Evaluation and the policy context: the European experience

Nicoletta Stame

Dr Nicoletta Stame is Professor of Social Policy at the University of Rome 'La Sapienza', Vice-president of the European Evaluation Society and co-founder and first president of the Italian Evaluation Association (AIV). This paper is based on her plenary address to the 2003 AES conference.

My talk today deals with the European experience with evaluation. A few points are necessary to frame what I am going to talk about.

First, I will be concerned with the development of a European tradition in evaluation, and I will consider two different instances of it:

- European evaluation as it was developed by various European countries;
- evaluation as it became understood inside the European Union context.

In both cases it is a matter of a diverse and multicultural political context.

Secondly, I will explore my topic against the international experience of evaluation, of which the Australasian one is an important part.

Thirdly, I will try to see how the European experience can contribute to an understanding of your theme, namely the politics of evaluation.

Fourthly, I will rely on the chronology of evaluation dissemination set out in the *International Atlas of Evaluation* (ed. Ray Rist, J.E. Furubo and R. Sandhal), which considers 25 countries, including Australia, New Zealand, and Italy. That chronology identifies three phases, and two main causes for the genesis of evaluation: social programs, and the public sector reforms known as New Public Management (NPM).

The three main phases are:

1. the 60s and 70s, when the first programs of the Great Society were launched in the USA, where evaluation was mandated by legislative sunset clauses. In this period only Canada joined the US.[1]
2. the 80s, when some countries within the Anglo-Saxon tradition started introducing NPM reforms in the public sector. Here the UK, Australia and New Zealand were foremost, and some Northern European continental countries followed suit.
3. the 90s, when evaluation was extended to many more countries, thanks to an external push coming from larger agencies, such as the EU for many European countries, or the World Bank or other international agencies for Third World countries like China, Korea and Zimbabwe. In this period the evaluation community has grown internationally, creating networks and links that make for much greater communication than in the past.

Old Europe and evaluation

Speaking in New Zealand, a 'new' country where you are celebrating the 21st anniversary of the AES, it is easy to think of ourselves through the stereotype of 'old Europe'. Indeed, from the point of view of evaluation, Europe has not been an innovator – our EES has been founded only in 1995. Why did evaluation come so late to our shores?

I see three main reasons:

a. A strong ideological tradition of state intervention in socio-economic affairs – the welfare state was born here. In Europe there could have been nothing like the debate on state intervention that accompanied the introduction of the War on Poverty and opened the way to evaluation. In Europe nobody doubts (at least until recently) the need for public provision of social services and for the implementation of public policies: the latter are goods in themselves, which do not need evaluating. The highly ideological debates refer to the amount of social spending (inputs) rather than to outputs or outcomes.

b. An administrative tradition, at least in continental countries that have administrative law, according to which the public servant is judged on the legitimacy of his/her actions, not by their results. Procedure comes first, outcomes come if they may.

c. A strong presence of trade unions in the public sector that are more interested in workers' rights (which they feel are threatened by evaluation!) than in citizens' needs.

This mix has produced a malfunctioning of the state, and at the same time an inability to reform.

This situation has evolved in the 90s. The UK, home of Lord Beveridge and of the universalistic model of the welfare state, has moved to public sector reforms that have also been adopted in other continental countries, with various ideological blending (from the French left-wing government of Rocard to the Italian centre-left governments of the 90s, to other conservative governments). In the same period, the evaluation community in the other countries had come to maturity, and this had made room for dialogue and debate among those people in Europe who became concerned with evaluation. It was clear that we could learn from each other. The European Evaluation Society was founded in 1995.

The first wave: Social Programs

Program evaluation started in the US with the Great Society. What was created then were new political tools, new ways of acting to tackle social problems, and a new practice, that of evaluation.

Just to remind you briefly of some main points:

- Programs can be understood as actions for the purpose of obtaining a change, and they have to be implemented with given means in order to obtain intended results within a given deadline.
- They require evaluation in order to know whether they were effective.
- A method for evaluation had to be worked out. Some thought it was just a question of methodology, but things were not that easy: positivist and constructivist paradigms opposed each other, the former proposing variable analysis (following the teachings of the Bureau of Applied Social Research) and experimentalism (Campbell), the latter proposing qualitative analysis and actors' involvement (Stake's responsive evaluation, Guba and Lincoln's fourth generation evaluation). And Patton tried to tame the 'methodological dragon'. Later on (in the late 80s) this problem was addressed by way of multi-method approaches (Greene and Caracelli). Others spoke of a logic of evaluation (Scriven).
- From the beginning, evaluation has been dogged by the 'black box' problem: programs were conceived as a causal sequence, but often without explaining why a certain result should be the effect of a program seen as its cause. Attention was first given to this problem by Chen and Rossi (1989), and especially by Carol Weiss (1997).

The main legacy of this period is therefore twofold:

- the paradigm debate should be mastered by some form of pluralism of methods
- the black box problem needs an elaboration, which has subsequently taken the form of the theory-based approaches.

Nothing of the kind was present in Europe at this time. The 70s were a decade of great structural reforms (in Italy, a reform of the health system and social security): all hopes rested on the political decision, and there was an assumption that implementation would follow. At the same time, programs would have been considered partial, fragmentary, not up to the situation.

Not having participated in the first wave, and in its debates, the European evaluation community is thus not aware of the sufferings of the beginning. So when it later entered the field, there was an expectation of finding ready-made methods and techniques. Analogously, it was not aware of the need to avoid certain errors, which were repeated (one is not vaccinated against committing the same errors). For instance, there was scarcely any knowledge, not only of the first paradigm wars between the positivist and constructivist approaches to evaluation, but also of their overcoming through Patton's sweeping critiques of the methodological dragon, though multi-method approaches and the 'paradigm of choice' (Patton, 1986). This means that we are still waging a rearguard war between qualitative and quantitative methods. This has been reinforced by the EU, which initially imposed a strong quantitative imprint, while qualitative methods, for all the discourse on methodological pluralism, still have to fight for acceptance.

The second wave: New Public Management

At the beginning of the 80s, with a growing fiscal crisis in welfare states, a new challenge emerged in industrialised states which had expanded social policies: that of reducing public spending.

A new way of thinking about the role of the state emerged, which posed different questions for evaluation. This was the idea of the New Public Management (NPM), developed first in Australia and New Zealand, but also in some European countries (Pollitt and Bouckaert, 2000).

At the beginning of the 80s, with a growing fiscal crisis in welfare states, a new challenge emerged in industrialised states which had expanded social policies: that of reducing public spending.

The characteristics of NPM can be summarised as follows:

- There is a change in the role of the state, 'steering, not rowing' (Osborne and Gaebler, 1992). Not all state functions are equally important: the specific function of the state is to lead, to give orientation, not to execute.
- Therefore, many functions can be outsourced to private or third sector agencies.
- What regulates this relationship is the 'principal/agent' concept, according to which there are distinct roles and responsibilities for the 'principal' (the state agency), which is interested in the outcome of an intervention, and the 'agent', (the contractor), who is responsible for the output of an action.

This system poses new problems for evaluation:

- New evaluation criteria become prominent, beyond effectiveness: see in particular the UK elaboration of the concept of 'value for money', meaning that there must be a direct return for every penny spent. In practice, this has meant a growing attention to efficiency.
- A new distinction comes into currency: that between accountability and learning. Both are new concepts. While it could be said that a knowledge function was present from the beginning, owing to the social science background of the first evaluators, the learning function draws attention toward the ability of actors and stakeholders to benefit from the evaluation. On the other hand, having identified the agent more clearly, s/he is held accountable to the principal for his/her actions.

Two main problems have been raised with regard to this new phase. The first is a corollary of the principal/agent concept: the separation between the leading role of the principal and the executive role of the agent. This separation can be interpreted under many different lights, but I would distinguish between its 'adversarial' and 'collaborative' aspects.

It displays its **adversarial** aspect when principal and agent behave as two separate agencies, with conflicting goals. On the one hand, the principal aims at getting a given result by the simple means of disbursing a sum of money, without having to engage in a fatiguing administrative task. On the other hand, the agent aims to keep the contract going, and the money coming in, often being more interested in the maintenance of his/her own income stream than in the program goals. This can be seen when agents win contracts for something they are well versed in, irrespective of the policy goals. This conflict can have negative repercussions for evaluation:

- Line agencies or contractors (agents) feel responsible only for the output they are expected to deliver, not for how it leads to the outcome. A consequence of this is that agents easily accept monitoring of their activity (which is what they have contracted to deliver), but not evaluation, which would call them to test for something they feel to be outside their control. As a result, they will develop strategies of resistance to evaluation.
- Ministries (principals) are mainly interested in the process of contracting out, since they leave the action to the agent, hoping that a good contract will produce a good product. When it comes to evaluation, they praise ex-ante evaluation, and are not interested in ex-post evaluation, often preceding on the false assumption that good specification of requirements in contracts leads to good results, in the same way as the program designers' bias that a well designed program will bring good results.

The combined result of these two tendencies is actually to undermine evaluation.

However, that same relationship could display **collaborative** aspects. This situation could be studied through Robert Behn's concept of 'democratic mutual accountability' (Behn, 2001), according to which a sharp separation between principal and agent can be counter-productive: the principal should be responsible for placing the agent in the best position to act , the agent should feel responsible for the outcome.

In order to achieve this, it would be necessary to bring principal and agent closer. They should share goals, not have their own complementary (indeed, conflicting) ones. Both (not only the principal) should contribute to program elaboration, goal definition, etc. They should understand that they can both benefit from evaluation, the principal by following the various phases of the intervention, and not only the ex-ante evaluation; the agent by considering process evaluation as a force for empowerment by building his/her capacity to address the present, and other future, situations in which he/she is engaged. Picciotto summarised this in discussing international development evaluation, advocating that the partners (which is already a collaborative concept), i.e. donor agencies and beneficiaries, should 'share objectives, have distinct accountabilities, have reciprocal obligations' (AJE, 2003, p. 232)

The second problem was that the principal/agent concept can only deal with single activities, but it is unable to account for mixed interventions, even though the latter represent the great majority of programs (Turner and Washington, 2002).

What happened in Europe during this second wave?

- Some European states introduced systems of NPM, and evaluation as a consequence of this. In some cases this led to an exaggeration: see the 'audit explosion' in the UK, as exposed by Michael Powell (1997).
- The EU introduced programs for social and territorial rebalancing. The first such programs were called 'Poverty' (reminiscent of the US 'War on Poverty'). Other important programs were the 'Integrated Mediterranean Programs', which have been pivotal for mandating evaluation.

The third wave: Evaluation Diffusion

During the 90s evaluation has become widespread all over Europe, with a combination of the two previous trends. In almost every nation there were public sector reforms introducing some aspect of NPM (e.g. in France through the left-wing Rocard government, in Italy through the centre-left government), meeting with greater or lesser resistance. At the same time, in the social, public health, employment, environmental, education and other sectors, political systems became more open to working with programs, which in due course became more and more complex, integrated and multidimensional.

In these instances, there is a growing interest in getting methods and techniques from the outside, but also in the development of original approaches. National characteristics thus came into play, with the UK and the Nordic countries more linked to the Anglo-Saxon debate, and other countries building on their own cultural traditions.

But a crucial spur to evaluation has been the external push coming from the EU Structural Funds (social funds for human resources and employment, for territorial rebalancing and social cohesion, and for rural development) which have represented a great mobilisation of financial and human resources, and have required monitoring and evaluation of their results. This push has brought with it a special evaluation style, which initially has had a greater impact on countries of the third wave of evaluation institutionalisation, but is also at work in countries of the second wave.

Indeed, the EU has developed a complex system of multi-level governance that is a strange mix of social programs and principal/agent principles, of which a specific architecture of evaluation is a crucial element. To understand how all this works, it is necessary to remember that the EU is a federal system, which has both similarities with and differences from other federal systems such as the USA, Canada or Australia. As in the latter cases, there is a division of competencies among levels (in the EU we have: EU, state, region, municipality), and a devolution of many tasks to the lower levels. What distinguishes the European experience, however, is the legacy of its history. The EU was born after centuries of wars between the states that now compose it, which now have decided to live in peace: European unity is a gradual process of unification, by which individual states give up bits of their sovereignty in order to build a new entity. Diversity and multiculturalism[2] are contributory elements, assets to be maintained through an original model, which I would describe as follows.

The *policies* which the European Union applies establish broad goals: territorial re-equilibrium and social cohesion, integrated rural development, human resources development (along the axes of employability, entrepreneurship, flexibility and equal opportunity). All these broad goals cover multiple dimensions of reality.

The *decision-making process*:

- is characterised by incremental negotiations about conflicting interests among states, or between states and the EU Commission (especially when their governments are on the other side of politics from the EU commissioners);
- is limited to financial contribution: the main conflict of interests hinges on allocations to the states;

- tends to incorporate the many vested interests of beneficiaries, implementers, etc. which, by the way, do not want to be checked by a non-political activity like evaluation.

This system of *multi-level governance* is characterised by the motto ‘the EU states goals, not means’; the latter are established at the lower levels. In brief, the process is:

- the EU establishes general goals and allocates money to the states,
- the states establish specific/intermediate goals and allocate money to the regions (or provinces, or municipalities, or specific sectors), and
- the lower levels decide on programs and interventions, and it is here that the money is spent.

The rule that regulates the relationships among these levels is the principle of ‘subsidiarity’: the higher level does not do what the lower level can do.

An *evaluation logic* follows from all this:

- the EU is more interested in financial evaluation (how the money is spent) than in effectiveness of the interventions – at least this is what happened in the beginning;
- an evaluation hierarchy has been instituted corresponding to the multiple levels of governance: evaluations have been mandated at the EU, state and local level in order to assess the corresponding level of spending. In the beginning, it was mainly a matter of commissioning evaluations, now it is a matter of creating evaluation units.

Analogous with what happened with the principal/agent principle under NPM, European subsidiarity can develop adversarial aspects as well as collaborative ones. There have been **adversarial** aspects when the lower level did not want the EU or the state to intrude into their own affairs: a new regionalism and localism that fought for extreme devolution. But it is possible to develop **collaborative** aspects, when the top is concerned with helping the lower levels do what they can do best, with latent resources, etc.: what we call ‘active subsidiarity’.

The following table can illustrate the range of options open to the actors in this system of governance.

	PRINCIPAL/AGENT	SUBSIDIARITY
adversarial	■ principal responsible for outcome, agent responsible for output ■ do not share goals: conflict of interests ■ agent accepts control, but not evaluation ■ principal is only interested in ex-ante evaluation	■ strict vertical division of competencies ■ lower level does not want higher to meddle with it: extreme devolution ■ top level only interested in financial control, not in effectiveness
collaborative	■ principal and agent share goals ■ principal and agent are both interested in success ■ agent accepts evaluation of how output leads to outcome MUTUAL ACCOUNTABILITY	■ the top is concerned with what the bottom can do (helps exploit latent resources) ACTIVE SUBSIDIARITY

Current European predicaments for evaluators: remedies, alternatives

So far we have seen the premises (complexity of dimensions, plurality of levels). Now let’s look at the problems. What do evaluators do when the top level states goals, not means? Based on the table above, one could see two alternatives.

The first corresponds to the adversarial mode. The top is only interested in assessing whether goals were achieved, not in how they were achieved (variety is admissible, but not relevant):

- the emphasis is on broad indicators of goal achievement;
- they are used in a pre-post verification logic, not even an experimental one, because there is no identification of a program (experimental) situation vs a non-program (control) situation.

The limitation of this approach is that there is neither learning (little understanding of the process) nor accountability, because the link between the agent/implementer and the principal/EU is too loose.[3] Consequently, the higher level is a passive recipient of information.

The second alternative corresponds to the collaborative mode. The top level is interested in understanding what works better, where and why. Therefore, it has to develop ways of understanding differences. In this case, it is constantly concerned with the lower levels, in line with the options of mutual accountability and active subsidiarity. Here we would find a learning organisation approach, in which all parties were involved in actively producing information for evaluation.

However, this is not the framework by which the EU users of evaluation judge the evaluations that they receive up the ladder. Their usual grounds for dissatisfaction with the quality of evaluations are low-quality data, evaluators who are too politicised and not independent, and evaluations that are not clear about the program logic – ex-ante evaluations are not good, the logic of action is missing, there is no evaluability assessment, etc. In other words, evaluations are considered inadequate because they neither perform the task of generalising results (summative evaluations) nor offer suggestions for improvement (formative evaluations). All this is usually attributed to the complexity and ambiguity of programs that 'seldom have well-specified or quantifiable objectives' and to 'poorly-developed monitoring systems' (Summa and Toulemonde, 2002, p. 417).

Two main remedies are proposed:

a) work with the *program logic*. Various models are proposed of the vertical links between general goals (at the EU level), intermediate goals (state levels) and operational goals (local, intervention level). This would make it possible to understand causal links between what happens at the various levels;

b) establish *best practices*. Identify the best application of planned interventions (actions, services, etc.).

These remedies to poor-quality evaluation are at odds with the policy of 'stating goals not means'. Instead of proposing how to account for the complexity and variation that is implied in the latter, they consider them as an accident to be overcome through models and generalisations.[4] Taking up Patton's classification of programs and evaluations as ranging from linear through systemic to complex[5], we could say that what is proposed is a 'linear' model of evaluation in a 'systemic' situation.

My contention would not be that the evaluations are bad[6], but that if evaluations have to match the requirements of the multi-level governance of the EU, they need to take other paths to be relevant. These paths should fit the two main streams that have merged in the European evaluation tradition: social programs and NPM, and they would be an alternative to the proposed remedy.

Theory-based evaluation for complexity

The first alternative deals with the program logic: it refers to the social program tradition and its developments. Take the MEANS guides, which distinguish between the 'hierarchy of objectives' and the 'logical diagram of expected impacts' (vol. 1, p.93 and 95):

- The former works from the top down, establishing a cascade of objectives: the results of the higher level are the goals of the lower one.
- The latter works from the bottom up. The assumption is: to get certain results you have to do certain things: if you do *a* then you will get result *b*, which will have impact *c*.

The logic behind these models is that there is only one theory of how things get done – the right theory – and that the program is articulated into a series of virtual linear chains moving from the results of local intervention, to the effect on performance of national programs, up to the impact of EU policies.

However, to be in tune with the complexity problems of the EU multi-level governance, we could elaborate what I would call a 'theory-based evaluation for complexity'. I see two main versions of this:

- using Carol Weiss's (1997) approach, we could say that among implementers of the big policies there are many theories: let's see what mechanism worked in a specific situation; let's ask stakeholders, implementers, etc. This approach is likely to be more friendly toward the various stakeholders who implement, and benefit from, any European program;
- using Pawson and Tilley's (1997) realistic evaluation, we could say that the outcome depends on the combination between a given mechanism (incentives, regulations, persuasion, providing services or training) and the context[7]
- how people embedded in different situations decide to use them. Although there is argument against the idea of programs as 'change agents', seen as something imposing change from above; in the realistic evaluation concept, by contrast, programs are seen as opportunities that the beneficiaries may decide to take, and hence as facilitators of a change that remains in the hands of actors. The combination of mechanism-context-outcome will tell us 'what worked better, where, in what circumstances and why'.

While the hierarchy of objectives is separate from context, and assumes that a given tool always works in the same way, these theory-based approaches start from complexity of aspects and multiplicity of contexts, and assume that what will work is always a combination of tools. Let me take the example of European employment policy.

The EU goal is raising the employment rate. The state goal is to improve matching between labour supply and demand. The local intervention consists of creating employment centres. There are however, various problems:

- context: each site has a different labour market (tight/slack; manual/clerical jobs available);
- tools: there are many tools for that goal; there can be a different combination of tools in each site;
- theory: it should explain why a given tool works better in a given context, e.g. where are a high proportion of employment occurs in the informal economy, the primary need may be for measures to identify those opportunities.

'Best' practices or 'good' practices?

The logic of 'best practice' is an attempt at establishing that all situations are alike (it is possible to generalise), and that some actors are just better than others. Invariably, some places are always better (Italian example: the province of Emilia-Romagna) and some always worse (Southern Italy in general). If the latter are not up to the former, it is their fault.

In my opinion, the opposite is true. Nothing can be considered best for all situations, and hence generalisable. But there are different situations, and something that has shown to be good somewhere perhaps could be studied and adapted/imitated somewhere else. All this could be attained if there were a continuous interaction between lower and higher levels of the hierarchy, if good practices were reflected in good theories, and there were learning about it. But nothing of the kind is possible inside the existing institutional hierarchy of the EU system of evaluation, where

- lower levels are expected to do only monitoring, not real evaluation; and
- higher levels do not receive contributions from below.

Such a link between theory and practice would allow for realisation of the learning organisation principle, in which mutual accountability would prevail. At the lower level, if people knew what they were doing, what to expect, and how they could contribute to the outcome, then they would be more favourably disposed to evaluation, and understand that it is for their good. At the higher level, theories received from below would be more grounded, and better suited to understanding a complex and diverse situation.

To conclude

Contrary to what happened up to ten years ago, the European scene is now an integral part of the international evaluation community. Individual European countries may follow their particular paths, but at the same time they are influenced by a new governance model, the multi-level governance of the EU. And while all countries have worked out their particular ways of adapting program evaluation and NPM practices, the EU multi-level governance system, which operates in all countries, is a particular mix of both traditions, social program evaluation and NPM.

Contrary to what happened up to ten years ago, the European scene is now an integral part of the international evaluation community. Individual European countries... are influenced by a new governance model, the multi-level governance of the EU.

I have proposed that the European evaluation community try to overcome its current predicaments by:

- building on the evolution of program evaluation, through 'theory-based evaluation for complexity', and
- building on the evolution of the NPM concept, especially its links with the learning organisation and mutual accountability.

The whole international evaluation community faces problems linked to complex programs (an evolution of programs) and multi-level governance (an evolution of NPM). I hope the European experience may be of some interest also to others, if it works as a benchmark, not as a model: we live in a global world, we share problems, we are different.

References

Behn, R., 2001, *Rethinking democratic accountability,* Washington, DC: The Brookings Institution.

Chen, H. and Rossi, P., 1989, ‘Issues in the theory-driven perspective’, in *Evaluation and Program Planning,* vol. 12, n. 4

Furubo, J.E., Rist, R. and Sandhal, R, eds., 2002, *International Atlas of Evaluation*, New Brunswick, NJ: Transactions Publishers.

Furubo, J.E., and Sandahl, R., 2002, ‘A diffusion perspective on Global Developments in Evaluation’, in Furubo, J.E., Rist, R. and Sandhal, R, eds. *International Atlas of Evaluation*, New Brunswick, NJ: Transactions Publishers

MEANS collection, 1999, *Evaluating Socio-economic Programmes*, 6 vols., Luxemburg: Office for Official publications of the European Communities.

Osborne, D. and Gaebler, T., 1992, *Reinventing Government*, Reading, Ma: Addison-Wesley.

Patton, M., 1986, *Utilization-focused evaluation*, Newbury Park, Ca : Sage.

Pollitt, C. and Bouckaert, G., 2000, *Public Sector Reform*, Oxford: Oxford University Press.

Powell, M., 1997, *The Audit Explosion*, Oxford: Basil Blackwell.

Ryan, B., 2003, ‘Death by evaluation? Reflections on monitoring and evaluation in Australia and New Zealand’, *Evaluation Journal of Australasia*, vol. 3 no. 1

Summa, H. and Toulemonde J., 2002, ‘Evaluation in the European Union: addressing complexity and ambiguity’, in Furubo, JE, Rist R. and Sandhal, R., eds. *International Atlas of Evaluation*, New Brunswick, NJ: Transactions Publishers.

Turner D. and Washington S., 2002, ‘Evaluation in the New Zealand Public Management System’, in Furubo, JE, Rist R. and Sandhal, R., eds. *International Atlas of Evaluation*, New Brunswick, NJ: Transactions Publishers.

Weiss, C., 1997, ‘Theory-based evaluation: past, present and future’, in Rog, DJ, and Fournier, D, *Progress and Future Directions in Evaluation*, New Directions for Evaluation, no. 76, San Francisco, Jossey-Bass.

Endnotes

1 In fact, there are European countries that claim to have established evaluation at this time: Germany, with its Länder system that imitates the American federalism, a strong facilitator of the first social experiments; Sweden, which imitated the US with its program of ‘The strong society’ (Vedung, 1997, p. 27).

2 In Europe, multiculturalism does not refer to a problem of integrating minorities: it is the very texture of the European society that is multicultural.

3 Ryan (2003, p. 13) raises a similar point in an article on monitoring and evaluation in Australia and New Zealand.

4 The same happened with program evaluation in the US when Weiss, Cronbach, Patton, among others, put the policy context at the centre of evaluation, as against the positivist tradition that considered it as a ‘threat to validity’.

5 I refer to Patton’s keynote speech to the 2003 AES conference [text not available owing to technical malfunction – *Eds.*].

6 Perhaps one could see here a similarity with what Ryan says of the Australian situation: ‘Australia wanted too much evaluation too quickly’ (2003, p. 7).

7 One could guess that the great popularity of realistic evaluation owes much to its focusing on differences in context, which is central to the European perception.