



ILO-INTI: PROMOTING EQUALITY IN DIVERSITY: INTEGRATION IN EUROPE

*(A project implemented by the International Labour Office and partners
with the financial support of the European Union)*

PRACTICE EVALUATION WORKING GROUP THIRD MEETING GENEVA, 15-16 SEPTEMBER 2005

SUMMARY REPORT AND CONCLUSIONS

1. Participants

Mohammed Abdul Aziz, European Network Against Racism (ENAR), London
Dimitria Clayton, Center for Immigration, North Rhine Westphalia, Solingen
August Gächter, Centre for Social Innovation, Vienna
Sarah Isal, Runnymede Trust, London
Patrick Taran, Lisa Simeone, ILO, Geneva
John Wrench, European Union Monitoring Centre on Racism and Xenophobia, Vienna

Memo items

See profiles database at www.wisdom.at/ilo.

Homework done since the February meeting

Per commitments made during the Vienna meeting, all documents, papers and references were circulated. Practice profiles to be reviewed during this meeting; see below.

2) Overview of Purpose and Expectations, including Summary of First Meeting in February and Work Done Since Then

During the Vienna meeting in February, INTI Project components were reviewed by the Evaluation Working Group, and the Terms of Reference for their task were refined – to develop evaluation indicators and methodology for selection of “best practices.” Existing approaches to evaluation of social practice were identified and reviewed, and initial proposals were considered for criteria, process, approaches, etc.

A discussion delimiting the Group’s parameters concluded that it should focus on anti-racism and anti-discrimination measures, while noting the continuum between them and measures more broadly promoting integration. The mandate was determined as follows:

- Establish useful approach to evaluate practice in the field of anti-discrimination
- Identify useful practice to look at, and identify relevant databases that exist
- Identify premises/sources for content of evaluation package
- Look at specific indicators and evaluation criteria, and come up with a methodology, including positive and negative assessments, that can be applied to anti-discrimination
- Carry out initial testing of the methodology on existing practices

Examples of evaluation theory, existing databases, and useful good practices were shared, and it was noted that most of the suggestions for criteria collected during the meeting were

listed in the “EU Guide to Procedures and Methodology.” The suggestion was made that this resource could be used as a starting point for the development of the Group’s indicators. Indicators and evaluation criteria were then agreed upon.

It was agreed that the second meeting would be held to work through models and proposals to come up with a working framework, which would later be refined and tested. The third meeting would then review initial application experience, consider refinements and finalize an evaluation “package.” It has been proposed that this package be adopted by working group participants in their expert roles and disseminated as a working model.

3) Review of Materials and Models

The group looked at the report from the Vienna meeting and agreed that it reflected the conclusions reached by the group. The variety of materials circulated at and since the last meeting were then discussed.

Patrick suggested that it might be useful to include in the final report or guide an annotated bibliography of all the materials used in putting these materials together. It was observed that however hard we all looked for models to inspire our project from, none of us had actually come across anything similar to what we are attempting to do. This is both positive as it proves the real usefulness of our exercise and challenging, as it implies more “pioneering” than initially expected. One of the differences that was noted between what exists and what we are trying to achieve is that other projects are looking at outcome indicators (examining the results of what is being done) whereas we are focusing on process evaluation (process of what is being done).

A discussion followed on what we mean by anti-discrimination project. In other words, what makes a project anti-discriminatory? A number of views were exchanged on that topic including:

- One would want to go to any organisation with our final report and be able to assess whether this organisation is doing anti-discrimination work or not, regardless of whether it claims to do it or not.
- In that sense, the most important criterion is who the work is actually targeting
 - Does it target majority group or those who are potential perpetrators of discrimination?
 - Or does it target the minorities or potential victims of discrimination, such as for instance those that empower black and minority ethnic communities to use and implement the legislation?

This led to the discussion around the selection criteria, i.e. what would we be choosing to evaluate? It was felt that in order to agree on selection criteria, we first needed to agree as a group on a definition of anti-discrimination activities and integration activities.

All members of the group were then asked to spend 5 minutes individually and come up with their definition of anti-discrimination activities.

Answers were:

- Any practice that intentionally or unintentionally resulted in the reduction of the scope and/or intensity of discrimination
- Any expressed intent of anti-discrimination.
- Any practice that aims to target the issue of unjustified differential treatment of groups, individuals and institutions in a specific and direct manner
- Practice mainly or essentially directed at the majority group
- If practice addresses the minority, or potential victim group, needs to have a reason.
- Any practice that covers the typology of discrimination developed by J.Wrench

- Any practice that aims to prevent discrimination (and not just redress)
- A practice which includes a reference to the values of equality (implicit or explicit)
- A practice that attempts to achieve measurable change.

4) Main Criteria and Indicators, Measures and Means of Measurement

Working on criteria for the evaluation of anti-discrimination projects it was understood the criteria should be useful beyond our immediate purpose. Lawmakers should be able to enshrine them in law, and it should be possible to base an ILO recommendation on them.

The first issue was which cases to include in the database and which to leave out. After vigorous discussion three criteria were shortlisted of which the last was subsequently eliminated again:

1. Practices explicitly or implicitly targeting discrimination defined as unjustified differential treatment. The focus might be on perpetrators, on outcomes, on types of discrimination (including legal discrimination), or on intent and effect. Euphemisms and synonyms shall be taken into account.
2. Explicit reference to identification of target groups. These could be either those at risk of discriminating or, if empowerment is intended, those at risk of being discriminated.
3. Discarded: Quantifiable change in the scope or intensity of discrimination is attempted.

Both criteria must be met. Activities should qualify under these criteria regardless of whether they targeted behaviour or attitudes. A reference to equal treatment or equal opportunities was not included on grounds of narrowing the field too much.

Criterion no. 3 was discarded because of its close relation to no. 1 and because an emphasis on precise measurement might in itself discourage action. It was pointed out that safety at work campaigns in the 19th century never had to be legitimated by their quantifiable impact. It was further pointed out that the results of preventive work, for instance, are hard to measure. The two criteria adopted only cover anti-discrimination work, not good-relations-work or other aspects of integration. It was decided to deal with integration later, in effect at the next meeting.

It was understood that the two criteria constituted the **RELEVANCE** item in the evaluation model.

The next item from the model was **USEFULNESS**. Renaming it **CATEGORY** was considered when it emerged the item was primarily about classifying cases. A debate on the precise meaning of “acknowledged approach” led at first to a three-way distinction: Is it a tried and tested approach (replication), is it an untested but not a new approach (practice innovative), or is it an innovative idea? [How do we rank these three?] The problem remained that information from outside the project description would be necessary in order to make a judgement.

Patrick Taran suggested to provide such outside information in the form of a list of roughly 60 items from two or three texts that are meant to cover a large part of the range of project approaches. This list will be distributed and will need to be amended. This means that at the next meeting a final decision about the contents of the list or about using it at all will have to be taken.

After discussion it was decided to delete the **USEFUL** category from the scheme. Whilst the question is useful for finding out whether the case fits into an existing category or whether we need a new one, this can be clarified beforehand, outside the scheme. This leaves seven criteria /categories in the scheme.

COHERENCE. The strength of the link between needs, aims, and activities should be the relevant criterion. Operationally the question would be how much reference the three project elements make to each other.

Having a timetable for the activities and their impact assessment will be part of coherence. All reference to outcomes was moved to the IMPACT item.

The original “sufficiency” was renamed **ADEQUACY**. Reference was made to OHCHR’s guidelines for national human rights monitoring bodies that require inspection and enforcement to be joined in the same organisation. This principle should also apply to anti-discrimination activities where the monitoring function within an organisation should also be given the enforcement function. Adequacy here refers solely to the adequacy of the intervention. Other aspects, such as adequacy of staffing and of resources was referred to VIABILITY. There were few ideas on the sort of data that would serve to indicate the degree of adequacy. Roger Zegers de Beijl had already worked on the problem when he was at the ILO. The EU directives against discrimination might offer starting points.

EFFECTIVENESS. In the course of the discussion this came to be understood as measurable output. How much of the planned activities had actually been carried out?

IMPACT. Activities are means to an end. Impact was understood to measure the progress made towards that end, towards the stated aims. One question was when to assess how much progress had been made. Should this be done right at the end of an activity, a year later, five years later? This would largely depend on the terms set for the project. It was urged that impact measures should be of two kinds. First there should be information on how much change from the baseline data was achieved. Second, one would want to know whether the “ethnic penalty” was reduced. [The latter requires considerable amounts of data and trust in adequate application of complex statistical techniques.]

VIABILITY. This was meant to cover the adequacy of resources including the quantity and quality of staff. [Would staff development also belong here?] No measures were developed at the meeting.

SUSTAINABILITY and follow-up. This would include reviewing of content and execution, learning from the reviews, and the interconnectedness of planning, management and resources. Self-evaluation and learning should be institutionalised, for instance in the form of periodic reviews of the objectives and redrafting of plans. A clue to whether this was done in the past is how much information is currently available on past aims, activities and results.

[If we decide to use any subjective criteria, more than one person will have to evaluate each case.]

5) Provisional Evaluation Model

ITEM	DEFINITION
Relevant	Explicit or implicit aim of targeting unjustified differential treatment? Explicit reference to target groups: either majority population or the empowerment of groups at risk of discrimination.
Coherent	Does the measure establish a clear link between the needs, the aims, and the activities implemented? Specifically, how much reference is being made between activities and needs and objectives? Is there a timetable covering the entire activity?
Adequacy	Are the components stated that are deemed to be required for the activity to be effective and to have the desired impact? (e.g. a company anti-discrimination policy will only be effective if it is accompanied by some form of enforcement).

	Does top management “own” the project?
Effective	Output: Were activities carried out in the quantities and the quality that had been planned, i.e. does the output match the intentions?
Impact	Outcomes: Is the aim being met or furthered?
Viability	Equal to “adequacy of resources” Operational: Management, resources, capacity, staffing Contextual: Support and restraints: legal, institutional, policy, politics; PEST (Political, Economic, Social, Technological) in project planning.
Sustainability	Follow up and learning, flexibility and adaptability? Replicability (scope for emulation).

The expression “project” should refer to an activity with a beginning and an end, i.e. the introduction of anti-discrimination activities that are themselves meant to continue indefinitely or special review, planning and reform activities within an open-ended activity.

[At a later stage it will be necessary to test which of the items in the model tend to be correlated.]

Integration criteria are still missing.

6) Practical Testing of the Model with Existing Profiles

Three pairs were formed that were each given the task of reviewing two arbitrarily chosen case profiles from the publicly accessible database at www.wisdom.at/ilo. Five of the six reviews were completed:

Denmark, counsellors: The project would rank zero on anti-discrimination but may still be a useful integration activity. There is not enough information available to make a conclusion on how *useful* the project is. There is little information that would allow judgement on the *output* and *impact* of the project. It is, for instance, unclear how the activity impacted the level of employment of the target groups. Indicators for assessing the *effectiveness* of the activity are missing. There appears to be no clear link between the *objectives* and the *outcomes*. The project employs a lot of resources and appears to be *viable*. *Sustainability* is maintained through the annual repetition of the project.

Germany, Volkswagen: The information is sufficient for the most part but linkage and impact information is missing. Because of the preventive approach the impact will be difficult to quantify. Relevant? Yes, the aim of combatting unjustified treatment clearly stated. Useful? Yes, several approaches are used that all seem to be well established. Coherent? Yes, but no timetable provided. Adequate? The necessary basic elements appear to be there. It is however unclear how exactly they will be implemented (e.g. what exactly is included in the training program?) Effective? The activity appears to have had some impact in the way of awareness raising.

France, Continent hypermarket: In some contexts it is apparently difficult to acknowledge discrimination. We must take this into account in the RELEVANCE criteria. Which evaluation item does the mimicking by other stores in the complex belong to? François Vourc’h should have some additional information on the case.

The project appears to be *relevant* as it implicitly addresses the issue of discrimination. Nevertheless, integration activities play a key role in its implementation. The project is *useful*. The *coherency* of the project is limited by the lack of clarity on the links between the different activities. The *impact* of the project is somewhat easier to measure. There are some indicators on the number of people employed as a result of the project, effect on rates of delinquency, etc. There is mention of the necessity to address the impact of the project through employee surveys. It is however unclear if these surveys were ever carried out. The

sustainability of the project was made evident by, for instance, the fact that other businesses in the area took on this approach. It was therefore *replicable*.

The criteria appear to be useful for a clear description of the project, making it possible to indicate its limitations. Overall it appears to be a sophisticated response to “discrimination”. *Output data* presented at the end of the report indicate that some efforts were made by the company itself to make some judgements about the impact of the activity. *Impact*, in the sense of preventing discrimination or awareness raising in this case should/could also be measured by consulting the target groups through, for instance, surveys of the personnel.

The ability to assess the project’s *sustainability* is somewhat limited by the difficulty of finding management, staff who have some clear memory of what had happened in years prior to the implementation of the project. This also poses a problem for efforts at data collection. It may indeed be a positive sign, that the outcomes of the project have been successfully “mainstreamed” and it is therefore unnecessary to remember them. Concerns about interruptions in *institutional memory* through company takeovers and personnel changes highlight the necessity of proper documentation of the project.

The question was raised, if perception should be an item or a criterion in the evaluation. It might help taking into account the very different starting points between east and west, south and north in Europe.

Ireland, Anti-racist Workplace Week: It is not enough to report on what you did without clearly linking it in some way to the activity’s expected outcome and impact. It is necessary to differentiate more clearly between *output* and *outcome*. The Irish project focussed a great deal on output – the outcome remained however unclear. The descriptors in both the Irish and Danish projects are not adequate. It is necessary to have more specific information on *quantitative* and *qualitative descriptors*.

United Kingdom, Diversity Policy at the British Council: The project is *relevant* and appears to be *useful*. The aims and needs are clearly stated. However, the description of the activities is not detailed enough. The background information on British legal requirements that would help to make sense of the case information is largely missing. Further, the given information did not permit judgement on whether the approach was tried and tested or novel, and information on the actual content of the activities was insufficient. This negatively affects the assessment of the *coherence* of the project, as there is not enough information provided which clearly links the activities to aims. A timetable is provided but no milestones to measure progress. The activities are untested but not necessarily *innovative*. It is difficult to determine the *adequacy* of the activities based on the available descriptors. In this context it is necessary to more clearly differentiate between “*adequacy*” and “*coherence*”.

Output target details were missing, and the information on the output milestones seemed incomplete. There is no mention of output but they immediately go on to talk about outcomes. The link between the activities’ objectives and outcomes is implicit but not clearly stated. The aim to improve recruitment is clear. There is not, however, enough information on how this aim is to be achieved. For example, there is no information about what exactly is to be done by when. This case study shows that output indicators can to some degree be considered outcome indicators. Quantitative indicators can be used to say something about a qualitative impact.

General Conclusions:

- The descriptive information provided is generally inadequate to make valuable judgements based on the 8 criteria that were used. It is therefore necessary to expand upon the descriptors and revise the descriptor matrix.
- More measurable indicators are necessary within the various categories (e.g. perception (?) indicators as well as statistical indicators). The category usefulness should be revised (or eliminated all together).

- The basic model works but needs to be refined. Descriptors for each type of criteria are needed.
- A more thorough discussion on the issue of “integration” should take place at the next meeting.

7) Identification of Improvements to Data Collection and Presentation

The nine items in the database structure – in addition to an abstract and to contact details – require some changes in definition in order to serve the purposes of the evaluation model.

1. **Needs identified**
2. **Aims and objectives:** Do they specifically address the needs? How is progress towards the aims to be measured or indicated?
3. **Activities:** Do activities correspond to the aims? How, and to what extent? Clarify that the description of the activities should also include reporting on the extent to which they were carried out. Where this is missing it needs to be made clear whether the information does not exist or is merely unknown to us.
4. **Project period**
5. **Sources of funding:** Is this an indicator of sustainability? Is it incorporated into mainstream budget?
6. **Involvement of migrants/minorities and social partners:** A value judgement, but realistically if not it will be less effective.
7. **Results achieved:** This section should cover specifically the impact, i.e. the progress towards the stated aims of the project. The mere output was covered in section 3. It is to be expected that reasonable impact information will often be unavailable. If so, this should be noted in the description.
8. **Innovative elements:** Difficult – more subjective
9. **Difficulties** encountered in implementation and scope for emulation. This question needs to be expanded to also address supports, not only obstacles.

We have to be very clear about any missing information: is it unavailable or is it merely unknown to us for the time being?

These changes will be implemented immediately. A more comprehensive revision needs to be considered (although any change to the database structure will cost money).

8) Methodology of/for what?

This section needs an explanatory intro.

Item	Definition	Indicator	Measure	Questionnaire item
The 7 or 8 items		What could express the definition best?	What is a measure of the indicator?	Which items on the questionnaire capture the required information?

9) Report

Nine chapters:

1. Introduction: Context and values (**Patrick**)
2. Process of elaboration – questions and challenges (**Drawn from minutes**)
3. Selection, or identification of focus (**John**)
4. The evaluation model itself: presentation of criteria and indicators for quality assessments, including process, outcome, and the relationship between process and outcome
5. Methodology of the evaluation/project user guide (**Wait until next meeting**)
6. The testing of the six examples (all type up notes)
7. Data problems of the national level: conclusions/recommendations (**Including ECRI on ethnic data**)
8. Partially annotated bibliography and resources (**Sarah Isal**)

10) Homework

What	Who
Type up discussion notes on 6 case studies (within 2 weeks)	All
Adapt case profile questionnaire	Patrick, August
Usefulness: Collate and distribute the 60 items	Patrick, August
Adequacy: Which data? Material on need for data on ethnic/national origin for inclusion in recommendations	ECRI
Bring know-how from the evaluation literature into the working group?	
Read up on integration	John
On-site visits	
Draft chapters for next meeting:	
Chapter 1	Patrick
Chapter 3	John
Bibliography	Sarah Isal

Before the next meeting, it was agreed that the Group should:

- divide up some of the seven chapters of the report.
- do more work on refining the indicators and measurements.

Patrick will transform the notes into a record of the workshop, hopefully by the middle of July.

11) Evaluation

We are rather late regarding the methodology. We had planned to have something by now to start testing. It would be useful if before the summer break we could get started on elements of the report.

We should meet in September even though we will not have had the discussion on integration by then. For the moment we will use the categories as they stand for the operational model. We need to move beyond the descriptors to refine the measurement and the indicators for each descriptor. They may be distinct, so we will need to use two extra columns on the right.

Note that some indicators will not be quantifiable measurements. The indicator will be in the first column and the measurement in the second. Also note that there may be a question of sensitivity if we analyse case studies by name in ways they were not warned about.

An ILO / INTI conference has been planned for 8-9-10 December in Brussels, which will include a presentation at the European Parliament.

Note that 2007 will be the European Year of Equal Opportunity for All.

10) Next Meeting

15-16 September 2005, Geneva.

Patrick Taran and Lisa Simeone will arrange local hosting of the meeting.

Summary Compiled by Lisa Simeone with notes contributed by August Gächter, Dimitria Clayton, Sarah Isal, John Wrench and Patrick Taran.

Contact: ILO International Migration Program, tel +41-22 799 8091 e-mail: taran@ilo.org