# AN **INTRODUCTORY MANUAL** FOR **ASSESSING** AND **EVALUATING INTERNATIONAL PROGRAMS** IN **EDUCATION**

Written by:
Dr. Roxana G. Reichman,
Vice-coordinator of DOIT
and leader of the assessment team

Dr. Manal Yazbak Abu Ahmad,
Assessment team member

# Table of Contents

# Introduction

Planning, developing and evaluating international programs is a very complex endeavor but the rewards for those interested in collaborating with colleagues from different institutions are immense, both on the personal and professional front, as well as on the level of the institutions involved in this process.

This handbook presents models for assessing and evaluating international programs in the field of education and proposes certain tips which can be helpful for assessing similar international programs. These tips are based on our experience as leaders of the team in charge of assessing a program on curriculum reform in the field of higher education on the topic of multicultural education and children's rights. This has been done as part of DOIT (Curriculum reform in multicultural education and diversity training), a program funded by the European Union.

In the spirit of this program, two Israeli researchers have collaborated in order to write this manual. Dr. Roxana G. Reichman has been the vice-coordinator of DOIT and the leader of the assessment team of the program. She works at Gordon Academic College, a multicultural Teachers' College in Haifa, Israel and after defining the terms and presenting the rational for assessment, she will focus on the theoretical issues related to assessment and evaluation of international programs in general, and on programs related to curriculum reforms in particular. Dr. Manal Yazbak Abu Ahmad has been an active member of the assessment team and she works at Sakhnin Academic College, an Arab Teachers' College in Sakhnin, Israel. Both authors have extensive experience in teacher education, both at the undergraduate and at the graduate level and they have been working together for DOIT for the last three years. They will present the case study of DOIT assessment team from their point of view. It is important to mention that in addition to the assessment team, another team was charged with the quality assurance process and their conclusions will be presented in a separate handbook.

In order to apply for funding in the framework of TEMPUS (Trans-European Mobility Scheme for University Studies), a consortium should be formed in order to enable the collaboration between at least eight institutions of higher education in five countries. DOIT was a consortium formed by 21 organizations in seven countries. Israel and Georgia were the two partner countries and the other countries that took part in this project were Austria, Estonia, Germany, the Netherlands and UK. Most of the organizations that took part in this program were universities and colleges, while the others were Four NGOs (one in Israel, two in Georgia and one in Estonia) and six Students Unions (three in Israel and three in Georgia).

This manual is divided into the following sections:

In the first chapter, four terms that will be used in the context of this handbook will be defined: *program, monitoring, evaluation and assessment*. The task of defining the first two terms is easier because there is quite a consensus about them. It will be shown that the other two terms have been defined in different ways in different languages and in different cultures and sometimes the terms are used interchangeably.

The second chapter will focus on presenting the main reasons for evaluating programs.

In the third chapter, models for assessing international programs will be presented and tips for evaluating such programs will be offered.

The fourth chapter will deal with assessing international programs in education in general, focusing on programs interested in developing, piloting and implementing curriculum reforms.

The fifth and last chapter will present a case study of the assessment team of DOIT, a program of developing, piloting and implementing curriculum reform in the field of multicultural education in higher education. This is not an evaluation of the program itself, but a self-evaluation of the assessment team, written from the point of view of two of its members.

# Chapter 1 Terms definitions

When writing a manual on assessment and evaluation of international programs, it is very important to be explicit about the ways in which different terms will be used.

**Program:** A set of activities or projects with the purpose of achieving a common goal. It includes several steps such as initiation, planning, development, implementation, assessment, dissemination and sustainability. In our context, a program is one that has been approved and funded by an external funding agency such as the European Commission.

**Monitoring:** The ongoing process of following the scope and pace of implementation of a given program and the changes that this implementation creates. Programs funded by the European Union undergo both a process of internal monitoring which is done by people involved in the program itself and a process of external monitoring done from time to time by the funding agency.

**Evaluation:** Any type of evaluation involves an explanation of the gaps that exist between what was proposed to be accomplished and what has actually happened as a result of a certain program. It always involves some sort of judgement about the quality of the process. According to Taylor (1993), "evaluation compares what has been accomplished (evidence) with what should have been accomplished (criteria) and then makes a judgement about how well it was done." Scriven (1991) defines evaluation as "the process of determining merit, worth or significance". Many researchers consider evaluation as a summative way of finding out the end result of a certain project/program/achievements, something like a final grade.

**Assessment:** Certain researchers choose the term assessment for describing a process of ongoing learning in order to improve a program, a project, etc. In their view, the assessment is formative, in comparison with the evaluation, which is summative.

It is important to mention, as it will be explained further on, that in some cases, the terms *evaluation* and *assessment* are used as synonyms. Such an approach was proposed by Hughes & Nieuwenhuis (2008). In "A Project Manager's Guide to Evaluation" they point out that there is a difference even in English between

the British English where this words are synonyms and the American English where assessment is closer to the evaluation of performance. Therefore, they suggest to use these two terms in an interchangeable manner, with the main purpose of constant improvement and not necessarily from the point of view of accountability. When using the present manual, the reader should therefore be aware of the fact that the two terms will be used alternatively, as synonyms. We stress the fact that for the purpose of this manual, the words *evaluation* and *assessment* will be used interchangeably, in spite of the fact that the literature presents differences of approach between the two terms. Since it was shown how researchers have pointed out that the use of these two terms is also a cultural issue, one additional argument will be presented.

The two authors of this handbook live and work in Israel, a country where there are two official languages: Hebrew (spoken by about 80% of the population) and Arabic (spoken by about 20% of the population). In Hebrew, like in British English, there is only one term (HAARACHA) for both of the English words *evaluation* and *assessment*. In Arabic, there is also only one term (TAKWEEM) used for both of the American English words. Therefore it is easier for us to decide to use either one or the other when talking about assessing and evaluating international programs and we made it obvious from the beginning (including in the title of this manual).

Still, in this chapter it is important to present the different opinions that exist in the literature regarding these two terms. This presentation will illustrate the differences of opinions that appear regarding these two notions, so that each one of the evaluators of any international program will be able to choose his or her own terms and to decide which elements need to be stressed.

According to the Institute for teaching, learning and academic leadership (www.itlal.org), *assessment* is the process of measuring effectiveness in order to improve the process (formative) while evaluation is defined as "summative assessment". *Assessment* is presented as formative and emphasizing progress while *evaluation* is seen as summative, focusing on the mastery of competencies. Secolsky & Denison (2012) agree and claim that evaluation happens at a certain moment in time while assessment is a long term process of systematic collection and analysis of data regarding a certain program.

On the other hand, Peretz and Moskowitz (2014) propose an *evaluation process, a formative* approach which should include the following steps: defining the purpose of what should be evaluated, designing the evaluation questions, designing the evaluation methods (which include interviews, recording of meetings), planning the data collection and then collecting it during the program itself, analyzing the data and disseminating the findings. One of the examples they give regarding the specific program they evaluated (DOIT) was that once a problem occurred (time was lost during one of the meetings, there was no meeting of the members in the hotel lobby and therefore some members arrived later than others, translations were not complete, team members had professional disagreements, etc.), their immediate feedback allowed to take measures in order to improve the process. This approach is closer to what most researchers call *assessment* although in their article they use they word *evaluation*.

Another approach which is worth mentioning is the one presented by Brennan (2000). He doesn't only deal with the issue of formative versus summative, but also with the issue of stakeholders and he claims that assessment is internal, for a person's/group improvement (formative) while evaluation is for outsiders (*for example a funding agency*).

Whatever the terms chosen might be, there is a consensus between researchers that both the formative and the summative evaluation are important for improving the program that is being assessed and for being able to learn from this experience when dealing with other similar programs.

# Chapter 2   Rational for assessment and evaluation

There are two basic reasons for the evaluation and assessment of educational programs: one of them (and this is the one which will be stressed in this manual) is to strive to *continually improve* the programs by learning from the best practices as well as from the challenges and pitfalls.

The second one is inevitable and can't be overlooked since it has to do with the issue of *accountability*. Different agencies fund programs either at the local level (smaller funds), at the national levels (more significant funding) or at the international level. In the latter case, the one which is stressed by this manual, the funding is significant, many institutions from different countries are involved, and therefore it is obvious that the program coordinator has to be accountable to the funding agency and to show that the funds that have been allocated have been spent according to the proposal that had been approved.

If changes need to be made, and this can obviously happen when so many people are involved in a certain program for a couple of years, it is crucial to build trust with the funding agency, to let them know that changes have to be done, to ask for their approval a priori and not to allocate any funds in a manner which is not identical to what had been proposed unless specific approval from the funding agency had been given. A good advice for any program coordinator is to make the funding agency aware of the changes as soon as possible, to explain the reasons for making these changes and to get their approval. Our experience shows that the funding agencies, which are very experienced in this field, understand the need to make some changes in the original proposal and are willing to approve those changes as long as they are reasonable.

Evaluation is important because *during* the implementation it can assist the program coordinator and the team leaders to correct their actions on issues that need special attention and it is also important *after* the implementation when the evaluators are able to see the big picture and to better understand the impact of the program. A final reason for conducting assessment is that the lessons learned from the best practices as well as from the possible failures can help funding agencies to make practical decisions regarding future programs and can help the program leaders to learn from their own experience when applying for funding for future programs.

The members of the assessment team should familiarize themselves with the relevant literature in the field and it is the responsibility of the team leader to make sure that all members have enough knowledge on the subject. The leader of the assessment team can't always choose the team members and in any case, the leader doesn't know all the members and their level of expertise. Therefore, if needed, the leader is the one who has to provide the team members the relevant information and in the process to find out their strengths and weaknesses.

Each model or group of models has its own advantages and disadvantages, and it is not the purpose of this manual to prefer one model over another. Therefore, one team can choose one evaluation model, another team can choose a different model and both can be as successful, regardless of the model which has been selected. It is important though that the team members are aware of the limitations of each model in order to be able to justify their choice to the other members of the consortium.

## 2a. Models of evaluation

The literature on evaluation models presents over one hundred different models which Worthen, Sanders & Fitzpatrick (1997) have divided into six categories: *objectives-oriented, management-oriented, consumer-oriented, expertise-oriented, adversary-oriented, and participant-oriented evaluation*. This chapter will focus on the most known models which have been vastly discussed by different researchers and which are considered the most useful for evaluating international programs in general, and international programs in the field of education, in particular.

The **objectives oriented evaluation** was pioneered by Tyler (1942) who considered that the role of the evaluating team is to determine to what extent a program has met its objectives as well as the reasons for not having achieved part of the objectives. This approach has been called objectivist rationalist approach or the *goal driven* evaluation and it was less interested in the reasons why a certain program succeeded. Tyler (1942) identified four main questions that must be asked in order to assess effective instruction: what are the *educational goals, what learning methods* are preferable for achieving these goals, *how can these methods be organized* and how can the *effectiveness* of these learning experiences be evaluated? The answers to these four questions would result in a program that includes the curriculum objectives, the type of

students' behavior that should be encouraged and the way this behavior occurs not only in educational settings but in the students' lives.

It was only almost three decades later that Scriven (1967) introduced the terms *formative* and *summative* evaluation as well as *intended* and *unintended* outcomes when he suggested the *goal free methodology*. In his view, evaluation should be interested in intended and unintended outcomes and the assessment can start in this way, without telling the assessment team in advance what were the objectives of the program. The evaluators are supposed to assess what has been going on without any preconceived ideas and then move back to the *goal-based* approach only if needed. This assessment should take into account criteria such as *relevance, validity, reliability, objectivity or credibility, importance , scope and efficiency*.

Patton (1997) argued that most of the evaluations are useless because the results don't have any impact, and therefore he suggested that the evaluation report should be written *having in mind the stakeholders* who would be interested in reading and learning from it and he proposed to use qualitative methods for evaluation purposes.

Stake (2005) proposed the *responsive evaluation* which focuses on the following aspects: *antecedents* (conditions existing prior to curriculum evaluation, students' prior knowledge or interest, learning environment and traditions of the institution), *transactions* (the encounters between students and teachers and between students and students) and outcomes (assessment of the short term and long term impact of instruction). The focus is on the way different people (students, teachers and the community) respond to assessment. This model stresses more the program activities than the results by presenting different points of view of all the people involved without assessing the causes of success and/ or failure. It seems that qualitative research methods are better fit for this model. In depth interviews or focus groups with the consortium members can tell evaluators a lot about the way in which the program and the process are perceived.

Provus (1971) proposed the *discrepancy model* (DEM) which looks at the gap (discrepancy) between what was intended to be achieved and what was actually achieved. It is *based on standards* which have been accepted by all the members before the program started. This model deals with the process (the

behavior) and with the *product* (outcomes), but unfortunately it disregards certain outcomes which may be very important just because they were not included in the objectives of the proposal. This approach has been accepted in education as being influenced by Tylor's model, and in the US it is a highly controversial political issue, where the Republican party adopted it as a way to fund educational programs, while the Democratic party asks for a more holistic approach that looks at a child as a whole and not necessarily in comparison to a standard norm of behavior, knowledge, etc.

The CIPP model (*Context, Input, Process, and Product*) was influenced by Provus' DEM model and uses the *systems theory*. Stufflebeam (1985) created this comprehensive framework for evaluation of programs and projects and he focuses especially on long term improvement and on sustainability. This framework implies multiple observers and informants, multiple procedures for gathering data (both quantitatively and qualitatively) and an independent assessment and feedback by stakeholders. The four parts of this model ask questions such as: what needs to be done? *How* should it be done? *Is it being done? Did it succeed?* In this model, success is measured in terms of effectiveness and sustainability.

Kirkpatrick's (1994) *four level model* of successive evaluation of training effectiveness implies building information from the first level (*reaction*, based on the perceptions of the participants) to the second level (*learning*, based on pretest and post test), then to the third level (*transfer*, which has to do with changes in the learners' behavior as a result of training) and finally to the fourth an last level (*results*, which is defined from the point of view of the stakeholders).

Other models that can be used come from the field of organizational theories and the most common one is SWOT, where the evaluators examine the internal *strengths and weaknesses* of an organization (in our case of a program) as well as the external *opportunities* and threats. This is a helpful way of assessing the people, the process and the outcome of a program from a different point of view. This model has been initiated having in mind for profit organizations (businesses, etc) but today it is used extensively in schools, institutions of higher education and NGOs. It is therefore a very useful tool in assessing international programs funded by the European Commission although its origin is still unclear. Some researchers claim that the author of this model is

Humphrey from Stanford University although he himself has refused to take credit for it.

The system *approach* to evaluation perceives organizations as goal oriented systems which act in order to achieve certain goals. Littlejohn (2008) writes that "the parts of a system must behave according to guidelines and must adapt to the environment on the basis of feedback" (p.39) and therefore the assessment provides the necessary feedback for constant improvement. When dealing with consortiums formed by faculty members of different institutions of higher education, this can prove to be a challenging mission for several reasons.

Colleges and universities are considered classic examples of what organizational theories call "*loosely coupled systems*" (Orton & Weick, 1990). These systems are defined by individual elements or people with high autonomy relative to the larger system, and therefore actions in one part of the system doesn't have to influence other parts of the system. The forces of specialization are stronger than the common culture or the common identity, and therefore it is more challenging to lead a consortium formed of faculty members of different institutions, when in fact each single institution is a loosely coupled system. Finding a common goal is therefore not easy and this is one of the main jobs of the program coordinator as well as of the leader of the assessment team.

The most modern approaches to evaluation propose the *learning-orientated method* (Peter Van der Knaap, 2006). This approach seems very appropriate to educational organizations and follows in the same line with researchers such as Sergiovanni (2007) who stress the aspect of *transformational leadership* that encourage the creation of *learning communities*.

Finally, there are also a few models which come from the field of education itself and which also stress the importance of a *community of learners*. McMilland and Chavis (1986) found four key factors that create a community: *membership, influence, fulfilment of individual needs* and *shared events and emotional connections*.

The principles of *learning communities and communities of practice* (Bonk, Wisher and Nigrelli (2004) fit these four key factors because the members of the learning community feel they belong to the group, they want to learn and work together, they are willing to express their views and to influence the direction in

which the group is going and to share common experiences. The fact that people learn to work together with colleagues from different institutions and even different countries can bring hope that in institutions of higher education it is possible to work in an interdisciplinary manner with the purpose of advancing students' learning, research and service to the community.

The common denominator of all models is that information should be collected in a systematic manner, both in a quantitative way (surveys) and in a qualitative way (interviews, observations, focus groups, analysis of documents etc.) in order to be able to make decisions regarding the quality of the program and the changes that must be made.

# Chapter 3   Assessing international programs: theoretical aspects and practical tips for leaders

The most important thing is to start the process as soon as possible. The best idea is to involve the evaluators from the very beginning, meaning during the time of writing the application. If this is not possible, at least once the program has been approved, it is important to create an assessment team which will follow and observe the process from the very beginning.

The application package for funding agencies such as the European Commission always includes a special section regarding assessment and in the application this job is delegated to a specific institution. In spite of this, it is crucial that all the institutions which are part of the consortium are involved in the assessment, in one way or another. It can be as simply as deciding how a certain part of the program will be evaluated, but different institutions can have one or more members as part of the assessment team. In the case of educational programs for example, although only certain institutions are officially responsible for developing courses, all the other members take part in the process. Similarly, the institution in charge of the assessment will lead the process, but all members interested in taking part in the process should be encouraged to do so. No members who express their will to take part in the process should be excluded.

Once the assessment team has been established, its members should make a decision about the main reasons for their evaluation. Are they going to focus on the issue of *accountability* or on the issue of *improvement*? Are they going to explain to the whole consortium the way the evaluation is going to be done, or are they going to start by watching how the different groups work and only then decide on the criteria? In order to answer these and similar questions, the evaluators should familiarize themselves with *theoretical issues* related to assessment so that they can decide on a specific approach or a combination of approaches. Hughes & Nieuwenhuis (2008) show that certain organizations give the evaluator the role of "*mentoring*" the teams and making sure they reach their objectives. This is an internal evaluator who may work independently or in collaboration with an external evaluator.

We believe that the leader of the assessment team must start with the premises that not all the members of the consortium have read the application and even if they have, they might have forgotten or overlooked certain issues. Therefore, the first thing that has to be done is to make sure that everyone is aware at least of what the team has committed to achieve in the application package. In addition to that, the team can choose to add certain aspects or to focus on specific elements which have not been part of the original application but which seem very important at the present time.

The very fact that the role of the assessment team has been decided in the application package prevents a lot of problems later on, although other problems can occur and the team members must be prepared to make changes. The good news is that their initial role has been established and when the members agreed to the application, they also agreed to be evaluated in the way which appears in this application. On the other hand, since most programs funded by the European Commission last for two to three years, some changes are inevitable.

Certain institutions or at least certain members of the consortium may not be able to perform the task which they had initially agreed to do and in this case it may be role of the assessment team to mentor those who need mentoring. If this is not possible, then the program coordinator should delegate this task to other members of the consortium. In such a case, the evaluation team should be aware of the changes that have been made and should assess the change that has been made and the quality of the work that is done by other persons/institutions than the ones that were supposed to do it.

The role of the assessment team is to inform the program coordinator of the problems that occur and to ask his or her help in making a decision regarding the changes that have to be done. While some coordinators might be afraid to report to the funding agency problems of this sort because they are not sure what might be the implications in terms of financial penalties, our experience has taught us that the opposite is true. If the program coordinator is upfront and keeps the lines of communication with the funding agency open, solutions are found to deal with the challenges that occur in the process. After all, the funding agency does not have any interest to penalize the program, but rather to help the partners find solutions in order to "get things done" in the best possible way and to justify their investment.

The solutions may be very creative: either someone else will do that part of the program which has been written in the application package, or these specific funds can be transferred to other part of the program which needs additional funds that what had been previously expected. Similarly, a certain coordinator might decide that one aspect of the application works much better than expected and therefore he or she might suggest the funding agency to transfer funds from one section to the other. The role of the assessment team will therefore be to make sure that this incident is reported, not for the sake of what Hughes & Nieuwenhuis (2008) call "shaming and naming", but for the purpose of learning from this challenge and from the way in which the problems have been solved.

This issue brings up the ethical aspect of evaluation which has to do with *confidentiality* and with the measures that are taken in order to assure it. When evaluating an international program formed of several partner institutions, it is quite easy to recognize who the assessment team evaluates, even if no names are mentioned. It is therefore very important to find creative ways to evaluate without shaming anyone while also making sure that the program achieves all its goals.

Even when delicate situations occur, the assessment team should remember that their main role is to evaluate the program, to stick to issues related to content and to process and not to focus on personal issues. The team must evaluate issues such as achieving the objectives and goals that have been agreed upon, relevance, efficiency, implementation, dissemination and sustainability.

It is worth saying a few words about each one of these terms:

1. **Objectives** are the desired or *intended achievements* as a result of a certain intervention or change within a given period of time.

2. **Outcomes** are the *achieved results* involving benefits to the beneficiaries of the program and they are defined in terms of activities or outputs that can be described and observed. The outcomes should be written in terms of the **SMART** model: linked to *specific* objectives that can be *measured*, that are *achievable, relevant* and done in the amount of *time* which was defined. In an international consortium it is important to take into account the fact that the national interests can sometimes be different for the different countries that

participate in the consortium. Therefore the outcomes must be defined in a way that will create a win win situation for all those involved.

3. **Relevance** of a program has to do with the extent to which the established goals relate to the problem that has been identified. It refers to students' learning experiences (*personal relevance*) or to the connection the problem has to real–world problems and contexts (*life relevance*).

4. **Efficiency** has to do with the process of evaluating the *use of resources* (time, money) needed in order to make sure that the activities that were proposed achieve the intended results. The team has to establish concrete criteria in order to make sure that there are no better or cheaper ways to achieve the same goals.

5. **Effectiveness** deals with the *practical benefit* of the program according to the proposal that was funded. The questions that have to be asked are about the *impact* of the program on the population, the *recipients* and the *beneficiaries* of the program or in another words the *level of results* from the actions taken.

6. **Sustainability** Funding agencies such as the European Commission wish to have a significant impact not only during the time period when the program is implemented, but rather to make sure that the positive *results of the program will continue even after the funding stops*.

The assessment team has to decide according to which criteria each one of these six elements will be evaluated.

Another important decision that the assessment team should make regarding the scope of their evaluation is related to the leadership styles of the main leaders of the program: the program coordinator/manager as well as the leaders of each of the teams, including the leader of the assessment team. Is the leadership style conducive to learning, to team building, to encouraging collaboration or does it interfere with the process and hinders the advancement of the program?

As previously stated, it is the role of the team leader to make sure the team members are familiar with the literature in this field and therefore the main styles of leadership will be presented below.

The main leadership styles found in the research literature are: *authoritative* style, *democratic* style and *hands off (laissez faire)* style. The leaders who adopts the first style make most of the decisions themselves starting from the premise that they have the knowledge, the authority and the responsibility. Such leaders do not encourage initiatives from the part of the group members and lead by "carrots and sticks" methods of praising what they consider positive behavior and sanctioning negative behavior. In theory, there might be a place for any of the three leading styles in dealing with specific tasks and environments. It must be said upfront however that the programs funded by the European Commission are based on collaboration between several institutions in different countries and therefore this leadership style is counter-productive for this type of situations.

The *democratic* leaders delegate authority while at the same time providing guidance whenever this is needed based on their experience. They have a moral purpose, share responsibility, encourage initiatives and debates regarding the goals of the evaluation, the scope, the decision making process regarding success or failure of a certain program or any other issue that might be relevant. They are transparent and understand the cultural, social and political aspects which may help or hinder the implementation process. (Sofer, 2011). In our view, such a leadership style is optimal for those interested in leading a consortium of faculty members from different institutions of higher education and for those interested in leading evaluation teams of such programs. Consortium leaders should remember that the participants are autonomous people who have knowledge, experience and understanding of the material, who are motivated and therefore ir ideas can enrich the program and improve it. If the leaders try to impose their views on such people, their resistance will grow and they might even leave the program and cause it to fail altogether.

The *hands off/laissez* faire leadership style, as it can be easily seen from its name, involves a passive role of letting other people make all the decisions and not getting involved in the process. The pretext used by such leaders is that they delegate and share responsibility, but in fact they give up their role as leaders and the members of the organization feel lost or find another leading

figure that they can trust. Our experience in evaluating international programs has taught us that such projects are always initiated by people who have invested a lot of time and energy in the application process and therefore chances are that these leaders do not belong to this third category of leadership style.

When talking about program evaluation at the national level, many times the government has a significant influence and control over the change that is implemented. When assessing international programs, this issue becomes much more complex and sometimes even problematic since different countries and different institutions have their own ways of understanding the problems that have to be addressed and the issues that have to be stressed.

There are several levels of understanding assessment: at the most basic level the evaluators should ask themselves basic questions such as: Did the program *delivered the outcomes* that they had committed themselves to deliver as part of the application? If so, *to what extent* did they do it? Were there any *other effects* (good or bad) that the program had on the participants which were not part of the goals presented in the application process?

On a more complex level, the assessment should take into account not only *quantitative* elements (how many products have been created or in educational terms how many courses have been created, how many books have been written, how many people have implemented the change), but also *qualitative* elements regarding the level of the products, their impact on the population and the issue of sustainability. How many people have been affected by the new product, *in what way* and what are the *long term plans* regarding the specific product that the program? These questions are more difficult to answer.

On an even higher level, questions can be asked regarding the way in which the decisions were made, who made them (top down or bottom up), the leadership styles of those in charge and the level of ownership that all the partners felt in the process. Obviously, these questions are even more difficult than the ones mentioned above. It is job of the assessment team and especially of the leader of this team to define for himself or herself and for the team the level and depth of the assessment.

# Chapter 4 Assessing international programs in education/ curriculum reforms

Curriculum reforms are never easy to implement in any educational setting and this is especially true when dealing with institutions of higher education. These institutions are generally considered to be very conservative and sometimes even a bit detached from the reality on the ground. The well-known metaphor of the "ivory tower" represents the way in which higher education is often perceived. In terms of organizational behavior theories, as it was explained in Chapter 2, the institutions of higher education are *loosely coupled systems* and it is very difficult to impose change. The only way to minimize their resistance is to share a common vision and to act as partners and not top down, especially in international settings.

Given this fact, it is easy to understand that if curriculum reforms are difficult to implement in education in general and in higher education in particular, it is even more difficult to implement changes in a consortium which consists of several institutions, from different countries, with different visions and different backgrounds, serving different types of populations. In order to overcome these challenges, leaders must act with determination and with sensitivity to personal differences, to cultural differences and to political issues.

Any consortium interested in development and assessment of curriculum reforms, must do that based on a solid theoretical background. There are many models of attempting to introduce curriculum reforms and the goal of this handbook is not to present all models, nor to propose one ideal model which must be followed by those willing to start this type of endeavor. Rather, it is to offer several options and to invite the readers to go and search for models of their own, which will fit their purpose.

Prideaux (2003) distinguishes between what he calls "the three states of the curriculum". First, there is the planned curriculum which is intended and developed by designers who have in mind several goals and look for the best ways to introduce them through the new curriculum which they proposed.

Second, there is *the delivered curriculum* which is taught by those who have been introduced to the planned curriculum and they are supposed to transmit it to their students in one way or another.

Finally, there is the *experienced curriculum* which reveals what has really been learned by the students. In the best situation, it fits the planned curriculum and the delivered curriculum and there is alignment between the three. In the worst case, there is no connection between the three, either because the teachers did not properly understand the goals of the planned curriculum, or because they were not able to translate it into student understanding.

Institutions or teams interested in curriculum reform should find ways to distinguish between the three states of the curriculum, and assessment teams should find ways to evaluate each one of these three aspects and think ahead of possible pitfalls in order to be able to avoid them.

We strongly believe that it is a mistake to start by developing the curriculum and only after this part of the program is done, to think about the issue of curriculum evaluation. Rather, we believe that it is better to work a priori with the teams and make sure that they are aware of the criteria for assessment and that they agree with it. Curriculum evaluators should not work top down and impose their criteria on the curriculum planners; on the contrary, they must involve the developers from the beginning and make sure that they are aware of the ways in which the curriculum will be evaluated. If the curriculum designers feel partners in this process and if they agree to the evaluation process, they will be more open and more inclined to cooperate with the assessment process.
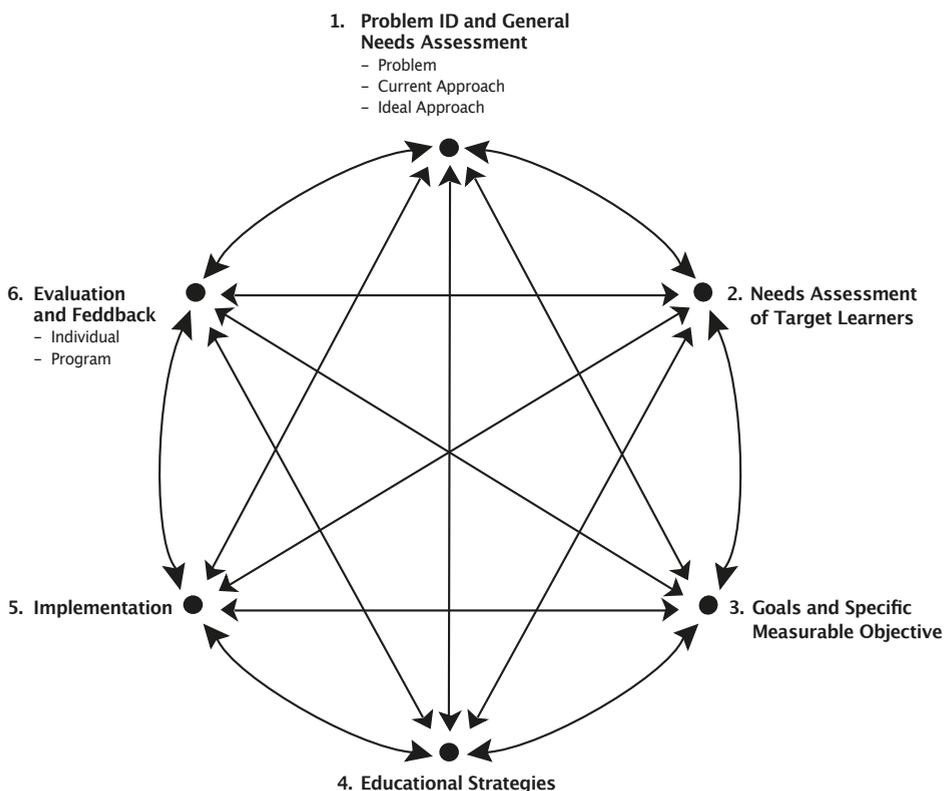
Sheephan & Kirklaud (2005) propose the *six steps approach to curriculum development*. This is a logical and systematic approach which should be followed in order to ensure the quality of the curriculum and its impact on the learners (see the figure on page 21).
The first three steps are *problem identification, needs assessment of targeted learners and finally goals and specifically measurable objectives*.

Our experience has taught us that identifying the problems is not always easy when institutions from different countries are involved in the process, and it is even more difficult to agree on the goals and on the specific objectives, but this part has to be covered during the application process when the partners come together and decide on the program they want to develop. Still, there are several challenges that must be overcome:

1. Sometimes, the application is written by members of one institution and the other partners join without fully understanding what is involved in the process. Therefore, it is strongly advised that the initiator of any application for international programs should start the collaboration process from the beginning and make sure that the objectives are agreed upon by all members of the consortium.
2. Sometimes, even after the application has been sent and before it has been approved, members of certain institutions back off, other staff members replace them and they are not entirely aware of the implication of the international project which has been approved. In such cases, it is the duty of the leader of the coordinating institution to make sure that everyone is "on board" before starting the process.
3. Finally, sometimes the participants agree on the goals at the beginning, but later on they change their mind because of personal reasons that don't necessarily have something to do with the program itself.

## Overview: A Six-Step Approach



1. **Problem ID and General Needs Assessment**
   – Problem
   – Current Approach
   – Ideal Approach

2. **Needs Assessment of Target Learners**

3. **Goals and Specific Measurable Objective**

4. **Educational Strategies**

5. **Implementation**

6. **Evaluation and Feddback**
   – Individual
   – Program

Once the three challenges mentioned above have been taken care of, we believe that one may move to the next three steps proposed by Sheephan & Kirklaud (2005):  The fourth step is to decide which educational strategies (*content and method*) should be used. These strategies should provide the means by which curriculum objectives are achieved. Group discussions should be encouraged in order to be able to choose the most creative strategies that fit the needs of the audience which has been identified in step two.

Once this is done, the next step is i*mplementation* which includes the identification of the resources need in order to be able to implement the program in a successful way. Finally, the last step is *evaluation and feedback*, a stage when the effectiveness of the program is assessed and recommendations are made.

There is a consensus among researchers that assessment is a critical part of any curriculum development because it creates the link between *learning outcomes, content, learning and teaching activitie*s. Boud (1998) suggests that assessment should be related to its impact on learning and he emphasizes the importance of a well done assessment when he says that "students can, with difficulty, escape from the effects of poor teaching, but they cannot escape the effects of poor assessment". In his view, the assessment can be either formative or summative and this is one more case where we can see that the separation of terms between evaluation and assessment. The most important thing for him is that the assessment must focus on encouraging learning and checking the progress made towards achieving the learning outcomes.

Boud and Falchikov (2006) argue for the development of assessment tasks that on the long run promote the students' ability to learn how to learn. This goal can be achieved only if the students are able to gradually make judgments about their own learning.

Brown (2001) suggests that the evaluation team asks several questions regarding the assessment task: What are the *outcome*s to be assessed? What are the *capabilities/skills* in the outcomes? Is the method relatively *efficient in terms of student time and staff time?* What *alternative*s are there? What are their *advantage*s *and disadvantag*es? Do the specific assessment tasks *match the outcome*s and skills? Are the criteria *appropriate?*

This chapter has focused on several models of evaluation without choosing one model over the other. On the contrary, leaders of assessment teams are encouraged to expand their readings on other types of models such as: management oriented, consumer oriented, expertise oriented and participant oriented assessment. Several researchers (Worthen et al, 1997) advocate the use of an "eclectic approach" which combines elements of different approaches in order to better fit the specific program that has to be evaluated.

## Chapter 5   Case study: The assessment team of DOIT (international program for curriculum reform in multicultural education at the higher education level)

The TEMPUS DOIT program was launched three years ago, in October 2012. Its main objectives were stated in the application and they included developing, piloting and implementing curriculum reform in higher education in the field of multicultural education and children's rights. In order to achieve these goals, a consortium of faculty members, students unions and NGOs belonging to over 20 institutions was formed. The members of this newly formed organization would be committed to the same vision as it was expressed in the application and would be willing to collaborate with international participants.

International teams have developed innovative courses as well as out of the classroom activities that promote inter–cultural relations. Then, the program was piloted in the partner countries (see page 2) and faculty workshops were given in order to train other faculty members to implement the courses, thus having a domino effect. The new courses were integrated in curriculum of the participating institutions, research was conducted and the results were disseminated at national and international conferences in the participating countries.

After inaugurating the first consortium meeting that took place in Tbilisi, Georgia, the DOIT professional members chose to be divided into six teams that focused on the following issues: Children Rights, Pedagogical Approaches, Identity Formation, Students' Activities in Georgia and Israel, Ethnic Groups in Georgia and Israel, and Curriculum Assessment. Prior to the consortium meeting, the team topics were posted in the Portal, and the DOIT members were asked to write their names under their preferred topic. However, by the time all the DOIT TEMPUS members arrived to Tbilisi, a lot of members had not registered for the teams, so they were given a second chance to choose where they wanted to be. This turned out to be a problem later on since the teams were not diverse enough and the number of team members was not balanced.

The Curriculum Assessment team consisted of Prof. Susan Jackson from the University of Birkbeck in London, Dr. Roxana Reichman, Gordon Academic College, Mr. Yaser Awad, Sakhnin College for Teacher Education, Dr. Manal Yazbak Abu Ahmad, Sakhnin College for Teacher Education, Dr. Izabella Petriashvili, Tblisi State University, Dr. Nino  Sozashvili from Telavi University, Dr. Tina Gelashvili from Samtskhe-Javakheti State Teaching College in Akhaltsikhe and Mr. Giorgi Gakheladze from Ilia State University.

During the first team session, each member spoke about his/her background and vision regarding this project. Then the work was divided between the group members and a deadline was set. The first phase started with building a Curriculum Glossary that was due before the second consortium meeting (April 30, 2013).  During the second phase the team members started designing a suitable syllabus template for the other groups to use. Several Skype meetings were arranged to discuss the team tasks and progress and the team constantly corresponded via email. In addition, when the Israeli team members came back to Israel, the program coordinator asked the DOIT team from Ben Gurion University to give the Israeli Curriculum Assessment Team a workshop on the Bologna process ECT system (the European Credit Transfer and Accumulation System). The members of this team included Prof. Ronen Peretz and Ms. Hannah Moskovitz and they were in charge of the quality assurance of the program.

The syllabus templates used by all the DOIT members were checked, and then a comprehensive syllabus template that all the teams could use was prepared. This template and the Curriculum Assessment tools were posted on the DOIT portal for all the group members to see. The team leader invited all the members to read it and to make comments on this tool in order to minimize resistance at a later stage.

During the second consortium meeting that took place at the University of Koblenz-Landau in Germany in February 2013, the team's mission was to make sure this process was followed. The Curriculum Assessment team members circulated among the groups while they were in action and handed them a checklist to make sure they included all the required items in their syllabus. The assessment team has presented the rubrics to the entire consortium and has asked for everyone's input, but still certain teams were defensive toward the

assessment team. In spite of the fact that a consensus had been previously reached regarding the criteria, some teams still opposed the fact that the assessment team was there while they were working. In addition, unsurprisingly, all the teams had a problem translating their institution local credit points to the ECT system, so it was left it to each Higher Education Institution to fill it. Also, at this stage, one of the teams was still struggling due to the largest number of team members.

It is recommended that when building teams, several considerations should be taken into account:

1. If it is possible, each team should have representatives from all the institutions or at least from every country that takes part in the consortium.
2. Each team should have members who are specialists in their fields as well as members who can learn from this experience.
3. The sizes of the teams should be more or less equal. This allows the team leader to do a better job and prevents ego issues between the leaders of different teams.
4. Although it is inevitable that many people tend to stick with people they already know, the whole point of collaboration is to meet new people, to try to influence them and to be open to be influenced by them. Therefore, the program coordinator and the team leaders should make an effort to prevent the creation of teams with members that represent the same institution.

During the next meeting which took place in Lintz, Austria in June 2013, the teams finalized their syllabi and they were asked to share them with the entire consortium. By this time, three groups have already finished developing their syllabi and the rest were to hand them in by July 31 at the latest. It is noteworthy to mention that there was a personal conflict between two team leaders about the "borders" that needed to exist, in their view, between the scope of the work that these two teams were expected to perform.  After the set deadline, all the teams except one have posted their syllabi on the portal.

At this point, the outcome of the Curriculum Assessment Team work resulted in a comprehensive Syllabus Template in addition to a Curriculum Assessment Tool which could help educators in higher institutions around the world to write their own syllabus.

Furthermore, the leader of the assessment team has adapted a pre-post Multicultural Education questionnaire and shared it with the members of the

quality assurance team. This survey was to be used in the courses that were piloted in different institutions in order to check whether there was any change in the students' attitudes as a result of taking the courses. The results were analyzed by the quality assurance team and posted on the portal so that all consortium members would be able to use them for their own research. The data collected shows that the students and faculty members who taught these courses found them important and relevant. The faculty members especially appreciated the flexibility of the courses (the fact that they could use the whole course or only certain modules) and their practical applications. The results also showed that the students who took part in these courses had a positive attitude toward the topic of multiculturalism and children's rights to begin with and therefore the changes between the pretest and the posttest were not very significant. However, several changes in the students' attitudes have been evident.

It is important to note that at the end of the program, the DOIT TEMPUS team accomplished much more than it was originally written in the application package. The plan which has been approved and funded required each institution of teacher education that took part in DOIT to implement four of the courses which would be developed. In addition to that, the European partners and the Israeli non-teaching colleges were required to pilot only two of the developed courses. In fact, a total of 58 full courses were piloted and other 39 existing courses use modules of the courses developed as part of the program (a total of 97 courses!). These courses were given both at the undergraduate and at the graduate level.

A major accomplishment of the DOIT program has been achieved as far as sustainability is concerned. New educational programs have become part of the curriculum in different institutions of higher education that participated in the program. Some of the courses that have been developed are compulsory and others are elective and some other courses use modules of the courses which have been developed by the partners in the consortium. It is the firm belief of the coordinator, Dr. Rhonda Sofer, and of the vice coordinator and leader of the assessment team, Dr. Roxana G. Reichman, that this is due to the shared vision among the key stakeholders and faculty in promoting change in the form of curriculum reform. Research has shown that this element is crucial in order to ensure implementation on the long run (Labagnon et al, 2010, Eskerod & Jepsen, 2013). Faculty members of this consortium have collaborated in the field of research and have presented their work at national and international

conferences. This dissemination process has been on going and hasn't ended once the program officially ends, in August 2015.

Three years of hard work and collaboration ended with an international conference that lasted for four days, included 180 presentations from 20 countries. This conference took place from June 29 till July 2, 2015 at the institution that coordinated the DOIT program, Gordon Academic College, in Haifa, Israel. A major achievement of the program is the type of relationships which have been created between the faculty members of the institutions that took part in the program. This kind of relations is difficult to measure in a quantitative way, but it can be assessed using qualitative means such as interviews with the faculty members. The leaders of DOIT believe that they will continue to collaborate in the future and in fact some of them have already started working on research and on other initiatives that have resulted in new applications which have been submitted to the European Commission and which include some of the members of this consortium.

## Challenges

One should mention a couple of challenges faced along the way. First, the group formation was optional, and this resulted in not having diverse DOIT members in this team. Right on the next day of the first consortium meeting, the problem was addressed and one European member joined the team which had included only Israeli and Georgian members. Unfortunately, this member later reported that he preferred to work in the same team with the other partner from his home institution. Hence, the assessment team didn't include any of the European partners and once this problem occurred, Prof. Sue Jackson was asked by the team leader and by the program coordinator to become a member of the Assessment team as well as of the Quality Assurance Team which had been her original task. Luckily, Prof. Jackson showed a lot of flexibility and both teams learned a lot from her experience.

The second challenge was related to the use of the ECT system which turned out to be a difficult process for many DOIT members, especially the Israeli ones, since it is totally different from the grading system that exists in institutions of higher education Israel.

Third, during the first consortium meeting, some team members had difficulty communicating in English and they needed simultaneous translation which was provided by other team members who didn't have prior experience with this

type of work. Therefore, it took some time for other members to translate and the frequent interruptions constantly affected the flow of the team work. Hence, some team members were replaced and thus the problem was solved.

We highly recommend constructing more diverse international teams, so that each team includes representatives from all participating countries. Also, a good level of competency in English language should be required from all the members since such projects are based on international collaboration.

The fourth challenge was at the beginning when some team members did not open up and claimed that there were no problems at all regarding multiculturalism in their country and only at a later stage, confidence was gained and people opened up and started discussing the delicate issues in their countries. The leader of the assessment team had a crucial role in building trust among all the members because without achieving this, it would have been impossible to continue as a team.

A weakness of the program was that one of the participating institutions was not able/willing to perform the task that they have committed themselves to do. Luckily, the consortium members who were affected by this incident did not decide to give up that specific task, but rather they volunteered to do it on their own, thus insuring the continuation of the program according to the application. It is worth mentioning that once the problem became obvious, the program coordinator decided to involve the funding agency as soon as possible in order to allow the TEMPUS office in Israel and in Brussels to take the necessary measures. At no point in time was this issue kept secret from the funding agency but at the same time, most consortium members who were not affected by this problem did not find out about it and the confidentiality was therefore carefully protected. The reaction of the funding agency was positive and this is the place to once again strongly suggest to ask for their advice whenever problems arise and to trust their experience and their good will to solve the issues in a positive way.

Finally, we recommend to keep the lines of communications open at all times between the program coordinator, the vice coordinator, the team leaders and the members of the consortium and to take full advantage of technological means such as a portal, emails, skype meetings, video conferences and, last but not least, consortium meetings. It is our experience that this approach creates personal and professional ties which continue beyond the scope and the time of the program.

# References

Bonk, C. J, Wisher, R & Nigrelli, M. (2004) Chapter 12. Learning Communities, Communities of practices: principles, technologies and examples in Littlton, Karen, *Learning to Collaborate*. Nova. USA.

Boud, D. & Falchikov, N (2006). Aligning assessment with long term learning. *Assessment and Evaluation in Higher education*, v 31, 4, p. 399–413.

Boud, D. (1998). *Assessment and learning – unlearning bad habits of assessment*. Presentation at the conference "Effective assessment at University", U. Of Queensland, November 1998.

Brennan, R.L (2000). (Mis)conceptions about generalizability theory. *Educational Measurement*; Issues and Practice, 19, 1, 5–10.

Brown, G. (2001). Assessment series 3. *Assesment: a guide for lecturers*. LTSN Generic Center: York.

Gilmore, T. N., Hirschhorn, L & Killy, M. (1999). Challenges of leading and planning in higher education. CFAR, p.1–14, Philadelphia, PA.

Hughes, J. & Nieuwenhuis, L. (2008). *A project manager's guide to evaluation*. Evaluate Europe Handbook Series Volume 1.
http://www.pontydysgu.org/wp-content/uploads/2008/02/ EvaluateEurope Volume1final.pdf

Institute for teaching, learning and academic leadership. www.itlal.org. Retrieved on June 28, 2015.

Kirkpatrick, D. (1994) Techniques for evaluating training programs
http://iptde.boisestate.edu/FileDepository.nsf/bf25ab0f47ba5dd 785256499006 b15a4/693b43c6386707fc872578150059c1f3/$FILE/Kirkpatrick_96.pdf

Littlejohn, S. W. (2008). *Theories of human communication*, 3rd ed. Belmont, CA: Wadsworth Publishing Company.

McMilland & Chavis (1986). Sense of community: a definition and theory, *Theory and concepts*, v 14, 1, p. 6–23.

Orton, J.D. & Weick, K.E. (1990). Loosely coupled systems: a reconceptualization. Academy of Management Review, vol. 15, 2, p. 203–223.

Patton, M. (1997). *Utilization-focused evaluation: The new century text*, 3rd ed. Thousand Oaks, Ca.

Peretz, R & Moskovitz, H. (2014). *Assessment and evaluation in DOIT*. Paper presented at the DOIT consortium meeting, Landau. (p. 455–461).

Prideaux, D. (2003). Curriculum design. *BMJ*, Feb 1, 326, 268–270.

Provus, M. (1971). *Discrepancy evaluation*. Berkeley, CA: McCutcheon Publishing Co.

Scriven, M. S. (1991) Pros and cons about goal free evaluation. *American Journal of Evaluation*, v 12, 1, p 55–62.

Scriven, M.S. (1967). The methodology of evaluation. In R.E. Stake (Ed), *Curriculum evaluation*. Skokie, IL: Rand McNally.

Secolsky, C & Denison, D.B. (2012). *Handbook on measurement, assessment and evaluation in higher education*, 2nd ed. NY: Routledge.

Sergiovanni, T. (2007). Rethinking leadership: *a collection of articles*, 2nd ed, Corwin Press, Thousand Oaks, Ca.

Sheephan & Kirklaud (2005) Six steps approach to curriculum development. http://www.oucom.ohiou.edu/fd/Workshops%20for%20RPAC/ED2%206 Step% 20Process%20of%20CD.pdf
Retrieved on July 1, 2015.

Sofer, R (2011). A model for structuring relations in a Tempus international partnership: a narrative presentation, in N. Davidovitch (ed), *International ties in higher education and academic teaching*, Ariel University Center of Samaria, Ariel.

Stake, R. E. (2005). Responsive evaluation. In S. Mathison (Ed) *Encyclopedia of evaluation*, p. 376–379. Thousand Oaks, Sage.

Stufflebeam, D.L. & Skinfield, A.J. (1985). Objective oriented evaluation: The Tylerian tradition. *Systematic evaluation*, v 8, p. 69–87.

Stufflebeam, D. L. (1971). *The relevance of the CIPP evaluation model for educational accountability*. Paper presented at the Annual Meeting of the American Association of School Administrators, Atlantic City, NJ, Feb. 1971.

Taylor, C.L. (1993). Foundation concepts essential for evaluation: criteria/evidence/judgement. http://edis.ifas.ufl.edu/PD022

Tyler, R. W. (1942). General statement on evaluation. *Journal of Educational Research*, 35, p. 492–501.

Van der Knapp, P. (2004). Theory–based Evaluation and
Learning: *Possibilities and Challenges Evaluation*, v 10 (1) p. 16–34.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). Program evaluation: *Alternative approaches and practical guidelines* (2ⁿd ed.).
New York: Longman.

**Dr. Roxana G. Reichman** is an expert in Comparative Education and Educational Administration, Organization and Policy (PhD. from State University of New York at Buffalo, 1996). Dr. Reichman has published in the field of higher education, educational reform, as well as in the field of multicultural teacher education. She has been vice coordinator of Tempus Program DOIT since 2012 and, as expert in curriculum development and assessment, she has been leading the curriculum assessment team of this program. In addition to that, during the last two years she has also been a member of the assessment team of the Tempus program LLAF (Lifelong learning in Applied Fields).

During the last two decades, Dr. Reichman has worked for ten years in the Department of Educational management at Haifa University and for 19 years at Gordon Academic College of Education, teaching undergraduate and graduate students, pre service and in service teachers and administrators. She has held several administrative positions for GACE Undergraduate and Graduate programs in teacher education, including Academic Head as well as Head of the MEd program in Learning, Instruction and Mentoring. She acts as Erasmus institutional coordinator of faculty exchange as well as students and administrative staff exchange.

**Dr. Manal Yazbak Abu Ahmed** has her M.A. in English Literature and her PhD in Educational Administration. For six years, she has been co-lecturing a joint intergroup collaborative e-learning course on "Dealing with Diversity" between the Arab students at Sakhnin College and Jewish students at David Yellin College in Jerusalem. Furthermore, she teaches a TEC course entitled: "Exploring Culture through English Literature" with five different Arab and Jewish colleges in Israel. Dr. Yazbak Abu Ahmad is an expert in e-learning collaboration and teaching strategies in a multicultural environment and international relations, and she has started teaching a "Global understanding" online course with the University of Michigan during her Fulbright year in 2014-15. Finally, she has been an active and integral member of the Curriculum Assessment and the Children Rights teams in TEMPUS DOIT.