

**EVALUATION IN THE HUMAN RIGHTS
EDUCATION FIELD: GETTING STARTED**

**by Felisa Tibbitts
HREA, 1997**

WHY DO RESEARCH AND EVALUATION?

Pure research and applied research (such as evaluations) are simply means for us to better understand what we are doing, how we are doing it, and the results of our efforts. The first beneficiaries of research and evaluation are those responsible for carrying out a particular program or activity.

*research
vs.
evaluation*

The major difference between research and evaluation is in their goals. Evaluative efforts are undertaken to assess “effects” or outcomes, as measured against predetermined criteria. When such research takes place in the ‘development’ phase of a program, it is called ‘formative evaluation’ and the results are used to improve the program. After the program is in place and we want to know whether or not it is measuring up to its original goals, then a ‘summative evaluation’ may take place.

Whether you use research to inform the conception of a particular program at its outset, or whether evaluation is carried out following the initiation of a program, this information is ideally designed to help you, the implementor, and funders, if applicable, understand if you are achieving your intended goals — whether it be in the classroom, in a teacher training, in the use of a new materials, or in your overall program. Evaluation can help you make adjustments to your activities so that you can achieve better results.

WHY IS EVALUATION INTIMIDATING?

Evaluations involve making informed judgments, but there is variation in how these judgments are made and applied. One basic distinction exists between ‘formative’ and ‘summative’ evaluations. Formative evaluations take place when a program is still in process, and it is understood that the results will initiate changes to produce better results in the future. Examples of formative evaluations are assessing students in the middle of the year and giving them feedback so that they can work on improving their weak points, getting feedback on the first day of a teacher training that is then used to reshape the agenda for the following day, and looking at a three-year program after its first year of operation to see how well it is meeting its original declared goals.

*formative
evaluation*

Summative evaluations take place at the end of a set of prescribed activities to see if the goals originally set have been reached. This might be an end-of-the-year exam for students, the overall evaluation of a teacher training, and an evaluation of a three-year program as it is ending. Although these results can also be instructive for the actors involved, it is also true that such “high stakes” evaluations can also have impact on the educational actors and outside audiences; summative results can influence how a student is ranked in his or her class, if trainings continue to be sponsored, or if program funding is renewed.

*summative
evaluation*

‘Feedback’ is an informal strategy for learning what others think or feel about an activity that you have conducted. The goal of feedback is not to ‘give a grade’ but to help directly revise a product or service. Feedback is often organized by the person conducting the task, rather than through a third party, such as an independent evaluator.

WHO CAN CONDUCT EVALUATIONS?

Some assessments can be conducted “in house,” meaning that they can be organized by existing staff. Other times, specialists are hired to either conduct the evaluations or train staff to do so. Funders frequently require that independent evaluators be used for program evaluations, since such people are supposed to embody both expertise and impartiality in terms of findings.

In practice, most teacher trainings are evaluated by the organizers, with input from the participants. Teachers evaluate the success or failure of a human rights education lesson by trying them out in the classroom. Evaluation specialists are often called in for materials pretesting and program evaluation, since data collection is more complex and has high requirements of impartiality; however, even with the involvement of an outside expert, those implementing the projects can be closely involved with the data collection and analysis.

Regardless of whether regular staff or specialists are used to develop evaluation procedures, you will have to follow certain steps, such as deciding what you want to evaluate (target goals), choosing measurements, systematizing your data collection methods, analyzing the results and feeding these back into the activities being studied. Also, regardless of who carries out the evaluation, they should meet the following standards:¹

*evaluations
steps*

- **utility:** The evaluation should serve the information needs of the intended users, including, where applicable, those organizing the activities, those participating in them, and funding oversight agencies.
- **feasibility:** The evaluation should be realistic, prudent, diplomatic and frugal.
- **propriety:** The evaluation should be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by the results.
- **accuracy:** The evaluation should reveal information sufficient to determine the merit of the program being evaluated.

*evaluations
standards*

WHERE CAN I FIND EVALUATION SPECIALISTS?

If you have decided that you need a specialist to assist you in evaluation, there are several places where you might look for personnel.² There are some experienced evaluators who have worked internationally, and even in the human rights education field. You might consider using such people for high profile projects, or high-cost projects that have been supported by international funders. You also might consider hiring an international expert for a once-only evaluation. You must weigh the anticipated benefits of using such a specialist against their probable expense, including travel costs.

For an ongoing project, with low-stakes outcomes, you might consider hiring an international specialist to work closely with a local researcher in key parts of the evaluation project. Such a specialist could assist a local evaluator in project design, training local personnel, and analyzing the results, while most of the data collection would be carried out locally. This experience of linking an international specialist with a local evaluator also has a nice side effect of enhancing local capacity-building in the evaluation field.

As an alternative, you could simply use local educational evaluators. Many time, educational evaluators can be found in local colleges or universities. If you end up linking with a local university, you may also be able to find doctoral candidates who would be interested to research your project as part of their thesis. Their involvement is likely to enhance the project, without incurring additional costs. Whether you use local personnel or an international specialist, talk to people who have had projects similar to yours and ask who they used. Get recommendations. You might ask for reports done previously by the evaluator candidates to make sure that they are knowledgeable about your field,

¹ These standards were developed by the Joint Committee on Standards for Educational Evaluation in the USA for program evaluation (Sanders, J (1994) The Program Evaluation Standards (Thousand Oaks, CA: Sage Publications, Inc.), but are applicable to all forms of educational evaluation. See Annex.

² Portions of this section are adapted from P. B. Campbell (1982), Evaluating Youth Participation: A Guide for Program Operators. New York: National Commission on Resources for Youth.

and that their research approach and working style are close to yours. Make sure that the candidates are people with whom you can easily work. When you have narrowed down the candidate field, you might ask for an evaluation design before making a final decision. If necessary, get assistance in evaluating the quality of the design.

The contract with the evaluator should include a Scope of Work Statement, that includes a description of the evaluation questions to be addressed; the data collection planned, including sources, sites, and instruments; a timetable for these activities; and a schedule of reports and meetings.³

When conducting evaluations, the general rule of thumb is to spend between 5%-10% of your total program budget. However, the labor and material expenses of the retained specialist are not the only costs; you must also think about the time that your own staff will have to spend in preparing documents, attending meetings or collecting data as part of the evaluation. The real costs of evaluation will include all these items, and should be factored in when planning.

HOW DO I GET STARTED? CAN YOU GIVE EXAMPLES?

It may be easiest to make suggestions according to the type of human rights education activity being evaluated: lessons in the classroom, teacher trainings, materials testing, and program evaluation. Since program evaluation is the area where questions most frequently arise in the human rights education field, relatively more attention will be devoted to it in this report. Keep in mind that entire books have been written about these kinds of evaluations. This report is only a primer.

Program evaluation

Program evaluations are common in human rights education projects, particularly when they are financially supported by outside agencies. Usually, it is a program, rather than the non-governmental organization in which it is based, that is evaluated. However, such evaluations can end up closely linked with the overall operations of the organization. This is because programs often reflect the decisionmaking, communication, problem-solving and public relations practices of the organization in which they are based.

Program evaluations may be either formative or summative in nature. A formative program evaluation will collect data mid-point or in an ongoing manner for the project, in order to provide information that will enable the organizers to 'reform' or 'redirect' their program mid-course so that it is more effective. A summative evaluation is once-only, and will attempt to document the degree to which the program — now in its final phase — successfully reached the goals it set out to achieve. Both forms of program evaluation benefit from both an independent evaluator and the active participation in documentation of those most deeply involved in conducting the project.

Before undertaking a program evaluation, you might want to know about various models for program evaluation:⁴

*program
evaluation*

³ Morris, L.L. and Fitz-Gibbon, C.T. (1978). *Evaluator's Handbook*. Beverly Hills and London: Sage Publications p. 27.

⁴ *Ibid*, p. 2.

Table 1 MODELS FOR PROGRAM EVALUATION

- **Goal-oriented evaluation:** Evaluations assess student progress and the effectiveness of educational innovations.
- **Decision-oriented evaluation:** Evaluations facilitate intelligent judgments by decision makers.
- **Transactional evaluation:** Evaluations depict program processes and the value perspectives of key people.
- **Evaluation research:** Evaluations focus on explaining educational effects and devising instructional strategies.
- **Goal-free evaluation:** Evaluations assess program effects based on criteria apart from the program's own conceptual framework.
- **Adversary evaluation:** Evaluations present the best case for each of two competing interpretations of the program's value. Both sides having access to the same information about the program.

In recent years, a field of research called “action research” has emerged. In this type of research, data is collected, interpreted and applied for organizational self-renewal exclusively by the practitioners in an organizational setting, such as a school.⁵ In a related, “participatory research” model, the evaluator is the coordinator of the project with responsibility for technical support, training and quality control, however, the study is conducted jointly with the practitioners.⁶ Either approach can be used for formative, but not summative, evaluations.

Different models involve smaller or larger sets of practitioners in the workplace, ranging from key individuals, a small collaborative group, or the entire staff. Who is involved will reflect the goals for the research itself. For example, research that aims to improve an entire school culture might involve every member of the faculty, as well as students and operational staff; research that is geared to improve classroom teaching in a subject might involve only teachers in that discipline.

Ideally, all evaluations should borrow a main principle of action research, that of involving practice-based decision makers. ‘Stakeholders’ — those either directly involved for carrying out human rights education programs, or people with a vital interest in the program — should at least be involved in a consultative way in establishing the evaluation questions. They may also become involved in data collection, interpretation and reporting.

⁵ See Calhoun, E.F. (1994). *How to Use Action Research in the Self-Renewing School*. Alexandria: Association for Supervision and Curriculum Development.

⁶ Cousins, J.B. and Earl, L.M. (1992). “The Case for Participatory Evaluation” in *Educational Evaluation and Policy Analysis* (Winter, Vol. 14, No. 4), pp. 397-418.

All program evaluations must address the topic of data collection. Data collection methods can include both quantitative and qualitative elements. Quantitative indicators of success relate to measurable outcomes; qualitative relate to subjective outputs, like quality and attitudes. There are also indicators of systemic impact, such as increasing the frequency of human rights teaching in the classroom.⁷ Often, evaluations combine both kinds of measurements. Here are some examples of quantitative and indicators of success that can be used in assessing human rights education projects.

quantitative measures: numbers of trainings, number of participants

qualitative measures: quality of training

systemic measures: in-service trainings adopted by the Ministry of Education

quantitative: number of books printed, and disseminated

qualitative: quality of materials produced

systemic: books authorized by the Ministry

quantitative: staff hired

qualitative: most qualified staff hired

systemic: procedure for hiring staff

quantitative: number of staff attending internship program, length of program

qualitative: impact of internship on professional development of staff

systemic: local in-service is carried out by staff attending internships

Programs have first-, second- and even third-order effects. First-order effects, which are typically looked at in evaluations (partly because they are easier to document) are those activities that are carried out directly by the executors of the program. A first-order effect of a HRE program might be whether the promised number of trainings took place and books were produced. Evaluations of second-order effects look at the impact of the program features, that is, whether there is any changed practice in target groups. Are the books being used by teachers, and have they affected classroom practice? A third-order effect that might naturally follow is changed knowledge, skills or attitudes in the children as a consequence of changed classroom environments. These kinds of impact evaluations are important to keep in mind, but will not be discussed in this report.

HOW DO I GET STARTED IN PLANNING A PROGRAM EVALUATION?

*planning a
program
evaluation*

You might begin by asking yourself the following questions. Your answers will guide your initial steps.

1. Purposes

What are the purposes of the evaluation?

Are the purposes understood by and acceptable to all concerned?

2. Motivations

Why is the evaluation being undertaken now?

3. Participants

Which staff or stakeholders will be involved in the evaluation?

What will be the nature of the involvement of various participants?

Will there be a representative planning team or steering committee?

Is the evaluation likely to be seen as threatening to any of the participants? How can any perceived threat be minimized?

⁷ Partly adapted from Bethel, G. Workshop presentation on Effective Program Management, Education for an Open Society Conference, September 1995.

4. Evaluation roles

Who will carry out the technical aspects of the evaluation?

- an Evaluator, an independent person involved in data collection and judgment?
- a Facilitator, one involved in assisting in the evaluation but not in judgment?
- a Consultant, a person involved in limited aspects of evaluation, like assisting in interviewing?

5. Intended audiences

Are the audiences of the evaluation clearly defined?

6. Area/Issue to be evaluated

What project objectives will be evaluated: achievement of certain goals; processes; systems?

What are the indicators of success?

7. Collection of information

What methods of data collection will be used?

Available methods include:

- observation (structured or unstructured);
- interviews (structured or unstructured)(individual or focus group);
- questionnaires;
- documentary analysis of reports, records, and other written materials;
- content analysis of curriculum materials;
- reports of informal discussions and conversations;
- achievement tests (criterion and norm-referenced);
- diaries and self reports;
- audio and video tape recordings.

Are there appropriate safeguards to ensure that the information is valid and reliable?

8. Feasibility of methods used to collect information

Is the conceived evaluation realistic in terms of time, personnel and finance available?

Are particular resources needed to make the evaluation more effective (e.g., secretarial assistance, equipment, paper, postage, working space, printing production)?

What is the time span of the evaluation?

9. Judgments

What are the procedures for the analysis of information?

Are there appropriate safeguards to validate the information?

10. Release of information

Who will have control over what is collected and reported?

What procedures will govern the collection and release of this information?

Who will have the right to reply to, correct and validate reports of the views and activities of individuals and groups?

Will all, or only part, of the information be released?

11. Reports

Reports typically include a summary; a table of contents; a list of tables and figures; purposes; methods; findings, and discussion. Recommendations might also be included.

What form (content, style, format) will the evaluation be reported in?

Will negative aspects of the program be reported, and to whom?

Are different reports for different groups necessary?

12. Outcomes

Is it possible to see or predict outcomes from the evaluation?

What steps have been taken for ensuring that the evaluation feeds into the appropriate decision-making processes?

Are the participants aware from the beginning of the possible outcomes?

CAN YOU TELL ME MORE ABOUT DATA COLLECTION METHODS?

*data
collection
methods*

If research design and analysis is the “head” of a program evaluation, data collection is the arms, legs and bodies of the rest of the work. The differences between qualitative and quantitative forms of measurement have already been presented. We can also compare these overall approaches in the following way.

Qualitative research methods...

- begin descriptively
- are not initially quantifiable
- use a small sample size
- are open-ended
- include perspectives of people studied, and also the researcher
- are process- rather than product-oriented
- are context oriented
- in purest form, is non-interventionary

*quantitative
vs.
qualitative
measurements*

Quantitative research methods...

- use pre-defined categories
- are quantifiable
- use a larger sample size than qualitative research projects
- are closed-ended
- use well defined methods of analysis
- employ defined variables
- are product-oriented
- are decontextualized

Few research projects are either exclusively quantitative or qualitative in approach. Methods may be combined within a single research framework. Exceptions would be case study, or ethnographic, research, which typically relies on qualitative interviews and observational data, and experimental/quasi-experimental studies that focus on statistical comparisons.

The majority of program evaluations rely on surveys and interviews for the collection of information. For this reason, it is worth focusing on the **design and presentation of questions**. The questions that you ask should obviously be core to the operation of your program or its intended impact. When selecting question topics, think about those that are essential to your core program goals, and which...

- will provide useful information
- will lead to decisions or reduce uncertainty
- are most likely to uncover effects
- you have time to answer
- can be answered⁸

For example, if your human rights education project is operating out of regional centers, you might decide to conduct an internal study to examine the operations of the regions in relation to the central office. You might want to focus on administrative and policy areas which you can influence, such as communication, decisionmaking and budgeting. Such a study might identify both informal and formal mechanisms that are in use, evaluate their effectiveness, and look for ways to improve the overall operation of the organization.

Alternatively, such a human rights education project might decide that it is interested to learn how well the regional centers are serving as resources for their local communities. In this case, standardized mechanisms might be established for documenting the provision of such local services, as trainings, use of the resource library, dissemination of materials, programming, and other forms of technical assistance. Also, outreach might be done to the community to obtain information about their use of the center, and ways in which it could better meet their needs.

Once you have established the overall goals for your questions, you will then need to create the questions themselves. These questions may be closed-ended, that is, with answer options already provided, or open-ended, with the respondent able to frame the answer in the way that she or he likes. These are some examples of open- and closed-ended questions.

Open-ended question format.

How have you benefited most from participating in this program?

Closed-ended questions format⁹

- Write a statement. Ask the respondent to agree or disagree.

I have benefited from participating in this program.

___ Yes ___ No

⁸ Davis, B.G. and Humphreys, S. (1983). Evaluation Counts: A Guide to Evaluating Math and Science Programs for Women, report to the National Science Foundation, pp. 10-12.

⁹ Inspiration taken from Fink, A. and Kosekoff, J. (1985). How to Conduct Surveys. Newbury Park: Sage Publications, pp. 23-40.

- Ask the respondent their degree of agreement or disagreement

I have benefited from participating in this program

1=definitely agree; 2=probably agree; 3=neither agree nor disagree;

4=probably disagree; 5=definitely disagree

Other kinds of “category scales” are:

1=frequently; 2=sometimes; 3=almost never

1=strongly approve; 2=approve; 3=undecided; 4=disapprove; 5=strongly disapprove.

- Develop a scale with statements at either end. Ask the respondent to place themselves on the scale.

I have benefited
from the program

I have not benefited
from the program

1 _____ 2 _____ 3 _____ 4 _____ 5

- Write a question, with the variable an evaluative element rather than degree of agreement. Ask the respondent their indicate their views with a rating.

My participation in the program has been

1=very useful; 2=somewhat useful; 3=not very useful.

- List items. Ask the respondent to rank order the items.

Rank the following program components according to their degree of usefulness to you. The top-ranked should be assigned the number 1 and the lowest ranked the number 4.

___ formal training

___ materials received

___ trying new methods in the classroom

___ informal discussion with colleagues

- List items. Ask the respondent to rate each item using a scale provided.

Rate each of the following program components, using the following scale:

1=very useful; 2=somewhat useful; 3=not very useful.

___ formal training

___ materials received

___ trying new methods in the classroom

___ informal discussion with colleagues

Typically, more open-ended questions are asked during the exploratory phases of research, and more closed-ended during confirmational periods. The main advantage of the closed-ended question format is that it provides a uniform frame of reference to use in determining answers to questions. Open-ended questions have the advantage of allowing the respondent great freedom in framing the answers. Open-ended questions have the disadvantage of taking more time to analyze.¹⁰

If you are including questions in a survey, you should be aware that common practice is that **questionnaires** should not take more than an hour to answer. When you are reviewing your questions, check to see that they are short and clear, and avoid unclear phrases, abbreviations or jargon. Pretest the questionnaires to see how long they take to fill out, to make sure that all possible responses are accounted for in the answers for each closed-ended question, and that none of the answers are overlapping. Pretesting should also confirm that there are no hidden biases, or leading questions.

Include convenient ways to collect the questionnaires (either by prepaid mail, or by picking them up personally). Sometimes questionnaires are not administered personally, but are sent and returned in the mail. This is the case, for example, when one uses sampling procedures. Although this is an efficient means of large-scale data collection, it is possible that misinterpretation can take place unless the questions are well constructed and focused.

questionnaires

Table 2 . COMMON MISTAKES WHEN INTERVIEWING

- Testing respondents and not the materials
- Teaching instead of testing
- Lack of proper introduction/motivation of the respondent
- Communicating non-verbally a defensive attitude, thus preventing free expression of criticism
- Talking more than listening
- Judging what respondents say as “stupid” or “incorrect”(expressed by laughing or facial gestures)
- Asking leading questions Being in a hurry
- Not being able to “change shoes” (empathize)
- Influencing or inhibiting the thinking of the respondent
- Forcing/leading the respondent to be formal.

In **interview** situations, the researcher faces the same challenges in designing proper questions. Most interview protocols, or procedures, use open-ended questions. A structured interview protocol allows for no variance from the pre-designed questions. A semi-structured interview protocol, which is quite popular, has a preset list of questions; however, there is some latitude on the part of the researcher to offer spontaneous follow-up questions that allow for the pursuit of unexpected lines of discussion.

interviews

¹⁰ Weisberg, H.F. and Bowen, B.D. (1974) An Introduction to Survey Research and Data Analysis. San Francisco: W.H. Freeman and Company, pp. 49-50.

In program evaluations, interviews are often held with key program staff and those associated with the implementation of the project (for example, key personnel from the Ministry of Education and district education staff). Focus group interviews might take place with other informants, who are too numerous to interview individually. In a human rights education project, these groups might include teachers, students, police officers, or other target groups for the program.

Conducting interviews sounds like an easy thing to do, but it is equally easy to make mistakes that will limit the information that is collected and alienate the interviewee. Table 2 contains some common pitfalls to avoid.

Classroom-based Assessments

*classroom
based
assessment*

Students need not be exempt from assessment simply because they are taking part in a human rights lesson. However, these evaluations should reflect the multifaceted goals intended for students (intellectual, skill and affective-values development) as well as the diverse pedagogical methods used (individual work, small group work, project work, discussion). Whenever possible, the teacher should not simply give a mark, but include constructive comments that note the strengths of the students' work as well as areas for improvement.

The areas of student development that might be assessed are:

- understanding of content, remembering basic factual material and grasping the meaning of concepts;
- skills of analyzing problems, skills of understanding the perspectives or points of view of other groups;
- attitudes, motivation or interest
- application, action and generalization¹¹

Most educators who teach human rights education use a combination of assessment techniques for capturing these various learning domains and pedagogical methods. Table 3 contains a sample marking system that incorporates participation in group work and discussions, results of cooperative projects, and written exercises and tests. Grades for project work and participation in classroom discussions might be given by classmates, as well as the teacher. Also, students might conduct self evaluations for their contributions to group work.

Most teachers are already familiar with standard assessment techniques, such as administering tests and marking essays. Giving marks for non-traditional classroom activities, such as small group work, is more challenging. Sometimes the teacher does not feel that he or she has sufficient information to assess the participation and cooperative behavior of individual students in group work.

¹¹ Torney-Purta, J. (1989) 'Issues of Evaluation' in D. Hicks and M. Steiner (eds) *Making Global Connections* (Edinburgh: Oliver and Boyd), p. 167.

Table 3. SAMPLE PLAN FOR MARKS FOR ONE TRIMESTER OF CLASSES

- 25% Marks for each group activity (1 per week)
- 40% Written tests and homework assignments
- 25% Project work (1 per trimester)
Graded for design, execution, and educational value for the student.
- 10% Participation and contribution to classroom discussions

Teachers could allow for student self-evaluation and constructive peer evaluation. These methods will help to strengthen a students' reflective process and encourage more self-direction in learning. When there are differences between the results of self evaluations, peer evaluation, and the teachers' assessment, these differences can be discussed and evaluation procedures adjusted. Below are some sample criteria for appraising small group work. These features can be used as a checklist, or with ratings: 1=good; 2=fair; 3=poor.

Table 4. APPRAISAL OF WORK IN SMALL GROUPS¹²

- keeps the purpose or task in mind
- cooperates with other members of the group
- works without disturbing others
- is courteous to all group members
- completes a fair share of the work
- helps find ways to improve group work

Examination of other artifacts of students work, such as reports, maps, and artwork, may be done to appraise learning and identify points to clarify in instruction. Comparison of samples gathered at the beginning and end of a term or unit may be used to assess student progress. In cooperative activities, teachers might give marks for individual students but also for whole groups or pairs.

Teachers in human rights education classrooms are often also concerned about how to fairly appraise the affective qualities of students. The teacher might decide to leave a students' personality charac-

¹² Michaelis, J.U. (1988). *Social Studies for Children*, 9th ed. Englewood Cliffs, NJ: Prentice Hall, p. 388.

teristics and value system ungraded. An alternative is to try to apply the criteria listed below, using the teacher's assessment or children's self-evaluation as the basis for marks.

Table 5. APPRAISAL OF OPEN-MINDEDNESS¹³

- considers new ideas and activities
- tries new ways to do things
- puts facts before feelings in discussions
- changes conclusion in light of new facts
- bases judgments on fairness to everyone
- considers all sides of an issue
- recognizes stereotypes and prejudice

Table 6. SELF-APPRAISAL FOR DEVELOPING VALUES¹⁴

How do you rate yourself on the items below?
(A for very good, B for good, C for fair and D for very poor)

- respect for others
- interest in others
- listening to others
- sticking to the job
- sensitive to others' needs
- fair judgment of others
- cooperating with others
- thinking before acting
- being honest
- being frank
- helping others
- admitting errors

Finally, the teacher should always integrate evaluation with instruction. A teacher's informal obser-

¹³ Ibid.

¹⁴ Michaelis, J.U. (1992). *Social Studies for Children*, 10th ed. Englewood Cliffs, N.J.:Prentice Hall, p.377.

vation of the different learning activities, combined with formal assessments, provide valuable information about the students' use of concepts and expression of attitudes. This information can be the basis for the tailoring of lessons in order to help students reach the many learning goals of the human rights education classroom.

Evaluation of Teacher Trainings

Evaluation of in-service training programs is practically standard procedure. Typically, assessments are organized through an anonymous written evaluation form that is distributed daily and/or at the end of a training. Informal feedback can also be given orally in whole group meetings.

If questionnaires are used, organizers often include a combination of closed-ended and open-ended questions that ask participants to indicate their overall degree of satisfaction about the training and its utility for their classroom teaching.

Table 7. SAMPLE IN-SERVICE EVALUATION FORM

Section 1. Operational aspects of the training: Please check (√).

- | Excellent | Good | Sufficient | Problematic |
|-------------------|------|------------|-------------|
| 1. Meeting rooms | | | |
| 2. Accommodations | | | |
| 3. Food | | | |
| 4. Transportation | | | |

Section 2. Experiences within the training

5. Using the scale below, rate how useful the overall seminar was for:
 1=very useful; 2=somewhat useful; 3=not very useful
 - learning about key human rights documents, principles and mechanisms for protection? ____
 - becoming familiar with activity-based methodology? ____
 - learning specific human rights-related activities that can be applied in the classroom? ____
6. How useful were the individual sessions?
 1=very useful; 2=somewhat useful; 3=not very useful
 - [List the titles of the individual sessions]
7. What was most valuable to you in the training?
8. What was the least useful aspect of the training?
9. Suggestions for ways that the training might be improved?

Section 3. Follow-up to the Training

10. How do you expect to apply what you have learned in the training in your classroom or school?

It may be desirable to include a question (such as number 10, above) concerning the participants' intended follow-up activities for infusing human rights education into their classroom, and to follow-up with participants within a four- to six-month period. This follow-up can serve not only as a final 'evaluation' of the training, but as a kind of support to the teacher who may appreciate being reminded of the enthusiasm that he or she felt way back in the training itself.

Field testing of materials

*field testing
of materials*

To field test means that you try out materials in a naturalistic setting to see if they function as intended by the authors or sponsors. Field testing is like a dress rehearsal before a play is performed for the public, excepting that there is time built in to make adjustments to the text.

There will naturally be discrepancies between what the creators intended for a lesson, and the ways that these texts are interpreted and applied by the practitioner. On the basis of field testing, researchers — in collaboration with teachers — gather information about how the text is applied (or misapplied) in order to make changes in the materials that clarify confusions; elaborate on detail; offer more supportive guides for teachers and students; and enhance the chances that the text will be used in the classroom as intended by the author.

Field testing is essential for human rights education texts, especially if activity-based methodologies are not commonly practiced. Moreover, if materials have been translated or adapted from abroad, a trial run to ensure cultural relevance is essential. Of course, field testing can be done not only for text, but also visual and auditory materials. However, this section focuses on the field testing of text, since it is more complex. Most of the criteria, methods and organization of data collection will apply to the testing of other human rights education materials.

Criteria.

*criteria
for
field
testing*

Text specialist Tim Hunt¹⁵ has worked with Romanian publishers in helping them to develop criterion for assessing proposed school textbooks for publication. Table 8 contains criterion can be incorporated into feedback forms that a teacher fills out after every lesson, a written report that a teacher fills out after using a larger portion of the textual materials, or interviews that are conducted by researchers with the field testing teachers. Each criterion could receive a rating — for example, a rating between 1 (unacceptable) to 5 (excellent), with 3 (acceptable). However, it is essential to collect details on the teacher's opinion of the text, so include open-ended questions.

Table 8. SAMPLE CRITERION FOR ASSESSING TEXTBOOKS	
1. Conformity to curriculum	To what extent does the text conform to curriculum guidelines?
2. Content	Is the content accurate and valid?

¹⁵ Hunt, T. presentation on "Effective interpretation of curriculum into print materials", Education for an Open Society Conference, September 1995.

3. Level of language

Is the language used in the text accessible to the pupils for whom it is intended?

4. Pedagogical method

Is the methodology suited to the children's age and level, and are the exercise and test materials equally useful?

5. Presentation and design

Is the quality appropriate in terms of: page layout; size and style of type used; general readability; spacing, margins, clarity of impression?

6. Illustrations

Are the illustrations relevant and appropriate in terms of: quality of execution: style, relationship with text, accuracy, use of color?

7. Originality

What elements of originality or creativity of approach are there; any special features of particular appeal to pupil or teacher?

8. Quality of materials

By how much, if anything, does the book exceed the minimum requirements for text paper, cover material and binding?

9. Teacher support

What help is given by way of a teacher's guide or additional materials in; the methodology of the teaching sequences; background information; test, evaluation and extra exercise materials; answer keys.

In one human rights education project, field testing made use of both weekly feedback forms, and an end-of-the-year report. The weekly forms included the following six questions concerning the teachers' use of the draft text, the success of the lesson, and suggested modifications.

Table 9. SAMPLE QUESTIONS FOR WEEKLY FEEDBACK FORMS

1. Was the activity used as it was presented in the text? Was it modified in any way?
If so, how?
2. Was the lesson completed?
3. Was the lesson successful? Comment.
4. Suggestions for changes.
5. Is the activity realistic for the classroom?
6. Special events that took place in the classroom:

At the end of the year, the field testing teachers completed an overall evaluation form that contained many of the criterion listed by Tim Hunt: language and concept clarity, the age appropriateness of the content, production quality, use of the materials in the overall school curriculum, teacher preparation time required, and other information necessary to support the teacher using the materials.

Methods.

Field testing research typically entails a variety of data collection methods, including the self-reporting of cooperating practitioners, individual and focus group interviews (with both teachers and students), and observations of the materials in use. Other, creative forms of data collection can also happen, such as the review of student work, videotaping of text in use, and the application of attitudinal questionnaires. With the exception of the self-reporting done by practitioners, data collection and analysis are usually conducted by trained researchers. These researchers have the skills and neutrality required for complex field testing processes.

Ideally, the data collection, analysis and modifications of the human rights materials will be a collaborative process between three sets of actors: the practitioner, the researcher and the text developer. The practitioner has the “view from the ground” about the effectiveness of the draft materials, and will make useful suggestions for modifying the text. In a school setting, practitioners might include students as well as teachers. The researcher role is critical for training teachers in data collection, developing instruments, skillfully eliciting suggestions from practitioners and analyzing results. The text developer needs to be informed in an ongoing manner of the results of the field testing, so that material alterations can be made expediently. However, it is ideal if the text developer and researcher conduct classroom observations together, so that the author can see with his or her own eyes how their materials are being used.

Classroom observations of a teacher field testing portions of a draft text might involve the following steps:

Ask the teacher to read through the unit or chapter, and then teach it her or his own way to the students.

Observe the teacher’s method of instruction and make note of the following points:

*field
testing
methods*

- clarity of presentation and logical sequencing
- use of examples and relevance to students' local culture and situation
- use of blackboard and other visual aides
- use of vocabulary
- involvement of students: do they ask questions, does the teacher ask them?
- what are the questions about?
teacher's assessment of student understanding
- does the teacher correlate the lesson with the student's previous knowledge?

Remember that under such observations, it is not the teacher that is being evaluated, it is the text. If there are differences between the way that the author conceived a lesson and the way that it is being carried out — differences that observers feel do not improve upon the lesson — then the burden is on the text developer to improve the draft in ways that will facilitate lessons being carried out as originally intended.

Organization of data collection.

***organization
of data
collection***

A successful field testing project is as equally dependent upon management skills as upon scientific expertise. Field testing of a full-year student text can take up to two years, including preparation for the field testing, data collection, analysis and recommendations for changes. If necessary, however, more expedient forms of field testing can be designed.

The following activities will typically take place in a full-term field testing project:

- selection of field testing teachers and/or schools
- development of field testing instruments and strategies
- preparation of field testing teachers for their roles
- development of a system for ongoing distribution and collection of written data (e.g., weekly fiches, trimester reports, attitudinal questionnaires).
- regular visits to classroom for interviews and classroom observations
- regular feedback to the text developer
- periodic meetings between the field testing teachers, researcher and possibly the text developer.
- formal compilation of results, with recommendations to text author

Of course, field testing should be tailored to reflect the degree of changes that you are willing to make. For example, do not invest large numbers of teachers in an elaborate two-year field testing program if you intend to make only minor editorial changes.

The results of such research will be rich and even surprising. Text developers will learn which topics and methods engage children most, and which teachers find most difficult. Practitioners will offer not only editorial comments, but suggestions for deletions (for example, to shorten a lengthy text) as well as additions (such as a glossary of terms). Benefits for field testing will extend beyond the modifications to the materials themselves. Participation in such an endeavor can be an extraordinary professional development experience for the teachers, as well as the text authors. The results can also be used to advertise the materials when they are ready for publication.

Field testing is usually time consuming, however. As an alternative to field testing, or as a preliminary review before field testing, you might ask a human rights text specialist to review your draft text. This reviewer might look for the following items in human rights materials:

Table 10. SAMPLE CRITERIA FOR HUMAN RIGHTS TEXT REVIEW

- reference to local, national and international documents and mechanisms guaranteeing human rights
- the use of culturally relevant examples
- overall representativeness in text and illustrations of gender, ethnic groups, majority and minority culture members
- the avoidance of stereotyping in language and illustrations
- the presence of local issues of concern
- avoidance of jargon or polemical language
- sensitivity to the material conditions of the classroom.

With some minor modifications, these principles can be used for non-text materials and materials that will be used in a non-school setting.

THIS IS A LOT OF INFORMATION TO TRY AND DIGEST, AND I STILL HAVE MANY QUESTIONS. IS THERE A BOOK THAT I CAN READ TO LEARN MORE?

There is a bibliography at the end of the report, that includes some basic evaluation materials. You might want to check out your local library, keeping in mind that you will need to be discriminating about the many models for evaluation that are out there.

ARE THERE ANY OTHER THINGS THAT I SHOULD KEEP IN MIND WHEN ORGANIZING EVALUATIONS IN THE HUMAN RIGHTS EDUCATION FIELD?

Yes. If an evaluation involves individuals from different nationalities or cultural backgrounds, you might want to think explicitly about cross-cultural issues. Cross-cultural differences obviously affect the ease and clarity of the communication of ideas. Double-check that key ideas are understood by all parties. Cultural differences can also influence the formality of language used, and ideas about how often and under what circumstances communications between the evaluator and program people should take place. It is best to clarify expectations about roles, responsibilities, and channels of communication at the beginning of the project.

***cross
cultural
issues***

Different countries also have different practices when it comes to evaluation and monitoring. In many countries, research and evaluation is conducted exclusively by trained experts, with little collaboration from practitioners, excepting for filling out closed-ended survey forms. Evaluation can be seen as highly threatening. In cultures with this context for evaluation, a qualitative researcher will need to justify the use of alternative data collection methods and work hard to establish a collaborative working relationship with local practitioners. Teachers and others will also need to be convinced that formative evaluations can contribute to their own professional development. Trust will need to be established.

Cultural differences also exist in the education, human rights and management fields themselves. Under these circumstances, change must be measured against local — not non-local — standards. Questions that a non-native evaluator external to a system might ask are:

- What is the starting point for change?
- What is the local capacity for making change?
- What are the constraints?
- Are there any special understandings of human rights problems?
- What are the local traditions in terms of program/organizational management (hierarchical vs. less hierarchical practices)?

A final point regarding the ethos of evaluation, an ethos that is even more imperative given that we are talking about the human rights education field. All those involved in evaluations or research of any kind should be honest, constructive and sensitive in their communications. Work should be conducted collaboratively whenever possible. Data and analysis might be shared early, and feedback from practitioners incorporated into the evaluation. Under circumstances where there are irresolvable differences of opinions between the evaluator and the educational agents, the latter can include their personal interpretations in an annex to the report that will be read by third parties.

BIBLIOGRAPHY

- Bethel, George (1995). "Effective Program Management." Workshop Presentation at Education for an Open Society Conference, Budapest. September.
- Calhoun, E.F. (1994). *How to Use Action Research in the Self-Renewing School*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Campbell, P.B. (1982) *Evaluating Youth Participation: A Guide for Program Operators*. New York: National Commission for Youth.
- Cousins, J.B. and Earl, L.M. (1992). "The Case for Participatory Evaluation" in *Educational Evaluation and Policy Analysis* (Winter 1992, Vol. 14, No.4), pp. 397-418.
- Davis, B.G. and Humphreys, S. (1983). *Evaluation Counts: A Guide to Evaluating Math and Science Programs for Women*. Report to the National Science Foundation.
- Fetterman, D.M. (1989). *Ethnography Step by Step*. Newbury Park: Sage Publications.
- Fink, A. and Kosekoff, J. (1985). *How To Conduct Surveys*. Newbury Park: Sage Publications.
- Fowler, F.J., Jr. (1985). *Survey Research Methods*. Beverly Hills: Sage Publications.
- Herman, J.L., Morris, L.L. and Fitz-Gibbon, C.T. (1987). *Evaluator's Handbook*. Newbury Park: Sage Publications.
- Hunt, T. (1995). "Effective interpretation of curriculum into print materials." Presentation at Education for an Open Society Conference, Budapest, September.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*. Thousand Oaks, CA: Sage Publications, Inc.
- Light, R.J., Singer, J.D. and Willett, J.B. (1990). *By Design*. Cambridge, MA: Harvard University Press.
- Madaus, G.F., Scriven, M.S. and Stufflebeam, D.L.(1991). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston and the Hague: Kluwer-Nijhoff Publishing.
- Michaelis, J.U. (1988). *Social Studies for Children*, 9th ed. Englewood Cliffs, NJ: Prentice Hall.
- Michaelis, J.U. (1992). *Social Studies for Children*, 10th ed. Englewood Cliffs, NJ: Prentice Hall.
- Morris, L.L. and Fitz-Gibbon, C. T. (1978). *The Evaluator's Handbook*. Beverly Hills and London: Sage Publications.
- Smith, N.L. (1991). "The Context of Investigations in Cross-Cultural Evaluations" in *Studies in Educational Evaluation*, Vol. 17, pp. 3-21.
- Torney-Purta, J. (1989). "Issues of evaluation" in D. Hicks and M. Steiner (eds), *Making global connections*, Edinburgh: Oliver & Boyd, pp. 163-170.
- Weisberg, H.F. and Bowen, B.D. (1974). *An Introduction to Survey Research and Data Analysis*. San Francisco: W.H. Freeman and Company.
- 1 These standards were developed by the Joint Committee on Standards for Educational Evaluation

in the USA for program evaluation (Sanders, J (1994) *The Program Evaluation Standards* (Thousand Oaks, CA: Sage Publications, Inc.), but are applicable to all forms of educational evaluation. See Annex.

2 Portions of this section are adapted from P. B. Campbell (1982), *Evaluating Youth Participation: A Guide for Program Operators*. New York: National Commission on Resources for Youth.

3 Morris, L.L. and Fitz-Gibbon, C.T. (1978). *Evaluator's Handbook*. Beverly Hills and London: Sage Publications p. 27.

4 Ibid, p. 2.

5 See Calhoun, E.F. (1994). *How to Use Action Research in the Self-Renewing School*. Alexandria: Association for Supervision and Curriculum Development.

6 Cousins, J.B. and Earl, L.M. (1992). "The Case for Participatory Evaluation" in *Educational Evaluation and Policy Analysis* (Winter, Vol. 14, No. 4), pp. 397-418.

7 Partly adapted from Bethel, G. Workshop presentation on Effective Program Management, Education for an Open Society Conference, September 1995.

8 Davis, B.G. and Humphreys, S. (1983). *Evaluation Counts: A Guide to Evaluating Math and Science Programs for Women*, report to the National Science Foundation, pp. 10-12.

9 Inspiration taken from Fink, A. and Kosekoff, J. (1985). *How to Conduct Surveys*. Newbury Park: Sage Publications, pp. 23-40.

10 Weisberg, H.F. and Bowen, B.D. (1974) *An Introduction to Survey Research and Data Analysis*. San Francisco: W.H. Freeman and Company, pp. 49-50.

11 Torney-Purta, J. (1989) 'Issues of Evaluation' in D. Hicks and M. Steiner (eds) *Making Global Connections* (Edinburgh: Oliver and Boyd), p. 167.

12 Michaelis, J.U. (1988). *Social Studies for Children*, 9th ed. Englewood Cliffs, NJ: Prentice Hall, p. 388.

13 Ibid.

14 Michaelis, J.U. (1992). *Social Studies for Children*, 10th ed. Englewood Cliffs, NJ: Prentice Hall, p.377.

15 Hunt, T. presentation on "Effective interpretation of curriculum into print materials", Education for an Open Society Conference, September 1995.

