K9). Therefore, these variables were selected for the IFP instrument. However, the variables 'income', 'number of promotions', and 'alumni' career satisfaction' have somewhat lower levels of I-CVI and CVR (table K9). Respondents found these variables relevant but difficult to measure. Measurement problems were related to the definitions of the variables. According to respondents alumni will not provide the right answer to items related to income/salary as it is a sensitive and too private topic. They recommended asking for a percentage increase in salary and not asking for salary. Further, according to respondents changes in offices or office décor could not be considered as a promotion. They recommended excluding changes in offices or office décor in the definition of promotion used in this study. Furthermore, respondents argued that satisfaction with career has no link with impacts of IFP. Satisfaction with career could also be caused by external factors not related to IFP. Respondents therefore recommended to create a new indicator (alumni satisfaction) and to use 'alumni satisfaction with IFP' as variable for this indicator. The variables 'income/salary', 'number of promotions' and 'alumni career satisfaction' were therefore revised and submitted again for evaluation in the second round.

Indicator 5: Type of conducted activities in the area of education
When asked about relevance of indicator 5, three respondents indicated that the indicator is relevant and five respondents indicated that the indicator is very relevant to measure outcomes and impacts of IFP (table K1). The value of content validity index of indicator 5 meet the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

Five variables related to indicator 5 were further evaluated by respondents: financial and resources activities, human resources activities, activities related to students' background conditions, activities related to community/parents' involvement, and instructional activities. Seven respondents found the variables 'financial and resources activities', 'students' background conditions activities', and 'community/parents' involvement activities' relevant to very relevant. Six respondents found the variables 'human resources activities' and 'instructional activities' relevant to very relevant. However, respondents do not agree on the measurability of the variables (table K10). Two variables 'financial and resources activities' and 'human resources activities' were rated measurable to very measurable by four respondents. Three variables 'students' background conditions activities', 'community /parents' involvement activities', and 'instructional activities' were rated by five respondents measurable to very measurable.

At least six respondents indicate that the five variables are essential to be included in the final instrument: financial and resources activities (six respondents); human resources (seven respondents); students' background conditions activities (seven respondents); community /parents' involvement activities (six respondents); and instructional activities (seven respondents). Details about the frequency distribution of responses for each of these variables are presented in appendix K.

Content validity index and ratio analyses indicate (see table K11 in appendix K) that the variables related to indicator 5 have somewhat lower levels of I-CVI (measurement) and CVR. Most variables were found relevant but not measurable. None of the five variables met the levels of I-CVI (I-CVI ≥ 0,78) for relevance and measurability and the level of CVR (CVR ≥ 0,75) set for this study. For example the variable 'financial and resources activities' was found relevant (I-CVI ≥ 0,78). But I-CVI (measurement) for the variable 'financial and resources activities' was 0,50 and CVR was 0,50. These values are below the level set for this study. Measurement issues were related to the definitions of the variables. According to respondents it will be difficult for alumni to influence the financial and human sector of education in developing countries. The educational systems in those countries are too centralized. According to respondent (R10) "in most developing countries that I visited

financial and human resources are decided by central government/ministries and not by schools". Overlaps were also noticed by respondents between the five variables. It was therefore recommended to making three groups of variables (educational management activities, voluntary activities and professional/operational activities) rather than the previous five groups of activities related to indicator 5. According to respondents, management activities can also be related to financial and human resources. Further, a respondent mentioned that "alumni can conduct voluntary activities in the community in order to sensitize parents to be involved in schools". Respondents also agreed that the first five variables do not take into account alumni that will be engaged in teaching or research (professional activities). These three new variables were therefore submitted for evaluation in the second round.

Indicator 6: Changes caused by conducted activities
When asked about relevance of indicator 6, six respondents indicated that the indicator is relevant and two indicated that the indicator is very relevant to measure impact of IFP (table K1). The value of content validity index of indicator 6 (see table K1) meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.
One variable related to indicator 6 was further evaluated by respondents: alumni' perceived impact (see table K12). All respondents rated the variable relevant (one respondent) to very relevant (seven respondents). Concerning measurability, five respondents rated the variable very measurable and three respondents rated the variable somewhat measurable. Seven respondents recommended including the variable in the final instrument and one respondent was undecided. Details about the frequency distribution of responses for this variable are presented in appendix K.
Content validity index and ratio analyses (see table K13) indicate that the variable 'alumni' perceived impact' has acceptable levels of I-CVI relevance (1,00) and CVR (0,75). However, I-CVI measurement (0,63) is somewhat low. Respondents found the variable relevant and essential for selection in the instrument. However, the variable was found somewhat difficult to be measured. Respondents mentioned that changes caused by activities conducted by alumni will be difficult to measure because of the influence of external factors. This variable was again submitted to respondents in round two in order to understand why respondents agreed to include the variable in the instrument while the variable has a low measurement I-CVI.

Summary results first round Delphi
From the initial 23 variables, eleven were retained for the IFP instrument, five were revised, and seven variables were deleted in the first Delphi round (table K14). Further, eight new variables and a new indicator (alumni satisfaction) were recommended by respondents (table K15). The new variables include: religion of alumni and country of birth and residence of alumni (related to indicator 1), year graduation (related to indicator 2), function description (related to indicator 3), educational management activities, voluntary activities, professional activities (related to indicator 5), and alumni satisfaction with IFP (related to new indicator 7). Therefore, one new indicator and 13 revised as well as new suggested variables were submitted to respondents for evaluation in the second Delphi round. The following section provides the results of this evaluation.

5.1.2 Results second Delphi-round
This section provides results of evaluation conducted in the second Delphi round. Respondents were asked to rate the relevance and measurability of the revised and recommended variables according to the same Likert-scales used in the first round. They were

also asked to indicate whether the variables should be included in the final instrument. As seven respondents participated in the second round, variables selected for the IFP instrument should have the following values: I-CVI ≥ 0,78 for relevance and measurability and CVR ≥ 0,99 (Lawshe, 1975; Lynn, 1986 in Polit & Beck, 2006). All variables which do not meet these criteria were deleted. Details about the frequency distribution of responses for this variable are presented in appendix K.

Indicator 1: Socio-biographical backgrounds alumni

Six out of seven respondents found the variable 'religion of alumni' not relevant for this study (see table K16). As one respondent mentioned "it will be impossible to attribute changes in the community to the religion of the alumnus". Further, the variable was not found measurable (difficult to measure) because some alumni will not be willing to provide the right information. As expressed by a respondent, "religion is a sensitive topic in some countries". Further, all respondents found the variable 'religion' not essential to be included in the instrument. Concerning the variable 'country of birth/residence', all respondents found the variable very relevant. They found the variable also measurable and also agreed that the variable should be included in the final instrument. As a respondent mentioned country of birth of alumni could allow country comparisons.

Content validity index and ratio analyses (table K17) indicate that the variable 'religion' has lower levels of I-CVI relevance (0,14), I-CVI measurement (0,00) and CVR (-1,00). The levels of I-CVI (I-CVI ≥ 0,78) for relevance and measurability and the level of CVR (CVR ≥ 0,99) set for this study were not met by this variable. Therefore, religion was not selected in the final instrument. The variable 'country of birth/residence' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (1,00), and acceptable level of CVR (1,00). Therefore, the variable was selected for the IFP instrument.

Indicator 2: Educational backgrounds alumni

The variable 'year of graduation' was found relevant and measurable by all respondents. Respondents also agreed that it should be included in the final instrument (table K18). They argued that perception of impact by alumni is related to the year of graduation. According to respondents perception of impact of IFP will be less important after a long period of time after graduation because of context factors. As a respondent expressed "a perception of impact two year after graduation is different to a perception ten years after graduation".

Content validity index and ratio analyses indicate (table K19) that the variable 'year of graduation' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (1,00), and acceptable level of CVR (1,00). Therefore, the variable was selected for the IFP instrument.

Indicator 3: Employment

The variable 'function description' was found relevant by five respondents and very relevant by two respondents (table K20). Six respondents found the variable measurable and one was undecided (do not know). Further, all respondents agreed to include the variable in the final instrument. The variable 'applicability knowledge' was again found very relevant (scale 4) by all respondents (table K20). All respondents argued that it should be included in the instrument. Similar responses were also observed in the first Delphi round. Asked again about measurability of this variable, four respondents found the variable measurable (scale 3) and three respondents rated the variable very measurement (scale 4). However, they argued that the variable is measurable only if means are available in order to check responses provided by alumni through country site visits and interviews with alumni employers. Further, all respondents agreed that the variable 'applicability knowledge' is essential to be included in the final instrument.

Content validity index and ratio analyses indicate (table K21) that the variable 'function description' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (0,86), and acceptable level of CVR (1,00). Therefore, the variable 'function description' was selected for the IFP instrument. Further, analyses indicate (table K21) that the variable 'applicability knowledge' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (1,00), and acceptable level of CVR (1,00). Therefore, applicability of knowledge was selected for the IFP instrument.

Indicator 4: Career progress

After revision, all respondents agreed that the variable 'income/salary' is a relevant variable for the present study (table K22). However, four out of seven respondents found the variable difficult to measure because alumni will not be willing to provide the right information. As a respondent mentioned it will be difficult to get access to secondary data in countries and to get the right data on salaries as it is a sensitive and private topic. Only three respondents agreed that the variable should be included in the final instrument. After revision, the variable 'number of promotions' was found very relevant by all respondents (table K22). Three respondents found this variable measurable (scale 3) and four respondents rated the variable very measurement (scale 4). Further, all respondents agreed that number of promotions is essential to be included in the final instrument.

Content validity index and ratio analyses show (table K23) that the variable 'income/salary' has lower levels of I-CVI measurement (0,43) and CVR (-0,14) compared to the levels set for this study. Therefore, this variable was excluded from the final instrument. Further, analyses indicates (table K23) that the variable 'number of promotions' has acceptable levels of I-CVI relevance (1,00), ICVI-measurement (1,00), and acceptable level of CVR (1,00). Therefore, the variable was selected for the IFP instrument.

Indicator 5: Type of conducted activities in the area of education

The variables 'educational management/policies activities', 'voluntary activities', and 'professional activities' were found very relevant (scale 4) and measurable (scale 3) by all respondents (table K24). According to them these variables should be included in the final instrument. According to one respondent "with a master or PhD degree, we should expect that some alumni will occupy management/decision making positions". Some respondents argued that alumni are expected to conduct voluntary activities. As one respondent said "IFP expects alumni to be active in their own community. It is normal to assume that they will conduct voluntary activities". A respondent mentioned that the variable 'professional activities' is relevant as alumni that conduct research activities can achieve different changes compared to alumni that occupy management positions. According to respondents alumni responses should be checked through country site visits. Further, all respondents agreed that the three variables are essential to be included in the final instrument.

Content validity index and ratio analyses indicate (table K25) that the three variables have acceptable levels of I-CVI relevance (1,00), I-CVI measurement (1,00), and acceptable level of CVR (1,00). Therefore, they were selected for the IFP instrument.

Variable related to indicator 6: Changes caused by conducted activities

The variable 'alumni perceived impact' was found again relevant and recommended by all respondents to be included in final instrument (table K26). Six out of seven found the variable measurable (scale 3) if means are available through country site visits. According to one respondent "perception of impact will be a subjective appraisal and should be treated as such, no matter how easy it is measured". One respondent argued that the variable is not measurable as it will be time consuming and therefore will involve lot of means that organizations may

not have. According to him "controlling what alumni report is time consuming and organizations may not have enough resources to support such investigations".

Content validity index and ratio analyses show (table K27) that the variable 'alumni perceived impact' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (0,86), and acceptable level of CVR (1,00). Therefore, this variable was selected for the IFP instrument.

Variable related to indicator 7: Alumni satisfaction with IFP
When asked about relevance of indicator 7, all respondents indicated that alumni satisfaction is very relevant (I-CVI = 1,00) to be used during evaluation of IFP. This value of content validity index of indicator 7 meets the selection criteria (I-CVI ≥ 0,78 for relevance) set for this study.

The new variable 'alumni satisfaction with IFP' was found relevant and measurable by all respondents (table K28). They also agreed that it should be included in the final instrument. As a respondent mentioned measuring satisfaction after the fellowship program could provide relevant information for the improvement of the program. According to him "alumni have obtained their degree and could provide all kind of information without fear for their fellowship".

Content validity index and ratio analyses indicate (table K29) that the variable 'alumni satisfaction with IFP' has acceptable levels of I-CVI relevance (1,00), I-CVI measurement (1,00), and acceptable level of CVR (1,00). Therefore, this variable was selected for the IFP instrument.

Summary results second round Delphi
From the initial 12 variables evaluated in the second round, ten were retained for the IFP instrument and two variables were deleted (table K30). The deleted variables in the second rounds are income/salary and religion of alumni.

5.1.3 Summary step 2 results
Evaluation of variables conducted in the second step resulted in the final selection of seven core indicators and 21 variables with I-CVI values higher than 0,78 and CVR values higher than 0,75 (table K31). Eleven variables were selected in the first Delphi round and ten variables were selected in the second round. Selected indicators are: (1) socio-biographical backgrounds of alumni (three variables), (2) educational backgrounds of alumni (four variables), (3) career progress (four variables), (4) employment (five variables), (5) type of activities in the area of education (three variables), (6) changes caused by conducted activities (one variable), and (7) alumni satisfaction (one variable). These indicators and related variables were used to develop the draft of the overall instrument. When asked how these retained variables could be measured, respondents recommend the following methods: use of data gathered by CHEPS, surveys with alumni, use of country sites visits in countries where alumni are active (document analysis - observations), alumni interviews, alumni employer interviews, and focus group discussion with alumni. These recommendations are in line with measurement methods mentioned in literature (see Aguirre International, 2004; CIDA, 2005; Ford, 2004; Rotem et al., 2010).

5.2 Results step 3: Evaluation overall instrument
This section provides the results of evaluations of the draft of the overall instrument developed after step 2. The purpose of the evaluation was to measure content validity of the overall instrument. Respondents were therefore asked to indicate items that should be deleted or included in the instrument. As six respondents participated in the evaluation, all respondents should agree in order to include or exclude an item from the instrument (Lawshe,

1975). The section provides first the items that were deleted or included in the instrument. Further the section depicts the frequency distribution of the responses for each question included in questionnaire 3 (see appendix H) and the Fleiss (1971) multiple rater kappa statistics in order to estimate the degree of agreement between respondents.

5.2.1 Deleted or included items
When asked whether items should be excluded from the draft of the instrument, all respondents recommended to exclude the items related to data sources about changes that alumni bring about in countries/communities. According to them alumni are not able to answer such items. As expressed by a respondent data on changes bring about by alumni are not systematically collected by national statistical organizations because evaluation practices are scarce in developing countries. This respondent also argued that fellowship agencies do not systematically collect such data. Further, respondents recommended two items to be included in the instrument: strategic overall management and work experience. Strategic overall management can be used as an item for the variable 'management/policy activities'. 'Work experience' can be used as an item for factors that could influence promotion. Furthermore, respondents recommended offering possibility to alumni to formulate comments and critiques on the IFP programme (for example what could be improved in the IFP programme according to alumni' experience).

5.2.2 Content validity of overall instrument
When asked whether the instrument contributes to measure the outcomes of IFP all respondents agreed that the instrument contribute in measuring the outcomes of IFP. Five respondents agreed with the statement (scale 4) and one respondent strongly agreed with the statement (scale 5). However, four respondents out of six indicate that the instrument could somewhat (scale 3) contribute in measuring the impact of IFP. One respondent agreed that the instrument can measure the impact of IFP and one respondent was undecided (don't know). These results support earlier findings during orientation interviews where all interviewed experts argued that it will be difficult to measure the 'real' impact of IFP. According to respondents changes in the area of education in countries could be difficult to be attributed to IFP alumni only due to the external factors that could interact with activities that alumni could conduct and because it was impossible to create control groups. However, respondents mentioned that one way to have evidence of impact is to conduct intensive country site visits beyond alumni' perceived impact responses. Asked whether the instrument is content valid, all respondents agreed that items included in the instrument are valid measures of the selected variables. Table 5.1 below depicts the frequency distribution of the answers of the respondents to the three questions included in questionnaire 3.

Table 5.1. Frequency distribution answers respondents (n=6)

Items	Frequency distribution responses					
	1	2	3	4	5	Don't know
The instrument contributes in measuring the outcomes of IFP				5	1	
The instrument contributes in measuring the impacts of IFP			4	1		1
The instrument is content valid				6		

1= Not at all/ 2= Not much/ 3= Somewhat/ 4= Considerably/ 5= To a very high extent

Based on data from table 5.1, the multiple rater Fleiss kappa computed as followed.

	1	2	3	4	5	Pi
item 1	0	0	0	5	1	0,667
item 2	0	0	4	1	0	0,367
item 3	0	0	0	6	0	1,000
Totaal nij	0	0	4	12	1	
Pj	0	0	0,222	0,667	0,056	

$$P_i = \frac{1}{n(n-1)} \left(\sum_{j=1}^{k} n_{ij}^2 - n \right) \text{ and } P_j = \frac{1}{Nn} \sum_{i=1}^{N} n_{ij} \text{ (n = 6; k = 5; N = 3)}$$

$$\bar{p} = \frac{1}{N} \sum_{i=1}^{N} P_i = (0{,}667 + 0{,}367 + 1{,}000)/3 = 0{,}678$$

$$\bar{P_e} = \sum_{j=1}^{k} P_j^2 = 0^2 + 0^2 + 0{,}222^2 + 0{,}667^2 + 0{,}056^2 = 0{,}497$$

$$k = \frac{\bar{P} - \bar{P_e}}{1 - \bar{P_e}} = (0{,}678 - 0{,}497)/(1 - 0{,}497) = 0{,}36$$

The computed value of the multiple rater Fleiss kappa was 0,36. This value of kappa indicates a fair agreement between respondents according to Landis & Koch (1977). This means that agreement observed between respondents is greater than the expected chance agreement.

5.3 Results step 4: Field test in Senegal

In this step, the refined instrument (based on recommendations from step 3) was submitted to four alumni graduated and active in the area of education in Senegal for evaluation. Respondents were first asked to indicate the extent to which questions in the instrument are easy to follow, easy to understand, and easy to answer according to a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Further, they were asked to indicate the time needed for completion and to indicate the questions that were difficult to answer. They were also offered the opportunity to provide personal comments. However, only one out of four alumni filled in the questionnaire. Table 5.2 depicts the responses of this respondent.

Table 5.2. Frequency distribution answers respondent (n=1)

	Frequency distribution responses				
	1	2	3	4	5
The questionnaire is easy to follow				1	
The questions are easy to understand				1	
The questions are easy to answer				1	

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

Table 5.2 shows that the respondent agrees (scale four) that the questionnaire is easy to follow and that the questions are easy to understand and to answer. However, this respondent indicated that the questionnaire was time consuming to fill out as it was an English version. He recommended a French version of the questionnaire for alumni from French speaking countries. Because of the low response rate (25%), data was not statistically analysed by computing Fleiss kappa statistics in order to measure the strength of agreement between respondents. Therefore, results from step 4 should be interpreted cautiously. Further, no

conclusion could be made about the approximate time needed to fill out the questionnaire as only one respondent fill out the questionnaire.

5.4 Summary of all results

Based on orientation interviews followed by two Delphi rounds of evaluation of indicators and variables, 7 core indicators and 21 variables were found appropriate to measure outcomes and impacts of IFP. The seven indicators included: (1) socio-biographical backgrounds of alumni (three variables), (2) educational backgrounds of alumni (four variables), (3) career progress (four variables), (4) employment (five variables), (5) type of activities in the area of education (three variables), (6) changes caused by conducted activities (one variable), and (7) alumni satisfaction (one variable). An overview of indicators and variables retained for the IFP instrument are included in appendix G as well as data collection methods that could be used during evaluation of IFP. As an example, table 5.3 below provides the variables that were retained for indicator 2 and data collection methods that could be used to gather data on these variables.

Table 5.3. Variables related to educational backgrounds variables and data collection methods

Variables	Operational definition	Data collection methods
Education degree obtained	Highest education degree obtained through participation in the IFP program	- Use of a control group if feasible or before and after training approach - Tracer study with survey questionnaire (or interviews)
Place of study	Country and University where alumni obtained the degree	- Review of conducted formative evaluations (e.g. review of data gathered by CHEPS)
Kind/content of study	Fields of study followed by alumni Kinds of study followed by alumni such as teacher training, curriculum development, etc.	- Review of records of fellowship completion -Testing of knowledge and skills integral to teaching and learning program - Assessment of ability to practice selected skills and apply particular knowledge in different settings
Year of graduation	year when alumnus/alumna obtained his/her degree	-Baseline assessment where fellows indicate their perceived competence level

The IFP instrument developed after selection of these indicators and variables (see appendix I) was found content valid by respondents with a multiple rater Fleiss kappa value of 0,36. This means that agreement between respondents is greater than chance agreement. Further, this instrument was found appropriate to measure outcomes of IFP. However, the instrument on its own cannot measure the 'real impact' because of attribution issues.

6. Conclusions, discussions, and recommendations

The previous chapters have provided the problem statement and research questions (chapter 1), the findings of literature review (chapter 2), the list of pre-selected indicators and variables (chapter 3), the research methodology (chapter 4), and the results of evaluations conducted in order to establish content validity of the IFP Ed-instrument (chapter 5). In this final chapter, the conclusions based on the findings of the study are presented and discussed. Further, recommendations for use and further development of the IFP Ed-instrument are made. In the last section, the limitations of the study are explained and avenues for further research are proposed.

6.1. Summary of the research process and problem statement

The main aim of this study was to develop an instrument that could measure the impact of the International Fellowships Program (IFP) in the area of education. IFP was launched in 2001. The program has provided fellowships to more than 4300 academics from 22 developing countries or territories (IFP, n.d.a). Since its launch, different formative evaluations have been conducted by the Center for Higher Education Policy Studies (CHEPS) in order to measure whether or not the program has been implemented successfully. However, to the time this research was conducted, the impact of the program has not yet been evaluated. This pilot study was therefore conducted on behalf of the Center for Higher Education Policy Studies (CHEPS). Therefore, the central research question was formulated as follows:

How to develop a valid and reliable instrument that can measure the impact of the International Fellowships Program (IFP) in the area of education in its selected countries?

To answer this central question, the following three additional questions were answered first:

4- *What frameworks are available in the literature for the assessment of the impact of educational programs and to which extent are these frameworks appropriate for the assessment of the impact of IFP?*

5- *What kinds of outcome and impact indicators are necessary to assess the impact of IFP in the area of education?*

6- *How could the quality in terms of validity and reliability of the developed instrument be tested?*

To address these questions, this study was performed in an exploratory way as in literature little is known about the assessment of the impact of fellowship programs (Boeren, 2005; Norad, 2005; Searle et al., 2006). Qualitative and quantitative data were collected through semi-structured interviews and mails questionnaires with academics and practitioners specialized in the field of education as well as experts in the field of impact assessment. Collected data were analysed qualitatively and quantitatively.

Answers to the three research questions are presented and discussed separately in the following three sections. Findings related to these three questions are used to answer the main research question.

6.2 Research question 1

Question 1: What frameworks are available in the literature for the assessment of the impact of educational programs and to which extent are these frameworks appropriate for the assessment of the impact of IFP?

The first question investigates the extent to which frameworks developed by social scientists for the assessment of educational programs are applicable for the IFP program. The findings of this study show that different frameworks are available for the assessment of educational programs. Examples of such frameworks (see chapter 2) are the Scheerens' quality

framework, the OECD framework, and the Kirkpatrick four-level framework (Kirkpatrick, 1996; OECD, 2004; Scheerens, 2004). However, analyses conducted in chapter three show that a modified Kirkpatrick framework is appropriate to assess the outcomes and impacts of IFP at individual, institutional and system level. The use of the modified Kirkpatrick framework in order to evaluate fellowship programs is supported by literature (Aguirre International, 2004; CIDA, 2005; DFC, n.d.; Rotem et al., 2010). For example, a modified Kirkpatrick framework was used during evaluation of the ATLAS/AFGRAD fellowship program and the Canadian Francophone Scholarship Program (Aguirre International, 2004; CIDA, 2005). Further, a modified Kirkpatrick framework is used by the Danish International Development Agency (DANIDA) for their monitoring and evaluation system (DFC, n.d.).

The modified Kirkpatrick framework used in this study as conceptual framework (see chapter three) in order to pre-select indicators and variables comprises three components rather than the four components of the original four-level Kirkpatrick framework. The first component (the so-called reaction level) of the original Kirkpatrick four-level framework is not used in this study as this level produces information that are not useful for the assessment of the impact of fellowship programs (Aguirre International, 2004; Rotem et al., 2010; Zinovieff, 2008). The three components used in this study are the learning component, the behaviour component, and the results component.

The learning component includes indicators and variables appropriate to measure the knowledge, skills and attitudes gained by trainees during the fellowship program. Learning variables are related to the objectives of IFP at individual level. These variables are important to be measured as without evidence of learning it will be impossible to attribute alumni' impacts to fellowship programs (Aguirre International, 2004).

The behavior component includes indicators and variables appropriate to measure whether IFP alumni are able to apply the acquired knowledge, skills, and attitudes in institutions and organizations where they are active and beyond institutions. Behavior indicators and variables are related to the objectives of IFP at institutional level. According to Krasulin et al. (1998), fellowships will contribute to capacity building in recipient countries only if alumni get the opportunity to apply their newly acquired knowledge or skills. Similarly, Boeren (2005) and Norad (2005) argue that fellows may have impact when they can fully apply what they have learned. The statements of these social scientists stress how important it is to be able to measure on which degree the acquired knowledge can be applied by the alumni.

The result component includes indicators and variables appropriate to measure outcomes and impacts of IFP. Results indicators are related to the objectives of IFP at individual as well as institutional and system (country) level (Enders & de Boer, 2002). Within this study, outcomes of IFP (see chapter 3) are activities that alumni conduct in order to bring about attitudinal and behavioural changes as well as changes in the functioning and performances of institutions and organizations where alumni are active. As mentioned in chapter three, impact is defined in this study as *changes in IFP countries as a result of activities conducted by individual or by a network of alumni*. Similar definitions of impact are used during evaluation of fellowship programs. For example, during evaluation of ATLAS/AFGRAD program, impact was defined by the evaluation team as: "*any change that occurred at institutional, sectoral, national, or regional level attributed to ATLAS/AFGRAD-sponsored training*" (Aguirre International, 2004, p.xvi).

6.3 Research question 2

Question 2: What kinds of outcome and impact indicators are necessary to assess the impact of IFP in the area of education?

The second question investigates the kinds of outcome and impact indicators that are necessary to assess the impact of IFP in the area of education. Findings of this study show that seven core indicators are appropriate to measure outcomes and impacts of IFP in the area of education. The seven indicators are: (1) socio-biographical backgrounds of alumni, (2) educational backgrounds of alumni, (3) employment, (4) career progress, (5) type of activities in the area of education, (6) changes caused by conducted activities, and (7) alumni satisfaction. To each of the seven indicators at least one variable could be allocated with in total 21 variables elaborated in this research (see table 6.1). The variables have content validity indexes (I-CVI) values higher than 0,78 and content validity ratio's (CVR) values higher than 0,75. These values indicate a high degree of agreement (Lawshe, 1975; Lynn, 1986 in Polit & Beck, 2006) among the consulted experts concerning the relevance and measurement of the individual variables selected for each of the seven indicators. The values also indicate that selected variables are content valid measures of outcomes and impacts of IFP. An overview of variables related to these seven indicators is depicted in table 6.1.

Table 6.1. Indicators and variables selected for the IFP Ed- instrument

INDICATORS	VARIABLES
Indicator 1: Socio-biographical backgrounds	- Gender - Age - Country of birth and residence
Indicator 2: Educational backgrounds	- Education degree obtained through IFP - Place/country of study - Kind/content of study - Year of graduation
Indicator 3: Employment	- Place of employment - Kind of employment institution - Sector of employment - Applicability knowledge/skills acquired through IFP - Function description
Indicator 4: Career progress	- Numbers of promotion - Job position - Responsibility in work - Locus of decision
Indicator 5: Type of activity	- Voluntary activities - Educational management activities - Professional/operational activities
Indicator 6: Change caused by conducted activities	- Alumni' perceived impact
Indicator 7: Alumni satisfaction	- Alumni' satisfaction with IFP

Indicator 1 *(socio-biographical backgrounds of alumni)* includes variables related to gender of alumni, age of alumni, and countries of residence of alumni. These variables allow comparisons between the performances of male and female alumni and comparisons between countries that participate in the program. Selection of such variables is in line with recommendations of literature (Prennushi, Rubio & Subbarao, 2001; Roche, 1999; World Bank, 2004) which argue that indicators should provide data that allow wherever possible international comparability and disaggregation (World Bank, 2004). Disaggregation helps to understand the differences between groups, gender and areas (Prennushi et al., 2001; Roche, 1999).

Indicator 2 *(educational background)* relates to the degree obtained by alumni through participation in the IFP program and points out the content of the study followed by alumni. This indicator includes variables found appropriate to measure whether learning was acquired by alumni during fellowship program. The use of educational variables lends support to earlier studies on fellowship programs (CSC, 2009; Norad, 2005; Strömbom, 1989; Teichler, 1991) where such variables were used to measure the amount of knowledge and skills gained during training received by alumni. Measurement of educational variables is recommended in literature because scholars consider these variables as important determinants of career progress (Heslin, 2005; Lortie-Lussier & Rinfret, 2005; Melamed, 1995; Polk & Armstrong, 2001; Turban & Dougherty, 1994; van der Sluis & Poell 2003; Wayne et al., 1999; Whitely et al., 1991). The use of educational variables to predict career progress of alumni lends support to studies conducted in organization settings. These studies show a positive correlation between educational levels and promotion (Whitely et al., 1991; Polk & Armstrong, 2001) and between educational levels and position occupied (Lortie-Lussier & Rinfret, 2005; Melamed, 1995; Polk & Armstrong, 2001). The importance of educational background variables is also emphasized by consulted experts. For example, respondents argue that the variable content of study followed by alumni is important to measure because it helps to understand whether the acquired knowledge will be applicable in the context of developing countries.

Indicator 3 (*employment)* and indicator 4 (*career progress)* include variables related to employment of alumni (e.g. sector of employment of alumni after participation in IFP) and variables related to career progress of alumni (e.g. number of promotions after participation in IFP). Indicators three and four are related to the behavior component of the conceptual framework as well as to results at individual levels. Findings of this study show that the selected variables are appropriate to measure outcomes and impacts of IFP. According to respondent R6 *"outcomes of IFP will depend on the sector of employment of alumni but also on the position and function that alumni occupy".* Similar to R6, respondent R4 said that *"changes will depend on the level of employment of alumni but also on their visions on education".* According to R10 *"IFP will be more effective when IFP alumni will occupy high positions in the decision making structure".* The statements of these consulted experts stress how important it is to be able to measure employment and career progress of alumni in order to measure outcomes and impacts of IFP. Selection of employment and career progress variables is supported by earlier studies on fellowship programs (CSC, 2009; Norad, 2005; Strömbom, 1989; Teichler, 1991) which confirm that employment and career progress variables were used during evaluation of fellowship programs. Scholars argue for example that behaviour of alumni in terms of applying knowledge and skills gained during participation in the IFP program can only be measured if alumni are employed (Aguirre International, 2004; Krasulin et al., 1998; Norad, 2005). Career progress variables (e.g. locus of decision) for example could implicitly indicate whether alumni have authority in order to influence changes in organizations and institutions where they are active (Aguirre International, 2004).

Indicator 5 (*type of activity conducted by alumni*) is related to the professional as well as voluntary activities that IFP alumni conduct in the area of education. This indicator attempts to measure the outcomes of IFP. Indicator 5 is related to results at institutional as well as system level. Findings of interviews show that variables related to indicator 5 are relevant and measurable as outcomes of IFP. These findings lend support to Kottmann & Enders (2009) who argue that IFP alumni conduct professional as well as voluntary activities. Further, these findings support earlier studies on fellowship programs (Aguirre International, 2004; CIDA, 2005; CSC, 2009; Norad, 2005) where similar variables were used to evaluate activities that alumni conduct in organizations and countries.

Indicator 6 (*changes in the area of education*) measures the changes caused by activities conducted by IFP alumni in the area of education. This indicator attempts to measure the impact of IFP. Indicator 6 is related to results at institutional as well as system level. Findings of this study show that impact of IFP should be measured on alumni perceptions about changes caused by their activities in the country. According to respondent R10 for example *"impact should be measured on alumni perceptions with the focus on the relation between degree obtained [master or PhD] and what alumni do in the country".* These findings support earlier evaluation studies on fellowships programs (ADB, 2007; Aguirre International, 2004; CIDA, 2005; CSC, 2009; Norad, 2005; Rotem et al., 2010; World Bank Institute, 2010) where impact of programs were measured based on alumni perceptions.

Indicator 7 (*alumni satisfaction*) relates to the degree to which IFP alumni are satisfied with the IFP program. Findings of this study indicate that satisfaction is a variable that can produce data relevant for the improvement of the IFP program. According to consulted experts, alumni satisfaction should not be limited to training received but should include the whole process from selection till employment after graduation. These findings lend support to fellowship program evaluation studies (CSC, 2009; Norad, 2005) where satisfaction variables were used to measure alumni degree of satisfaction with the fellowship programs.

6.4 Research question 3

Question 3: How could the quality in terms of validity and reliability of the developed instrument be tested?

The third question investigates how the quality in terms of instrument validity and reliability of the IFP Ed-instrument could be tested. Findings of this study (see chapter 4) show that only the content validity of the IFP Ed-instrument could be established because of the scope of the present study. Four steps were therefore used to develop a content valid IFP Ed-instrument.

During the first step, 11 indicators and 25 variables were pre-selected from literature. Selection from literature is emphasized by scholars (Aladwani & Palvia, 2002; Clark & Watson, 1995; DeVellis, 2012; Haynes et al., 1995; Lawshe, 1975; Moore & Benbasat, 1991; Radhakrishna, 2007; Sendjaya, 2003; Wynd et al., 2003) as the review of literature is an important step to ensure content validity of instruments. According to Lynn (1986) in Wynd et al. (2003) a comprehensive literature review is needed in the developmental stage of any new instrument.

During the second step, the content validity of all pre-selected indicators and variables was evaluated by consulted experts. The importance of such an evaluation is stressed in literature in order to estimate content validity of individual variables (DeVellis, 2012; Moore & Benbasat, 1991; Sendjaya, 2003). According to DeVellis (2012, p.100), review of items for example maximize content validity of instrument. Findings of Delphi study conducted in the second step (see chapter 5) show that seven indicators and 21 variables are content valid measures of outcomes and impacts of IFP.

During the third step, content validity of the overall IFP Ed-instrument was estimated by consulted experts. This way of proceeding is in line with literature (Grant & Davis, 1997; Lynn, 1986 in Wynd et al., 2003; Sendjaya, 2003) which recommends assessing the content validity of whole instruments. According to Grant & Davis (1997, p.272) the whole instrument should be evaluated by experts in order to find out whether "all dimensions of the content domain are included in the instrument" and for addition or deletion of items. Findings (see chapter 5) show that the developed IFP Ed-instrument is content valid and appropriate to measure the outcomes of IFP. However, it was found that it will be difficult to measure

impact of IFP with the IFP Ed-instrument because of attribution issues such as the influence of external factors and because of the lack of control groups.

During the fourth step, the IFP Ed-instrument was field tested with IFP alumni in Senegal. The importance of field test is emphasized in literature. According to scholars (DeVellis, 2012; Moore & Benbasat, 1991; Radhakrishna, 2007; Sendjaya, 2003) field testing of instruments with an appropriate sample is important to establish its construct or criterion validity as well as its reliability. The aim of the field test in this study was to check the length and clarity of the instrument and not to establish construct validity and reliability of the developed instrument. However, findings of this study are not appropriate to draw conclusions about the clarity and length of the IFP Ed-instrument because of the low response rate (only one respondent out of four) observed during the evaluation.

6.5 Summary of conclusions: developed IFP Ed-instrument
In summary, this study found that a modified Kirkpatrick framework is appropriate to assess the impact of IFP in the area of education. Further, four steps were found appropriate to develop the IFP Ed-instrument: pre-selection of variables from literature, selection of variables by experts followed by development of instrument, evaluation of content validity of the instrument, and field test of the instrument. Results of this study show that seven core indicators with 21 variables are appropriate to be used to develop the IFP Ed-instrument. An overview of all selected indicators and variables as well as their operational definitions and data collection methods is presented in appendix G.

As an example in order to clarify the operationalization process, table 6.2 and table 6.3 provide the variables retained for indicator 2, the operational definitions of these variables, the items that were retained for the variables and data collection methods that could be used in order to gather data on these variables.

Table 6.2. Operational definitions educational backgrounds variables and data collection methods

Variables	Operational definition	Data collection methods
Education degree obtained (level study)	Highest education degree obtained through participation in the IFP program	- Use of a control group if feasible or before and after training approach - Tracer study with survey questionnaire (or interviews)
Place of study	Country and University where alumni obtained the degree	- Review of conducted formative evaluations (e.g. review of data gathered by CHEPS) - Review of records of fellowship completion
Kind/content of study	Fields of study followed by alumni Kinds of study followed by alumni such as teacher training, curriculum development, etc.	-Testing of knowledge and skills integral to teaching and learning program - Assessment of ability to practice selected skills and apply particular knowledge in different settings -Baseline assessment where fellows indicate their perceived competence level
Year of graduation	year when alumnus/alumna obtained his/her degree	

Table 6.3. Items pre-selected for educational backgrounds variables

Variables	Items
Education degree obtained through IFP	Whether alumni have obtained one of the following degrees: Master degree, Doctorate/PhD degree, or other degree
Place/country of study	Whether alumni have studied in own country, abroad in own continent or abroad in another continent
Kind/content of study	Whether alumni have done one or more of the following studies in the IFP program: Curriculum development - Education evaluation and monitoring - Education assessment - Special education- Instructional material development - Teacher training - Education management - E-learning - Media application in education - Human resources in education - Accreditation of education
Year of graduation	Year alumni have been graduated

With these findings the IFP Ed-instrument was developed. The instrument is composed of five parts:
- Part 1 includes questions related to alumnus/alumna socio-biographical and educational backgrounds.
- Part 2 includes questions related to the employment of alumnus/alumna and the work environment within which he/she is active.
- Part 3 includes questions related to activities (professional or voluntary) conducted by alumnus/alumna in his/her institution or community/country in the area of education.
- Part 4 includes questions related to the impact of IFP in the institution or community/country where alumnus/alumna is active.
- Part 5 includes questions related to the satisfaction of alumnus/alumna with his/her participation in the IFP program and his/her career.

An overview of the final IFP Ed-instrument is included in appendix I. Results of this study show that the IFP Ed-instrument is appropriate to measure the outcomes of IFP. However, the instrument, without control groups and country site visits, is limited to measure on its own the impact brings about by IFP in the selected countries.

6.6 Recommendations on how to use the IFP Ed-instrument
From the study findings, three recommendations on how the IFP Ed-instrument should be used are made in this section.

6.6.1 Time-lags between measurements with IFP Ed-instrument
In order to reduce influence of context factors on data generated with the IFP Ed-instrument, short time-lags should be used between measurements with the IFP Ed-instrument. If possible learning indicators (knowledge, skills and attitudes) should be measured before and immediately after training (Kirkpatrick, 1996; Rotem et al., 2010; DFC, n.d.). According to Rotem et al. (2010) learning indicators are easy to assess during training and relatively difficult to measure the longer the intervals after training. Based on the formative evaluation of IFP (Enders & de Boer, 2002) and on recommendations of consulted experts, the frequency of 1-3 years is recommended within this study. According to respondents, attribution of impacts to IFP alumni will become more complicated with long time lags because of the influence of context factors. The use of short time lags between measurements facilitates the

measurement of changes and facilitates the attribution of these changes to the intervention (Innovation Network, 2005).

6.6.2 Data collection process with IFP Ed-instrument

Data collection with the IFP Ed-instrument should be conducted by the Center for Higher Education Policies Studies (CHEPS) because of their experience in collecting data in the context of IFP. CHEPS has developed the evaluation framework of IFP and has conducted several formative evaluations since 2002 (Enders & de Boer, 2002; Enders et al., 2005). Local stakeholders such as IFP country partners should be involved in data collection processes. Involvement of IFP country partners will empower them for further assessments and will create a sense of ownership (Maredia, Byerlee & Anderson, 2000; Roche, 1999; Prennushi et al., 2001). During data collection, effort should be made to involve alumni not directly active in the country (alumni who stay abroad) in order to point out their contribution in the area of education of their own countries (CIDA, 2005). This is supported by an alumnus (respondent R5) who does not return in its country after graduation. According to him "*it will be important to include alumni who stay abroad during the assessment because what I see here I forward it at home. So, I indirectly influence the educational area in my country*". Further, in order to facilitate data collection, it is recommended to develop different versions (in term of language) of the IFP Ed-instrument. Lack of a French version was one of the reasons of the low response rate in Senegal according to respondent S1.

If resources are available, it is recommended to collect data before and after graduation of alumni in order to allow a before-and-after analysis as recommended by (Kirkpatrick, 1996). In order to implement such an approach, CHEPS should make an effort to up-date their database in terms of current contact information on alumni phone numbers, emails addresses, name employers, and country of residence.

Data collected with the IFP Ed-instrument should be cross checked through country site visits in order to have evidence of impact as the instrument measures impact based on alumni perceptions. Examples of methods that could be used to cross check responses of alumni include: interviews with alumni employers or supervisors, document reviews, group meetings with alumni, individual interviews with alumni or association, institutional visits, internet search for impact in countries not visited by evaluators, and interviews with program officers (ADB, 2007; Aguirre International, 2004; CIDA, 2005; Rotem et al., 2010).

6.6.3 Data analysis and report of findings

Data collected with the IFP Ed-instrument should be analysed by CHEPS because the center has already built lots of experience in analysing data included in the different formative evaluations (see Enders & de Boer, 2002; Enders et al., 2005). CHEPS should also be responsible for the report of findings because the center is already responsible for reports about the different formative evaluations since 2002 (see Enders et al., 2005; Enders et al., 2007; Enders, Kottmann & Deen, 2006; Kottmann & Enders, 2009; Enders & de Boer, 2002). The findings should be reported to the appropriate audiences involved in the IFP such as IFP secretariat, international country partners, placement universities, the British Council, the Dutch Higher Education Organization (NUFFIC), strategic university partners, the international advisory committee, and alumni associations (Enders & de Boer, 2002).

6.7 Reflection

In this section limitations of the study are provided and recommendations avenues for further research and how to further develop the IFP Ed-instrument are made.

6.7.1 Limitations of the study

The most important restriction of this study relates to the unavailability of construct (or criterion) validity and reliability estimates for the developed IFP Ed-instrument. These estimates were not established because of the scope of the study. For further improvement of the developed IFP Ed-instrument it is recommended to use field test studies in order to test its validity and reliability (credibility or consistency). The field tests could also be used to evaluate the clarity (in terms of easy to follow, easy to understand, and easy to answer) and the length of the IFP Ed-instrument. Clarity and length of the instrument were not established within this study because of the low response rate during the field test in Senegal.

Another shortcoming of this study is related to the relatively low number of respondents who participated in the development process of the IFP Ed-instrument. Only 10 experts participated in the orientation interviews. From these ten experts, only eight participated in first Delphi-round, seven participated in the second Delphi round and six participated in the third step (validation of overall instrument). Non-response bias can therefore be a limitation of the present study since non-responses between the two Delphi-rounds (number of respondents vary between rounds) were observed and responses on some statements were missed (during evaluation of items). Further, because of the small sample sizes the selected experts may not be fully representative of the target population of experts in the field of impact assessment of fellowship programs. It is therefore required to be cautious when drawing conclusions based on the findings of this study.

6.7.2 Avenues for further research

This study is a pilot study intended to understand how to develop an instrument that could assess the impact of IFP in the area of education in its selected countries. Much more studies in order to refine and verify the findings of this study are therefore recommended.

This study was limited to the area of education. However, from earlier formative evaluations conducted by CHEPS, it was pointed out that IFP alumni are active in diverse areas. Therefore, further research should examine the possibility of using the findings of this study to other areas where alumni could be active. For that purpose, the specified list of indicators and variables can be used as starting point for the development of indicators in other areas where alumni are active. Further, another avenue of research could be to investigate the feasibility to use school effectiveness research variables as benchmarks for the evaluation of outcomes of IFP in the area of education because these variables are positively associated with educational achievement and attainment.

An important avenue of research is to investigate how control groups could be used in order to facilitate attribution of observed changes to IFP alumni. Another important avenue of research could be to investigate how to address the issue of low response rate of alumni observed during evaluation of fellowship programs particularly from older cohorts. Low response rate from alumni was observed in this study and was also notice as a concern by Aguirre International (2004) during evaluation of the ATLAS/AFGRAD fellowship program and by the World Bank Institute during evaluation of the Joint Japan/World Bank Graduate Scholarship Program (JJ/WBGSP) (World Bank Institute, 2007).

References

Asian Development Bank (ADB) (2007). ADB's Japan funds: Japan scholarship program: Special evaluation study. Reference number: SES: REG 2007-11. Retrieved February 9, 2011, from http://www.oecd.org/derec/adb/39506843.pdf

Aguirre International (2004). Generations of quiet progress: The development impact of U.S. Long-Term University training on Africa from 1963 to 2003. USAID bureau for economic growth and trade, Education office. Washington, D.C. Retrieved February 9, 2011, from http://pdf.usaid.gov/pdf_docs/PNADB130.pdf

Alliger, G., & Janak, E. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology, 42(3), 331–342.*

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50, 341–358.*

Aladwania, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web Quality. *Information & Management, 39*(6), 467–476.

Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. Retrieved July 25, 2009, from http://www.asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html

Aoki, A., Bruns, B., Drabble, M., Marope, M., Mingat, A., Moock, P., …Yang, H. (2001). Education. Draft for comments in chapter 19, Poverty Reduction Strategy Papers Source Book. Retrieved April 24, 2009, from http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/5798_chap19.pdf

Baker, J. L. (2000). *Evaluating the impact of development projects on poverty: A handbook for practitioners.* Washington, D.C.: The World Bank.

Bakewell, O., & Garbutt, A. (2005). The use and abuse of the logical framework approach. Stockholm, Swedish International Development Cooperation Agency (SIDA). Retrieved February 10, 2009, from http://www.intrac.org/data/files/resources/518/The-Use-and-Abuse-of-the-Logical-Framework-Approach.pdf

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52, 1-26.*

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning, 27, 341–347.*

Boeren, A. (2005). A Bird's eye view: Lessons from evaluations of international cooperation programmes in higher education, research and manpower development. Paper prepared for the Nuffic conference 'A Changing Landscape – making support to higher education and research in developing countries more effective', The Hague, the Netherlands, 23-25 May 2005.

Boeren, A., Bakhuisen, K., Christian-Mak, A.M., Musch, V., & Pettersen, K. (2008). Donor policies and implementation modalities with regard to international postgraduate programmes targeting scholars from developing countries, VLIR-UOS and Nuffic. Retrieved February 12, 2011, from http://www.vliruos.be/downloads/Donor_policies_Nuffic.pdf

Boissiere, M. (2004). Determinants of primary education outcomes in developing countries: Background paper for the evaluation of the World Bank's support to primary education. Washington, D. C.: The World Bank Operations Evaluation Department. Retrieved May 1, 2010, from http://www.worldbank.org/oed/education/documents/education_primary_determinants_paper.pdf

Campbell, S.M., & Cantrill, J.A. (2001). Consensus methods in prescribing research. *Journal of clinical pharmacy and therapeutics, 26, 5-14.*

CIDA (2005). Evaluation of the Canadian Francophonie Scholarship Program (CFSP), 1987-2005: A need for reorientation. Canadian International Development Agency, Evaluation Division Performance and Knowledge Management Branch, Gatineau, Quebec, December 2005. Retrieved February 9, 2011, from http://www.oecd.org/derec/canada/37778524.pdf

CIDA (1997). Guide to gender –sensitive indicators. Canadian International Development Agency, Hull, Quebec, August 1997. Retrieved February 12, 2011, from http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUImages/Policy/$file/WID-GUID-E.pdf

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319

Clemens, M., Radelet, S., & Bhavnani (2004). Counting Chickens When they Hatch: The Short- Term Effect of Aid on Growth. Center for Global Development, Working Paper Number 44. Retrieved June 10, 2009, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=567241

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20(1), 37-46.*

Commonwealth Scholarship Commission (CSC), (2009). Evaluating commonwealth scholarships in the United Kingdom: Assessing impact in key priority areas. Commonwealth Scholarship Commission in the UK, London, UK, June 2009. Retrieved February 12, 2011, from http://cscuk.dfid.gov.uk/wp-content/uploads/2011/03/evaluation-impact-key-report.pdf

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine, 119(2), 166.e7-166.e16*

Danida Fellowship Centre (n.d.). Effect measurement. Danish International Development Agency fellowship centre. Retrieved April 10, 2011, from http://dfcentre.com/about-us/effect-measurement/

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Los Angeles: SAGE

DeVon H. A., Block M. E., Moyle-Wright P., Ernst D. M., Hayden S. J., Lazzara D. J., … Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship 39*(2), 155 –164.

Dooley, D. (2001*). Social research methods (fourth edition)*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Enders, J.,& de Boer, H. (2002). *The IFP evaluation framework: Approaches and instruments for data and information gathering*. University of Twente: Center for Higher Education Policy Studies.

Enders, J., & de Boer.H. (n.d.). *The IFP Evaluation Framework. Volume 2: Instruments for data and information gathering*. University of Twente: Center for Higher Education Policy Studies.

Enders, J., Kottmann, A., & Deen, J. (2006). *The International Fellowship Program mid-term experiences and outcomes: An interim report*. University of Twente: Center for Higher Education Policy Studies.

Enders, J., Kottmann, A., & Leisyte, L. (2007). *The International Fellowship Program experiences and outcomes: An update report*. University of Twente: Center for Higher Education Policy Studies.

Enders, J., Theisens, H., & Westerheijden, D.F. (2005). *The International Fellowship Program: Target group definition, outreach and selection; pre-academic training, placement, and post-graduate experience of fellows; early career of alumni.* University of Twente: Center for Higher Education Policy Studies.

European Commission (2002). Guidelines for the use of indicators in country performance assessment. The European commission, development policy and sectoral questions, Economic co-operation and PRSP process, Brussels, December 2002. Retrieved April 20, 2012, from http://ec.europa.eu/development/icenter/repository/guidelines_indicators_cpa_en.pdf

Edelfelt, R. A. & Reiman, A. J. (2004). *Careers in education.* New York: The McGraw-Hill

Eseryel, D. (2002). Approaches to evaluation of training: Theory & practice. *Educational Technology & Society, 5(2), 93-98.*

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76(5), 378-382.*

Ford, D.J. (2004). Evaluating performance improvement. *Performance Improvement, 43(1), 36-41.*

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education (eight edition).* New York: McGraw-Hill

French Ministry of Education (2010). The state of education 2010: Costs, activities, results. 29 indicators on the French education system. Retrieved May 15, 2012, from http://www.education.gouv.fr/cid53863/l-etat-ecole-indicateurs-sur-systeme-educatif-francais.html

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research, 64*(1), 119-157.

Garland, R. (1991). The Mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2, 66-70.* Retrieved June 07, 2009, from http://marketing-bulletin.massey.ac.nz/V2/MB_V2_N3_Garland.pdf

Gibson, F., Cargill, J., Allison, J., Begent, J., Cole, S., Stone, J., & Lucas, V. (2006). Establishing content validity of the oral assessment guide in children and young people. *European Journal of Cancer, 42*(12), 1817-1825.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8(4), 597- 606.*

Grant, J. S., & Davis, L. L. (1997). Focus on quantitative methods: Selection and use of content experts for instrument development. *Research in Nursing and Health, 20, 269-274.*

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33*(1), 64-86.

Havelock, R. G., & Zlotolow, S. (1995). *The change agent's guide (2nd ed.).* Englewood Cliffs, NJ: Educational Technology Publications

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32(4), 1008-1015.*

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238–247.

Heneveld, W., & Craig, H. (1996). *Schools count: World Bank project designs and the quality of primary education in Sub-Saharan Africa.* World Bank technical paper number 303. Africa technical department series. Washington, DC: The World Bank. Retrieved April 26, 2010, from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1996/01/01/000009265_3961214183318/Rendered/PDF/multi_page.pdf

Heslin, P. (2005). Conceptualizing and evaluating career success. *Journal of Organizational Behavior, 26, 113 –136.*

Hill, K. Q., & Fowles, J. (1975). The methodological worth of the Delphi forecasting technique. *Technological Forecasting and Social Change, 7*, 179-192

Holton, E. F., III (1996). The flawed four level evaluation model. *Human Resource Development Quarterly, 7(1), 5–21.*

Hsu, C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment Research & Evaluation, 12*(10), 1-8. Retrieved from April 10, 2011, from http://pareonline.net/pdf/v12n10.pdf

IFP (n.d.a). About IFP. Retrieved September 15, 2011, from http://www.fordifp.org/AboutIFP.aspx

IFP (n.d.b). Our vision. Retrieved September 15, 2011, from http://www.fordifp.org/AboutIFP/Vision.aspx

IFP (n.d.c). The IFP model. Retrieved September 15, 2011, from http://www.fordifp.org/AboutIFP/TheIFPModel.aspx

IFP (n.d.d.). History. Retrieved September 15, 2011, from http://www.fordifp.org/AboutIFP/History.aspx

IFP (n.d.e). Outcomes: An overview. Retrieved September 15, 2011, from http://www.fordifp.org/AboutIFP/Outcomes.aspx

Innovation Network (2005). Logic model workbook. Washington D.C.: Innovation Network, Inc. Retrieved February 10, 2009, from http://www.innonet.org/client_docs/File/logic_model_workbook.pdf

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*(12), 1212-1218.

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. *Journal of Advanced Nursing, 53(2), 205-212.*

Kirkpatrick, D. (1996). Great Ideas Revisited. Techniques for Evaluating Training Programs. Revisiting Kirkpatrick's Four-Level Model. *Training and Development, 50(1), 54-59*

Kottmann, A., & Enders, J. (2009). Alumni of the International Fellowships Program: Executive summary 2008. May 13, 2009. Center for Higher Education Policy Studies (CHEPS), Enschede, University of Twente.

Kottmann, A., & Enders, J. (2011). Alumni of the International Fellowships Program: Main findings from the survey 2011. Final report, august 2011. Center for Higher Education Policy Studies (CHEPS), Enschede, University of Twente. Retrieved June 11, 2012, from http://www.utwente.nl/mb/cheps/publications/

Kumar, K. (2006). Conducting mini surveys in developing countries. Washington, DC: USAID. Retrieved Mart 10, 2010, from http://pdf.usaid.gov/pdf_docs/PNADG566.pdf

Kusek, J. Z., & Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. Washington: The World Bank. Retrieved October 10, 2009, from http://www.oecd.org/derec/worldbank/35281194.pdf

Lamont, M. (2002). The evaluation of scholarship programs. Report prepared for the Paul and Daisy Soros Fellowships for New Americans Foundation, and to be presented at the conference/workshop on "strengthening Nationally Competitive Scholarships," Rockefeller Foundation's conference and study center, Bellagio, Italy, November 6-9, 2002. Retrieved February 20, 2011, from http://www.wjh.harvard.edu/~mlamont/bellagio.pdf

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1), 159-174.*

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Litwin, M. S. (1995). *How to Measure Survey Reliability and Validity*. London: Sage Publications.

Lortie-Lussier, M., & Rinfret, N. (2005). Determinants of objective and subjective success of men and women. *International Review of Administrative Sciences*, 71(4), 607-624.

Madaus, G. F., Scriven, M. S., & Stufflebeam, D. L. (1983). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Dordrecht, The Netherlands: Kluwer-Nijhof.

Maredia, M., Byerlee, D., & Anderson, J. (2000). Ex post Evaluation of Economic Impacts of Agricultural Research Programs: a Tour of Good Practice. Paper presented to the workshop on the future of impact assessment in the CGIAR: needs, constraints, and options, Rome, Italy, May 2000. Retrieved October 10, 2009, from http://impact.cgiar.org/pdf/99.pdf

Melamed, T. (1995). Career Success: The Moderating Effect of Gender. *Journal of Vocational Behavior,* 47, 35-60.

Merriam, S. B. (1995). What can you tell from an N of 1?: Issues of validity and reliability in qualitative research. *PAACE Journal of Lifelong Learning, 4, 51- 60.*

Mikkelsen, B. (2005). *Methods for development work and research: A new guide for practitioners (2nd edition).* New Delhi: Sage Publications.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2(3), 192-222.*

Mullen, P. M. (2003). Delphi: Myths and reality. *Journal of Health Organization and Management, 17(1), 37-52.*

Neuman, W. L. (2000). *Social research methods (fourth edition): Qualitative and quantitative approaches.* Needham Heights, MA: Allyn & Bacon.

Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. Oxford: Pergamon

Norad (1999). The logical framework approach (LFA) : Handbook for objectives-oriented planning. Norwegian Agency for Development Cooperation, Oslo, Norway, January, 1999. Retrieved June 15, 2009, from http://www.norad.no/en/tools-and-publications/publications/publication?key=109408

Norad (2005). Evaluation of the Norad fellowship programme. Evaluation report 1/2005. Norwegian Agency for Development Cooperation, Oslo and the Hague, July, 2005. Retrieved February 20, 2011, from http://www.norad.no/en/tools-and-publications/publications/publication?key=109531

Norad (2009). Evaluation of the Norwegian programme for development, research and education (NUFU) and of Norad's programme for master studies (NOMA): Evaluation report 7/2009. Norwegian Agency for Development Cooperation, Oslo, Norway, January, 2010. Retrieved February 15, 2011, from http://www.norad.no/en/tools-and-publications/publications/publication?key=156957

OECD (2002). Evaluation and aid effectiveness: Glossary of key terms in evaluation and results based management. OECD, the Development Assistance Committee (DAC) Working Party on Aid Evaluation, Paris, France. Retrieved June 23, 2012, from http://www.oecd.org/development/peerreviewsofdacmembers/2754804.pdf

OECD (2004). OECD Handbook for internationally comparative education statistics: Concepts, standards, definitions and classifications. Paris: OECD. Retrieved June 23, 2012, from http://www.esds.ac.uk/international/support/user_guides/oecd/Educhandbook.pdf

OECD (2011), Education at a Glance 2011: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/eag-2011-en

OECD (2012). OECD Indicators of educational systems. Retrieved June 23, 2012, from http://www.oecd.org/education/highereducationandadultlearning/49338320.pdf

OECD (2012), Education at a Glance 2012: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/eag-2012-en

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42(1), 15-29.*

Örtengren, K. (2004). A summary of the theory behind the LFA method: The logical framework approach. Retrieved January 20, 2009, from http://unpan1.un.org/intradoc/groups/public/documents/un/unpan032558.pdf

Owen, J. M. (2007). *Program Evaluation: Forms and approaches*. New York: the Guilford Press

Patton, M.Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage

Pennycuick, D. (1993). School effectiveness in developing countries: A summary of the research evidence. Education research paper, Department for International Development (DFID) serial No.1., London, UK. Retrieved April 28, 2010, from http://ageconsearch.umn.edu/bitstream/12857/1/er930001.pdf

Phillips, J. (1996). ROI: The search for best practices. *Training and Development, 50 (2), 42–46.*

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health, 29*, 489–497.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467.

Polk, O.E., & Armstrong, D.A. (2001). Higher education and law enforcement career paths: Is the road to success paved by degree?. *Journal of Criminal Justice Education*, 12(1), 77-99

Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing, 41(4), 376-382.*

Prennushi, G., Rubio, G., & Subbarao, K. (2001). Monitoring and Evaluation. Draft for comments in chapter 3, Poverty Reduction Strategy Papers Source Book. Retrieved April 24, 2009, from http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/4480_chap3.pdf

Radhakrishna, R. B. (2007). Tips for developing and testing questionnaires/instruments. *Journal of Extension, 45*(1), 1-4.

Rajeev P., Madan M. S., & Jayarajan K. (2009). Revisiting Kirkpatrick's model: An evaluation of an academic training course. *Current Science, 96(2), 272 – 276.*

Ramos, T. B., Caeiro, S., & de Mello, J. J. (2004). Environmental indicator frameworks to design and assess environmental monitoring programs. *Impact Assessment and Project Appraisal, 22(1), 47-62.*

Roche, C. (1999). *Impact Assessment for Development Agencies: Learning to Value Change*. Oxford (UK): Oxfam GB

Rotem, A., Zinovieff, M. A., & Goubarev, A. (2010). A framework for evaluating the impact of the United Nations fellowship programmes. *Human Resources for Health, 8(7),* 1-8. Retrieved February 25, 2011, from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856526/pdf/1478-4491-8-7.pdf

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15(4),* 353-375.

Sauvageot, C. (1997). Indicators for educational planning : A practical guide. International Institute for Educational Planning/UNESCO, Paris. Retrieved May 15, 2012, from http://unesdoc.unesco.org/images/0010/001034/103407e.pdf

Scheerens, J. (1991). Process indicators of school functioning: A selection based on the research literature on school effectiveness. *Studies in Educational Evaluation, 17,* 371-403.

Scheerens, J. (2000). Improving school effectiveness. Fundamentals of educational planning-No. 68. International Institute for Educational Planning/UNESCO, Paris. Retrieved April 22, 2010, from http://unesdoc.unesco.org/images/0012/001224/122424e.pdf

Scheerens, J. (2001). Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement, 12(4), 359-384.*

Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational evaluation, assessment, and monitoring: A systemic approach*. New York (NY): Taylor & Francis

Scheerens (2004). The quality of education at the beginning of the 21st century. Paper commissioned for the EFA Global Monitoring Report 2005, the Quality Imperative. Retrieved April 20, 2010, from http://unesdoc.unesco.org/images/0014/001466/146697e.pdf

Scheerens, J., & Hendriks, M. A. (2004). Benchmarking the quality of education. *European Educational Research Journal, 3*(1), 101-399. Retrieved July 4, 2012, from http://www.wwwords.co.uk/eerj/content/pdfs/3/issue3_1.asp

Searle, N. S., Hatem, C. J., Perkowski, L., & Wilkerson, L. (2006). Why invest in an educational fellowship program? *Academic Medicine, 18*(11), 936-940. Retrieved March 12, 2011, from http://www.ncme.in/Resource-Material/Different%20Models%20of%20Faculty%20Development%20in%20Medical%20Education/Medical_education_fellowships_across_NA.pdf

Sendjaya, S. (2003). Development and validation of servant leadership behavior scale. Retrieved June 24, 2012, from http://www.regent.edu/acad/global/publications/sl_proceedings/2003/sendjaya_development_validation.pdf

Shavelson, R. J., McDonnell, L., & Oakes, J. (1991). What are educational indicators and indicator systems? Practical Assessment, Research & Evaluation, 2(11). Retrieved June 8, 2012, from http://pareonline.net/getvn.asp?v=2&n=11

Shyamsundar, P. (2002). Poverty- Environment Indicators. The World Bank environment department paper No.84. Washington, D.C. Retrieved April 24, 2009, http://siteresources.worldbank.org/INTEEI/811099-1115813815502/20486522/PovertyandEnvironmentIndicators2002.pdf

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education, 6, 1-21.*

Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly, 13*(2), 147–165.

Straub, D., Boudreau, M. -C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems, 13*, 380-427.

Strömbom, M. (1989). Evaluation of fellowships awarded to developing countries: What do the studies tell? *Higher Education, 18(6), 707-724.*

Stufflebeam, D.L. (2000). The CIPP model for evaluation. In Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Eds.), Evaluation models (2nd ed.). (Chapter 16). Boston: Kluwer Academic Publishers.

Teichler, U. (1991). Evaluation of the EC training fellowship programme based on a fellows' questionnaire survey. *Scientometrics, 21(3), 343-365.*

Turban, D. B., & Dougherty, T. W. (1994). Role of protégé personality in receipt of mentoring and career success. *Academy of Management Journal, 37(3), 688-702*

Krasulin, B., Ouedraogo, L. –D., & Quijano, R. (1998). Fellowships in the United Nations system. Joint Inspection Unit, JIU/REP/98/1, Geneva, 1998. Retrieved, March 3, 2011, from http://esa.un.org/techcoop/fellowships/Downloads/RelatedReports/report_JIU_1998.pdf

UNESCO (n.d.). Gender-sensitive education statistics and indicators: A practical guide. Training material for workshops on education statistics and indicators. United Nations Educational, Scientific and Cultural Organization, Division of Statistics. Retrieved November 10, 2010, from http://www.uis.unesco.org/Education/Documents/unesco-gender-sensitive-edstats-indicators.pdf

UNESCO (2000). Education for all: Status and trends 2000: Assessing learning achievement. United Nations Educational, Scientific and Cultural Organization, Paris, France. Retrieved April 26, 2010, from http://unesdoc.unesco.org/images/0011/001198/119823e.pdf

UNESCO (2002). Education for all: Is the world on track? EFA Global Monitoring Report 2002, UNESCO, Paris. Retrieved April 25, 2010, from http://unesdoc.unesco.org/images/0012/001290/129053e.pdf

USAID (1991). A training impact evaluation methodology and initial operational guide. Prepared for: Agency for International Development, Africa Bureau. Creative Associates International, Inc., Washington, D.C., October, 1991. Retrieved February 23, 2011, from http://pdf.usaid.gov/pdf_docs/PDABD539.pdf

USAID (1995). Capturing the results of 30 years of AFGRAD training: Evidence of change in individuals and institutions across Africa. Final evaluation report. USAID Project No. 698-0455, December, 1995. Retrieved February 23, 2011, from http://pdf.usaid.gov/pdf_docs/PDABM695.pdf

Van der Sluis, E. C., & Poell, R. (2003). The impact on career development of learning opportunities and learning behavior at work. *Human Resource Development Quarterly, 14(2), 159 – 80.*

Wayne, S. J., Liden, R. C., Kraimer, M. L., & Graf, I. K. (1999). The role of human capital, motivation, and supervisor sponsorship in predicting career success. *Journal of Organizational Behavior, 20, 577-595.*

Whitely, W., Dougherty, T. W., & Dreher, G. F. (1991). Relationship of career mentoring and socioeconomic origin to managers' and professionals' early career progress. *Academy of Management Journal, 34(2), 331-351.*

W.K. Kellogg Foundation (2004). Logic Model Development Guide (Updated January 2004). Battle Creek, Michigan: W.K. Kellogg Foundation. Retrieved February 10, 2009, from http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf

World Bank (1996). Performance Monitoring Indicators. A Handbook for Task Managers. Washington D.C.: World Bank, Operations Policy Department. Retrieved April 25, 2009, from http://web.coneval.gob.mx/rw/resource/coneval/eval_mon/1331.pdf

World Bank (2004). Poverty monitoring guidance note 1: Selecting indicators. The World Bank, April 2004. Retrieved November 10, 2010, from http://siteresources.worldbank.org/INTPAME/Resources/Selective-Evaluations/NoteIndicators_eng_Apr04_doc.pdf

World Bank Institute (2007). Joint Japan/World Bank graduate scholarship program (JJ/WBGSP): Tracer study VII. World Bank Institute, may 2007. Retrieved February 25, 2011, from http://siteresources.worldbank.org/INTWBISFP/Resources/551491-1108589837615/tracer_studies7.pdf

World Bank Institute (2010). Joint Japan/World Bank graduate scholarship program (JJ/WBGSP): Tracer study VIII. World Bank Institute, may 2007. Retrieved February 25, 2011, from http://siteresources.worldbank.org/INTJJWBGSP/Resources/Tracer_Study_VIII_2010.pdf

Worthen, B. R., & Sanders, J. R. (1987). *Educational Evaluation: Alternative approaches and practical guidelines*. New York: Longman

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change, 40, 131-150.*

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*(5), 508-518.

Yin, R. K. (2003). *Case Study Research: Design and Methods (third edition).* Thousand Oaks, California: Sage Publications, Inc.

Yu, G. (2007). Research evidence of school effectiveness in Sub-Saharan Africa. EdQual working paper. Effectiveness No. 2. Retrieved April 26, 2010, from http://www.edqual.org/publications/workingpaper/edqualwp7.pdf

Zinovieff, M. A. (2008). Review and analysis of training impact evaluation methods, and proposed measures to support a United Nations system fellowships evaluation framework. Prepared for the WHO's department of human resources for health on behalf of the UN Task Force on impact assessment of fellowships. Retrieved February 25, 2011, from http://esa.un.org/techcoop/fellowships/SFOMeeting/ParticipantArea/BackgroundDocuments/6_REVIEW%20report%20FINAL%20.pdf

Appendices

Appendix A: Indicators included in quality and OECD frameworks

Indicators included in input component of Scheerens' quality framework

Table A1. Financial and material resources at system and school level (Scheerens, 2004, p.90)

System level financial and material resources indicators
- Proportion of Gross Domestic Product spent on education
- Educational expenditure per student
- Proportion of public and private investments in education
- Total expenditure on programs and special facilities for disadvantaged students
- State provision of ancillary services
- Household expenditure and public subsidies to parents
- Percentage of spending on salaries for administrative personnel
- Percentage of spending on pensions for educational personnel
- Percentage of spending on salaries for teachers
School level financial and material resources
- Proportion of the school's budget that is acquired through other than public funding
- School building facilities
- Classroom equipment (furniture, computers, etc.)
- School supplies like pencil and paper, chalk board, flipchart
- Availability of textbooks in the major school subjects
- Basic services like separate toilets for girls and boys, water, electricity, heating, telephone, provision of ancillary services, regarding nutrition, health and transportation

Table A2. Categories of teachers' indicators (Scheerens, 2004, pp.91-92)

Categories of teacher indicators
1- Teacher background characteristics
2- Teacher professional knowledge and skills
3- Teacher working conditions
4- Teacher morale and status
5- Staff to student ratios

Table A3. Indicators related to categories of teachers indicators (Scheerens, 2004, pp.91-92)

Indicators related to teacher background characteristics
- Age, sex, and ethnicity distribution
- Full-time/part-time distribution
- Certification/license status
- Formal qualifications
- Year of experience
- Language
- Health, specifically HIV
- In-service training history

Indicators related to teacher professional knowledge and skills
- General knowledge
- Content knowledge
- Knowledge about pedagogical and didactic strategies
- Knowledge about students
- Beliefs and attitudes about teaching
- Flexibility in adapting teaching repertoire

Indicators related to teacher working conditions
- Salaries (relative to other professionals)
- Working time
- Average class-size
- Merit based incentives
- Other incentive policies
- Career structures
- Teacher training/certification requirements
- Teacher autonomy
- Standards-based teacher appraisal
- Secondary working conditions (e.g. vacations)
- Exposure to external inspection

Indicators related to teacher morale and status
- Opinions about career and job mobility
- Teacher morale
- Perception about being needed by society
- Perceived status as a teacher
- Appreciation of general working conditions
- Appreciation of the work situation at school of current employment
- Job mobility
- Sense of political efficacy

Indicators related to staff to student ratios
- System level student teacher ratio
- School level student teacher ratio
- Support staff student ratio (system and school level)
- School managerial "overhead" relative to the number of students

Table A4. Students background characteristics (Scheerens, 2004, p.93)

General student background characteristics
- General intelligence or scholastic aptitude - Socio-economic status - Mother's level of educational attainment - Gender - Ethnicity
Student background characteristics associated with specific situational constraints
- Discrepancy between language spoken at home and language at school - Distance a student has to walk to school - The amount of out of school time a student has to spent on labor - Whether the students has had a meal when arriving at school - Place to study at home - Number of books in the home - Malnutrition - Ill health/HIV

Indicators included in process component of Scheerens' quality framework

Table A5. Examples of system level process indicators (Scheerens, 2004, p.83)

Process indicators defined at the level of national education systems
- Teaching time per subject - Total hours of instruction per year, for specific grade levels in primary and secondary education - Opportunity to learn, in terms of expert ratings of test curriculum overlap - The locus of decision-making in education, by education level (this indicator shows at which administrative level decisions in sub-domains of education – curriculum, personnel management – instruction, resources – are made with a certain degree of autonomy) - School autonomy (this indicator is actually included in the concept of locus of decision making) - Education standards by level [e.g. targets like increased completion rates, percentage of students scoring at or above a particular achievement level]. - Whether or not formal examinations are taken at the end of each school category - The degree of categorization and formal streaming at secondary level - The evaluation capacity of the system (defined as a quantification of the occurrence and intensity of various evaluation forms, such as national assessment programs, examinations, school inspection, an educational management information system etc.) - The magnitude and diversification of an educational support structure in the country (possibly comprising a curriculum development unit, ICT services, school counseling, an educational assessment and testing unit etc.) - The division of private, government dependent and public schools - Incentive based policies to stimulate school performance - The degree to which school choice is free

Table A6. Process indicators at school level included in the quality framework (Scheerens, 2004, p.87).

Process indicators at school level
1- Community involvement
2- School financial and human resources
3- Achievement oriented policy
4- Educational leadership
5- Continuity and consensus among teachers
6- Orderly and safe climate
7- Efficient use of time
8- Opportunity to learn
9- Evaluation of pupils' progress
10- Ratings of teaching quality

Table A7. Overview of variables related to process indicators at school level (Scheerens, 2004, pp.87-88)

Variables related to community involvement
- The degree of actual involvement of parents in various school activities (the teaching and learning process, extra-curricular activities and supporting activities)
- The percentage of the total annual school budget that is obtained from the local community
- The amount of discretion local school boards have in the conditions of labour of teachers

Variables related to school financial and human resources
- Average years of teachers' experience per school
- School level pupil teacher ratio
- Average class size per school
- Proportion of formally qualified teachers per school
- School managerial "overhead" (principal and deputy-principal fte per 1000 students)

Variables related to achievement oriented policy
- Whether or not schools set achievement standards
- The degree to which schools follow (education) careers of pupils after they have left the school
- Whether or not schools report achievement/attainment outcomes to local constituencies

Variables related to educational leadership
- The amount of time principals spend on educational matters, as compared to administrative and other tasks
- Whether or not principal's appraise the performance of teachers
- The amount of time dedicated to instructional issues during staff meetings

Variables related to continuity and consensus among teachers
- The amount of changes in staff over a certain period
- The presence or absence of school subject-related working groups or departments (secondary schools)
- Frequency and duration of formal and informal staff meetings

Variables related to orderly and safe climate

- Statistics on absenteeism and delinquency
- Ratings of school discipline by principals, teachers and pupils

Variables related to efficient use of time
- Total instruction time and time per subject matter area
- Average loss of time per teaching hour (due to organization, moving to different rooms, locations, disturbances)
- Percentage of lessons "not given", on an annual basis

Variables related to opportunity to learn
- Teacher or student ratings of whether each item of an achievement test was taught or not

Variables related to evaluation of pupils' progress
- The frequency of use of curriculum specific tests at each grade level
- The frequency of use of standardized achievement tests
- The actual use teachers make of test results

Variables related to ratings of teaching quality
- Quality of instruction as rated by peers (other teacher)
- Quality of instruction as rated by students

Table A8. Overview of teaching and learning variables (Scheerens, 2004, p.88)

Effective teaching variables
Main teaching factors
- Opportunity to learn
- Structuring and scaffolding (cognitive structuring)
- Stimulating engagement (motivational structuring)
- Climate aspects: task orientation – mutual respect – orderliness – safety
- Monitoring and questioning
- Feedback and reinforcement
- Modeling learning and self-regulation strategies
- "authentic" applications
- Adaptive teaching
Learning strategies of students
- Overt: engaged learning time – student use of resources – cooperative learning
- Covert: self-regulatory capacity – auto-control – meta-cognitive "actions" – learning styles

Indicators included in outcome component of Scheerens' quality framework

Table A9. Overview of educational outcome indicators (Scheerens, 2004, p.80)

Main categories of outcome indicators	Sub-categories
Output indicators	Achievement measures - Subject matter based - Literacy (reading, mathematical, scientific) - Competencies (e.g. learning to learn)
Outcome/attainment indicators	Attainment measures - Graduation rates - Proportion of students graduated without delay - Drop-out rates - Class repetition rates
Impact indicators	Social participation rates - (for each attainment level) % of employed at a certain job level - % of unemployed - (for lower school levels) % of enrolled in follow-up education - Degree of social participation (social capital) - Adults literacy rates - Average income, for each attainment level; earnings differential - Skills shortages and surplus

Indicators included in context component of Scheerens' quality framework

Table A10. Societal conditions relevant to education (Scheerens, 2004, p.95)

Societal conditions /contextual conditions of education systems
- Demographic developments - The labour market, e.g. shortages and surplus in certain sectors - The general state of the economy - Relevant cultural aspects - The institutional infrastructure - The general health situation in a country - Disasters of nature and war

Table A11. Variables related to cultural aspects (Scheerens, 2004, p.97)

Cultural aspects within educational system
- The status of teachers as perceived by the general public - Appreciation of education and being educated - Expectations about pedagogical functions of the school (e.g. educating for good citizenship, moral education, teaching democracy) - Cultural embedded interpretations relative to authority and educational leadership

Table A12. Variables related to demographics aspects (Scheerens, 2004, p.97)

Demographics aspects within educational system
- The supply and demand of teachers in a country - The proportion of teachers over 50 years old - The gender composition of the teacher force per school level - Percentage of students in school outside the age ranges for grade levels

Table A13. Variables related to institutional infrastructure aspects (Scheerens, 2004, p.97)

Institutional infrastructure aspects within educational system
- Degree of formalization of teacher working conditions - Formalization of teacher, student and parent rights (e.g. free school choice) - Formal monitoring and inspection of schools - Rules and enforcement of rules with respect to teacher absenteeism - Regulations with respect to private tuition by teachers in the public service - Anti-corruption measures in education - Framework for delivering and assessing the curriculum

Table A14. Organizational infrastructure of local community (Scheerens, 2004, p.98)

Organizational infrastructure of the local community
- The existence of a school board in which the local community is represented - The availability of a local or regional education resources centre (which, among others, might offer ICT facilities to the schools in the community) - The "openness" of local companies and industry to work with schools and offer students opportunities for site visits or specific training opportunities - The role of the community in financing the school; in-kind support

Table A15. Overview of local cultural conditions (Scheerens, 2004, p.99).

Local cultural conditions
- Parents' values concerning school participation of their children - Discrepancy between indigenous knowledge and "school knowledge" - Discrepancy between local perspectives on authority and ideas on active participation of students during lessons - Culturally constructed meaning of school inputs

Indicators included in responsiveness component of Scheerens' quality framework

Table A16. Areas of responsiveness to context at system and school level (Scheerens, 2004, p.101)

Areas for describing responsiveness to context at system level
- The availability of an institutional infrastructure for curriculum development
- Enforcement mechanisms that monitor curriculum development and implementation
- Liaison functions of educational authorities and societal organizations
- Analysis and research units that try to predict the demands of the labor market
- Dual systems in vocational education
Areas for describing responsiveness of school towards the local community
- External contacts of school management
- "School marketing policies"
- Active role of the school in acquiring parental involvement
- "authentic" teaching examples involving representatives from the local community

Indicator included in OECD Education at a Glance 2012

Chapter A indicators: the output of educational institutions and impact of learning

Table A17. Indicators included in chapter A (OECD, 2012, p.25)

The output of educational institutions and the impact of learning
A1- To what level have adults studied?
A2- How many students are expected to finish secondary education?
A3- How many students are expected to finish tertiary education?
A4- What is the difference between the career aspirations of boys and girls and the fields of study they pursue as young adults?
A5- How well do immigrant students perform in school?
A6- To what extent does parents' education influence access to tertiary education?
A7- How does educational attainment affect participation in the labour market?
A8- What are the earnings premiums from education?
A9- What are the incentives to invest in education?
A10- How does education influence economic growth, labour costs and earning power?
A11- What are the social outcomes of education?

Chapter B indicators: Financial and human resources invested in education

Table A18. Indicators included in chapter B (OECD, 2012, p.213)

Financial and human resources invested in education
B1- How much is spent per student?
B2- What proportion of national wealth is spent on education?
B3- How much public and private investment in education is there?
B4- What is the total public spending on education?
B5- How much do tertiary student's pay and what public support do they receive?
B6- On what resources and services is education funding spent?
B7- Which factors influence the level of expenditure?

Chapter C indicators: Access to education, participation and progression

Table A19. Indicators included in chapter C (OECD, 2012, p.317)

Access to education, participation and progression
C1- Who participates in education?
C2- How do early childhood education systems differ around the world?
C3- How many students are expected to enter tertiary education?
C4- Who studies abroad and where?
C5- Transition from school to work: where are the 15-29 year-olds?
C6- How many adults participate in education and learning?

Chapter D indicators: Learning environment and organization of schools

Table A20. Indicators included in chapter D (OECD, 2012, p.423)

Learning environment and organization of schools
D1- How much time do students spend in the classroom?
D2- What is the student-teacher ratio and how big are classes?
D3- How much are teachers paid?
D4- How much time do teachers spend teaching?
D5- Who are the teachers?
D6- Who makes key decisions in education systems?
D7- What are the pathways and gateways to gain access to secondary and tertiary education?

Appendix B: Indicators used during evaluation of fellowship programs

A- Indicators and variables used during evaluation of the ATLAS/AFGRAD program

Three types of indicators were used to evaluate the ATLAS/AFGRAD program: before the training indicators, about training indicators (during training), and after returning (after training) indicators (Aguirre International, 2004).

Table B1. Before training indicators used for the ATLAS/AFGRAD program

Type of indicator	Variables	Operational definitions variables
Employment status	1- Employment fellows 2– Organizational type (kind of organization) 3– Sector of employment 4– Job title 5– Function	1- The number of fellows, who were employed, not employed, studying before selection in program. 2-Type of organizations where fellows were employed 3- Sector of employment where fellows organizations were active (e.g. education – agriculture …) 4- Fellow job title before training 5- Fellow main activity in the job (function) (such as project management, research, teaching)

Table B2. About training indicators used for the ATLAS/AFGRAD program

Type of indicator	Variables	Operational definitions variables
Education	1- Degree level 2- Field of study	1-Highest degree obtained during training 2- Kinds of study followed by alumni such as teacher training, curriculum development, etc.
ATLAS/AFGRAD contribution to alumni knowledge	ATLAS/AFGRAD contribution to specific alumni Knowledge, Skills and Attitudes	Alumni perception whether any specific knowledge skills, and attitudes have been acquired from the academic training

Table B3. Level 3 after training indicators used for the ATLAS/AFGRAD program

Type of indicator	Variables	Operational definitions variables
Applicability of acquired knowledge skills, and attitudes	Applicability of acquired knowledge skills, and attitudes at institutions or at other areas (family, in community, in other professional relationships)	Alumni perception about applicability of acquired knowledge skills, and attitudes in work situation (e.g. institution) where they immediately worked after returning or to other areas such as within family, in community, in other professional relationships

Table B4. Level 4 and 5 after training indicators used for the ATLAS/AFGRAD program

Type of indicator	Variables	Operational definitions variables
Employment status and career	1- Employment type 2- Organizational type 3- Level of authority and responsibility 4- Career progression	1- Number of employed alumni 2- Kind of organization where alumni are active (or have been active) 3- Whether responsibilities has increased, stayed the same, or decreased since return from training 4- Alumni perception about impact of degree obtained on career path
Income	Income of alumni after graduation	Degree to which alumni attribute present income to participation in ATLAS/AFGRAD
Alumni activities (e.g. achievements, discoveries, contributions…)	Activities that alumni conduct after graduation	Examples of specific activities (contributions, achievements or discoveries) made by alumni, that are directly linked, in their view, to the knowledge skills, and attitudes acquired during academic training.
ATLAS/AFGRAD contribution to country	ATLAS/AFGRAD contribution to: - Institutional changes - Alumni association - Higher level (sectoral, national, regional or international levels) impacts	- Alumni perception about institutional changes (performance, productivity, quality…) that can be attributed to the acquired knowledge skills, and attitudes during ATLAS/AFGRAD - Alumni perception about contribution to alumni association - Alumni perception about changes caused at sectoral, national, regional or international levels that can be attributed to participation in ATLAS/AFGRAD program

B- Indicators and variables used during evaluation of the CFSP program

Two kinds of outcome indicators and two kinds of impact indicators were used to evaluate the CFSP program: immediate outcomes, medium-term outcomes, expected impact indicators, and unexpected impact indicators.

Table B5. Immediate outcome indicators used in CFSP evaluation framework

Type of indicator	Variables	Operational definitions variables
Effectiveness education received by alumni	Effectiveness of education received in Canada	Alumni perceptions of the effectiveness of their Canadian education in terms of meeting expectations (up to date training, quality of training etc.)
Benefits academic training	Alumni personal benefits from academic training received	Alumni perceptions of the benefits of academic training in terms of new knowledge, skills and attitudes

Table B6. Medium-term outcome indicators used in CFSP evaluation framework

Type of indicator	Variables	Operational definitions variables
Employment	1- Employment status 2- Access to higher positions	1- Alumni perceptions of the usefulness of the degree in accessing to appropriate jobs 2- Proportion of alumni employed or occupying higher positions appropriate to their education
Career advancement	Career advancement of alumni after graduation	Proportion of alumni progressing in career due to acquired knowledge, skills, and attitudes in Canada
Applicability knowledge, skills, and attitudes	Use of knowledge, skills and attitudes in home institutions or organizations	Percentage of fellows able to apply acquired Knowledge, skills and attitude
Organizational capacity building	Fellows effects on their organizations performance	Alumni perception of their effects on their organizations performances

Table B7. Expected and unexpected impact indicators used in CFSP evaluation framework

Type of indicator	Variables	Operational definitions variables
Expected impact indicators: CFSP contribution to recipient country	CFSP Contribution to sustainable development in recipient countries	Alumni conducted activities that contribute to sustainable development of their countries
	CFSP Contribution to poverty reduction policies	Evidence of alumni contribution to poverty reduction policies
Unexpected impact indicators: Diaspora contribution to home country	Diaspora (alumni who do not return to home country) contribution to the development of CFSP recipient countries	Evidence of Diaspora contribution to the development of home country

C- Indicators and variables used during evaluation of the JJ/WBGSP program

Three types of indicators are used to evaluate the JJ/WBGSP program: output indicators, outcome/impact on individual alumni, and outcome/impact on country (World Band Institute, 2010).

Table B8. Output indicators used in the JJ/WBGSP evaluation framework

Type of indicator	Variables	Operational definitions variables
Degree attainment	1- Degree obtained 2- Place of study 3- Field of study	1- Number of fellows who completed the JJ/WBGSP program and obtained their degree. 2- Country and University where alumni obtained the degree 3- Kind of study followed by alumni
Effectiveness and relevance of training	Effectiveness and relevance of the training received by alumni	Alumni perception of usefulness of the overall academic program, as well as specific courses and research undertaken, networks forged and experiences shared.

Table B9. Outcome/Impact indicators used in the JJ/WBGSP evaluation framework

Type of indicator	Variables	Operational definitions variables
Alumni personal benefits of the program	Alumni individual benefits from the JJ/WBGSP program	Alumni's perception of the personal benefits of the program in terms of skills recognition, progression in the same job, mobility across jobs, and higher income
Alumni residence	Alumni residence status(rate of return in home country)	Number of alumni who returned to home country
Employment status and career	1- Type and level employment 2- Sector of employment 3- Type of positions occupied	1-The number of alumni who were employed 2- Type employer of alumni 3- Positions occupied by alumni
Alumni contribution to socio economic development of countries	Alumni contribution to the socio-economic development of home country or other developing country	Alumni perceptions of the relevance of knowledge, skills, and attitudes for the country needs (for the development of home country or other developing country)

D- Indicators and variables used during evaluation of the JSP program

Two types of indicators are included in the questionnaire: impact indicators related to JSP contribution to development of alumni and impact indicators related to JSP contribution to socioeconomic development of alumni countries (ADB, 2007).

Table B10. Impact indicators related to contribution of JSP on alumni own development

Type of indicator	Variables	Operational definitions variables
Education	1- Degree attainment (level of study) 2- Field of study 3- Place of study	1- Percentage of alumni who completed the program 2- Fields of study followed by alumni: alumni are asked to report their field of study 3- Country and University where alumni obtained his/her degree
Employment status and career progression	1- Employment of alumni 2- Sector of employment alumni 3- positions occupied by alumni 4- career advancement after obtaining degree	1-The number of alumni who were employed; not employed; or pursuing studies 2- Sector of employment of alumni (private, government, own business etc.) 3- Positions occupy by alumni after graduation 4- Evidence of career advancement in terms of promotions upon returning to organization or higher positions in a different organization.
Effectiveness of training received	Alumni perceptions of effectiveness of training	Alumni' perceptions of the effectiveness of their study in transferring knowledge and skills to students
Alumni personal benefits from JSP program	Alumni perceptions of the benefits received from the JSP	Alumni' perceptions of the effectiveness of the degree program for prospects for career progression and in improving alumni' network of academic and professional partnerships

Table B11. Impact indicators related to contribution of JSP in alumni countries development

Type of indicator	Variables	Operational definitions variables
Rate of return	Rate of return of alumni in home country	Percentage of alumni who work in home country; work in another ADB DMC; work in other developing countries; residing and/or working in more developed countries
Activities conducted by alumni	Activities that alumni conduct in own countries	Alumni responsibilities in organization where they employed in home country
JSP contribution to socioeconomic development of alumni home countries	JSP contribution to socioeconomic development	Alumni' perceptions on the relevance and usefulness of the knowledge and skills gained from their degree programs to the development needs of their home country

Appendix C: Pre-selection and operationalization of indicators and variables

The conceptual framework developed in chapter 3 served as basis for the pre-selection of indicators and variables. Pre-selection and operationalization of indicators and variables were conducted according to the following five rounds:

- Conceptual framework for pre-selection (described in chapter 3)
- Pre-selection criteria (described in chapter 3)
- Analysis of indicators and variables (described in chapter 3)
- Summary of indicators/variables according to conceptual framework
- Clustering of indicators
- Allocation of variables to indicators
- Operational definitions of variables

Pre-selection and operationalization of indicators and variables is particularly influenced by four evaluations of fellowships programs conducted by Aguirre International (2004) in order to evaluate the African Graduate Fellowship (AFGRAD) and Advanced Training for Leadership Program (ATLAS) (in the following referred to as ATLAS/AFGRAD program); by CIDA (2005) in order to evaluate the Canadian Francophone Scholarship Program (CFSP); by the World Bank Institute (2010) in order to evaluate the Joint Japan/World Bank Graduate Scholarship Program (JJ/WBGSP); and by the Asian Development Bank [ADB] (2007) in order to evaluate the Japan Scholarship Program (JSP).

Summary of indicators and variables

The following provides a summary of indicators and variables used during evaluation of these four fellowship programs described in chapter 2. Each cell in the following tables (C1 to C4) depicts an indicator (and related variables) used during evaluation of the respective program. For example, employment status and career were used as result indicator during evaluation of the ATLAS/AFGRAD program. Related variables to this indicator were: employment type, organizational type, level of authority and responsibility, career progression.

Summary ATLAS/AFGRAD indicators and variables

Table C1 outlines the different indicators and variables used during evaluation of the ATLAS/AFGRAD program according to the conceptual framework used in this study.

Table C1: Summary ATLAS/AFGRAD indicators and variables

Learning indicators	Behavior indicators	Result indicators
Education: degree level – field of study	Applicability of acquired knowledge, skills, and attitudes at institutions or at other areas (family, in community, in other professional relationships)	Employment status and career: employment type – organizational type – level of authority and responsibility - career progression
ATLAS/AFGRAD contribution to specific alumni Knowledge, Skills and Attitudes		Income
		Alumni activities (achievements, discoveries, contributions…)
		ATLAS/AFGRAD contribution to institutional changes, alumni association, and higher level (sectoral, national, regional or international levels) impacts

Summary CFSP indicators and variables
Table C2 outlines the different indicators and variables used during evaluation of the CFSP program according to the conceptual framework used in this study.

Table C2: Summary CFSP indicators and variables according to conceptual framework

Learning indicators	Behavior indicators	Result indicators
Effectiveness of training received	Use of knowledge, skills and attitudes in home institutions or organizations	Employment: employment status
Benefits academic training in terms of new knowledge, skills and attitudes		Career advancement: access to higher positions
		Organizational capacity building (fellows effects on their organizations performance)
		CFSP Contribution to sustainable development in recipient countries and to poverty reduction
		Diaspora (alumni who do not return to home country) contribution to the development of CFSP recipient countries.

Summary JJ/WBGSP indicators and variables
Table C3 outlines the different indicators and variables used during evaluation of the JJ/WBGSP program according to the conceptual framework used in this study. No indicator related to behavior/applicability in terms of applicability of knowledge, skills and attitudes by alumni was used.

Table C3: Summary JJ/WBGSP indicators and variables according to conceptual framework

Learning indicators	Behavior indicators	Result indicators
Degree attainment: degree obtained – place of study – field of study		Alumni personal benefits of the JJ/WBGSP program
Effectiveness and relevance of the training		Alumni residence status: rate of return
		Employment status and career: type and level employment– sector of employment – type of positions occupied
		Alumni contribution to the socio-economic development of home country or other developing country

Summary JSP indicators and variables
Table C4 outlines the different indicators and variables used during evaluation of the JSP program according to the conceptual framework used in this study. No indicator related to behavior/applicability in terms of applicability of Knowledge, skills and attitudes (KSA) by alumni was used.

Table C4: Summary JSP indicators and variables according to conceptual framework

Learning indicators	Behavior indicators	Result indicators
Education: degree attainment (level of study) – field of study – place of study		Employment status and career progression: employment of alumni - Sector of employment alumni – positions occupied by alumni - career advancement after obtaining degree - responsibilities in organization
		Alumni perceptions of the benefits received from the JSP program
		Rate of return
		Activities that alumni conduct in own countries
		JSP contribution to socioeconomic development

Clustering indicators and variables

As outlined above, each program uses different indicators and variables related to the three components of the conceptual framework used in this study. Therefore, all indicators and variables used by the four programs were first sorted per category of indicator according to conceptual framework. Each row in the following tables depicts similar indicators and variables that are used by the programs reviewed. Each row depicts therefore a certain type of indicator and variables related to this indicator.

Table C5. Learning indicators used by the four programs

ATLAS/AFGRAD	CFSP	JJ/WBGSP	JSP
Education: degree level – field of study		Degree attainment: degree obtained – place of study – field of study	Education: degree attainment (level of study) – field of study – place of study
	Effectiveness of training received	Effectiveness and relevance of the training received	
ATLAS/AFGRAD contribution to specific alumni knowledge, skills and attitudes	Benefits academic training in terms of new knowledge, skills and attitudes		

Table C6. Behavior indicators used by the four programs

ATLAS/AFGRAD	CFSP	JJ/WBGSP	JSP
Applicability of acquired knowledge, Skills and Attitudes at institutions or at other areas (family, in community, in other professional relationships)	Use of knowledge, skills and attitudes in home institutions or organizations		

Table C7. Results indicators used by the four programs

ATLAS/AFGRAD	CFSP	JJ/WBGSP	JSP
Employment status: employment type – organizational type	Employment: employment status	Employment status: type and level employment– sector of employment - -Alumni residence status (rate of return)	Employment status: employment of alumni - Sector of employment alumni – -Rate of return to home country
Career progression: level of authority and responsibility	Career advancement: access to higher positions	Career progress: type of positions occupied	Career progression: positions occupied by alumni - career advancement after obtaining degree – responsibilities in organization
Income			
		Alumni personal benefits of the JJ/WBGSP program	Alumni perceptions of the benefits received from the JSP program
Alumni activities (achievements, discoveries, contributions…)	Organizational capacity building (fellows effects on their organizations performance)		Activities that alumni conduct in own countries
ATLAS/AFGRAD contribution to institutional changes, alumni association, and higher level (sectoral, national, regional or international levels) impacts	- CFSP Contribution to sustainable development in recipient countries and to poverty reduction -Diaspora (alumni who do not return to home country) contribution to the development of CFSP recipient countries.	Alumni contribution to the socio-economic development of home country or other developing country	JSP contribution to socioeconomic development

Overview of pre-selected indicators and variables

All type of indicators (in different rows of above tables) were pre-selected for this study. To these indicators, measurable variables provided by the four programs reviewed (in rows) were allocated. However, identical indicators or variables used by the four programs are only once mentioned in order to avoid repetition and overlaps. For example the indicator employment status used by all fellowship programs reviewed as indicator of result is only once mentioned in this study. In the same line, the variable sector of employment (related to employment status) is once mentioned. Further, variables at school and classroom level used in the Scheerens' quality framework were pre-selected (see chapter 3) as outcomes variables in order to understand whether IFP alumni conduct such activities. In line with the conceptual framework used in this study, the following provides respectively learning indicators, behavior indicators, and results indicators pre-selected for the IFP Ed-instrument.

Table C8. Learning indicators and variables pre-selected for IFP Ed-instrument

Type of indicators	Variables
Education attainment	Degree obtained (level of study) – Field/kind of study – place of study
Effectiveness and relevance of training	Perception of effectiveness and relevance of training received during fellowship
Contribution of training to acquisition specific knowledge, skills, and attitudes	Perception of contribution of training received to acquisition of specific knowledge, skills, and attitudes

Table C9. Behavior indicators pre-selected for the IFP Ed-instrument

Type of indicator	Variables
Applicability of acquired knowledge, skills, and attitudes	Applicability of acquired knowledge, skills and attitudes at institutions/organizations or at other areas such as in community where alumni are active

Table C10. Results indicators pre-selected for the IFP Ed-instrument

Type of indicators	Variables
Employment status	Place of employment (rate of return/residence status) – Type of job – Sector of employment – Level of employment
Career advancement/progress	Job position – Number of promotions since graduation – Responsibility in current work – Locus of decision – Satisfaction with career
Income	Income of alumni after graduation
Conducted activities	Activities that alumni conduct after graduation: activities related to financial and resources issues, activities related to human resources issues - activities related to students' background conditions - activities related to community/parents' involvement - activities related to instructional issues
Contribution of fellowship to personal alumni	Personal benefits received by alumni from participation in fellowship program
Contribution of fellowship to alumni home country	Contribution of training received to alumni home country

Based on the conceptual framework and the selection criteria used in this study, ten core indicators and 22 variables were pre-selected in order to develop the IFP Ed-instrument. In addition to these ten indicators, one indicator (socio-biographical backgrounds of alumni) and three related variables (gender – age – marital status) were pre-selected as recommended by respondents during orientation interviews in order to international comparability and disaggregation. A total of 11 indicators and 25 variables were therefore pre-selected for the IFP Ed-instrument. The following provides the operational definitions of the pre-selected variables.

Operationalization of variables

Operationalization is the process of defining constructs (in this study: indicators and variables) in measurable terms (Neuman, 2000). One way of operationalization constructs is to build on definitions used by other scholars (Neuman, 2000, p.161) and not to start from scratch. Therefore, operationalization in this study was based on operational definitions of indicators/variables used by Aguirre International (2004) in order to evaluate the ATLAS/AFGRAD program; by CIDA (2005) in order to evaluate the CFSP program; by the World Band Institute (2010) in order to evaluate the JJ/WBGSP program; by the Asian Development Bank (ADB) (2007) in order to evaluate the JSP program. However, the operational definitions provided by these organizations were adapted to the specific context and objectives of IFP. Therefore, the following provides the operational definitions of variables pre-selected for the IFP Ed-instrument.

Table C11. Operationalization of pre-selected learning variables

Variables	Operational definition
Education degree obtained (level study)	Highest education degree obtained through participation in the IFP program
Place of study	Country and University where alumni obtained the degree
Field of study	Fields of study followed by alumni Kinds of study followed by alumni such as teacher training, curriculum development, etc.
Effectiveness and relevance of the training received by IFP alumni	IFP Alumni' perceptions on the relevance and usefulness of the knowledge and skills gained from their degree programs to the educational needs of their home country
IFP contribution to acquisition of specific knowledge, skills and attitudes by alumni	IFP Alumni perception whether any specific knowledge, skills, and attitudes have been acquired from the academic training

Table C12. Operationalization of pre-selected behavior variable

Variable	Operational definition
Applicability of acquired knowledge, skills and attitudes at institutions/organizations or at other areas such as in community where alumni are active	Alumni perception about applicability of acquired knowledge, skills, and attitudes at institutions/organizations where they immediately worked after graduation or to other areas such as within family, in community, in other professional relationships

Table C13. Operationalization of pre-selected results variables

Variables	Operational definitions
Place of employment (residence status)	Place where IFP alumni are active after graduation
Type of job	Percentage of alumni who are employed permanently of temporarily
Sector of employment	- Sector of employment of alumni (private – government – own business etc.)
Level of employment	Type of organizations where alumni are employed
Job position	Proportion of alumni occupying relevant/higher positions appropriate to education received
Number of promotions since graduation	Evidence of career advancement in terms of promotions upon returning to organization OR after graduation/due to acquired degree/knowledge, skills, attitudes
Responsibility in current work	Whether responsibilities has increased, stayed the same, or decreased since return from training
Locus of decision	Degree to which alumni are involved in decision taking processes
Satisfaction with career	Alumni satisfaction with career advancement after graduation
Income of alumni after graduation	Degree to which alumni attribute their income to participation in IFP program
Activities related to financial and resources issues,	Whether alumni conduct financial and material resources activities such as providing equipment to classrooms
Activities related to human resources issues	Whether alumni conduct human resources activities such as providing training to teachers
Activities related to students' background conditions	Whether alumni conduct activities related to students' background conditions such as promoting the use of local language in schools
activities related to community/parents' involvement	Whether alumni conduct activities that promote community/parents' involvement in schools
activities related to instructional issues	Whether alumni conduct activities that address instructional issues such as the use of new teaching methods during instructions
Personal benefits received by alumni from participation in fellowship program	Alumni's perception of the personal benefits of the program in terms of skills recognition, progression in the same job, mobility across jobs, and higher income
Contribution of training received to alumni home country	Alumni perception about educational changes caused to educational sector at institutional, community or national levels that can be attributed to their participation in IFP program

Table C14. Operationalization of pre-selected socio-biographical variables

Variables	Operational definitions
Gender	Whether alumni is male or female
Age	Age (in years) of alumni
Marital status alumni	Whether alumni is single or married

Appendix D: Interview guide

1- As specialist/expert in evaluation, in which developing country/countries have you ever conducted impact assessment studies OR evaluation studies?

2- Do you have any experience with the assessment of the impact of fellowships programs or any comparable program?

3- Do you see any similarities between this program and the IFP? Could you please elaborate? What are the differences?

4- Which kind of impact indicators are you familiar with? Could you mention some example of indicators that you already used (or have developed)?

5- In your opinion, are such indicators appropriate to assess the impact of fellowships programs in developing countries? Could you please elaborate?

6- In your opinion, are indicators included in existing education data sources such as UNESCO, OECD … (e.g. enrolment rate – drop-out etc.) appropriate to assess the impact of fellowships programs in the area of education in developing countries? Could you please elaborate?

7-As an expert in impact assessment, which kind of indicators do you hold necessary to be included in an instrument in order to assess the impact of a fellowship program in the area of education?

8- How these impacts indicators could be measured?

9- Suppose you were me, how would you proceed to develop the instrument that could be used to assess the impact of IFP?

10- How could the validity and reliability of this instrument be established?

11- How could control groups be used in order to find a counterfactual (in order to address attribution issues)?

12- You have been very helpful. Are there other things you would like to share with us to help us understand how to develop an appropriate instrument for the IFP program? Anything at all you would like to add?

Part 2: Questions to experts in the area of education

1- As expert in education, in which developing countries have you ever conducted educational activities?

2- Which kind of education indicators are you familiar with? Could you mention some example of such indicators?

3- In your opinion, are such indicators appropriate to assess the impact of fellowships programs in developing countries in the area of education? Could you please elaborate?

4- In your opinion, are indicators included in existing education data sources such as UNESCO, OECD … (e.g. enrolment rate – drop-out etc.) appropriate to assess the impact of fellowships programs in the area of education in developing countries? Could you please elaborate?

5- As an expert in education in developing countries, which kind of indicators do you hold necessary to be included in an instrument in order to measure the impacts of a fellowship program in the area of education in developing countries?

6- How these impacts indicators could be measured?

7- You have mentioned some outcome and impact indicators above. Where (in which sources) could data on such indicators be found?

8- Suppose you were me, how would you proceed to develop an instrument that could be used to assess the impact of IFP in the area of education?

9- How could the validity and reliability of this instrument be established?

10- You have been very helpful. Are there other things you would like to share with us to help us understand how to develop an appropriate toolbox for the IFP program? Anything at all you would like to add?

Appendix E: Questionnaire 1

International Fellowship Program (IFP)

The Ford Foundation International Fellowships Program (IFP) is an educational aid program that provides fellowships for graduates who lack access to post-graduate education and who are resident of one of the countries or territories selected by the program. IFP was launched in 2001. To date the program has provided fellowships to more than 4300 fellows from 22 developing countries. IFP fellows are selected among graduates from marginalized and disadvantaged communities. Fellows may apply at any University worldwide including Universities of the country of origin and they may choose to study in any academic discipline or field of study related to the following three subject areas: asset building and community development - peace and social justice - knowledge, creativity and freedom. With its program, IFP aims to enhance the fellows' capabilities in order to become change agents in their countries of origin.

This study focuses on IFP alumni graduated in the area of education. The aim of this study is to develop an instrument that could measure the outcomes and impacts of the International Fellowships Program (IFP) in its selected countries in the area of education.

Outcomes refer to activities that alumni or network of alumni conduct in the area of education in their countries and i**mpacts** refer to changes caused by the conducted activities in the area of education in the countries.

Instructions

We kindly ask you to assess the **relevance** and the **measurability** of the different indicators and variables in the questionnaire on the next pages.

With **relevance** we ask if the indicators/variables could provide a representative picture of the outcomes and impacts of IFP in selected countries AND whether the indicators/variables are appropriate to measure the outcomes and impacts of IFP.

With **measurability** we ask whether it will be easy (or not) to collect data on the indicator/variable at reasonable cost OR whether it will be possible to find data on the indicator/variable in international or national secondary sources in selected countries.

For each indicator:

- Please indicate first (*in the first cell*) how, in your opinion, the pre-selected indicator is relevant to measure the outcomes and impacts of IFP.

- Second, please indicate how, in your opinion, the related pre-selected variables are relevant to measure the outcomes and impacts of IFP.

- Third, please indicate whether the related pre-selected variables are measurable.

- Fourth, please indicate whether the variables are essential (or not) to be included in the final instrument.

- Finally, please suggest, if possible, relevant and measurable indicators/variables that are not included in the pre-selected list of indicators/variables. Please also indicate how relevant and measurable they are.

Name:
Organization:
Professional background:

INDICATORS AND VARIABLES
ALUMNI QUESTIONNAIRE

Please use the following answer categories for relevance and measurability:

1= Not at all relevant/measurable
2= Somewhat relevant/measurable
3= Relevant/measurable
4= Very relevant/measurable
5= Don't know

INDICATOR 1 : DEMOGRAPHY CHARACTERISTICS

Not at all relevant — Very relevant

1	2	3	4

Variables	Relevance					Measurable					To be included in final instrument?		
	Not at all relevant			Very relevant	Don't know	Not at all measurable			Very measurable	don't know	Yes	No	Don't know
	1	2	3	4	5	1	2	3	4	5			
1.1 Gender													
1.2 Age	1	2	3	4	5	1	2	3	4	5			
1.3 Marital status (single-married)	1	2	3	4	5	1	2	3	4	5			
Other (please specify): ………………	1	2	3	4	5	1	2	3	4	5			

INDICATOR 2 : EDUCATIONAL BACKGROUNDS

Not at all relevant — Very relevant

1	2	3	4

Variables	Relevance					Measurable					To be included in final instrument?		
	Not at all relevant			Very relevant	Don't know	Not at all measurable			Very measurable	don't know	Yes	No	Don't know
	1	2	3	4	5	1	2	3	4	5			
2.1 Education degree obtained													
2.2 Place of study (own country or elsewhere)	1	2	3	4	5	1	2	3	4	5			
2.3 Kind of study (teacher training, curriculum development …)	1	2	3	4	5	1	2	3	4	5			
Other (please specify): …………….	1	2	3	4	5	1	2	3	4	5			

<table>
<tr><td colspan="3" align="center">INDICATOR 3 : EMPLOYMENT
Not at all relevant   Very relevant
1   2   3   4</td><td></td></tr>
</table>

Variables	**Relevance**	**Measurable**	**To be include in final instrument?**
4.1 Type of job (permanent or temporary)	Not at all relevant Very relevant Don't know 1 2 3 4 5	Not at all measurable Very measurable don't know 1 2 3 4 5	Yes No Don't know
4.2 Place of employment (e.g. in own country or elsewhere)	1 2 3 4 5	1 2 3 4 5	
4.3 Kind of employment sector (e.g. public, private …)	1 2 3 4 5	1 2 3 4 5	
4.4 Level of employment (e.g. government level, local level …)	1 2 3 4 5	1 2 3 4 5	
4.5 Applicability knowledge/skills (degree to which alumni can apply knowledge/skills)	1 2 3 4 5	1 2 3 4 5	
Other (please specify): …………….	1 2 3 4 5	1 2 3 4 5	

<table>
<tr><th colspan="4">INDICATOR 4 : CAREER PROGRESS</th></tr>
</table>

	Not at all relevant		Very relevant	
	1	2	3	4
	☐	☐	☐	☐

Variables	**Relevance**					**Measurable**					**To be included in final instrument?**		
3.1 Income/salary	Not at all relevant			Very relevant	Don't know	Not at all measurable			Very measurable	don't know	Yes	No	Don't know
	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
3.2 Number of promotions since graduation (see definition below)	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
3.3 Job position	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
3.4 Responsibility in current work	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
3.5 Locus of decision (degree to which alumni are involved in decision taking processes)	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
3.6 Alumni' career satisfaction	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			
Other (please specify): ……………	1	2	3	4	5	1	2	3	4	5	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐			

- Within this study **promotion** refers to one of the following: significant increase in scope of responsibilities, significant increases in annual salary, changes in level in the employing company, changes in offices or office décor, and becoming eligible for bonuses, incentives, or stock plans" (Whitely et al., 1991 p.337)

INDICATOR 5: TYPE OF CONDUCTED ACTIVITIES IN THE AREA OF EDUCATION			

Not at all relevant   Very relevant

1   2   3   4

Variables	Relevance	Measurable	To be included in final instrument?
5.1 Activities related to **financial and resources** issues	Not at all relevant Very relevant Don't know 1 2 3 4 5	Not at all measurable Very measurable don't know 1 2 3 4 5	Yes No Don't know
5.2 Activities related to **human resources** issues	1 2 3 4 5	1 2 3 4 5	
5.3 Activities related to **students' background conditions**	1 2 3 4 5	1 2 3 4 5	
5.4 Activities related to **community/parents' involvement**	1 2 3 4 5	1 2 3 4 5	
5.5 Activities related to **instructional** issues	1 2 3 4 5	1 2 3 4 5	
Other (please specify): …………….	1 2 3 4 5	1 2 3 4 5	

- **Financial and material resources** at system and school levels refer among other things to factors such as student' educational expenditure, classroom equipment, and other school supplies (textbooks, pencil and paper etc.) (Scheerens, 2004).

- **Human resources** refer to variables related to teachers. The variables include among other: teacher professional knowledge and skills (e.g. knowledge about pedagogical and didactical strategies); teacher working conditions (e.g. average class-size); teacher morale and status (e.g. opinions about career and job mobility); teacher salaries etc. (Scheerens, 2004).

- **Students' background conditions** refer to contextual conditions in the home situation of students. Examples of such conditions include: differences in language spoken at home and in the school; long absences from school because of students being obliged to work or because of illness; students arriving at school without having had a meal etc. (Scheerens, 2004).

- **Community/parents' involvement** refers to the degree to which community/parents are involved in schools (Scheerens, 2004).

- **Instructional issues** refer to issues such as: the effective use of instructional time; the use of new teaching methods during instructions; the quality of instructional material; the quality of instructional leadership (practices that promote and support teaching and learning) etc. (Scheerens, 2004).

INDICATOR 6: CHANGES CAUSED BY CONDUCTED ACTIVITIES

	Not at all relevant		Very relevant	
	1	2	3	4

Variables	Relevance					Measurable					To be included in final instrument?		
	Not at all relevant			Very relevant	Don't know	Not at all measurable			Very measurable	don't know	Yes	No	Don't know
	1	2	3	4	5	1	2	3	4	5			
6.1 Alumni' perceived impact (changes caused by conducted activities according to alumni)													
Other (please specify): …………….	1	2	3	4	5	1	2	3	4	5			

Thank you very much for filling out this questionnaire.

Appendix F: Questionnaire 2

International Fellowship Program (IFP)

The Ford Foundation International Fellowships Program (IFP) is an educational aid program that provides fellowships for graduates who lack access to post-graduate education and who are resident of one of the countries or territories selected by the program. IFP was launched in 2001. To date the program has provided fellowships to more than 4300 fellows from 22 developing countries. IFP fellows are selected among graduates from marginalized and disadvantaged communities. Fellows may apply at any University worldwide including Universities of the country of origin and they may choose to study in any academic discipline or field of study related to the following three subject areas: asset building and community development - peace and social justice - knowledge, creativity and freedom. With its program, IFP aims to enhance the fellows' capabilities in order to become change agents in their countries of origin.

This study focuses on IFP alumni graduated in the area of education. The aim of this study is to develop an instrument that could measure the outcomes and impacts of the International Fellowships Program (IFP) in its selected countries in the area of education.

Outcomes refer to activities that alumni or network of alumni conduct in the area of education in their countries and i**mpacts** refer to changes caused by the conducted activities in the area of education in IFP countries.

Instructions

We kindly ask you to assess the **relevance** and the **measurability** of the different indicators/variables in the questionnaire on the next pages.

With **relevance** we ask if the indicators/variable could provide a representative picture of the outcomes and impacts of IFP in selected countries AND whether the indicators/variables are appropriate to measure the outcomes and impacts of IFP.

With **measurability** we ask whether it will be easy (or not) to collect data on the indicator/variable at reasonable cost OR whether it will be possible to find data on the indicator/variable in international or national secondary sources in the selected countries.

- Please indicate first how, in your opinion, the variables are relevant to measure the outcomes and impacts of IFP in selected countries.
- Second, please indicate whether the variables are measurable.
- Third, please indicate whether the indicators/variables are essential (or not) to be included in the final instrument.

Name:
Organization:
Professional background:

INDICATORS AND VARIABLES
ALUMNI QUESTIONNAIRE

Please use the following answer categories for relevance and measurability:

1= Not at all relevant/measurable
2= Somewhat relevant/measurable
3= Relevant/measurable
4= Very relevant/measurable
5= Don't know

INDICATOR 1 : DEMOGRAPHY CHARACTERISTICS			
Variables	**Relevance**	**Measurable**	**To be included in final instrument?**
1.1 Religion of alumni	Not at all relevant — Very relevant — Don't know: 1 2 3 4 5	Not at all measurable — Very measurable — don't know: 1 2 3 4 5	Yes No Don't know
1.2 Country of birth and residence of alumni	1 2 3 4 5	1 2 3 4 5	

- Religion refers to whether alumni is Christian, Muslim, or have other religion

- Country of birth and country of residence refers to the country where alumni are born and to countries where they are currently active

INDICATOR 2 : EDUCATIONAL BACKGROUNDS			
Variables	**Relevance**	**Measurable**	**To be included in final instrument?**
2.1 Year of graduation	Not at all relevant — Very relevant — Don't know: 1 2 3 4 5	Not at all measurable — Very measurable — don't know: 1 2 3 4 5	Yes No Don't know

- Year of graduation refers to the year when alumnus/alumna obtained his/her degree

INDICATOR 3 : EMPLOYMENT			
Variables	**Relevance**	**Measurable**	**To be include in final instrument?**
3.1 Function description	Not at all relevant — Very relevant — Don't know: 1 2 3 4 5	Not at all measurable — Very measurable — don't know: 1 2 3 4 5	Yes No Don't know
3.2 Applicability knowledge/skills	1 2 3 4 5	1 2 3 4 5	

- Function refers to the function occupied by alumni after graduation within institution or organization

- Applicability knowledge/skills: degree to which alumni can apply knowledge/skills

INDICATOR 4 : CAREER PROGRESS							
Variables	**Relevance**			**Measurable**			**To be included in final instrument?**
4.1 Income/salary	Not at all relevant	Very relevant	Don't know	Not at all measurable	Very measurable	don't know	Yes — No — Don't know
	1 2 3 4 5			1 2 3 4 5			
4.2 Number of promotions since graduation	1 2 3 4 5			1 2 3 4 5			

Revised definitions:

- Income/salary refers to the percentage increase in salary due to participation in IFP program

- Within this study **promotion** refers to one of the following: significant increase in scope of responsibilities, significant increases in annual salary, changes in level in the employing company, and becoming eligible for bonuses, incentives, or stock plans.

INDICATOR 5: TYPE OF CONDUCTED ACTIVITIES IN THE AREA OF EDUCATION							
Variables	**Relevance**			**Measurable**			**To be included in final instrument?**
5.1 Educational management activities conducted by alumni	Not at all relevant	Very relevant	Don't know	Not at all measurable	Very measurable	don't know	Yes — No — Don't know
	1 2 3 4 5			1 2 3 4 5			
5.2 Voluntary activities conducted by alumni	1 2 3 4 5			1 2 3 4 5			
5.3 Professional/operational activities conducted by alumni	1 2 3 4 5			1 2 3 4 5			

- Educational management activities refers to management (such as financial, human resources …) or policy related activities that alumni conduct in decision making positions

- Voluntary activities refers to voluntary activities that alumni conduct in communities such as education awareness campaigns, educational lobby etc.

- Professional/operational activities refer to all activities not related to management/policy activities and voluntary activities: e.g. teaching − research − curriculum − instructional material − consultancy activities.

INDICATOR 6: CHANGES CAUSED BY CONDUCTED ACTIVITIES			
Variables	**Relevance**	**Measurable**	**To be included in final instrument?**
6.1 Alumni' perceived impact	Not at all relevant Very relevant Don't know 1 2 3 4 5	Not at all measurable Very measurable don't know 1 2 3 4 5	Yes No Don't know

- Alumni perceived impact refers to alumni perceptions about changes caused by their conducted activities in institutions/organizations, communities and countries in the area of education and perceptions about impact of IFP on their personal life.

INDICATOR 7: ALUMNI SATISFACTION			
Variables	**Relevance**	**Measurable**	**To be included in final instrument?**
7.1 Alumni satisfaction with IFP	Not at all relevant Very relevant Don't know 1 2 3 4 5	Not at all measurable Very measurable don't know 1 2 3 4 5	Yes No Don't know

- Alumni satisfaction with IFP refers to satisfaction of alumni with the IFP program from selection till employment after graduation

Thank you very much for filling out this questionnaire.

Appendix G: Selected indicators/variables – data collection methods - items

The following provides successively the list of selected variables, the operational definitions of selected variables and data collection methods that could be used to gather data on these variables, and items related to selected variables used to develop the IFP Ed-instrument.

A-List of selected indicators and variables

Seven core indicators and 21 variables were selected for the IFP Ed-instrument

Table G1. Indicators and variables selected for the IFP Ed-instrument

Indicators	Variables
Indicator 1: Socio-biographical backgrounds	- Gender - Age - Country of birth and residence
Indicator 2: Educational backgrounds	- Education degree obtained - Place/country of study - Kind/content of study - Year of graduation
Indicator 3: Employment	- Place of employment - Sector of employment - Kind of employment institution - Applicability knowledge/skills - Function description
Indicator 4: Career progress	- Numbers of promotions - Job position - Responsibility in current work - Locus of decision
Indicator 5: Type of activity	- Educational management activities - Voluntary activities - Professional/operational activities
Indicator 6: Change caused by conducted activities	- Alumni' perceived impact
Indicator 7: Alumni satisfaction	- Alumni' satisfaction with IFP

B- Operational definitions selected variables and data collection methods

The tables below provide the operational definitions of the selected variables per category of indicator. For each table, the first column provides the variables, the second column includes the operational definition, and the last column provides the methods that could be used according to scholars in order to collect data related to the indicator.

The data collection methods follow recommendations provided by Kirkpatrick (1996); by the UN (Rotem et al., 2010); by Aguirre International (2004); by CIDA (2005); by World Bank Institute (2010); by the Asian Development Bank (ADB) (2007); and by DANIDA (DFC, n.d.).

Table G2. Operational definitions socio-biographical variables and data collection methods

Variables	Operational definition	Data collection methods
Gender	Whether alumni is male or female	- Review of conducted formative evaluations (e.g. review of data gathered by CHEPS)
Age	Age (month and years) of alumni	
Country of birth and country of residence	Country where alumni are born and countries where they are currently active	- Tracer study with survey questionnaire (or interviews)

Table G3. Operational definitions educational backgrounds variables and data collection methods

Variables	Operational definition	Data collection methods
Education degree obtained (level study)	Highest education degree obtained through participation in the IFP program	- Use of a control group if feasible or before and after training approach - Tracer study with survey questionnaire (or interviews)
Place of study	Country and University where alumni obtained the degree	- Review of conducted formative evaluations (e.g. review of data gathered by CHEPS)
Kind/content of study	Fields of study followed by alumni Kinds of study followed by alumni such as teacher training, curriculum development, etc.	- Review of records of fellowship completion -Testing of knowledge and skills integral to teaching and learning program - Assessment of ability to practice selected skills and apply particular knowledge in different settings
Year of graduation	year when alumnus/alumna obtained his/her degree	-Baseline assessment where fellows indicate their perceived competence level

Table G4. Operational definitions employment variables and data collection methods

Variable	Operational definition	Data collection methods
Place of employment	Place where IFP alumni are active after graduation	- Use of control groups if feasible or before and after training approach - Review of conducted formative evaluations (e.g. review of data gathered by CHEPS) - Tracer study with survey questionnaire (or interviews) - 360 degrees survey of alumni and coworkers using questionnaires and/or structured interviews and focus group discussion - Observation of job performance - Assessment of ability to practice selected skills and apply particular knowledge in different settings - Case studies of institutional success and failure - Survey or interview of trainees, trainees' bosses, trainees subordinate and others "who often observe trainees' behavior on the job" - Country site visits that includes alumni group meetings, individual interviews with alumni, individual interviews with alumni supervisors who could give impressions of performance before and after receiving training - Local institutions/organizations visit to meet decision-makers who might be able to comment on whether changes attributable to training occurred after alumni' return
Sector of employment	Sector of employment of alumni (public, private, own business etc.)	
Kind of employment institution	Type of organizations where alumni are employed	
Applicability of acquired knowledge, skills and attitudes at institutions/organizations or at other areas such as in community where alumni are active	Alumni perception about applicability of acquired knowledge, skills and attitudes at institutions/organizations where they immediately worked after graduation or to other areas such as within family, in community, in other professional relationships	
Function description	Function occupied by alumni after graduation within institution or organization	

Table G5. Operational definitions career progress variables and data collection methods

Variable	Operational definition	Data collection methods
Numbers of promotions	Evidence of promotions after graduation due to acquired degree/knowledge, skills, attitudes	- Use of control groups if feasible or before and after training approach - Survey or interview of trainees, trainees' bosses, trainees subordinate and others "who often observe trainees' behavior on the job" - Review of conducted formative evaluations (e.g. review of data gathered by CHEPS)
Job position	Proportion of alumni occupying relevant/higher positions appropriate to education received	- Tracer study with survey questionnaire (or interviews) - Review of personnel files concerning career progression and other HRM data
Responsibility in current work	Whether responsibilities has increased, stayed the same, or decreased since return from training	- 360 degrees survey of alumni and coworkers using questionnaires and/or structured interviews and focus group discussion - Tracer study with survey questionnaire (or interviews) based on the developed IFP Ed-instrument - Country site visits that includes alumni group meetings, individual interviews with alumni, individual interviews with alumni supervisors who could give impressions of performance before and after receiving training
Locus of decision	Degree to which alumni are involved in decision taking processes	

Table G6. Operational definitions type of activity variables and data collection methods

Variables	Operational definition	Data collection methods
Educational management activities	Evidence of educational management activities (e.g. financial, human resources etc.) or policy related activities that alumni conduct in decision making positions	- Tracer study with survey questionnaire (or interviews) - Review of conducted formative evaluations (e.g. review of data gathered by CHEPS) - Case studies of institutional success and failure - 360 degree survey of fellows and co-workers using questionnaires and/or structured interviews and focus group discussions - Meta-analysis
Voluntary activities	Evidence of voluntary activities that alumni conduct in communities such as education awareness campaigns, educational lobby etc.	- Country site visits that includes alumni group meetings, individual interviews with alumni, individual interviews with alumni supervisors who could give impressions of performance before and after receiving training
Professional/operational activities	Evidence of professional/operational activities (activities not related to management and voluntary activities) that alumni conduct such as teaching – research – curriculum – instructional material – consultancy activities.	- Internet search for alumni impact in countries not visited by assessment team (see Aguirre International (2004) for more details about this method)

Table G7. Operational definitions change caused by conducted activities variables and data collection methods

Variables	Operational definition	Data collection methods
Alumni' perceived impact	Alumni's perception of the personal benefits of the IFP program (in terms of skills recognition, progression in the same job, mobility across jobs, and higher income) AND Alumni perception about educational changes caused to educational sector at institutional, community or national levels that can be attributed to their participation in IFP program	- Use control groups if feasible or before and after training approach - Tracer study with survey questionnaire (or interviews) - Review of conducted formative evaluations (e.g. review of data gathered by CHEPS) - Consultation with relevant stakeholders - Case studies of institutional success and failure - 360 degree survey of fellows and co-workers using questionnaires and/or structured interviews and focus group discussions - Meta-analysis - Country site visits that includes alumni group meetings, individual interviews with alumni, individual interviews with alumni supervisors who could give impressions of performance before and after receiving training - Local institutions/organizations visit to meet decision-makers who might be able to comment on whether changes attributable to training occurred after alumni' return - Internet search for alumni impact in countries not visited by assessment team (see Aguirre International (2004) for more details about this method)

Table G8. Operational definitions alumni satisfaction variables and data collection methods

Variables	Operational definition	Data collection methods
Alumni' satisfaction with IFP	Degree to which alumni are satisfied with IFP program since the whole process from selection till employment after graduation	- Tracer study with survey questionnaire (or interviews) - Review of conducted formative evaluations (e.g. review of data gathered by CHEPS) - Country site visits that includes alumni group meetings, individual interviews with alumni, focus group discussions

C- Items pre-selected for the IFP Ed-instrument

Development of the initial items was based on a review of relevant literature related to education, fellowships programs and career development; on different recommendations made by respondents during orientation interviews; and on recommendations made by respondents during the Delphi-rounds.

Literature explored in this operationalization step include among others: Scheerens et al. (2003) (e.g. items related to locus of decisions and for the variables educational

management/policy activities) – Aguirre International (2004), CIDA (2005), World Bank Institute (2010), ADB (2007), Strömbom (1989), Teichler (1991) and Norad (2005) (e.g. employment items) - Edelfelt & Reiman (2004) (e.g. items related to education institutions, functions in educations – kind/content of study) – Kottmann & Enders (2009) and Norad (2005) (e.g. items related to responsibility and applicability acquired knowledge) – Enders & de Boer (2002) (e.g. career satisfaction items). Literature about career progress (e.g. Turban & Dougherty, 1994; Whitely et al., 1991; Waine et al., 1999; van der Sluis & Poell 2003; Heslin, 2005 etc.) were explored in order to find items that could be used for variables such as promotions and alumni satisfaction etc. Further, in order to take into account former evaluations conducted by CHEPS, appropriate items from the formative evaluation instruments were pre-selected. Examples include: items related to the variables education degree, place of study, place of employment and sector of employment (Enders & de Boer (n.d.). The following tables provide the items related to variables selected for the IFP Ed-instrument.

Table G9. Items pre-selected for socio-biographical backgrounds variables

Variables	Items
Gender	Whether alumni are male or female
Age	Month and year of birth alumni
Country of birth and residence	Alumni Country of birth and country of current residence

Table G10. Items pre-selected for educational backgrounds variables

Variables	Items
Education degree obtained through IFP	Whether alumni have obtained one of the following degrees: Master degree, Doctorate/PhD degree, or other degree
Place/country of study	Whether alumni have studied in own country, abroad in own continent or abroad in another continent
Kind/content of study	Whether alumni have done one or more of the following studies in the IFP program: Curriculum development - Education evaluation and monitoring - Education assessment - Special education- Instructional material development - Teacher training - Education management - E-learning - Media application in education - Human resources in education - Accreditation of education
Year of graduation	Year alumni have been graduated

Table G11. Items pre-selected for employment variables

Variables	Items
Place of employment	Whether alumni are employed: In home country (rural/remote area in own community) - In home country (urban area in own community) - In home country (rural/remote area NOT in own community) - In home country (urban area NOT in own community) - In the host country of postgraduate studies - In another country (not home country or host country)
Sector of employment	Whether alumni are employed in public sector, private sector or are self-employed
Kind of employment institution	Whether alumni are employed in one of the following institutions: Ministry of education – Parliament - Provincial/regional education institution - Sub-regional or inter-municipal education institution - Local education institution - Tertiary/higher institution - Schools (primary-secondary-vocational) - International/inter-governmental institution - Local/international NGO – self-employed
Applicability knowledge/skills acquired through IFP	Extent to which alumni is able to use the acquired knowledge/skills through IFP in the professional or voluntary work setting
Function description	Teaching in secondary education - Teaching in secondary vocational/technical education - Teaching in higher/tertiary education - Research in higher/tertiary education - Instructional material development - Curriculum development and improvement - Education quality control - School management - Instructional management - Human resources management - Finance management - Higher education institution management - Education assessment - Education monitoring and evaluation - Education lobby - Education research and development - Education policy development and implementation - Strategic education management

Table G12. Items pre-selected for career progress variables

Variables	Items
Numbers of promotion	Number of promotions alumni receive in the area of education following participation in the IFP program?
Job position/title	Job position/title of alumni before and following participation in the IFP program
Responsibility in work	Extent to which alumni have more responsibility in organization/institution following participation in the IFP program
Locus of decision	Extent to which alumni were/are involved in the decision making process within organization/institution before and following participation in the IFP program

Table G13. Items pre-selected for type of activity variables

Variables	Items
Educational management activities	Whether alumni conduct one or more of the following activities: Strategic overall management - Financial management - Material resources management - Human resources management - Quality control management - Pedagogical/instructional management
Voluntary activities	Whether alumni conduct one or more of the following activities: Education lobby in own community/country - Education awareness campaign (sensitization of people/parents) in own community - Promotion of the importance of education particularly in rural/remote areas - Promotion of higher education - Gender/minorities' advocacy in the area of education
Operation/professional Activities	Whether alumni conduct one or more of the following activities: Pedagogical activities - Didactical activities - Research activities - Curriculum development activities - Instructional material development activities - Consultancy/advisory activities

Table G14. Items pre-selected for change caused by conducted activities variable

Variable	Items
Alumni' perceived impact	Extent to which alumni think that participation in the IFP program has changed their personal lives AND extent to which alumni think that their conducted activities have brought about changes in the communities/countries where they are active

Table G15. Items pre-selected for alumni satisfaction

Variable	Items
Alumni' satisfaction with IFP	Extent to which alumni are satisfied with participation in IFP program from the selection process till employment after graduation

Based on these items an initial IFP Ed-instrument was developed and submitted to expert for evaluation during the third step.

Appendix H: Questionnaire 3

Purpose:
The aim of this questionnaire 3 is to evaluate the overall instrument (developed after pre-selection of items) for clarity of items and completeness (identification of items that need to be deleted or to be added). The aim of the instrument is to assess the outcomes and impacts of IFP in the area of education.

Definitions:
Outcomes refer to activities (professional/voluntary) that alumni or network of alumni conduct in the area of education in their countries and that are influenced by the knowledge/skills acquired during the IFP program.
Impacts refer to changes caused by the conducted activities in the area of education in the communities/countries or in the organization/institution where alumni are active.

Instructions

Please find below the instrument that could in our opinion measure the outcomes and impacts of IFP. We kindly ask you to read all items included in the instrument and to provide an **overall assessment** of the whole instrument (and **not to fill in** the instrument).
After reading all items:
- Please indicate (by highlighting in yellow), in your opinion, what items should be excluded from the instrument.

- Please indicate, in your opinion, what is missing in the instrument:
………
………
- Please indicate to what extent the instrument will contribute in measuring the outcomes and impacts of IFP in its selected countries and to what extent you think the instrument is valid.

1= Not at all/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5=Agree to a very high extent

	Not at all to a very high extent 1 2 3 4 5	Don't Know
The instrument contributes in measuring the **outcomes** of IFP		
The instrument contributes in measuring the **impacts** of IFP	1 2 3 4 5	
The instrument is valid	1 2 3 4 5	

Thank you very much for filling out this questionnaire.

IMPACT
OF
FORD FOUNDATION INTERNATIONAL
FELLOWSHIPS PROGRAM (IFP)
IN
THE AREA OF EDUCATION

OUTCOME AND IMPACT
QUESTIONNAIRE

ALUMNI QUESTIONNAIRE

Definitions:

Outcomes refer to activities (professional/voluntary) that alumni or network of alumni conduct in the area of education in their countries and that are influenced by the knowledge/skills acquired during the IFP program.

Impacts refer to changes caused by the conducted activities in the area of education in the communities/countries or in the organization/institution where alumni are active.

Instructions:

● The alumni questionnaire is composed of five parts:
- Part 1 includes questions related to alumnus/alumna socio-biographical and educational backgrounds. The questions are to be filled out by all respondents.
- Part 2 includes questions related to the employment of alumnus/alumna and the work environment within which he/she is active.
- Part 3 includes questions related to activities (professional/voluntary) conducted by alumnus/alumna in his/her institution or community/country in the area of education.
- Part 4 includes questions related to the impact of IFP in the institution or community/country where alumnus/alumna is active.
- Part 5 includes questions related to the satisfaction of alumnus/alumna with his/her participation in the IFP program and his/her career.

● It is recommended that all items in the questionnaire applicable to the respondent should be answered.

● It is also recommended to fill out the tables according to the following example:

> In which country have you obtained your education degree through the IFP program?
> *Please check (**in the first column**) the place applicable.*

	In own country
	Abroad in own continent *Please specify country*: …………………………………………………..
X	Abroad in another continent *please specify country*: Netherlands

The example above shows how the question should be answered if alumnus/alumna has studied abroad but in another continent.

PART 1: SOCIO-BIOGRAPHICAL AND EDUCATIONAL BACKGROUNDS

1 - SOCIO-BIOGRAPHICAL BACKGROUND ALUMNUS/ALUMNA

1.1 Family name:

………………………………………………………………………………………...

1.2 Given name(s):

………………………………………………………………………………………...

1.3 Current mailing address:

………………………………………………………………………………………...

1.4 Current e-mail address (es):

………………………………………………………………………………………...

1.5 Gender:

	Female
	Male

1.6 Month/Year of birth:

[ ][ ] / 19 [ ][ ]

1.7 Country of birth
………………………………………………………………………………………..

1.8 Country of current residence
………………………………………………………………………………………...

2- EDUCATION THROUGH THE IFP PROGRAM

2.1 What is your education degree obtained through the IFP program?
*Please check (**in the first column**) the degree applicable **AND** give information on the year of graduation.*

	DEGREE OBTAINED	YEAR GRADUATION
	2.1.1 Master degree	……………………………..
	2.1.2 Doctorate/PhD degree	……………………………..
	2.1.3 Other *Please specify degree:* …………………………	…………………………………

133

	2.2.1 In own country
	2.2.2 Abroad in own continent *Please specify country:* ……………………………………………………
	2.2.3 Abroad in another continent *Please specify country:* …………………………………………………

2.3 Please, provide information on the content of the study you have done in the area of education through the IFP program.
*Please check (**in the first column**) all studies applicable (more than one choice possible).*

	2.3.1 Curriculum development
	2.3.2 Education evaluation and monitoring
	2.3.3 Education assessment
	2.3.4 Special education
	2.3.5 Instructional material development
	2.3.6 Teacher training
	2.3.7 Education management
	2.3.8 E-learning
	2.3.9 Media application in education
	2.3.10 Human resources in education
	2.3.11 Accreditation of education
	2.3.12 Other *Please specify:* ……………………………………………………..

PART 2: EMPLOYMENT AND WORK ENVIRONMENT

The aim of the following questions is to get a better picture of your employment and work environment before and after participation in the IFP program.

3. PLACE OF EMPLOYMENT BEFORE AND AFTER PARTICIPATION IN IFP

3.1 Were you employed **BEFORE** participation in the IFP program?
*Please check (**in the first column**) all items applicable.*

	3.1.1 Yes
	3.1.2 No
	3.1.3 Other
	Please specify: ……………………………………………………………………..

3.2 Please, provide information on the place of your employment **BEFORE** participation in the IFP program.
*Please check (**in the first column**) all items applicable.*

	3.2.1 In home country (**rural/remote area** in own community)
	3.2.2 In home country (**urban area** in own community)
	3.2.3 In home country (**rural/remote area** NOT in own community)
	Please specify area: …………………………………………………………..
	3.2.4 In home country (**urban area** NOT in own community)
	Please specify area: …………………………………………………………..
	3.2.5 In another country
	Please specify country: …………………………………………………………
	3.2.6 Other
	Please specify: ……………………………………………………………………..

3.3 Are you currently employed (have been employed) **FOLLOWING** participation in the IFP program.
*Please check (**in the first column**) all items applicable.*

	3.3.1 Yes
	3.3.2 No
	3.3.3 Other
	Please specify: ……………………………………………………………………..

3.4 Please, provide information on the place of your employment **FOLLOWING** participation in the IFP program.
*Please check (**in the first column**) in the first column all items applicable.*

	3.4.1 In home country (**rural/remote area** in own community)
	3.4.2 In home country (**urban area** in own community)
	3.4.3 In home country (**rural/remote area** NOT in own community) *Please specify area*: ……………………………………………………………
	3.4.4 In home country (**urban area** NOT in own community) *Please specify area*: ……………………………………………………………
	3.4.5 In the host country of postgraduate studies *Please specify country*: …………………………………………………………
	3.4.6 In another country (not home country or host country) *Please specify country*: ……………………………………………………..........
	3.4.7 Other *Please specify*: …………………………………………………………………

4. JOB POSITION/TITLE AND FUNCTION BEFORE AND AFTER PARTICIPATION IN IFP

4.1 Please, indicate your job position/title **BEFORE** participation in the IFP program.

…………………………………………………………………………………………………...

4.2 Please, provide information on your function within your organization/institution **BEFORE** participation in the IFP program.
*Please check (**in the first column**) all items applicable.*

	4.2.1 Teaching in secondary education
	4.2.2 Teaching in secondary vocational/technical education
	4.2.3 Teaching in higher/tertiary education
	4.2.4 Research in higher/tertiary education
	4.2.5 Instructional material development
	4.2.6 Curriculum development and improvement
	4.2.7 Education quality control
	4.2.8 Secondary school management
	4.2.9 Instructional management
	4.2.10 Human resources management
	4.2.11 Finance management
	4.2.12 Higher education institution management
	4.2.13 Education assessment
	4.2.14 Education monitoring and evaluation
	4.2.15 Education lobby
	4.2.16 Education research and development
	4.2.17 Education policy development and implementation
	4.2.18 Strategic education management
	4.2.19 Other *Please specify:* ………………………………………………………………

4.3 Please, indicate your job position/title **FOLLOWING** participation in the IFP program.

………………………………………………………………………………………………………...

4.4 Please, provide information on your function within your organization/institution **FOLLOWING** participation in the IFP program.
*Please check (**in the first column**) all items applicable.*

	4.4.1 Teaching in secondary education
	4.4.2 Teaching in secondary vocational/technical education
	4.4.3 Teaching in higher/tertiary education
	4.4.4 Research in higher/tertiary education
	4.4.5 Instructional material development
	4.4.6 Curriculum development and improvement
	4.4.7 Education quality control
	4.4.8 Secondary school management
	4.4.9 Instructional management
	4.4.10 Human resources management
	4.4.11 Finance management
	4.4.12 Higher education institution management
	4.4.13 Education assessment
	4.4.14 Education monitoring and evaluation
	4.4.15 Education lobby
	4.4.16 Education research and development
	4.4.17 Education policy development and implementation
	4.4.18 Strategic education management
	4.4.19 Other *Please specify:* ……………………………………………………………

4.5 Did you reach a higher job position/title **FOLLOWING** participation in the IFP program?
*Please check (**in the first column**) the item applicable.*

	4.5.1 YES
	4.5.2 NO

4.6 If yes, do you think you would have reached this higher position/title without participating in the IFP program?

Yes, because

………………………………………………………………………………………………………...

No, because

……………………………………………………………………………………………………….

Other, because

……………………………………………………………………………………………………….

4.7 Various factors may play a role in getting a position. To what extent do you feel your current position is a result of the IFP program or other factors?
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

My current position/title is caused by:

	strongly disagree				strongly agree	Don't Know
	1	2	3	4	5	
4.7.1 Participation in IFP program						
4.7.2 Another factor *Please specify:* …………………………..	1	2	3	4	5	

5. SECTOR OF EMPLOYMENT BEFORE AND AFTER PARTICIPATION IN IFP

5.1 Please, indicate the sector of employment in which you worked **BEFORE** participation in the IFP program.
Please check (in the first column) all items applicable.

	5.1.1 Public sector
	5.1.2 Private sector
	5.1.3 Self-employed
	5.1.4 Other *Please specify:* ……………………………………………………………………

5.2 Please, indicate the sector of employment in which you worked (are working) **FOLLOWING** participation in the IFP program.
Please check (in the first column) all items applicable.

	5.2.1 Public sector
	5.2.2 Private sector
	5.2.3 Self-employed
	5.2.4 Other *Please specify:* …

5.3 Please, indicate the **institution** in which you were employed **BEFORE** participation in the IFP program?
*Please check (**in the third column**) all items applicable AND specify the institution.*

SECTOR OF EMPLOYMENT	INSTITUTIONS	
Public sector	5.3.1 Ministry of education	
	5.3.2 Parliament	
	5.3.3 Provincial/regional education institution *Please specify*: ………………………………………	
	5.3.4 Sub-regional or inter-municipal education institution *Please specify*: ………………………………………	
	5.3.5 Tertiary/higher institution (*e.g. University-professional vocational institution – polytechnics*) *Please specify*: ………………………………………	
	5.3.6 Schools (primary-secondary-vocational) *Please specify*: ………………………………………	
	5.3.7 Other *Please specify*: ………………………………………	
Private sector	5.3.8 International/inter-governmental institution *Please specify*: ………………………………………	
	5.3.9 Tertiary/higher institution (*e.g. University-professional vocational institution – polytechnics*) *Please specify*: ………………………………………	
	5.3.10 Schools (primary-secondary-vocational) *Please specify*: ………………………………………	
	5.3.11 Local/international NGO *Please specify*: ………………………………………	
	5.3.12 Other *Please specify*: ………………………………………	
Self employed	5.3.13 Please specify: ………………………………	
Other sector *(please specify):* ……..	5.3.14 Please specify: ………………………………	

Please, indicate the **institution** in which you were employed (are employed) **FOLLOWING** participation in the IFP program.
*Please check (**in the third column**) all items applicable AND specify the institution.*

SECTOR OF EMPLOYMENT	INSTITUTIONS	
Public sector	5.3.1 Ministry of education	
	5.3.2 Parliament	
	5.3.3 Provincial/regional education institution *Please specify*: …………………………………	
	5.3.4 Sub-regional or inter-municipal education institution *Please specify*: …………………………………	
	5.3.5 Tertiary/higher institution (*e.g. University-professional vocational institution – polytechnics*) *Please specify*: …………………………………	
	5.3.6 Schools (primary-secondary-vocational) *Please specify*: …………………………………	
	5.3.7 Other *Please specify*: …………………………………	
Private sector	5.3.8 International/inter-governmental institution *Please specify*: …………………………………	
	5.3.9 Tertiary/higher institution (*e.g. University-professional vocational institution – polytechnics*) *Please specify*: …………………………………	
	5.3.10 Schools (primary-secondary-vocational) *Please specify*: …………………………………	
	5.3.11 Local/international NGO *Please specify*: …………………………………	
	5.3.12 Other *Please specify*: …………………………………	
Self employed	5.3.13 Please specify: ………………………………	
Other sector *(please specify):* ……..	5.3.14 Please specify: ………………………………	

5.5 Do you think you would have worked in this type of sector of employment/institution without participating in the IFP program?

Yes, because

……………………………………………………………………………………………………...

No, because

…………………………………………………………………………………………………….

Other, because

………………………………………………………………………………………………….

6. RESPONSIBILITY AFTER PARTICIPATION IN IFP

6.1 Below you find two statements that refer to responsibility in your institution/organisation **FOLLOWING** participation in the IFP program.
Please check to what extent these statements are applicable to you.

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

After participation in IFP:

	strongly disagree 1 2 3 4 strongly agree 5	Don't Know
6.1.1 I have more responsibility within my professional activities	1 2 3 4 5	
6.1.2 I have more responsibility within my voluntary activities	1 2 3 4 5	

6.2 In case you have more responsibility do you think you would have more responsibilities without participating in the IFP program?

Yes, because

……...

No, because

……..

Other, because

……...

7. LOCUS OF DECISION BEFORE AND AFTER PARTICIPATION IN IFP

7.1 Below you find a number of statements that refer to your contribution in the decision-making process in your organisation/institution **BEFORE** participation in IFP.
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

Before participation in IFP program:

	Strongly disagree 1 2 3 4 strongly agree 5	Don't Know
7.1.1 I took decisions in full autonomy without consultation of higher level authority	1 2 3 4 5	
7.1.2 I took decisions in consultation with a higher level authority	1 2 3 4 5	
7.1.3 Decisions were always taken by higher level authorities in the organization	1 2 3 4 5	

7.2 If you agree on 7.1.3 than state what was your role in the decision making process.
Please check (in the first column) all items applicable (more than one choice possible).

	7.2.1 I prepare the proposal (draft the documents etc.)
	7.2.2 I provide information to the decision-maker
	7.2.3 I discuss the decisions to be taken with the decision-makers
	7.2.4 Other *Please specify:* ……………………………………………………………

7.3 Below you find a number of statements that refer to your contribution in the decision-making process in your organisation/institution **FOLLOWING** participation in IFP.
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

After participation in IFP:

	Strongly disagree strongly agree 1 2 3 4 5	Don't Know
7.3.1 I take decisions in full autonomy without consultation of higher level authority	1 2 3 4 5	
7.3.2 I take decisions in consultation with a higher level authority	1 2 3 4 5	
7.3.3 Decisions are always taken by higher level authorities in the organization	1 2 3 4 5	

7.4 If you agree on 7.3.3 than state what was your role in the decision making process.
Please check (in the first column) all items applicable (more than one choice possible).

	7.4.1 I prepare the proposal (draft the documents etc.)
	7.4.2 I provide information to the decision-maker
	7.4.3 I discuss the decisions to be taken with the decision-makers
	7.4.4 Other *Please specify:* ………………………………………………………….

7.5 Did/do you have more autonomy in taking decisions **FOLLOWING** participation in IFP?
Please check (in the first column) the item applicable.

	7.5.1 YES
	7.5.2 NO

Yes, because

……..

No, because

……..

Other, because

…….

8. PROMOTIONS

Within this study **promotion** refers to one of the following:

- Significant increase in scope of responsibilities;
- Significant increase in annual salary;
- Changes in level in the employing company;
- And becoming eligible for bonuses, incentives, or stock plans.

8.1 How many promotions have you received in the area of education **FOLLOWING** participation in the IFP program?
*Please check (**in the first column**) the item applicable.*

	8.1.1 No promotion
	8.1.2 One promotion *Please specify the kind of promotion*: ………………………………
	8.1.3 Two promotions *Please specify the kind of promotion*: ………………………………
	8.1.4 More than two promotions *Please specify the kind of promotion*: ………………………………

8.2 In case you have been promoted do you think you could have had this (these) promotion(s) without participating in the IFP program?

Yes, because

……..

No, because

……..

Other, because

…….

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

My promotions are caused by:

	Strongly disagree strongly agree	Don't Know
8.3.1 The degree obtained in the IFP program	1 2 3 4 5	
8.3.2 My job performance	1 2 3 4 5	
8.3.3 My work experience	1 2 3 4 5	
8.3.4 Other factors *Please specify*: ………………………	1 2 3 4 5	

PART 3: ACTIVITIES CONDUCTED AFTER PARTICIPATION IN IFP PROGRAM

The aim of the questions in part 3 is to find out the activities (influenced by your acquired knowledge/skills during the IFP program) that you have conducted in the area of education **following** participation in the IFP program.

9. APPLYING ACQUIRED KNOWLEDGE/SKILLS

9.1 Below you find a number of statements that refer to the applicability of your acquired knowledge/skills.
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

	Strongly disagree strongly agree 1 2 3 4 5	Don't Know
9.1.1 The acquired knowledge/skills are relevant for my work	1 2 3 4 5	
9.1.2 The acquired knowledge/skills match with the work environment of my country	1 2 3 4 5	
9.1.3 I am able to apply the acquired knowledge/skills within my professional work settings	1 2 3 4 5	
9.1.4 I am able to apply the acquired knowledge/skills within my voluntary work settings	1 2 3 4 5	

9.2 Please explain your choice in case you select "strongly disagree", "to some extent" or "uncertain".
……………………………………………………………………………………………………...

10. TYPE OF ACTIVITY

10.1 Please, inform on your major **type of activities FOLLOWING** participation in the IFP program. *Please check (**in the first column**) all items applicable (more than one choice possible).*

	10.1.1 Education Management/Policy activities → **go to question 11**
	10.1.2 Operation/professional Activities: *e.g. teaching-research-curriculum-instructional material-consultancy activities* → **go to question 12**
	10.1.3 Voluntary activities → **go to question 13**
	10.1.4 Other *Please specify*: ……………………………………………………………………………

11. MANAGEMENT/POLICY ACTIVITIES CONDUCTED IN THE AREA OF EDUCATION

The aim of the following questions is to have a picture of management/policy activities that you have conducted thanks to the knowledge/skills acquired through the IFP program.

11.1 Please, inform on the major type of **management/policy activities** that have been influenced by your acquired knowledge in your institution **FOLLOWING** participation in the IFP program. *Please check (**in the first column**) all items applicable AND specify the activity that you have conducted (more than one choice possible).*

	11.1.1 Strategic overall management (e.g. developing policies, plans etc.) *Please specify activity*: ……………………………………………………………………
	11.1.2 Financial management *Please specify activity*: ……………………………………………………………………
	11.1.3 Material resources management *Please specify activity*: ……………………………………………………………………
	11.1.4 Human resources management *Please specify activity*: ……………………………………………………………………
	11.1.5 Quality control management *Please specify activity*: ……………………………………………………………………
	11.1.6 Pedagogical/instructional management *Please specify activity*: ………………………………………………………………….
	11.1.7 Other *Please specify:* ……………………………………………………………………

11.2 In case you have conducted management/policy activities do you think you would conduct such management/policy activities without participating in the IFP program?

Yes, because

………………………………………………………………………………………………...

No, because

……………………………………………………………………………………………….

Other, because

……………………………………………………………………………………………….

12. OPERATIONAL/PROFESSIONAL ACTIVITIES

The aim of the following questions is to find out whether you have been able to conduct operational/professional activities thanks to the knowledge/skills acquired through the IFP program.

Operational/professional activities refer to all activities not related to management/policy activities: e.g. teaching – research – curriculum – instructional material – consultancy activities.

12.1 Please, inform on the major type of **operational/professional activities** that have been influenced by your acquired knowledge in your institution **FOLLOWING** participation in the IFP program. *Please check (**in the first column**) all items applicable AND specify the activity that you have conducted (more than one choice possible).*

	12.1.1 Pedagogical activities *Please specify activity*: ………………………………………………………………………
	12.1.2 Didactical activities *Please specify activity*: ………………………………………………………………………
	12.1.3 Research activities *Please specify activity*: ………………………………………………………………………
	12.1.4 Curriculum development activities *Please specify activity*: ………………………………………………………………………
	12.1.5 Instructional material development activities *Please specify activity*: ………………………………………………………………………
	12.1.6 Consultancy/advisory activities *Please specify activity*: ………………………………………………………………………..
	12.1.7 Other *Please specify:* ………………………………………………………………………

12.2 In case you have conducted operational/professional activities do you think you would conduct such operational/professional activities without participating in the IFP program?

Yes, because

………...

No, because

………..

Other, because

……….

13. VOLUNTARY ACTIVITIES IN THE AREA OF EDUCATION

The aim of the following questions is to find out whether you have been able to conduct voluntary activities thanks to the knowledge/skills acquired through the IFP program.

13.1 Please, inform on your major type of voluntary activities that you have conducted **FOLLOWING** participation in the IFP program.
*Please check (**in the first column**) all items applicable AND specify the activity that you have conducted (more than one choice possible).*

	13.1.1 Education lobby in own community/country *Please specify activity*: ………………………………………………………………..
	13.1.2 Education awareness campaign (sensitization of people/parents) in own community *Please specify activity*: ………………………………………………………………..
	13.1.3 Promotion of the importance of education particularly in rural/remote areas *Please specify activity:* ………………………………………………………………
	13.1.4 Promotion of higher education *Please specify activity:* ………………………………………………………………
	13.1.5 Gender/minorities' advocacy in the area of education *Please specify activity:* ………………………………………………………………..
	13.1.6 Other *Please specify activity:* ………………………………………………………………..

13.2 In case you have conducted voluntary activities do you think you would conduct such voluntary activities without participating in the IFP program?

Yes, because

…………………………………………………………………………………………………..

No, because

…………………………………………………………………………………………………..

Other, because

………………………………………………………………………………………………….

PART 4: IMPACT OF IFP IN THE AREA OF EDUCATION

The aim of the following questions is to find out whether the activities that you have conducted have brought about changes (**small or big**) in the organization/institution or communities/countries where you are active. In other words what do you think you have changed in the area of education within your institution or community/country after participation in the IFP program?

Within this study change can refer for example to: the building of a classroom in a remote/rural area – a number of girls having access to education – availability of instructional material in remote areas – revision of the curriculum - use of new pedagogical and didactical strategies by peers – provision of on-site/off-site trainings to teachers etc.

14. ACHIEVEMENT

14.1 Below you find a statement that refers to your achievements.
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

14.1.1 I have brought about changes in the area of education	Strongly disagree				strongly agree	Don't know
	1	2	3	4	5	

14.2 If you agree with this statement, please, indicate at what level the changes took place.
*Please check (**in the first column**) all items applicable.*

	14.2.1 Within institution/organization where you work (have worked)
	14.2.2 In community
	14.2.3 In country (national level)
	14.2.4 Other
	Please specify: ……………………………………………………………………

15. WRITE A STORY ABOUT IMPACT OF IFP ACCORDING TO YOU

15.1 Please, write (maximum 1A4) about the impact of IFP on your personal life (what have you personally achieved) **AND** the changes that you think to have brought about in your organization/institution or in your community/country. You can use the following questions as guidelines/suggestions.

15.1.1 What are your positive/negative personal achievements because of participation in IFP (changes caused by IFP on your personal life)?
15.1.2 Which factors have hindered you to achieve your personal goals?
15.1.3 What changes in the area of education (small or big) do you think to have brought about in your organisation/institution or your country/community and that could be attributed to your conducted activities?
15.1.4 What change do you expect and when do you expect the changes (if the changes are not yet visible)?
 15.1.5 To what extent can personal achievements and changes in your organization or in your community/country be (or not) attributed to your participation in the IFP program?

15.1.6 How do you contribute in addressing social injustice in the area of education in your country or community?

15.1.7 Please, indicate - when you have this kind of information – the data source (national or international) in which these changes can be found (are mentioned).

PART 5: SATISFACTION WITH IFP AND CAREER SATISFACTION

The aim of the following questions is to find out whether you are satisfied with your participation in the IFP program and with your career.

16. ALUMNUS/ALUMNA SATISFACTION

16.1 Below you find a statement that refers to your academic expectations.
Please check to what extent this statement is applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

16.1.1 The IFP program helped me to meet my academic expectations	strongly disagree 1 2 3 4 strongly agree 5	Don't Know

16.2 In case the program didn't help you to meet your academic expectations please write what you missed in your study.
……………………………………………………………………………………………………...

16.3 Below you find a number of statements that refer to the role of IFP in your personal life.
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

Participation in IFP:

	Strongly disagree 1 2 3 4 strongly agree 5	Don't know
16.3.1 Helped me to get a job	1 2 3 4 5	
16.3.2 Helped me to access higher positions	1 2 3 4 5	
16.3.3 Helps/helped me to reach my personal career goals	1 2 3 4 5	
16.3.4 Helps me to be successful in my organization	1 2 3 4 5	
16.3.5 Contributed to my personal development	1 2 3 4 5	
16.3.6 Influenced my way of working	1 2 3 4 5	
16.3.7 Influenced my way of thinking	1 2 3 4 5	
16.3.8 Broadened the scope of my knowledge	1 2 3 4 5	
16.3.9 Other *Please specify:* ……………………………	1 2 3 4 5	

16.4 To what extent are you satisfied with your **career** following participation in the IFP program?
Please check to what extent these statements are applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

I am satisfied with:

	strongly disagree strongly agree 1 2 3 4 5	Don't Know
16.4.1 The success I have achieved in my career.	1 2 3 4 5	
16.4.2 The progress I have made toward meeting my overall career goals.	1 2 3 4 5	
16.4.3 The progress I have made toward meeting my goals for income.	1 2 3 4 5	
16.4.4 The progress I have made toward meeting my goals for advancement.	1 2 3 4 5	
16.4.5 The progress I have made toward meeting my goals for the development of new skills.	1 2 3 4 5	
16.4.6 The progress I have made compared to my peers.	1 2 3 4 5	

16.5 How satisfied are you with the IFP program in general terms?
Please check to what extent the statement is applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

	strongly disagree strongly agree 1 2 3 4 5	Don't Know
16.5.1 I am satisfied with the IFP program	1 2 3 4 5	

16.6 To what extent have you contributed to making IFP visible (dissemination goals, content, and results of IFP) in general terms?
Please check to what extent the statement is applicable to you?

1= Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

16.6.1 I contribute in making IFP visible in my country/community	strongly disagree strongly agree 1 2 3 4 5	Don't Know

16.7 In case you have contributed in the visibility of IFP state than how do you contribute in making IFP visible in your country/community?
………………………………………………………………………………………………...

16.8 What do you see as strengths of IFP?
………………………………………………………………………………………………...

16.9 What could/should be improved (if any) in the IFP program in order to make alumni more successful in conducting activities in the area of education?

………………………………………………………………………………………………...

Thank you very much for filling out this questionnaire.

Appendix J: Questionnaire 4 (feedback field study)

Purpose:

The purpose of the field study is to improve our instrument based on your feedback. The aim of questionnaire 4 is to test the clarity of questions included in the instrument developed in order to assess outcomes and impact of IFP as well as the length of the instrument. The aim of the instrument is to measure the outcomes and impacts of IFP in the area of education.

- **Outcomes** refer to activities (professional/voluntary) that alumni or network of alumni conduct in the area of education in their countries and that are influenced by the knowledge/skills acquired during the IFP program.

- **Impacts** refer to changes caused by the conducted activities in the area of education in the communities/countries or in the organization/institution where alumni are active.

Instructions:

We kindly ask you to answer the questions below after reading the items included in the instrument.

1- Please indicate to what extent the questions are easy to follow, understandable and easy to answer.

1=Strongly disagree/ 2= Disagree to some extent/ 3= Uncertain/ 4= Agree to some extent/ 5= Strongly agree

	strongly disagree 1 2 3 4 strongly agree 5
The questionnaire is easy to follow	☐ ☐ ☐ ☐ ☐
The questions are easy to understand	1 2 3 4 5 ☐ ☐ ☐ ☐ ☐
The questions are easy to answer	1 2 3 4 5 ☐ ☐ ☐ ☐ ☐

2- How much time it takes you approximately to fill out the questionnaire? ………………….

3- Which questions did you find difficult to answer? (You can also insert your comments in the final instrument)

……………………………………………………………………………………………………..

4- Please indicate, in your opinion, what questions should be excluded from the instrument. (You can also insert your comments in the final instrument):

……………………………………………………………………………………………………

……………………………………………………………………………………………………

5- Please indicate, in your opinion, what do you miss in the final instrument:

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Thank you very much for filling out this questionnaire.

Appendix K: Frequency distributions, I-CVI, and CVR

Frequency distributions, I-CVI, and CVR first Delphi-round

1= Not at all relevant/measurable
2= Somewhat relevant/measurable
3= Relevant/measurable
4= Very relevant/measurable
5= Don't know

Table K1. Frequency distribution indicators and I-CVI values (N=8)

Indicators	Frequency distribution Relevance				I-CVI relevance	
	1	**2**	**3**	**4**	**Number of ratings of 3 or 4**	**I-CVI**
1- Socio-biographical backgrounds			8		8	1
2- Educational backgrounds				8	8	1
3- Employment				8	8	1
4- Career progress			8		8	1
5- Type of conducted activities			3	5	8	1
6- Changes caused by conducted activities			6	2	8	1

Table K2. Frequency distribution socio-biographical backgrounds variables (N=8)

Variables	Frequency distribution: Relevance					Frequency distribution: Measurability					Frequency distribution: Included in instrument		
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	Yes	No	Don't know
1.1 Gender (male/female)			1	7		1			7		8		
1.2 Age			1	7		1			7		8		
1.3 Marital status (single-in partnership-married-married by customary law-…	3	2		3		1	2	1	4		4	3	1

Table K3. I-CVI and CVR values socio-biographical backgrounds variables (N=8)

Variables	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
1.1 Gender (male/female)	8	1,00	7	0,88	8	4	1,00
1.2 Age	8	1,00	7	0,88	8	4	1,00
1.3 Marital status (single-in partnership-married-married by customary law-…	3	0,38	5	0,63	4	4	0,00

Table K4. Frequency distribution educational backgrounds variables (N=8)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
2.1 Education degree obtained				8				1	7		8		
2.2 Place of study (own country or elsewhere)				8					8		8		
2.3 Kind/content of study (teacher training, curriculum development …)			1	7				1	7		8		

Table K5. I-CVI and CVR values for educational backgrounds variables (N=8)

Variables	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
2.1 Education degree obtained	8	1,00	8	1,00	8	4	1,00
2.2 Place of study (own country or elsewhere)	8	1,00	8	1,00	8	4	1,00
2.3 Kind/content of study (teacher training, curriculum development …)	8	1,00	8	1,00	8	4	1,00

Table K6. Frequency distribution employment variables (N=8)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
3.1 Type of job	1	2	2	2	1	1	2	1	4		4	2	2
3.2 Place of employment			1	7					8		8		
3.3 Kind of employment sector				8			1		7		8		
3.4 level of employment				8					8		8		
3.5 Applicability knowledge/skills				8			4		4		8		

Table K7. I-CVI and CVR values for employment variables (N=8)

Variables	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
3.1 Type of job	4	0,50	5	0,63	4	4	0,00
3.2 Place of employment	8	1,00	8	1,00	8	4	1,00
3.3 Kind of employment sector	8	1,00	7	0,88	8	4	1,00
3.4 level of employment	8	1,00	8	1,00	8	4	1,00
3.5 Applicability knowledge/skills	8	1,00	4	0,50	8	4	1,00

Table K8. Frequency distribution career progress variables (N=8)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
4.1 Income/salary	1		3	4			2	3	3		6	1	1
4.2 Number of promotions since graduation	1		1	6			3		5		5	1	2
4.3 Job position				8			1		7		8		
4.4 Responsibility in current work				8			1	2	5		8		
4.5 Locus of decision (degree to which alumni are involved in decision taking processes)			2	6			1	4	3		7		1
4.6 Alumni' career satisfaction	1			6	1		2	1	4	1	6	1	1

Table K9. I-CVI and CVR values for career progress variables (N=8)

Variables	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
4.1 Income/salary	7	0,88	6	0,75	6	4	0,50
4.2 Number of promotions since graduation	7	0,88	5	0,63	5	4	0,25
4.3 Job position	8	1,00	7	0,88	8	4	1,00
4.4 Responsibility in current work	8	1,00	7	0,88	8	4	1,00
4.5 Locus of decision (degree to which alumni are involved in decision taking processes)	8	1,00	7	0,88	7	4	0,75
4.6 Alumni' career satisfaction	6	0,75	5	0,63	6	4	0,50

Table K10. Frequency distribution type of conducted activities variables (N=8)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
5.1 Activities related to financial and resources			2	5	1		2	2	2	2	6		2
5.2 Activities related to human resources			1	5	2		2	1	3	2	7		1
5.3 Activities related to students' background conditions			1	6	1		2	2	3	1	7		1
5.4 Activities related to community/parents' involvement			2	5	1		2	2	3	1	6		2
5.5 Activities related to instructional		1	1	5	1		2	2	3	1	7		1

Table K11. I-CVI and CVR values for type of conducted activities variables (N=8)

Variables	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
5.1 Activities related to financial and resources	7	0,88	4	0,50	6	4	0,50
5.2 Activities related to human resources	6	0,75	4	0,50	7	4	0,75
5.3 Activities related to students' background conditions	7	0,88	5	0,63	7	4	0,75
5.4 Activities related to community/parents' involvement	7	0,88	5	0,63	6	4	0,50
5.5 Activities related to instructional	6	0,75	5	0,63	7	4	0,75

Table K12. Frequency distribution changes caused by conducted activities variable (N=8)

Variable	Frequency distribution: Relevance					Frequency distribution: Measurability					Frequency distribution: Included in instrument		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
6.1 Alumni' perceived impact			1	7			3		5		7		1

Table K13. I-CVI and CVR values for changes caused by conducted activities variable (N=8)

Variable	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
6.1 Alumni' perceived impact	8	1,00	5	0,63	7	4	0,75

Table K14. Summary results Delphi round 1

Variables	Selected	Revised and submitted in round 2	Deleted
1.1 Gender	X		
1.2 Age	X		
1.3 Marital status			X
2.1 Education degree	X		
2.2 Place of study	X		
2.3 Kind of study	X		
3.1 Type of job			X
3.2 Place of employment	X		
3.3 Sector of employment	X		
3.4 Kind of employment institution	X		
3.5 Applicability knowledge		X	
4.1 Income/Salary		X	
4.2 Number promotions		X	
4.3 Job position	X		
4.4 Responsibility	X		
4.5 Locus of decision	X		
4.6 Alumni career satisfaction		X	
5.1 Activities Finance and Resources			X
5.2 Activities Human resources			X
5.3 Activities students' conditions			X
5.4 Activities community/parents			X
5.5 Activities instructional issues			X
6.1 Alumni' perceived impact		X	

From the initial 23 variables, eleven were retained for the IFP instrument and seven variables were deleted in the first Delphi round. Five variables were revised as recommended by respondents and submitted again for evaluation in the second Delphi round. Further, eight new variables and a new indicator (alumni satisfaction) were recommended by respondents. The new variables include: religion of alumni and country of birth and residence of alumni (related to indicator 1), year graduation (related to indicator 2), function description (related to indicator 4), educational management activities, voluntary activities, professional activities (related to indicator 5), and alumni satisfaction with IFP (related to indicator 7). Table K15 provides the lists of revised as well as new variables.

Table K15. Revised and new variables from first Delphi-round

Indicators	New and revised variables
Indicator 1: Socio-biographical backgrounds	Religion alumni (N) – country of birth/residence of alumni (N)
Indicator 2: Educational backgrounds	Year of graduation (N)
Indicator 3: Career progress	Income/salary – number of promotions
Indicator 4: Employment	Function description (N) – applicability knowledge
Indicator 5: Type of activity	Educational management/policies activities(N) – voluntary activities(N) – professional activities(N)
Indicator 6: Changes caused by conducted activities	Alumni perceived impact
Indicator 7: Alumni satisfaction (N)	Alumni satisfaction with IFP (N)

Note. N refers to new indicator or variable

Frequency distributions, I-CVI, and CVR second Delphi-round

Table K16. Frequency distribution new/revised variables related to indicator 1 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Religion	6		1			6	1					7	
Country of birth/residence of alumni				7					7		7		

Table K17. I-CVI and CVR values for new/revised variables related to indicator 1 (N=7)

Variables	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Religion	1	0,14	0	0,00	0	3,5	-1,00
Country of birth/residence of alumni	7	1,00	7	1.00	7	3,5	1,00

Table K18. Frequency distribution new/revised variables related to indicator 2 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Year of graduation				7					7		7		

Table K19. I-CVI and CVR values for new new/revised variables related to indicator 2 (N=7)

Variable	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Year of graduation	7	1,00	7	1,00	7	3,5	1,00

Table K20. Frequency distribution new/revised variables related to indicator 3 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Function description			5	2				6		1	7		
Applicability knowledge				7				3	4		7		

Table K21. I-CVI and CVR values for new/revised variables related to indicator 3 (N=7)

Variables	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Function description	7	1,00	6	0,86	7	3,5	1,00
Applicability knowledge	7	1,00	7	1,00	7	3,5	1,00

Table K22. Frequency distribution new/revised variables related to indicator 4 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Income/salary				7		4		3			3	4	
Number of promotions				7				3	4		7		

Table K23. I-CVI and CVR values for new/revised variables related to indicator 4 (N=7)

Variables	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Income/salary	7	1,00	3	0,43	3	3,5	-0,14
Number of promotions	7	1,00	7	1,00	7	3,5	1,00

Table K24. Frequency distribution new/revised variables related to indicator 5 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Educational management/policies activities				7				7			7		
Voluntary activities				7				7			7		
Professional activities				7				7			7		

Table K25. I-CVI and CVR values for new/revised variables related to indicator 5 (N=7)

Variables	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Educational management/policies activities	7	1,00	7	1,00	7	3,5	1,00
Voluntary activities	7	1,00	7	1,00	7	3,5	1,00
Professional activities	7	1,00	7	1,00	7	3,5	1.00

Table K26. Frequency distribution new/revised variables related to indicator 6 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Alumni perceived impact				7		1		6			7		

Table K27. I-CVI and CVR values for new/revised variable related to indicator 6 (N=7)

Variable	**I-CVI relevance**		**I-CVI measurement**		**Content Validity Ratio**		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Alumni perceived impact	7	1,00	6	0,86	7	3,5	1,00

Table K28. Frequency distribution new/revised variables related to indicator 7 (N=7)

Variables	Frequency distribution: **Relevance**					Frequency distribution: **Measurability**					Frequency distribution: **Included in instrument**		
	1	2	3	4	5	1	2	3	4	5	Yes	No	Don't know
Alumni satisfaction with IFP				7					7		7		

Table K29. I-CVI and CVR values for new/revised variables related to indicator 7 (N=7)

Variable	I-CVI relevance		I-CVI measurement		Content Validity Ratio		
	Number of ratings of 3 or 4	I-CVI	Number of ratings of 3 or 4	I-CVI	Ne	N/2	CVR
Alumni satisfaction with IFP	7	1,00	7	1,00	7	3,5	1,00

Table K30. Summary results Delphi round 2

Variables	Selected	Deleted
Income/Salary		X
Number promotions	X	
Alumni satisfaction with IFP	X	
Applicability knowledge	X	
Alumni perceived impact	X	
Religion		X
Country of birth/residence	X	
Year of graduation	X	
Function description	X	
Educational management/policies activities	X	
Voluntary activities	X	
Professional activities	X	

From the initial 12 variables evaluated in the second round, ten were retained for the IFP instrument and two variables were deleted.

Summary step 2 results

Table K31. Variables selected for the IFP instrument

Variables	I-CVI values relevance	I-CVI values measurement	CVR Values
1.1 Gender	**1,00 (n=8)**	**0,88 (n=8)**	**1,00 (n=8)**
1.2 Age	**1,00 (n=8)**	**0,88 (n=8)**	**1,00 (n=8)**
1.3 Country of birth/residence	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
2.1 Education degree obtained	**1,00 (n=8)**	**1,00 (n=8)**	**1,00 (n=8)**
2.2 Place of study	**1,00 (n=8)**	**1,00 (n=8)**	**1,00 (n=8)**
2.3 Kind/content of study	**1,00 (n=8)**	**1,00 (n=8)**	**1,00 (n=8)**
2.4 year of graduation	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
3.1 Place of employment	**1,00 (n=8)**	**1,00 (n=8)**	**1,00 (n=8)**
3.2 Sector of employment	**1,00 (n=8)**	**0,88 (n=8)**	**1,00 (n=8)**
3.3 Kind of employment institution	**1,00 (n=8)**	**1,00 (n=8)**	**1,00 (n=8)**
3.4 Applicability knowledge/skills	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
3.5 function description	**1,00 (n=7)**	**0,86 (n=7)**	**1,00 (n=7)**
4.1 Number of promotions	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
4.2 Job position	**1,00 (n=8)**	**0,88 (n=8)**	**1,00 (n=8)**
4.3 Responsibility in current work	**1,00 (n=8)**	**0,88 (n=8)**	**1,00 (n=8)**
4.4 Locus of decision	**1,00 (n=8)**	**0,88 (n=8)**	**0,75 (n=8)**
5.1 Educational management/policies activities	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
5.2 Voluntary activities	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
5.3 Professional activities	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**
6.1 Alumni perceived impact	**1,00 (n=7)**	**0,86 (n=7)**	**1,00 (n=7)**
7.1 Alumni satisfaction with IFP	**1,00 (n=7)**	**1,00 (n=7)**	**1,00 (n=7)**

Note. n=8 indicates that the variable was selected after the first round, n=7 indicates that the variable was selected after the second round