IMPACT MEASUREMENT REVISED: EVALUATING THE IMPACT OF LIFELONG LEARNING PROGRAMME IN SLOVENIA

Klemen Širok, University of Primorska – Faculty of management Koper, Slovenia
klemen.sirok@fm-kp.si

Gregor Petrič, University of Ljubljana - Faculty of Social Sciences, Slovenia
gregor.petric@fdv.uni-lj.si

ABSTRACT

European Union programmes in the field of education and training demand regular evaluation. However, due to several conceptual and measurement issues, the validity of evaluation results is put into question. The paper proposes to assess the programmes from a refined perspective of impact, calling for the clarification of some key methodological issues. Two groups of problems for the validity of evaluation research in the field of educational programmes are identified, one referring to causality and second to the absence of proper conceptualisation and operationalisation. The case of LLP impact evaluation in Slovenia is used to present the solutions to outlined problems.

Keywords: impact assessment, Lifelong Learning programme, evaluation, methodology, validity

1. Introduction

The Erasmus, Comenius, Socrates and Leonardo da Vinci programmes have been present in the Slovene education system since 1999. Between 2000 and 2006, more than 3,600 teachers and mentors, 4,000 students, 3,000 pupils, and 500 organisations participated in these actions of European education and training funding programmes in Slovenia. Since these programmes have been predominately funded by the European Commission (Mihelič Debeljak & Pajnič in Taštanoska, 2006; CMEPIUS, 2007), questions immediately arise regarding the use of funds and the achievement of the programmes’ stated goals. The same evaluation issues apply to the new European education and training programme - Lifelong Learning Programme (European
Parliament and the Council, 2006) – that presents the next generation European Community action programme in the field of education and training.

Assessing the impact of programmes and policies usually calls for experimental or quasi-experimental research designs. Due to several conceptual issues (i.e. vague understanding of impact) and measurement issues (i.e. ex-post evaluation context, or no opportunity to monitor or manipulate the amount and quality of obtained programme intervention), the validity of evaluation results of these programmes are put into question. Furthermore, it seems that the existing approaches to the evaluation of the above programmes are limited to the investigation of basic outputs and outcomes (e.g. funds, number of participants, etc.). We propose to assess the impact of programmes from a refined perspective of impact, calling for the clarification of some key methodological issues (i.e. causality, conceptual and measurement issues).

Our paper starts from a premise that in order to conduct a valid and reliable assessment of policy programme impacts, it is necessary to solve some fundamental conceptual and measurement issues, which arise when trying to assess empirically the true impact of a certain programme. We propose an approach that considers several of these issues and, through the empirical case of assessing the impact of the Lifelong Learning Programme (LLP) in Slovenia, point out and discuss several pitfalls that appear in assessing the micro impact of macro level policy programmes. The LLP impact evaluation in Slovenia also serves as an example, how to practically address presented methodological issues within the limitations posed by the evaluation context.

The text is organized as follows. The next section provides a summary of impact assessment in the European Community (EC) education policy. Section three discusses the concept of impact and two fundamental topics, vital for valid assessment: the issue of causality; and issues of conceptualization and operationalization. The forth section presents our research approach to measuring and explaining the impact in LLP impact evaluation in Slovenia, focusing on solutions applied to the outlined impact assessment pitfalls. We conclude the paper by outlining viable directions of further development of presented impact assessment approach.
2. Approaches to assess impact

In this section we examine a selection of general EC evaluations of education programmes and/or appertaining actions carried out in the period 2002 – 2008, covering the majority of EU countries (McCoshan et al., 2008; Bracht et al., 2006; Ernst and Young, 2003, 2006; Sahlin et al., 2005; ECOTEC, 2004; European Commission, 2001, 2002, 2004a; ECOSFERA, 2004; Frinault et al., 2004; CSES, 2004; Yellow Window Management Consultants, 2002; Deloitte and Touche, xxxx)\(^1\). In our analysis we focus on applied evaluation design and methodology and, to a lesser extent, evaluation findings.

\(^1\) The following evaluations were examined:
With respect to evaluation design, the majority of examined evaluations are structured in a similar way. First, evaluators examine and analyse principal programme documents. Findings of prior evaluations are also taken into consideration, usually followed by the presentation and analysis of general statistics. This typically forms the foundation for subsequent questionnaire development, presentation of programme background, analysis of financial data and/or programme end-users, analysis of action application, and analysis of programme/action projects (number and time span). Occasionally, the starting evaluation stages include interviews with important European Community education programme stakeholders, administrators and/or experts. Expert opinions help to identify programmes’ end-user needs or determine other relevant evaluation variables, including indicators or expected target values. The second evaluation stage usually consists of a web survey. Paper and pencil or telephone surveys (or telephone interviews) are rarely carried out. Common characteristics of analysed surveys are relatively low response rates and problems related to missing or low quality data. The majority of evaluations, thus, triangulate different stakeholder views as a strategy to provide reliable evaluation conclusions. Stakeholders covered in evaluations range from micro to macro levels: national agencies; programme/action coordinators; programme end-users, such as school or company management; local authorities; and partner organisations.

Next, the majority of evaluations attempt to gain insights and/or additional feedback with regard to carried out survey from relevant programme stakeholders. This is normally carried out using telephone interviews, site visits, or most frequently group interviews. Often the last evaluation step is evaluation (i.e. dissemination) seminar. It serves to identify or explore the content areas that demand special attention, or to present or identify cases of good practice usually on a national or institutional level.

The second and third stages the evaluation focus on end-users’ and primary stakeholders' opinions related to national policies and goal attaining mechanisms, grounds for their selection, or reactions from governments and administration to certain EU programmes. Most often evaluators seek to determine the quality of processes and procedures, including the initial programme/action implementation stages, effectiveness and/or efficiency of

implementation, and administration. In some cases evaluations seek out new or improved approaches to programme implementation, problem solving, and programme management. In doing so, evaluations focus on initial programme outcomes in relation to programme goals, or seek to determine programme goal relevance in relation to target group needs, policy context, or other programmes. Programme impact assessment is rarely performed to determine its effectiveness or efficiency is rarely performed\(^2\). If yet, mostly the effects of mobility on student/teacher professional development are determined and explained through greater extent of their international and intercultural competences, vertical or horizontal professional mobility, and international mobility. Impact assessment, thus, focuses predominantly on the individual level, finding impact in competencies, skills improvement, and language improvement. Impact on the organisational and policy levels is also assessed. These evaluations poorly explain the role of contextual factors on European Community action programmes in education and training on all levels: individual, organisational, and system/national.

The evaluation approaches that were examined encountered at least three sets of problems:

1. Vague and broad understanding of the term 'impact'.
2. Poor designs, not applying experimental or quasi-experimental research designs being prone to problems such as: (a) causality issues; (b) conceptual confusion; and (c) inconsistency relating to the unit of analysis, resulting in incoherent impact perception with regard to the individual, institutional, or system/national level.
3. Measurement problems generated by applying measuring instruments or data gathering methods that are biased due to direct socially desired indicators, predominately focusing on facts and opinions and often ignore knowledge, beliefs, and values.

Next we present an approach to the evaluation of micro-impact of macro-level policy programmes that considers these issues. Using the case of an empirical assessment of the LLP impact in Slovenia, we demonstrate a practical solution to presented methodological issues developed within the limitations posed by the evaluation context.

\(\text{\footnotesize \textsuperscript{2} As a special case, the recent EU evaluation (McCoshan, et al., 2008) comprehensively evaluates Socrates II, Leonardo da Vinci II, and eLearning programmes. The evaluation methodology exceeds the above examined evaluations, unequivocally uncovering relevant methodological issues relating to evaluation logic and sought programme outcomes, distinguishing between impact at the practical level and impact at the socio-economic level. This evaluation also relates programme goals stated in the EU decision to expected impact.}\)
3. Impact measuring approach

3.1. What is impact anyway?

Whilst the term impact frequently appears in evaluation documents, its meaning is not entirely clear (Weinwright, 2003). Since an exact definition of a concept is a precondition for any kind of measurement that tries to be valid, we first ask ourselves what exactly is meant by the concept of an impact. We believe that a fruitful starting point to answer such a question is to consider the ‘impact chain’. This idea is build upon definitions by Blankenburg (1995) and Weinwright (2003) who propose that impact refers to long-term and sustainable changes introduced by a given intervention in the lives of beneficiaries, related to the specific objectives, an intervention, or to unanticipated changes caused by an intervention.

Figure 1: The ‘impact chain’ from action to impact

In this chain impact is a sustained consequence of a certain action/policy, and is usually latently present but not directly observable. This chain helps us to quickly conclude that the majority of evaluations, which we examined, mostly focus their attention on the outcomes of an action, on the first part of the impact chain. The main reason for this probably lies in the difficulties of measuring impact within temporal/financial constrains, which often results in outcome focused evaluations and measurement portraying impact through directly countable policy outputs and outcomes (i.e. in terms of money spent, number of students/institutions involved, etc.). Considering the stated goals of the Lifelong Learning Programme (and many other EU policy programmes), it will be demonstrated later that relevant impact assessment for policy makers comes not only from investigating the first links of the impact chain (i.e. outputs and outcomes), but mostly through investigating the last part of the chain: impact as a latent, sustained consequence of a programme.

---

3 Outputs are countable units, and are the direct products of a programme or organisation’s activities. They are not the objectives of the organisation. Outcomes are the benefits or changes for intended beneficiaries. They tend to be less tangible and, therefore, less countable than outputs. Outcomes are usually planned and are, therefore, set out in an organisation’s objectives. Outcomes may be causally and linearly related; that is, one outcome leads
It is not sufficient to situate impact in the above presented chain to conduct a valid impact assessment. We suggest that impact is a very general construct and cannot be operated within its generality, but has to be situated in a certain context. Within (policy) evaluation, various stakeholders play major roles in determining such contexts and without examining context, impact assessment would be meaningless. Methodologically speaking it could be claimed that trying to assess impact is similar to trying to assess an attitude. Assessing attitude cannot be separated from the object towards which attitude is expressed (Bradburn, et al., 2004; Henerson, 1987). We can assess attitudes towards abortion or the death penalty, but not attitude per se. Similarly, we claim that assessing impact of an action in its generality is meaningless and, thus, needs to be situated in a certain context, which is framed with various parameters, such as: (a) the subjects or beneficiaries of the impact: individuals, organizations, or states; or (b) the substantive dimensions of impact. Is impact experienced in changed attitudes (i.e. toward other cultures), beliefs (e.g. about one's own employability), improved knowledge (i.e. communication competences), enrolment in activities (e.g. getting a job) or something else? For an evaluator such issues are further complicated by the possible contradictions of various stakeholder contexts. For instance, one can imagine the Leonardo da Vinci mobility projects. The basic understanding of employability evidently differs between pupils, the unemployed, young graduates, and vocational education and training experts. Furthermore, defining the domain of impact is often subject to normative issues, such as basic, normative policy documents (European Parliament and the Council, 2006) that rarely explicitly express the interests of certain impacts.

Situating impact in a certain context and substantial domain demands clarification of the above issues. Operatively, this means associating impact with some specific phenomena, as attitudes are connected with some specific object. More precisely, taking into account the interests of stakeholders, subjects of impact, substantive dimensions, and the logical reasoning on the whole impact chain might lead one to conclude that the impact of the Lifelong Learning Programme might be inter-organizational cooperation, individual tolerance, or end-user employability. The European dimension is the contextual domain for the case presented in the next chapter. Situating impact in a specific context thus means partitioning the impact on areas where latent, sustainable consequences are anticipated or reasonably expected. Such to another, which leads to another and so on, forming a linear sequence of if-then relationships (Weinwright, 2003).
partitioning results in a list of theoretical concepts (Meehan, 1994) that grasp the impact domain (i.e. cooperation, tolerance, employability), assuming that the action has an impact: greater cooperation of organizations involved in the action, more tolerance, and better employability of programme/action end-users. Evaluation theory as well as praxis (see Rossi et al., 2007) often approach impact measuring by emphasising appropriate research design, starting with quasi-experimental and ending with experimental evaluation approaches. Yet, theory neglects many of the above discussed methodological issues that precede research design decisions. In our research we propose that the evaluation process should approach seriously the question of how to measure appropriately complex and often latent concepts/variables that pertain to the impact domain.

Such reasoning raises two fundamental issues when trying to measure the impact of policy programmes: (a) the issue of causality between the policy action and the particular impact domain; and (b) the issue of conceptualizing and measuring a particular impact domain.

3.2. Causality issue

On the operational level, impact is most often understood as the difference between the state after the policy intervention and the state before the policy intervention (i.e. Davies, 1999; Leigh, 2009), where the amount of this difference is caused by the policy intervention. It could be claimed that the majority of existing approaches take the policy intervention as a cause (X) and the resulting phenomena (Y₁, Y₂, Y₃…) as a consequence of this cause.

Such a model assumes that X is the main cause of Y₁, Y₂ and Yₙ. This is legitimate, but a methodologically insufficient claim for true impact assessment. In the social sciences, phenomena (such as Y₁, Y₂,…, Yₙ) are usually determined by several factors and cannot be exclusively determined by single causes (Neuman, 2003). Experimental and quasi-experimental designs should, thus, take into account that some impact domain – Y₁ (i.e. employability) cannot be exclusively caused by the action (X), but can be caused by other
factors \( (X_1, X_2, \ldots, X_n) \), which should be controlled for in a empirical research. We propose that research design for assessing impact should be based on the following model:

\[
\begin{align*}
X & \quad \rightarrow \quad Y_1 \\
X_1 & \quad \rightarrow \quad Y_1 \\
X_n & \quad \rightarrow \quad Y_1 
\end{align*}
\]

This model seems to be best verified within a panel research design, where the difference between \( Y_1 \) in time after the action \( (Y_{1t1}) \) and \( Y_1 \) in time before the action \( (Y_{1t0}) \) can be attributed to the participation of beneficiaries in action \( (X) \), controlling for \( X_1, X_2, \ldots, X_n \). We, however, claim that an ex-post research design is also suitable to verify such a model, under specific assumptions: (1) \( Y_{1t1} - Y_{1t0} \) is not computed, but can validly and reliably be measured by respondent's subjective evaluation of this difference; and (2) the effects cannot be attributed to the (non)participation in action \( (X) \), yet \( X \) can be treated as an intensity of involvement/participation in an action. This seems to be a valid claim since educational actions (if not all policy actions) are a social process in which beneficiaries are involved in different ways. For instance, the beneficiaries of the Erasmus programme can formally study abroad, but not visit any lectures, whilst other beneficiaries can intensively participate at all lectures. In this sense, \( X \) seems to be more validly treated as a scale variable of intensity of participation rather than a nominal variable of (non) participation at the action. This claim, however, needs to be empirically verified in our further empirical work.

Independent of research design, each study should take into account as many relevant control variables as possible. Since evaluations are usually constrained by time and financial pressures, a careful selection of relevant control variables is needed. We shall present in our case that such selection cannot be trivial, but should be guided by a strong theoretical understanding of a particular impact domain in cooperation with primary evaluation users. For example, when investigating the impact of an action on the European dimension one should carefully examine literature on the European dimension to identify important control variables, such as duration of mobility, psychological involvement/attachment to activities.
organised within mobility (intensity of participation), or history of EU programme involvement.

We can conclude from the above discussion that two types of impact can be assessed on the empirical level. One is the ‘explanatory impact’, which is a value of the parameter pertaining to the influence of involvement in an action (either 0/1 or intensity of participation) on a particular impact domain and is usually expressed with a regression parameter, ranging from -1 to 1. This is usually verified with explanatory multivariate analytical procedures, such as regression analysis or path analysis. Second, is ‘measured impact’, which is a mean value of a variable measuring a specific impact domain, either as a difference between two states or the subjective evaluation of the difference.

3.3. Conceptualization and Operationalization

Whichever research design or impact understanding we choose, it is inevitable for a valid impact assessment to arrive at valid and reliable measures of particular impact domains: either a valid and reliable measure of $Y_{t0}$ and $Y_{t1}$ in a panel design, or a measure of subjective evaluation of $Y_{t1} - Y_{t0}$ in an ex-post design. In order to arrive at good measures of particular impact domains, it is most reasonable to follow the standard procedure of positive methodology and strictly follow the steps of conceptualization and operationalization (Neuman, 2003). In this process we discuss a few crucial issues that should be discussed and decided: how to select the appropriate concept definition in relation to programme goals; programme reality; selection of an appropriate impact measure; or proper impact level.

**Proper concept definition.** There is no valid measure without a proper theoretical definition of an impact domain (Neuman, 2003). For example, we cannot validly measure the impact of an action on the employability of individuals if we do not have a clear definition of employability. Moreover, a single concept has numerous definitions. In our approach, we support argument made by the majority of the evaluation theoreticians (Patton, 1997; Rossi, et al., 2004): the primary stakeholder(s) should be integrated into research process, adding the interest focus of research and end-user stakeholders to provide additional information on relevant, operative definitions of concepts. We would, however, wish to add that these should not be exclusive criteria. Proper definitions should also be judged to be essential (and not ostensive or nominal) or to imply some variability.
Selecting appropriate impact measuring instrument. Conceptualization issues directly influence the process of operationalization. The goal is to arrive at measures with high quality in terms of validity and reliability. In doing so, a chain of associated issues should be resolved: (a) identification of a suitable unit of analysis; (b) decision on the category of the impact domain in terms of whether it is attitude, belief, value, fact, knowledge, activity; (c) adopting existing measures to the field of impact domain and/or creating new measurement instruments; and (d) temporal and financial constraints in the data collection process (Neuman, 2003; Scherpenzeel & Saris, 1994). By taking these parameters into account, one arrives at a measurement instrument, which should have satisfactory face validity, or certainty that the measurement instrument is indeed measuring impact domain as it is defined. Furthermore, it should be clear that there is a suitable actor in the survey or content analysis (i.e. an individual beneficiary, organizational representative, or action documentation source), and that the items and scales are suitable for a particular category of impact domain. It should also be clear that the data collection process is optimal.

4. Implementing the alternative approach

In this section we present the implementation of the above presented approach to impact assessment, examining the case of individual mobility action of the Erasmus sub-programme of LLP.4 The alternative approach is being used to measure impact of the Lifelong Learning Programme in Slovenia, from 2007 to 2013. For the Erasmus case, we already have data from the first wave of data collection5.

From the onset, the principle of utilization-focused evaluation (Patton, 1997) was strictly followed by the active integration of primary evaluation users in the development and testing phases of the survey questionnaire. To this end, the Slovenian National agency established a working group consisting of members of the contracting authority and evaluators in order to determine the fundamental premises of ongoing evaluation. First, the LLP decision (European Parliament and the Council, 2006) was carried out in order to determine the structure of

4 We can briefly describe the Erasmus action as a study of the mobility of youth and academic staff. Student mobility lasts from at least three months to the maximum of one year. Teacher mobility, which lasts up to six weeks, has the purpose of teaching knowledge and methods transfer and enrichment of educational programmes. Mobility of other staff in higher education institutions lasts from a minimum of five working days to up to six week and has the purpose of additional training.

5 The plan is that the data for each LLP will be collected in more waves: at least three months and not more than one year after the end of the participation in the programme. The specificity of first wave data is that it collects information from respondents that participated in the LLP in one of the recent three years (2007 to 2009). This was however taken into account in our empirical assessment of impact.
programme goals. Next, the contracting authority was asked to list up to five goals that capture the essence of every evaluated action within the LLP programme. In the next stage, selected goals were examined and anchored to appropriate sociological concepts. Selected concepts were then reviewed by the contracting authority, which added points and areas of interest such as customer satisfaction and additional contextual variables. Lastly, the 'action-impact domain-measurement level' grid was formed, integrating the appropriate impact domain (i.e. employability) with the corresponding programme action (i.e. Leonardo da Vinci mobility projects) and measurement level (individual or organisational).

For the purpose of this paper we will present impact assessment in the domain of European dimension. This impact domain is especially complex and problematic, since little or no scientific literature exists regarding its definition, although this concept very frequently appears in the EU and national documents.

4.1. Resolving the conceptual confusion

The first step was to find a suitable theoretical definition and problems arose immediately. A common definition of the European dimension is ‘a principle present in education system enhancing understanding of wider European context and educational perspectives, opening horizons of global thinking and intercultural understanding’ (Walterová & Ježková, 1999, p. 26). This is clearly not a suitable definition for operationalization as it emphasizes the consequences of the European dimension (i.e. enhancing the understanding of European context and intercultural understanding), instead of focusing on the essence of these phenomena. The definition by Seebauer (2002) seems to be more precise as it points out that the European dimension includes both the dimension of cognition (knowledge of Europe) and the dimension of affection (relationship, attitude, and experience of Europe). Yet, it seems to be too broad. It also includes a wide variety of aspects: anthropological-existential; cultural; cognitive; emancipatory; participatory and effective; qualitative; economic; social; egalitarian; and communicative, as well as the aspects of protection, mobility, and security (Seebauer, 2002).

After a careful investigation of the existing literature we concluded that the European dimension is a complex theoretical construct and not yet a measurable concept, which should be partitioned into several dimensions (EuroBio, 2009):

1) Belief in a common and diverse European culture;
2) Sense of European identity;
3) Acknowledgment of values of European civilization (democracy, social rights, human rights);
4) Inclusion in the economic and social development of a community;
5) Awareness of the advantages and challenges of European Union; and
6) Knowledge of historical, cultural, economic and social aspects of European Union and its members.

These six European dimensions can be treated as individual concepts. Each can be operationalized with a separate measurement instrument.

4.2. Theoretical rationale for impact

In this phase we pay attention to the ‘activity-output-outcome-impact’ chain and identify good reasons why the participation in the Erasmus action (and the levels of its intensity) might influence the above six dimensions. It is possible to argue on the basis of social identity theory (Tajfel & Turner, 1984) that the belief in a common and diverse European culture, and the sense of European identity, would rise due to the participation in an action, as the individual experiences other environments and other social groups in which he/she is forced to retain positive self-image and relativize existing social and consequently also personal identity. It is expected that impact will be larger with more intensive and lengthy participation in the programme, as social identity becomes more stable and eventually transforms into personal identity. It is also possible to argue that with the intention to adapt to specific ‘European’ culture, an individual should involve himself/herself more actively in the environment. This includes participation in the economic and social development of community, building knowledge of common European culture, and thus becoming acquainted with historical, cultural, economic, and social aspects of the European Union and its member states.

In addition, the influence of participation in an action should be controlled by various factors, which intervene with an individual's confronting a foreign environment. Typical socio-demographic variables such as political orientation, education, and religion also play a role. Research into identity commonly takes into account the ethnicity, citizenship of parents, and connections with supranational environment. In our questionnaire we were very limited in the inclusion of control variables, since we were evaluating the validity and reliability of eight
other impact domains, which resulted in an already lengthy questionnaire, yet we were able to include at least several control variables such as study success, time distance from participation in LPP, age, gender, social class...

4.3. Measurement instruments

In the first wave of data collection we applied an ex-post design, which will later on in the research process be validated with the panel design. A decision was made to measure all six dimensions of the construct European dimension in terms of subjective evaluation of impact on a particular impact domain. We excluded the dimension of acknowledgment of values of European civilization from the analysis because we had not yet identified a measurement instrument that had satisfactory face validity.

For measuring the dimension ‘sense of European identity’ we used an instrument from Eurobarometer (European Commission, 2004), which is composed of a simple question: ‘People are different among each other in their attachment to their village, town, region, country and also Europe. To what extent do you consider yourself as an European?’ Values ranged from 1=not at all to 4=to a large extent. To assess an impact we simply added a question: ‘Would you say that you see yourself as European to larger extent due to your participation in the Erasmus action?’, which had nominal answers of 1=yes and 2=no. For the other five dimensions of the construct European dimension, no existing measurement instruments could be found, so we decided to develop them ourselves. They differ from measuring the dimension of the sense of European identity in three respects: (a) each measurement instrument is composed of a battery of items in order to check for reliability of the instrument; (b) individual items are positive and negative (measuring the absence of impact) in order to control for yes/nay-sayers and social desirability; and (c) the subjective impact assessment was not used in a separate question, but integrated in a single scale, except in the case of measuring acknowledgment of common European values. They are presented in the section 4.6 together with their descriptive statistics.

4.4. Data collection

Nowadays it seems that in such evaluation studies, which are based upon responses of individual participants in actions, a web survey is the preferred mode of data collection. This is not surprising due to low costs, good control, and short timings, whilst the problem of low
response rate still persists (Vehovar, et al., 2008). This is a pitfall, which we were especially worried about, since our pilot study was around half an hour long and put a significant time burden on respondents.

The data for the study were collected with a web survey in December 2009. We received a list of e-mail addresses of all participants (N=2279) in Erasmus programmes for previous years (2007-2009) and sent an e-mail invitation to these addresses with a link to web survey. Among the 2279 e-mail addresses, 24 of them were non-existent, whilst 482 individuals answered the entire web survey, which resulted in a high (relative to other web surveys) 21.4 per cent response rate. This might be a consequence of the fact that invitations to the students were sent from our organizational e-mail addresses, showing our affiliation with the academic institution, which might consequently initiate feelings of obligation in students.

Our obtained sample had the following socio-demographic structure: 72.7 per cent were female; average age was 24.5 (stdev=3.15); 60.2 per cent were middle class; 22.6 per cent were in a high socio-economic class; and 17.2 per cent were low class. 11.8 per cent of respondents studied natural sciences; 39.7 per cent social sciences; 21.8 per cent humanities; and 6.9 per cent technical sciences. The vast majority of students (63.5 per cent) considered themselves successful with their studies, while 27.8 per cent estimated that they were very successful. Obviously our sample was quite biased in terms of gender structure (although the population is also female dominant), which should be kept in mind whilst reading the results of statistical analysis.

4.5. Investigating the quality of measurement instrument(s)

Before the empirical assessment of the impact we evaluated the quality of the European dimension measurement instrument. The quality in measurement theory can be approached from various perspectives, whilst for the purposes of this case we shall focus only on issue of reliability. More specifically we used confirmatory factor analysis to estimate internal consistency and unidimensionality of the individual dimensions of the European dimension construct, which consisted of multiple item sets and also to estimate internal consistency and unidimensionality of the European dimension construct as a whole. Internal consistency was also estimated using Cronbach’s coefficient alpha, where the rule of thumb is that a alpha value of 0.6-0.7 indicates acceptable reliability, and that a value of 0.8 or higher indicates good reliability (Brown, 2006).
We first conducted confirmatory factor analysis\(^6\) on all indicators to verify if the proposed measurement instruments fit the data. For this purpose we used Lisrel 8.70 software (Jöreskog & Sörbom, 2004), which assesses whether the proposed factor structure fits the data and provides information on factor loadings, their statistical significance and also how much each item contributes to the measurement of the concept, and how large is the associated error variance (Brown, 2006). It needs to be noted that two concepts were inserted in the model with only single indicator, thus their factor loading was fixed to one and consequently error variance could not be estimated. One concept was the impact dimension of a “Sense of European identity”, which was measured with a single item, while the second measure was “Inclusion in the economic and social development of the community” The latter was measured with effect indicators, which compose an index and not with a scale, which would consist of effect indicators, that are caused by underlying construct (De Vellis, 1991). For this reason we a priori computed an index, which was inserted in the factor analysis as a single-item variable.

The initial factor model was a little above the limit of the acceptable fit ($\chi^2=344.98$, df=69, $p<.01$, RMSEA=0.091, RMR=0.14, GFI=0.96, AGFI=0.94), so we modified the model by taking into account the values of modification indices and inserted two error covariances: 1) between two indicators of impact on knowledge on EU that measured heightened knowledge on country, where action was taking place; 2) between an indicator of impact on awareness of EU advantages and knowledge on EU. Both error covariances are suggesting that some other common factors between the mentioned indicators might exist. Since substantive reasons can support this claim, we decided to accept this modification. The modified factor model demonstrates a moderate to good fit ($\chi^2=213.09$, df=67, $p<.01$, RMSEA=0.061, RMR=0.097, GFI=0.98, AGFI=0.97) to the data according to various criteria (i.e. Browne & Cudeck, 1993), altogether supporting the internal consistency and unidimensionality of each measurement instrument that consists of multiple item sets. The loadings of all indicators on their respective factors were statistically significant, with values ranging from .34 to .92 (see Table 1).

---

\(^6\) Since measures of majority of indicators are ordinal, we used matrix of polychoric correlations as an input for analysis. The analysis of multivariate normality shows some violation of this assumption, that's why we used WLS method to estimate the model.
Table 1: Factor loadings, mean score, standard deviation and Cronbach’s alpha of questionnaire items

<table>
<thead>
<tr>
<th>Impact dimension</th>
<th>Item</th>
<th>Factor Loading</th>
<th>Error variance</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common_c: Belief in a common and diverse European culture</strong></td>
<td>v10: Europeans speak different languages, yet we share common culture</td>
<td>0.78</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v11: There's more than competitive position to USA that unites Europeans</td>
<td>0.85</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v12: Respect to each other national traditions is what all Europeans have in common</td>
<td>0.74</td>
<td>0.45</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>v13: It is a fact that although having a European Parliament, European nations don't have much in common (R)</td>
<td>0.39</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td><strong>Advantage: Awareness of the advantages and challenges of European Union</strong></td>
<td>v2: Only when I returned from mobility, I became aware of numerous advantages of EU membership</td>
<td>0.72</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v3: Participation in mobility made me realize that EU countries working together can achieve more than working alone</td>
<td>0.87</td>
<td>0.24</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>v4: Through individual mobility I realized that EU offers numerous possibilities</td>
<td>0.9</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v5: At the mobility I have realized that idea on EU as a common state is sentenced to fail (R)</td>
<td>0.34</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge: Knowledge of historical, cultural, economic and social aspects of European Union and its members</strong></td>
<td>v6: History of destination country</td>
<td>0.38</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>v7: Social and economic situation in the destination country</td>
<td>0.43</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v8: History of EU</td>
<td>0.92</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>v9: Economic and political facts of EU</td>
<td>0.89</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td><strong>Identity: European identity</strong></td>
<td>v1: Would you say that you see yourself as European to larger extent due to your participation in the Erasmus action</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>inclusion: Inclusion in the economic and social development of the EU community</strong></td>
<td>v14: Index of items</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The factor loadings at two reversed items are rather low, suggesting that they are also measuring something else than intended. We would suggest to improve these indicators (i.e. soften them, as they demonstrate quite extreme attitudes) and add at least one more reversed
item per each scale. Next, the knowledge dimension might better perform if it would be partitioned in two subdimensions: knowledge on EU and knowledge on country of mobility. The error covariance term between two indicators already suggested that there is some other common factor between the first two indicators of Knowledge.

Anyhow, we can conclude that the measurement instruments for each dimension of the European dimension construct demonstrated satisfactory levels of internal consistency and unidimensionality, especially if we take into account the novelty of the instruments. Before investigating the descriptive statistics of impact measures, we also verified whether European dimension is a single, internally consistent latent variable, which could consequently be used as a unique measure of European dimension. For this reason, we performed second-order factor analysis (Bollen, 1989), where each dimension of European dimension is an indicator of a European dimension construct. In other words, we assessed the fit of the factor model, where second order latent construct “European dimension” is measured by first order latent constructs that correspond to five individual dimensions, which are further measured with specific measurement instruments. In comparison to the above confirmatory factor analysis, the $\chi^2$ of the second order factor model was surprisingly only a little larger ($\chi^2=228.69$, df=72, $p<.01$). Fit statistics thus demonstrate a moderate to good fit (RMSEA=0.067, RMR=0.010, GFI=0.98, AGFI=0.96) also for the second order factor model. The values of second order factor loadings range from 0.54 to 0.90 and are statistically significant, supporting the idea that European dimension is a internally consistent construct (see Figure 2).
Figure 2: European dimension construct
4.6. The amount of measured impact on European dimension

Following the definition of measured impact in section 3.2., we can now simply take factor scores or mean values\(^7\) of each dimension of the European dimension that was measured with multiple item sets and evaluate the intensity of impact of Erasmus mobility programme. We can even use Lisrel programme to compute and save factor scores for the single common factor of European dimension construct and investigate it. However, we shall first focus on individual dimensions of this construct. As mentioned a sense of European identity was measured with a single item, taken from Eurobarometer survey, which has the following distribution: Only 1.2 per cent of the students don’t have any sense of European identity; 14.0 per cent feel European only to small extent, whilst 46.2 per cent feel Europeans to some extent and 38.7 per cent feel European identity to large extent. The impact of programme on the sense of European identity that was measured with a simple indicator “Would you say that you feel more European due to your participation in mobility?”. 63.0 per cent of respondents attributed the heightened European identity to the participation in Erasmus action.

The impact on the dimension of inclusion in the economic and social development of the community was measured as an activity that started with participation in the Erasmus action. The introductory text stated: ‘Did you start with any of the following activities after returning from the action?’ The corresponding percentages are as follows:

Table 2: Inclusion in economic and social development of a community

<table>
<thead>
<tr>
<th>v14: individual items</th>
<th>% yes</th>
<th>% already before the action</th>
</tr>
</thead>
<tbody>
<tr>
<td>reading newspapers from other EU countries</td>
<td>22.9</td>
<td>24.4</td>
</tr>
<tr>
<td>publicly expressing your interest in international problems (i.e. ecological problems in EU)</td>
<td>26.2</td>
<td>26.5</td>
</tr>
<tr>
<td>subscribed to newspaper from other EU country</td>
<td>2.9</td>
<td>2.3</td>
</tr>
<tr>
<td>tried to establish a contact with people who express their interest in European problems</td>
<td>9.5</td>
<td>7.1</td>
</tr>
<tr>
<td>joining the debate on the future of EU</td>
<td>12.1</td>
<td>10.9</td>
</tr>
<tr>
<td>being active in international working groups dealing with EU socio-economic problems</td>
<td>8.5</td>
<td>4.1</td>
</tr>
</tbody>
</table>

\(^7\) We decided to take mean values for the ease of interpretation.
We computed\(^8\) a total index of this concept in order to assess an impact. The average value of the computed index (min=0, max=6) was 0.81 (stdev=1.15), meaning that respondents in our sample in average started with little less than one of the above activities after returning from the action. It should also be noted that the ‘measured impact’, or its average mean, is quite difficult to interpret. More precisely, the interpretation depends on the expectations of policy makers. Although the average of the above variable is rather low (0.81), its frequency distribution informs us that no impact can be observed on minority (46.2 per cent) of respondents, whilst 26.3 per cent of respondents started to involve themselves in one of the mentioned activities. 17.6 per cent of respondents participated in two stated activities, 10.2 per cent in three, and 3.9 per cent in four. These results might be very encouraging for policy makers with moderate expectations, whilst more strict would probably demand higher averages.

We now move to investigation of impact on dimensions of European dimension construct that were measured with multiple item sets. Belief in a common and diverse European culture was measured with the items that are listed in Table 3, while the introductory question was ‘Have your attitudes towards the following statements changed due to your participation in the action?’ with responses ranging from -2 (have more negative attitude) to 0 (no change in attitude) to +2 (have more positive attitude).

<table>
<thead>
<tr>
<th>Common_c: individual items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>v10: Europeans speak different languages, yet we share common culture.</td>
<td>0.35</td>
<td>0.99</td>
</tr>
<tr>
<td>v11: There's more than competitive position to USA that unites Europeans</td>
<td>0.54</td>
<td>0.89</td>
</tr>
<tr>
<td>v12: Respect to each other national traditions is what all Europeans have in common</td>
<td>0.36</td>
<td>0.96</td>
</tr>
<tr>
<td>v13: It is a fact that although having a European Parliament, European nations don't have much in common (R)</td>
<td>-0.20</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note: (R) indicates reversed worded item.

\(^8\) As mentioned the reliability of such a scale could not be estimated, but we could, however, speculate that such questions are reliable, since they only demand from respondents to recall some fact that occurred in the near past in their everyday lives. One could argue that all these items are prone to social desirability, since the professor-student relationship is implied in the survey process. This might stimulate students to create a more positive image of themselves. Although we did not control for social desirability in our study, we noticed that there were items with a small percentages of ‘yes’ answers, indicating that responses represented true values and were not biased by social desirability.
A computed mean index of impact on belief in a common and diverse European culture (M=0.26, SD=0.66) indicates that students expressed, on average, a slightly positive impact of action on their beliefs in European common culture (see Table 4).

The impact of action on awareness of advantages and challenges of the European Union was measured by a battery of items reflecting the subjective evaluation of change in beliefs due to participation in the Erasmus action.

Table 4: Awareness of the advantages and challenges of European Union

<table>
<thead>
<tr>
<th>Advantage: individual items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>- v2: Only when I returned from mobility, I became aware of numerous advantages of EU membership</td>
<td>3.08</td>
<td>1.11</td>
</tr>
<tr>
<td>- v3: Participation in mobility made me realize that EU countries working together can achieve more than working alone</td>
<td>3.31</td>
<td>0.98</td>
</tr>
<tr>
<td>- v4: Through individual mobility I realized that EU offers numerous possibilities</td>
<td>3.61</td>
<td>0.92</td>
</tr>
<tr>
<td>- v5: At the mobility I have realized that idea on EU as a common state is sentenced to fail (R)</td>
<td>4.11</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note: (R) indicates reversed worded item.

All four items were computed in a single index (scale from 1 to 5), with an average of 3.53 and standard deviation of 0.67. Frequency distribution shows that only 26.0 per cent respondents expressed no impact or negative impact of action on their awareness of the advantages and challenges of European Union.

The last dimension of impact domain considered the knowledge of historical, cultural, economic, and social aspects of European Union and its member states. Respondents were asked whether the participation in action helped them to improve their knowledge of these aspects. The values of items correspond to answers no(1), yes,a little (2), yes, a lot(3)
Table 5: Knowledge of historical, cultural, economic and social aspects of European Union and its members

<table>
<thead>
<tr>
<th>Knowledge: individual items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>- v6: History of destination country</td>
<td>2.55</td>
<td>0.56</td>
</tr>
<tr>
<td>- v7: Social and economic situation in the destination country</td>
<td>2.72</td>
<td>0.51</td>
</tr>
<tr>
<td>- v8: History of EU</td>
<td>1.45</td>
<td>0.63</td>
</tr>
<tr>
<td>- v9: Economic and political facts of EU</td>
<td>1.72</td>
<td>0.64</td>
</tr>
</tbody>
</table>

We computed an index ‘knowledge of EU’ as a mean of the above indicators. The average of the index was 2.11 and standard deviation 0.4, indicating that the impact of the action on risen knowledge is quite important. As already mentioned in this kind of investigation of measured impact, we unfortunately did not have clear criteria to define the strength of the impact, which is an issue that needs to be more seriously taken into account in the further research.

When policy makers are interested in the impact on European dimension as a whole, we can use factor score from second order confirmatory factor analysis and investigate its characteristics. Yet as factor analysis is performed on standardized variables, the mean value of factor score is 0, which makes it impossible to make any meaningful interpretation in relation to the strength of impact on European dimension in general. We can however use this factor score to evaluate the explanatory impact on European dimension in general.

4.6. Assessing explanatory impact on European dimension?

Depending on the interest of policy makers, one could estimate explanatory impact on each individual dimension of European dimension or on the concept of European dimension in general. In this paper we shall focus on the latter. The same procedure can also be used for the former.

In order to assess explanatory impact we will test an explanatory model, where factor score of European dimension is a dependent variable and intensity of participation an independent variable. In addition we included several control variables.

The intensity of participation in action was measured with several indicators, which are
treated as separate independent variables: length of participation in an action (in months), frequency of visiting lectures (scale from 1=never to 5=always), comprehending lectures (scale from 1=no comprehension to 5=excellent comprehension) and frequency of attending informal events (scale from 1=never to 5=always). The control variables were more or less intuitively added: age, gender, social class, study success, study field (1=nontechnical, 0=technical), year of participation. We used simple multiple linear regression analysis to estimate the effect of intensity of participation on the selected impact domain.9

Table 6: Estimated regression model10

<table>
<thead>
<tr>
<th></th>
<th>Beta</th>
<th>t-value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intensity of participation in an action</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of participation</td>
<td>0.14</td>
<td>2.32</td>
<td>0.02</td>
</tr>
<tr>
<td>Frequency of visiting lectures</td>
<td>0.04</td>
<td>0.73</td>
<td>0.46</td>
</tr>
<tr>
<td>Comprehending lectures</td>
<td>0.11</td>
<td>1.77</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Frequency of visiting informal events</strong></td>
<td>0.15</td>
<td>2.48</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Control variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.07</td>
<td>1.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Gender (0=male, 1=female)</td>
<td>0.14</td>
<td>2.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Social Class</td>
<td>0.08</td>
<td>1.42</td>
<td>0.16</td>
</tr>
<tr>
<td>Study Success</td>
<td>0.05</td>
<td>0.76</td>
<td>0.44</td>
</tr>
<tr>
<td>Study Field (0=technical, 1=nontechnical)</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.92</td>
</tr>
<tr>
<td>Years since participation</td>
<td>0.01</td>
<td>0.13</td>
<td>0.89</td>
</tr>
<tr>
<td>sig. (F)</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td>0.093</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proposed model manages to explain only 9.3 per cent of the variability of the impact on European dimension, but demonstrates a statistically significant fit to the data. Low explanatory power of the model is not surprising for this kind of impact domain, as it is difficult to expect that participation in mobility programme would significantly change such relatively stable phenomena as identity, beliefs and knowledge, which European dimension concept comprises. Nevertheless, two indicators of intensity of participation in an action

---

9 We used simple linear regression, since all independent variables are single-item measures, but a suggestion would be to develop multiple item sets also for independent variables and use all in a structural equation model to assess simultaneously a measurement and structural model.

10 We tested the standard assumptions of multiple linear regression and found of the following: 1) scatter plots between independent variables and dependent one do not indicate nonlinear associations; 2) Kolmogorov-Smirnov test demonstrated normal distribution of dependent variable and two ordinal independent one, while the test was not performed on dummy variables; 3) Breusch–Pagan proved that we cannot reject the null hypothesis of homoskedasticity and 4) No indicator of multicolinearity could be detected.
demonstrated a weak to moderate, and statistically significant effect on the dependent variable after controlling for several relevant factors. More precisely, the more time a beneficiary is involved in the programme, more likely the programme will have an impact on the European dimension ($\beta=0.14; p=0.02$). Similarly – and not entirely surprisingly - we could claim that more frequently a beneficiary visits informal events surrounding the programme ($\beta=0.15; p=0.01$), higher the probability of impact on European dimension. On the other hand, the frequency of visiting lectures ($\beta=0.04; p=0.46$) and the cognitive input ($\beta=0.11; p=0.08$), demonstrated no significant impact. The impact on European dimension is not significantly influenced by socio-demographic variables, except for the gender, where one could claim that females are more susceptible to impact on their European dimension ($\beta=0.14; p=0.02$). Interestingly the time distance from the participation doesn’t have any influence on the European dimension. We can summarize that the results of explanatory and measured impact are quite congruent, as the investigation of the means of individual dimensions of the European dimension concept similarly demonstrated a relatively weak to moderate impact on the rise of individuals’ European dimension.

5. Conclusion

We discussed in this paper fundamental conceptual and measurement issues to be tackled for a valid and reliable assessment of policy programmes’ impacts when trying to empirically assess the true impact of a certain programme. Although the paper omits the evaluation problems related to data gathering (e.g. questionnaire competition, low response rates of web surveying, end-user heterogeneity and diversity of their lifestyle), the social desirability problem of programme impacts, and selection of explanatory variables, it focuses on two main problem groups that are currently posing problems for the validity of evaluation research at the EU level in the field of educational programmes: poor causality; and the absence of proper conceptualization/operationalization. The purpose of this paper was to discuss these core issues and to provide solutions regarding the valid (twofold) understanding of impact, modelling the explanation of impact, and measuring impact in a valid and reliable way. The limited empirical insight into our alternative approach demonstrated that we are on the right track, but this track has opened many new, unexpected problems that still need to be solved.

Building upon proposed solutions applied in the case of LLP impact assessment in Slovenia, we conclude with the question, which evaluation aspects could be further improved in the future? Based on general evaluation theory (Rossi, et al., 2004), the development can take two
courses. As already mentioned several times, the established assessment of programme impacts anticipates at least the use of quasi-experimental approaches. Those, however, would demand setting up an information system which enables periodic monitoring of final beneficiaries from the time contact has been established with the National agency, throughout the participation in an action and the late period after the action was concluded. The second development direction anticipates further development of indicators that would enable closer monitoring of key functions necessary for a particular action to achieve desired impacts. Additionally, the factors which either essentially effect or explain the effects and consequentially the success and efficiency of the actions of European education and training programmes in Slovenia should be further explored. Here the established praxis of tight cooperation with various stakeholders should be further developed in order to identify and operationalize new and better impact factors. Particular attention should be paid to the factors that could explain the difference in the influence on individuals or organisations enabling the introduction of multi-level analysis.

6. References


Deloitte & Touche (YEAR) Final Evaluation of Leonardo da Vinci. Available at:


http://ec.europa.eu/dgs/education_culture/evalreports/education/2004/grundtvig/grundintanx_en.pdf and


EuroBio (2009) European dimension. Obtained through the Internet:

European Commission (2001) External evaluation of the European Centre for the Development of Vocational Training – Cedefop. Available at:

European Commission (2002) Programme for the enhancement of quality in higher education and the promotion of intercultural understanding through co-operation with third countries. Evaluation ex-ante. Available at:


