



National higher education evaluation systems: Methods for analysis and some propositions for the research and policy void

H.R. KELLS

Rutgers University, USA

Abstract. A review of the status of national higher education evaluation systems and an examination of the lack of research into the process and nature of policy choice which brought them into existence and shaped them. Proposals are made concerning methods to assist future policy choices and analysis of these systems and propositions posited to further our understanding and as a basis for further research.

Thirty years ago, when I started work on higher education evaluation by taking part in an external peer visit, I truly believed that it was all about the medieval ideals; magisters acting as true professionals, creating and communally enforcing, if not prescribed expectations for academic behavior, then at least reminding their peers about the normal, generally accepted, range of expected behavior and doing it in ways that encouraged improvement. For, as I saw it, this was *our* academic world and we must improve and protect it for ourselves, for our students and for future generations. However, just as the magisters in the late middle ages relinquished control over much of their domain in search of financial stability and continuity, university leaders in the late 20th century have too often, through either inaction or compliant overreaction, permitted the State and its agents to gain the initiative in and the control of quality, the setting of standards and the matter of reasonable assurance to our clients.

Organized national evaluation systems, in their most progressive examples, are useful, but they are *not*, by far, the most important aspect in a well developed culture of university self-regulation. They are, too often, examples of somewhat patronizing, expensive, and often quite political, activity in the name of accountability. Maurice Kogan (1983) and Guy Neave (1988) have vigorously protested about the most flagrant examples of this aspect of what they have called the 'new managerialism.' On this matter, many university leaders stand condemned as well through inactivity to safeguard their own houses against the absurdities put forth by the worst of the agency officials – those who would throttle even the most mature and dis-

tinguished universities with half-baked recipes and simplistic, wasteful and inflexible approaches.

Sadly, while ineffective and wasteful evaluation activities are the case within some countries, to compound the folly, documents and methods from one country are often passed along by ministry officials and their compatriots like whispers in a party game with essence and direction lost as the often ill-conceived and inappropriate message proceeds. Purposes and means in these proposals become ill-matched and even confused one with another. Simplistic, rote, sometimes partial and even some outrageously expensive procedures are bundled and carted across national boundaries and cultures to breed confusion and to sow the further seeds of distaste in the academic mouth for any of 'this evaluation nonsense' (see Sanders 1995 about Vroeijenstijn 1995; Tomusk 1997).

Twenty years ago in North America and the Caribbean my own relationship to this matter was one of advocacy, living as I did in a part of the world where the culture and understanding of management and evaluation was high. This has changed. My transformation through work in Europe, Latin America – and to a slight degree in Asia and Africa – to a role heavily laden with responsibilities for *ex post facto* damage control, cautionary questioning and the teaching, nay, endless reiteration of fundamental concepts, has become exhausting and disappointing. Country after country and institutions *ad nauseum* seem to fall for the same hucksterism and national or institutional self-promotion, about particular schemes of 'accreditation,' 'indicators,' 'quality management' and the like, without a fundamental understanding about the way, in a given institution, system or nation, to build a culture and relevant system of evaluation (see Kells 1995, 1995a). The idea seems often to be a searching for some relatively quick, magical solution, the adoption of something which *seems* (rarely do they *know*) to work in another place, country or institution. 'We must do it *now*, you don't understand, there are real *problems* here . . . etc . . .'

I recently returned from a two-nation visit in the Andes in which I found that one country which had copied part of an evaluation system from another country, now has seen three years go by after the appointment of a national accreditation commission but with *not one* program which has completed the review. In the other country, the leaders have chosen to use the available funds and talent trying to copy an accreditation system from another country, to the horror of all the universities, instead of attempting to address the known, national strategic investment and reform agenda – in effect, attempting to use, so to speak, a water pistol to fight a forest fire. Other neighboring countries are attempting to copy some expensive and totally irrelevant system from a well known European country – a country whose agency officials and professors

have been aggressively pushing its partial and very unique system across national and poorly understood cultural boundaries in several parts of the world. The illness is spreading. It is time to take stock and use some logical analysis and tools to assist in the unpacking of this complex and unfortunate situation. We need a better way to support the policy choices about national evaluation systems and more vigorous research about these matters.

Purposes and methods

In the following analysis I will attempt, using some tools from both the fields of management and social anthropology, to distill and analyse a wide range of experiences with national evaluation systems. By reviewing in a fairly systematic way both the positive and negative experiences of a wide range of countries and institutions, I will attempt to put forth a set of propositions worthy of further study and some questions and suggestions as a beginning base for policy choices about the establishment and further development of national systems – for most of the existing national evaluation systems are truly only partially developed, often partial in terms of the job which needs to be done and quite embryonic.

The methods I have employed are largely qualitative in nature and are guided by a set of quite useful tools developed both to analyse evaluation systems (Kells 1995b) and, very significantly, to analyse national cultural attributes (Hofstede 1991). The interface between the management and the social anthropological tools provides some power for the analysis of the policy base. The records for a decade of experiences in a wide range of countries were employed as a source of information for this study. For seven countries (Holland, Chile, Finland, Sweden, Colombia, the USA and one province in Canada), the accumulated records are rich and longitudinal extending over several years of on-site visits involving months of work by this writer, not as designer of the systems developed in the nations in question, but as policy analyst, questioner and trainer for participants in the systems – experiences rich with learning. The discussions and analysis in and about those countries ranged from those concerning policy choice and design, to procedures, implementation and, occasionally, reflexive redesign. In another thirteen countries (Argentina, Brasil, Ecuador, Denmark, France, Mexico, Norway, Philippines, Portugal, South Africa, Taiwan, the United Kingdom and Venezuela), my experiences and records range from somewhat to quite a bit less extensive, but are still rich enough to add meaningfully to the pool of information which could be arrayed and analysed in a grounded way.

This type of analysis requires time and distance from the intensive activities often required in the establishment and refinement of national and other

evaluation systems. In order to provide for this, I moved into the Copenhagen Business School in early 1996 for a three year period with the primary purpose of sifting through both my own and others' experiences with evaluation systems and with some particular emphasis on the relationship between the national cultural attributes and local circumstances of countries establishing such systems and the nature of the system which either was established or could, by logical extension of the attributes, have been expected.

Finally, it was hoped to be able to offer a few small pieces to the emerging but admittedly meager theory of self-regulation in higher education, which I first attempted to generate, to little obvious appreciation at the time (Kells 1992), and which was largely based upon studies with Robert Kirkwood (Kells and Kirkwood 1979).

Some background

The literature about organized, national evaluation systems before 1985 is almost totally consumed by reports about the conduct or reformation of the systems developed in the *United States* and the *United Kingdom*. The subelements in the multi-part system of each of these countries have been subject to an agonizing (the USA) and somewhat dizzying (the UK) sequence of attempted reforms, mergers, and trials which at this writing are still underway. The lack of recent continuity and/or the specific nature of these schemes have made interpretation by all but the most dedicated and clairvoyant of international observers almost impossible. A large literature about these efforts, mostly descriptive, fiercely provincial and almost always not generalizable to other settings is what we have, unfortunately, from these the most experienced countries in the world of evaluation. The extensive reports from these countries are seldom placed in a framework which would permit officials or analysts from other countries (save those who choose merely to copy the moves of the great nations), to set their own policies in light of needs, attributes and possibilities, while benefitting where possible from the experience of these more experienced nations.

In addition, since 1985, of course, the world of evaluation has expanded in most parts of the world (see Craft 1992, 1994) as part of the political and related changes occurring in most western countries and first, most notably, and thoughtfully initiated in very different ways, in *France* and *The Netherlands*. The ripples which started in these ponds have in the last 10 to 12 years been felt substantially in probably 40–50 countries wherein one or another kind of organized and more or less nationally-sponsored system has been developed (see Kells 1995a,b). The efforts have become nothing less than a movement and have, of course, generated the usual professional trappings

of numerous conferences, several new and dedicated journals, special meetings for agency executives, continuing 'tracks' at professional association conventions and an extensive formal and informal literature.

It is probably fair to state that this literature is overwhelmingly descriptive; largely story-telling by those whose hard work and/or political interest are deeply intertwined with the messages which are delivered about the system which the writers helped to develop. There are, of course, some serious essays; reflections which extend beyond personal experience, with warnings, musings and interesting questions (see, for example, Brunner 1993; Schwartzman 1993; and Vessuri 1993 for Latin America, and Bauer 1993; Cave, Kogan and Hanney 1993; Teichler 1993; van Vugt 1988, 1993 (with Westerheijden); and Yorke 1996 in Europe). However, serious and independent studies of either multi-nation or multi-institutional cases or those which step back to look at the policy choices which begot them and impacts discernible, are few. Some of these have been undertaken at an initial level and limited by the brevity of experience in the countries involved, by our colleagues in Oslo (see Stensaker and Karlsen 1996) and Enschede (see, for example, Westerheijden, Fredericks and Weusthof 1994). **What has not received adequate treatment are the vital policy-related studies which address the central questions with respect to national evaluation schemes.** These would seem to include: the nature of the policy choices; the impact of national cultural characteristics on the possibility of using certain approaches; the factors which seem to relate to the longevity of the system; the impact of the evaluations in light of the stated intentions for the system; unintended outcomes; and the extent of change in the culture and practice of evaluation within institutions either as a result of, or despite the existence of, the system. These things remain relatively untouched in and about the countries and systems we read about all the time.

The findings and some case examples

Before delving into the analysis of relationships between possibly influential factors and the development and presentation of a set of propositions, it is important to attempt to put the collective development of much of the national evaluation movement into perspective with some information across the nations which is related to the range and diversity of the national systems, the apparent relative maturity of the efforts, and the extent of development of a culture of evaluation in the settings concerned.

Range and diversity of national evaluation schemes

If one displays the general characteristics of the systems developed in the twenty or so countries indicated above against a simple taxonomy of general descriptors and in a matrix of evaluation behaviors against characteristics of any evaluation – purpose(s); initiator; unit of analysis; major procedures; framework for judgements and the like – the tools explained in some detail in Kells (1995a,b), then we can begin to describe what has occurred. Even if one discounts the placement of the various national evaluation schemes in a particular category (what, for instance, really is the dividing line between ‘government agency’ and ‘independent agency with government financing and eventual oversight’?), *there clearly is a wide range of types of evaluation schemes in terms of control and units of analysis*. In addition, there is quite a wide range in things such as what the system is called, whether there is a sanction-related decision made (such as in accreditation-type listings) and whether the process has any relationship to direct national budgetary support for regular budgets (as opposed to the occasional use of marginal incentive or development funds to carry out improvements). If one considers the above issues and considers them in light of several obvious country conditions, then the following summary statements can be made –

- there is a *wide range in sponsorship and control* of the national evaluation schemes: with about one-third of the twenty countries in this study clearly having government control or a high level of government control, of the scheme, with about one third having an independent agency with government financing; and, about one third with no government sponsorship or control, most usually in the form of an agency supported by the institutions themselves. The question emerges, therefore, about why these patterns occur; a question I attempt to address in the last sections of this paper.
- there is a *wide range of approaches* employed in terms of *unit of analysis* which is addressed in the system: with about 20% of the different country evaluation instances (about 40 different types of programs across the 20 countries) being directed at entire institutions; about 40% being directed at entire disciplinary programs or departments (including the full range of activities in them); and about 40% being directed at either just teaching and content (‘education’) or at research (usually the whole profile of research rather than the research of a given person). There also are 5 or so % of the individual approaches which could be referred to as meta evaluation, the review of the system of evaluation and quality control rather than of the substance of the institution.
- *most of the national systems which were reviewed in this study are, by any reasonable measure, only partial systems*. In terms of their national docu-

ments and the nature of the intentions sought, most systems are focusing in the initial decade on but one aspect, or at best (in a few instances) on two aspects of these complex institutions. Most approaches with an *institutional or institution-wide* unit of analysis obviously cannot and do not review the individual academic programs or individual administrative systems very well. Most systems which just look at 'education', by discipline or groups of related disciplines, do not look at how the overall programs which host them actually function or their many problems, nor do they expect the institutions to do it. In some of the countries, only a few selected disciplines have been reviewed or just the 'education' dimensions in a few. They are often rather simplistic reviews, which for many institutions (unless the institutions themselves begin to do the type of local, in-depth reviews which only they, acting independently and privately, can do) will make little impact regarding the, usually, stated purpose of improvement. It is certainly possible to do part of the public assurance task with the existing national systems, but not much, it seems, of the improvement task. For a discussion of this critical area, see the section on the ideal self-regulating institution in Kells (1995b) and some of the questions raised by Yorke (1996) and van Vugt (1988).

- there is, very positively, an emerging presence of organized, regular reviews of academic and even administrative service programs *within institutions* in some countries which are initiated, organized and funded *by the institutions in question*. These are clearly not national programs, of course, but in the long run they will become the basis of the institutional responses to such 'outside' national and other evaluations. Through them the basis of the culture of evaluation in the institutions will be developed and these internal schemes, not the national schemes, will yield most of the recognizable benefits – a most important trend. This is happening at some of the universities in eight of the twenty countries reviewed in this study, and in a total of 11 of the 40 instances of types of evaluation in the sample.

Growth of a culture of evaluation

If one reviews the experiences of those national evaluation schemes which have been in existence 5–10 years, then, as expected, one begins to see the evidence of the growth within the system in the attitudes about, the understanding, if not a great appreciation of the value of, evaluation. One sees some of the development of regularly employed techniques, if not adequate infrastructure, for evaluation and one sees the disappearance of professorial and managerial protest and the transformation of inquiries for help from basic, endlessly repeated, first-level questions to more advanced subjects. In

a few instances, one can recognize the ultimate measure of the advancement of the culture of evaluation; the beginnings of an institution-based system which is designed locally, operates on its own schedule and is designed to meet local, as opposed to national needs. This observer believes that most of the 20 countries are an additional 5–10 years away from an adequate enough level of evaluation culture in the institutions so that truly high level of local as well as national satisfaction with the national system will exist.

Purposes and means in evaluation schemes

When the country examples and the admittedly complex patterns are scrutinized, several things seem to this observer to be worthy of mentioning:

- beyond the generalizations previously reported (Kells 1995b) concerning the characteristics vs. behaviors matrix – (internal, highly improvement-oriented efforts of a cyclical, process-related nature vs. externally-directed, ‘one-off’ efforts with government or rationalization-related intentions) – it is clear that *most of the efforts started in the last few years, are clustered broadly across the middle of such spectra, with heavy emphasis on public assurance, on the external peer visit in the process and on academic, often just teaching and program content, as opposed to institutional and administrative matters.*
- Some countries, of course, have several kinds of evaluation efforts and therefore have representations in several parts of the characteristics vs behaviors matrix.
- Most national schemes are basically *new* efforts and they seem, as far as this observer can tell, to be evolving and moving somewhat toward the improvement/internal end of the spectra of behaviors, while keeping their basic character. They *are adding focus on improvement as they find out how to do it and as the culture of evaluation is developed.* This seems to be the case in 4 or 5 of the countries in question, certainly in the countries with a decade or more of experience.
- While there is a reasonable match between stated intention for the system and the procedure(s) selected to achieve it (for reference purposes, see what would seem to be an ideal mix of procedures to achieve certain purposes in Figure 2.1, Kells 1995b), *there are clearly several cases where countries say one thing and then do another with respect to this purpose/means matter.* The most common mismatch is the usual focus on improvement in the stated intentions of the system, *but then one finds an almost total lack of focus on self-evaluation processes, on assistance and training for such efforts, on any national expectation of the use of such processes, and for evidence of the institutional commitment to implement changes and for demonstration of impacts – improvements. Most of the*

national systems, in practice, focus most effort on peer visits and the reports thereof and/or on the institution's tables of often unanalysed data or 'performance indicators.'

Range of diversity and apparent quality in the system of institutions

One of the truly interesting findings from the general review of the experiences of the twenty countries concerns the matter of accreditation (in the Americas sense; that is, the use of a system of sanction-related, choice-related reviews of institutions or programs in light of both a set of pre-stated accreditation standards and the extent to which the unit being examined seems to be meeting its stated intentions) and concerning the implications of the range of quality and diversity in the system of institutions on the appropriate (at least theoretically) type of evaluation scheme to employ.

With respect to *accreditation*, it is clear that the presence of a wide range of diversity and apparent range of quality in the institutions in the national system, such as one finds in the Americas and recently in the countries of central and eastern Europe, requires that a choice-related and sanction-driven system be employed for the national evaluation scheme. The need for basic protection of the public in systems in which the types of institutions range from research universities to taxi-cab or 'garage' institutions is a primary consideration. Indeed, such a set of conditions requires a regulation approach with two steps, not just one. The first is some kind of culling step – an approval against a set of basic standards (concerning the institutions right to exist, to use the term university, to offer certain programs and degree levels) usually run by government and backed by a phalanx of lawyers and often called *licensing*. This step, periodically and cyclically run, ideally, for all institutions, is followed by and, indeed, permits the normal operation of the second, accreditation, step. Accreditation is a relatively weak regulation process compared to licensing, but it is also much more subtle and, if relieved of the basic culling and protection responsibilities, can be the mechanism which encourages and supports truly improvement-oriented efforts with respect to teaching, learning and research and the environments which support them (for more on this see Kells 1996).

If a national higher education system is very diverse and with a wide range of apparent quality, logic would suggest that it should use a system (following licensing) with choices, culling, and sanctions. If, as in much of western Europe, the system of institutions is *not* diverse and the range of apparent quality is narrow and the level fairly high, these things are not needed – and, obviously, the initial 'licensing' has been very carefully done by the State. It is surprising how many of those struggling to set up national evaluation systems, have not begun to think this matter through. They tend to copy a

scheme from another country. A number of countries have attempted to use, for instance, the well-known systems of Chile, The Netherlands or the USA, without analysing their own basic needs. What is worse, they often copy but part of the system forgetting that it was designed to meet specific national needs and to do so with *several* complementary policies and procedures. In addition, of course, the more diversity a national system of institutions has, the more need there is for differentially applied standards and the less likely it is that one can move relatively quickly to employ a meta-evaluation scheme, which work best for relatively uniform and relatively mature systems and cultures of evaluation (for more on this, see Kells 1997).

Some positive and some controversial examples

While it is tempting to identify country examples, it is less than useful to a longitudinal process of growth and development of these systems. Rather, anonymity will remain. There is the chance that I have it all wrong in a given case, that I have extracted too much out of context or drawn the lines too severely. My intention is to illustrate and to expand knowledge, not to damage or to create lists of good or bad places. As I see it, sustainable change comes about through relatively private discovery, not through public embarrassment.

Some very positive national evaluation system achievements

- Clearly three or four of the national schemes (in both Europe and Latin America) have *designed their systems carefully and then sustained the nature and the control of the scheme* long enough so that the programs and institutions concerned have experienced several cycles of activity and both the institutions and the system officials have been able to become comfortable with the scheme. They have been able to build some of the needed culture and infrastructure of evaluation within institutions and even to encourage institutions to begin taking control of the system through the design of internal, institutionally-based regular program reviews. In at least one country in Europe, a major review effort was made after 5 years of experience to upgrade the national system and to build-in improvements without altering the basic approach, the logic of which for the country in question has been confirmed.
- Several national systems have invested in the *training of peer visitors which has paid off in the quality and consistency of the work of visiting teams*.
- In one European country, government-sponsored agencies encouraged *self-evaluation experiments* within institutions for more than a decade, and in South America and in two cases in Africa, such sponsored exper-

iments, over two or three years, significantly assisted in the building of a beginning level of evaluation culture and a firm belief on the part of some academics that they can use such systems for their own purposes. In the case of the European nation in question, a book resulted from some of the experiments and the long term prospect for the national system is that it will be anchored in the recognized responsibility of each institution itself to have a plan for self-regulation and to implement it faithfully, continually and completely.

- In another European country, *a full range of research, teaching/program and institutional reviews were developed over a ten year period with a careful balance being enacted between institutional interests and government concerns*. The idea, at least in 1995, was to move the total system through several more cycles and then to offer to those institutions mature enough to benefit (in terms of evaluation experience) the chance to control their own reviews, balanced and checked only on the meta-level by the government ministry concerned.
- In several countries the decision was to *start the development of the system and to learn-by-doing during a 'protected' or 'experimental' year or two, conducted for volunteer programs or other units*, as opposed to protracted considerations of methods, standards and detailed models and indicators. This has clearly been shown to be a way to defuse fear and opposition (particularly to external peer visits) and to build support for the system and some culture of evaluation, at least in the parts of the universities subject to review. This same logic has proven to be correct in systems developed within institutions as well.
- In one South American country, the national evaluation secretariat, while developing its own scheme, used some of its own *resources to bring respected, informed consultants to help the flagship institutions in the country develop, in a very private way over a four year period, their own internal evaluation systems*. Two of these institutions are now into their third or fourth review cycles and setting the example for institutions all over the country and beyond.
- Several national schemes have wisely set the policy that the *executives of the system* should be energetic, mature, respected academics with some evaluation interest and experience who are totally committed to developing a relevant and effective system. This stands in stark contrast to the cases where politically appointed people with superficial experience and a fraction of the respect and commitment are employed or where a change in government means even the successful officials must go.

Some controversial examples from the experience of national evaluation systems

- Clearly we must start with the reluctance of university leaders (usually elected, and some might say, understandably reluctant) in several of the countries *to band together and assume the design and even the control of the national system*. In several nations, while cooperation was offered by the universities at the last minute, the window of opportunity was lost and the system, developed by largely government initiative, was clearly external to the universities, little able to affect the building of a culture of evaluation in the institutions; a system ‘pasted on’ to the university world.
- In one of the most notorious cases, a ministry in a European country, in an effort to increase research output and to raise faculty salaries, conducted an evaluation (really just a totalling up) of research output by professors in recent years. They counted the publications, drew an arbitrary line of an acceptable number and paid salary bonuses to those professors with more than the acceptable number, with no payments to those with less. They then compounded the madness, all conducted under the guise of improvement, by sorting the institutions by overall productivity level and publishing a league table about it all, doing damage all the way around. This was matched recently in Asia when a new, reform-minded minister decided to quickly evaluate and grade, publicly, all universities.
- Several problems *have been created by wandering experts* on all of the continents, usually by those who know but one system and ‘sell’ it – pushing a round peg into a square hole, situationally, in terms of history and setting, and culturally as well. Several such unfortunate situations have occurred in Central and Eastern Europe and several are still afoot in Latin America. In some examples, this has resulted in the deterioration of proposed ‘accreditation’ processes into a type of ‘one-off’ licensing step for institutions seeking the right to exist. We have seen a melange of partially developed systems wherein but a fraction of the job has been accomplished and the standards, the basis of decision, in these transition countries have sometimes been controlled by the most traditional members in the local academy. The evaluation systems have been rushed into existence, often with an international funding agency and its chosen consultants looking breathlessly over the shoulders of the needful local institutions.
- In another country, in the absence of expressed university leadership, a new government-sponsored evaluation agency with the mandate to review educational programs of the institutions, decided that it, the agency, should conduct part of what is usually the institution’s self-

evaluation, namely to seek the reactions of employers and other clients to the program. The result is a system which many feel has held back the development of the culture and infrastructure of evaluation within the institutions and has caused battles to occur between the agency and the institutions.

- One country has seen the government-initiated evaluation effort adopt, on recommendation of an outside ‘expert’, essentially, a system which had been developed in another country (a planned society; high culture of consensus; high level of quality and low range of diversity in the system; and, a high need to rigorously and rapidly look at every program on a four year cycle; etc.) and then use it in the recipient society which is different in every aspect mentioned and which has a set of institutions vastly different in circumstances and need – a recipe for disaster.
- A nation with vast developmental needs and a reform process slowly underway, has seen its evaluation council, in an effort to impress the government with its ability to ‘do the job’, attempt over the past two years to import a kind of ‘audit’ system (usually employed, logically, with highly mature institutions with a relatively high level culture of and experience with evaluation) into a system wherein experience, culture of evaluation and actual quality are either low or very diverse at best. The casting of this very questionable die will unfortunately be played out in the years to come.
- A country with a fine reputation in the area of evaluation, in general, has seen two conditions emerge as a result of the development of its multiple-aspect approach. First, in an effort to use evaluation to ‘target’ core funding for research, it has used public grading of disciplines at institutions and, as many report, has seen the system stimulate a range of unintended changes in the institutions as they participate in an almost obsessive attempt to secure the funding. Furthermore, the perceived cost in time and money for the relatively unarticulated, multiple systems of evaluation, caused a serious reaction on the part of many participants.
- In South America, several conditions have emerged. Two were mentioned previously – the country which has taken three years to develop a model with not one program yet accredited in a country desperately needing choices and firm review; and the country which knows its major systemic problems, but has chosen, instead of acting on the reforms, the development of an accreditation system which is likely to have relatively little impact. To those one must add the experience in another country where a new reform-minded government has begun to grade the programs of institutions by testing the graduates on a single, rapidly-developed disciplinary examination. This flies in the face of so many technical difficulties

and conceptual faults (cause and effect dimensions; impact of the input variable, etc) as to almost not be worthy of comment.

One could go on with this listing of examples, but perhaps enough has been said. *What remains to be adequately studied in all of this is the nature of the the driving forces for the policy making process.* Whether, in the foregoing circumstances, the choice makers would have *used* a properly developed policy base, if it existed and was available to them, is an open question. It is, however, possible to begin to establish a foundation, useable in many countries, for policy making about analysing the needs of a given country and then matching it with the nature and specific aspects of a national evaluation scheme which could possibly serve well that given country.

Choosing a system: Circumstances of the higher education system and national cultural attributes vs. the types of national evaluation approaches which could possibly serve well

It is often axiomatically stated by many concerned professionals that the evaluation system must be appropriate for the setting. This somewhat idealistic, but surely logical, thought, ought to lead to a thorough examination of *the nexus between national circumstances and possibilities for policy and method.* Too often, however, it leads to a politically nationalistic, more psychological interpretation; namely, if *we* choose or design it, it is *ours*, we will be committed to carry it out and, therefore, it will be correct. And basically, this simple logic seems to have controlled some of the policy making efforts concerning the development of national evaluation systems.

In some cases, we have seen the development of a system which seems to 'fit' with both the general conditions of the higher education setting, what one can call *the national circumstances* (size of the system; maturity of the institutions; range of apparent quality; range of diversity; as well as number of programs; and, critically important historical, political and economic dimensions) *and the general national inclinations – the general cultural attributes, the collective mental programming* spoken of by Hofstede (1991). Used together, these general factors should lead one to be able to select an appropriate approach for a national evaluation system. In a general sense, such choices are often also described as concerning systems which are: public and 'accountable' vs. private and internal; 'now' and 'one-off' vs. long term and incremental; toward identifying the 'best' institutions or programs vs. attempting to help every institution with the possibility for all to benefit and develop; or conducted with a strong government 'hand' or a 'tough' approach vs. controlled by the institutions; using one set of expectations or criterion levels vs. several or many sets; encouraging the scheme to be a 'battlefield

for resources and reputation' vs. 'pulling together' to develop the country; as being 'tough' vs. concerned about how to be 'tough' and how 'tough' to be; etc. There would, however, seem to be a need to *sort out these dimensions in light of the national circumstances and the national cultural attributes*.

In too many cases it is almost as if the need, politically, to establish a system, almost any system, and almost regardless of the cultural and other circumstances, seems to have driven the policy making. The extent to which any of the factors which are obviously potentially influential have been meaningfully considered in the countries which have developed national systems is generally not known. This writer, for example, knows these matters but for a small number of the twenty countries being used in this study. Detailed policy analysis via in-depth' case analysis are required. We can, however, begin the formulation of the questions and some propositions and make them more readily available to choice makers and their staff to assist, at one level, the policy making process.

Choosing in light of national circumstances

Surely, the notion of taking into account the general national circumstances and local characteristics, such as age, maturity, size and complexity of the system to be evaluated are matters for intelligent staff work, wisdom in application and courage of the choice makers. The issues involve, at this level, the old notion of institutional research support, or 'decision support systems.' This is vital to policy or other choice making but it is nothing new or particularly ingenious. For instance, with respect to establishing national evaluation schemes, it makes little sense to try to review hundreds of programs in a brief period of time, and/or when no one knows how to conduct the reviews, and/or when institution-level reviews are needed first because systemic problems plague the institutions. Similarly, it makes little sense to select a technique which is applicable for mature systems and apply it to embryonic ones, or to select one part of a two part interdependent system from one setting and apply it as if it could do both tasks in another setting. Surely the better national agencies and institutions have mastered this kind of analysis. Sadly, many others have not. Significantly, though, what has not been adequately researched and discussed is the more subtle matter of national cultural influences. Herein lies the potential to assist greatly the policy making about establishing national evaluation schemes.

Awareness and understanding of cultural attributes

It seems clear from the analysis of the twenty country cases that in some instances, the system developed would seem to be a logical extension of the

general national characteristics; a kind of extension of the often quoted logic that **education is a mirror of society** and one of its purposes is to socialize participants to be able to function in the ways society requires. Evaluation must fit into that relationship. An obvious example would be the inclusion of public comparisons, a high concern for uniformity of method and reproducible, measurable dimensions and a high concern for accountability and firm assurance of quality by the evaluation scheme, in a *relatively elitist* society which values distinctions between people, institutions, and the like, and wishes, for instance, to ‘target’ public subsidy to the best units in the system. Another example is the adoption of a wide range of approaches and highly internal foci in evaluation with much confidentiality and privacy for some aspects of the system in very *egalitarian* systems, such as in the Nordic countries, where, universally, everyone is considered to be precisely equal in status in society, must be treated accordingly and would be usually be permitted to use a method of their own choice in evaluation. These things are seen to some extent across the sample and certainly very strongly in some of the countries.

What is interesting are the instances wherein the policy makers have chosen a system which seems to be not at all naturally related to the characteristics generally attributed to their culture. A general example would be as follows: a highly politicized society (which usually does not have consensus in its policy making; which often does not see tasks completely accomplished; which does not favor any public acknowledgement of dysfunction; and, with a relatively large number of highly diverse institutions and a wide range of quality) – for such a society to select a system developed by, and which was designed to fit, the opposite kind of country – with, for instance, these characteristics: a highly consensual culture with a national acknowledgement that they have a penchant for detail and proper completion of all tasks, with a small, selective system of institutions with a narrow range of quality and which developed a short-cycle, all-programs-reviewed evaluation scheme which employs publicly available reports. Yet, strangely, this very example of cultural mismatch has been attempted by several nations. Were they thinking about these matters? What conditions and concerns, perhaps caused these factors to be disregarded in the policy making? Were the people concerned aware of them at some level but just silly or unwise?

Thanks to the work of **Geert Hofstede (1991)** and others we are now able to lift this kind of relatively simplistic analysis to a higher plane and in more useful ways to support policy making. *While not a panacea and certainly still requiring judgement and wisdom*, it is step forward for us. I have taken Hofstede’s work on defining several **key, common, national cultural dimensions** (actually spectral or generally graduated dimensions) which he has applied to

Power Distance	the extent to which the less powerful members of institutions and organizations within a country expect and <i>accept</i> that power is distributed unequally. Roles; authoritarian values; salary range in an organization; centralization, consultation, and privilege; equality; power and position; pluralism based on majority vote; power sharing; roles of employees are concerned
Uncertainty Avoidance	the extent to which individuals seek to avoid uncertainty; the relative level of intolerance in society. Extent to which: difference is seen as dangerous, human rights are important; protest is acceptable. It concerns relativism, empiricism, views of youth and of institutions, tolerance for ambiguity, regionalism and internationalism, and support of the political process
Maculinity/Femininity	extent of distinctness in gender roles in society. Reflects relative focus on consensus, living to work, quality of work life, modesty, sympathy for the weak, view of failing, etc
Individualism/Collectivism	extent to which ties between individuals are strong, whether loyalty to the group is strong, and whether the rights of the group outweigh the rights of individuals.

Figure 1. Hofstede's four major national cultural attributes

many countries, and I have related them as best I can to the **characteristics of and choices with respect to national evaluation schemes.**

We must begin with *Hofstede's uniformly applicable cultural dimensions*, which are four in number: *power distance; uncertainty avoidance; masculinity/femininity and individualism* (see Figure 1).

Clearly, the evidence put forth by Hofstede (see his country data by characteristic, 1991) dramatically points out that countries demonstrate very large

differences on these attributes, attributes which have a significant number of relevant dimensions concerning organizational life, classroom behavior and societal views of institutions, professionals and leaders – all aspects which are relevant to the world of higher education institutions in a given country and to the nature of systems which serve them such as national evaluation systems.

A dramatic example from the extensive national data presented by Hofstede would be the differences between highly masculine societies with a high power distance (such as Mexico) and those which have low power distance and are feminine cultures (such as Denmark). Mexico's ratings on masculinity and power distance are four to five times higher than those of Denmark. Its uncertainty avoidance figure is 3–4 times higher. One would have to conclude that the type of organizations and the type of management schemes which might work best for them and the way that evaluation schemes should be organized in these two countries would and should be very different. For instance, with regard to evaluation in Denmark, it would seem that both a wide range of professionals and leaders of institutions would have to be included in the design of the evaluation system (because of the low power distance), that there would have to be a variety of approaches possible and that this would be respected by the authorities (because there is low uncertainty avoidance) and that improvement-orientation, which provides for upgrading of every unit rather than selecting the best (femininity scores are very high in Denmark) must be used. In Mexico, the system *could* conceivably (I didn't say *was*) be designed centrally, imposed on the institutions and it would be accepted and respected; the scheme could be fairly uniform and mandatory and it could even target rewards for good places and rationalize and retrench others with general societal support. In Denmark, while there is acceptance of the fact that a national evaluation scheme can exist (the rights of the group and central steering are respected), there is much more grousing about it than one would expect to find in Mexico (individualism scores are higher in Denmark).

In Figure 2, I propose how, in general, one could apply some of the ideas put forth by Hofstede to help steer the choices about national evaluation schemes. I have taken the first three of the four, admittedly not mutually exclusive, dimensions put forth by Hofstede and tried to ascribe the most logical potential characteristics of national evaluation systems to the extremes of the Hofstede spectra. I hope this initial attempt to provide a link between large national cultural inclinations and some of the choices required in developing a national evaluation system will lead to much more discussion on this topic. I truly believe that some of these links are strong and that the potential for the construction of a relatively simple and useful guide *for the first levels of choicemaking* in policy development is large. Perhaps an expert

High Power Distance	High Avoidance of Uncertainty	Masculinity
<ul style="list-style-type: none"> ● heavy use of objective measures, performance indicators ● use of ranking and comparison ● heavy acc'tability ● public reporting ● system designed by people in power, centrally ● peer review highly important ● powerful offices and figures rarely evaluated 	<ul style="list-style-type: none"> ● standards and firm even prescriptive, objective criteria ● use of ranking and comparison ● favor the best ● tendency toward one high standard, one model ● deadlines and firm schedules for review ● take long time to design system ● short perspective 	<ul style="list-style-type: none"> ● money is important; link eval. to funds where possible ● use sanctions; make choices ● favor the best ● use of performance measures ● focus on rules, schedules and compliance ● self-eval is more difficult; public acknowledgement of dysfunction rare
↕	↕	↕
Low Power Distance	Low Avoidance of Uncertainty	Femininity
<ul style="list-style-type: none"> ● many professionals help design the system ● all levels and parts are included ● self-eval, internal focus ● all get evaluated ● improvement is the focus ● national system best based on university-based culture and internal eval. systems 	<ul style="list-style-type: none"> ● different approaches are permitted ● long term perspective; cycles used ● learn-by-doing; 'get started'; protection ● low reliance on standards ● high focus on achievement of stated intentions and on use of guild expertise 	<ul style="list-style-type: none"> ● few if any sanctions; some incentives used ● all are helped; 'best' are <i>not</i> favored ● little public reporting ● training provided ● change to come through 'discovery' and acquired 'psych. ownership'

Figure 2. Probable appropriate characteristics of national evaluation systems arranged by major cultural dimensions

system might be developed to assist with such first choices. In essence, if the relationships put forth in Figure 2 are as useful as I think they are, then these, coupled with other reasoned choices which flow from national circumstances (as opposed to cultural attributes, such as size, maturity and complexity of the system, range of apparent quality and diversity, and the like), should be able to assist in the development of, and even the review of past choices about, national evaluation systems. It does, of course, place some faith in the country data presented by Hofstede, which came largely from a massive study conducted by the IBM corporation with middle managers. My own reactions to the data by country seem, in most cases, to coincide with my experiences in many of the countries.

Let us examine how this approach might aid in the development of national evaluation schemes. If one considers a country such as Sweden with a highly feminine culture, a low tendency to avoid uncertainty and a low power distance, then, as expected, one finds a country which has fostered improvement-oriented, longer term approaches, with local university-based evaluation systems of a wide variety as the basis of the national system. There are no lists of performance indicators, no one model imposed and no use of centrally based standards. The object is assurance to the public via continuing, generally dispersed, yet centrally steered, efforts to make a good system of universities better. The evaluation system which has slowly but continuously, been fostered over the last 20 years in Sweden is one which 'fits' the culture and the circumstances. Denmark is similar, and has also developed in recent years a central system to review 'education' aspects of selected disciplines. When, during a recent, voluntary, university-invited, multi-national team visit to review the strategic development and to audit the quality systems of a Danish university, a prominent British team member expressed his dismay that the Danish university permitted each faculty unit to design its own quality system, he was expressing the British tendency to want high accountability and public comparison; he didn't understand Denmark.

If one looks at national evaluation systems which have floundered somewhat, which have not been readily accepted or have been recognized to need reform, one can begin to understand the situation if one uses the Hofstede cultural attribute scales. For instance, in Colombia, Ecuador and the Philippines, all cultures with high masculinity and with high power distance (Colombia also has high uncertainty avoidance), the national systems have developed slowly, have been criticized for being partial (not containing needed strong licensing dimensions and active basic public assurance mechanisms). We find on examination that all three have copied part of the USA accreditation system and have tended to hope that universities would on their own develop control mechanisms and use the accreditation system to provide reform and

public assurance. With a few exceptions, it hasn't happened. The culture, and the circumstances which have developed naturally in it, demand a more **accountable, centrally developed, objective, sanction-oriented system.**

When performance indicators were suggested for the Netherlands in the mid-eighties, a broad scale opposition was based, I believe, on both *cultural* considerations (high femininity; high individualism; low power distance and high cooperation across institutions and between institutions and government based on a high culture of consensus and acceptance of central steering) and on *circumstances* (**high average quality; low range of diversity; small number of institutions**). Very simply, the indicator schemes were not needed and the wise Dutch planners knew it. They have since successfully fostered a consistent rise in the culture of evaluation in the institutions, and the national system, owned and conducted by the institutions, is celebrating continuity and international acclaim. But when this system is attempted in Portugal, the Hofstede data suggest that the national attributes don't match well on three of the four scales and, also importantly, the national circumstances are vastly different. In Costa Rica where it has also been suggested one might predict a better match on the basis of some of the scales but not in terms of national circumstances.

Clearly, the Hofstede national cultural attributes scheme cannot stand alone. It is a useful tool which, when coupled with a wise and informed analysis of the local circumstances, can help bring a stronger set of choices to a given country. It is not a panacea. It does not explain all the variance in national evaluation systems but it brings some potential help. In the end, of course, no system will be stronger than the wisdom of the choicemaking and the nature of the continuous, hard and good work of a talented staff and policy/decision board. One can lead a horse to water

Some propositions

In light of the analysis in this study, a few things can be put forth by way of a set of propositions, beginning efforts at theory building, which can be considered in future studies and which might supply some assistance to policy makers. They are propositions which one can consider along with those presented in the early nineties by this writer (see Kells 1992, Chapter 7). These would seem to be transportable across national boundaries, while most specific evaluation schemes do *not* travel well across these boundaries –

- ***there is a fairly strong relationship between the national cultural attributes (power distance; uncertainty avoidance; masculinity/femininity; individualism/collectivism) of a country and the type of national evaluation system one can and should build there.*** This stands as comple-

mentary to other important factors such as the national circumstances (size; complexity; diversity and the like). The nexus between these two factors is the critical dimension in policy making about national and other evaluation systems.

- *the relative influence of national circumstances and national cultural attributes* is probably as follows: the circumstances affect the initial statement of purposes, the unit of analysis, the size of the evaluation scheme, the relationship to matters of national reform and whether a phased development of the system is needed. The cultural attributes probably influence the nature of the process and procedures within the system, the relative openness, whether there is flexibility in the scheme, the use of standards, indicators and, of course, the real focus or purpose of the system.
- *the wider the range of apparent quality and diversity in the higher education system, the more there is a profound need for a national regulation system with two parts* – (1) a basic, accountable, inclusive (all institutions), cyclical, rigorously-enforced national licensing scheme to decide which institutions and programs may exist or continue to exist, to decide which ones may be universities and the like, all in a basic, legally-enforced effort to protect the public, and (2), to permit the second stage, perhaps called accreditation in such countries, to function well and to achieve its intentions.
- *the larger the set of basic, systemic problems in a country or in a given institution, the more important it is to implement country-wide or institution-wide reforms in these areas (often, support levels; credentialed staff and other human resource issues; capital and other related investments) before, or at least during, the development of a national evaluation scheme. Not to do so places an impossible burden on the evaluation system* which generally cannot provide such reforms.
- *the lower the level of the culture of evaluation in a national system or in an institution, the more important it is to operate with an 'experimental,' learn-by-doing and protected period, before attempting to implement some evaluation scheme across the system.* The members of the professional corps, or at least a fraction of them, must learn how to do evaluation by trying it, in any scheme, and they must have a chance to develop their culture (attitudes, values, techniques, infrastructure) of evaluation.
- *in most settings it is important to have the national system encourage institutions to move marginal resources to pay for the evaluation scheme and the recommended improvements.* In some settings, particularly where central actions regarding evaluation are acceptable, the national government might consider providing a development fund to assist with this, at least on a matching basis.

- *in most countries, although it is certainly more possible in some than in others (due to cultural factors), in the long run, the heart of the evaluation culture and the evaluation systems must be owned by, designed by and conducted by institutions and they must be internal to the institutions.* If this is achieved, national systems complementing these conditions will then have an opportunity to thrive primarily with respect to the matter of public assurance, will meet less resistance, and can eventually move more readily to some reliance on meta methods.
- *the matters included in Figure 2 concerning cultural influences on evaluation processes may also be stated in such propositional format.*

The foregoing propositions and the more descriptive findings in this study, it is hoped, will help to stimulate work on what seems to this observer to be a large policy and research void. The void must be filled because national evaluation systems will not die off in most countries. All national evaluation systems *must evolve*, particularly as evaluation efforts are developed in useful ways *within institutions*.

This study owes much to many dedicated people who are developing national and institution-based systems and who opened their world to this analyst over the years. For this I am very grateful.

References

- Bauer, M. (1997). 'Changing contexts for quality assessment in higher education: the Swedish case'. Paper at a local conference, Amsterdam.
- Brunner, J.J. (1993). 'Experiencias latinoamericanas: el caso Chileno', in Vessuri, Hebe (ed.) *La Evaluacion Academica*. Vol 2, Paris: CRE-UNESCO, pp. 37–49.
- Cave, M., Hanney, W. and Kogan, M. (1993). *The Use of Performance Indicators in Higher Education*. London: Jessica Kingsley.
- Craft, Alma (ed.) (1992). *Quality Assurance in Higher Education*. London: Falmer.
- Craft, Alma (ed.) (1994). *International Developments in Assuring Quality in Higher Education*. London: Falmer.
- Hofstede, G. (1991). *Cultures and Organizations: Intercultural Cooperation and Its Importance for Survival*. London: Harper Collins.
- Kells, H.R. (1992). *Self-Regulation in Higher Education*. London: Jessica Kingsley.
- Kells, H.R. (1995). 'Building a national system for higher education evaluation: lessons from diverse settings', *Higher Education in Europe* 20(1/2), 18–26.
- Kells, H.R. (1995a). 'Creating a culture of evaluation and self-regulation in higher education organizations', *Total Quality Management* 6(5/6), 457–467.
- Kells, H.R. (1995b). *Self-Study Processes*. 4th Edition. Phoenix, Arizona: Oryx (also available in Spanish from Fondo Editorial, U. Catolica del Peru, Lima; in Japanese from Toshindo in Japan).
- Kells, H.R. (1996). 'Higher education evaluation systems for Latin America: An analysis of recent experiences and formulation of a generalized model', *Higher Education Policy* 9(3), 239–253.

- Kells, H.R. (1997). 'The implications of increased institutional diversity for higher education evaluation', in Herbst, M., Latzel, G. and Lutz (eds.), *Wandel im Tertiären Bildungssektor*. Zurich: Hochschulverlag AG an der ETH Zürich, pp. 43–51.
- Kogan, M. and Kogan, D. (1983). *The Attack on Higher Education*. London: Kogan Page.
- Neave, G. (1988). 'On the cultivation of quality, efficiency and enterprise: An overview of recent trends', *European Journal of Education* 23(1/2), 7–23.
- Schwartzman, S. (1993). 'Experiencias Latinamericanas: el caso Brasileño', in Vessuri, H. (ed.), *La Evaluación Académica*, vol 2. Paris: CRE-UNESCO, pp. 29–36.
- Stensaker, B and Karlsen, R. (1996). 'Evaluation in higher education in Norway', in Smeby, Jens-Christian (ed.), *Evaluation of Higher Education in the Nordic Countries*. Stockholm: Nordic Council.
- Sanders, C. (1995). 'All you need is love', *Times Higher Education Supplement* 12 May, p. 11.
- Teichler, U. (1993). 'Beneficios y peligros de la evaluación', in Vessuri, H. (ed.), *La Evaluación Académica*, 1. Paris: CRE-UNESCO, pp. 28–46.
- Tomusk, V. (1997). 'External quality assurance in Estonian higher education: Its glory, take-off and Crash', *Quality in Higher Education* 3(2), 173–181.
- van Vught, F. (1988). 'New Autonomy in Higher Education', in Kells, H.P. and van Vught, F. (eds.), *Self-Regulation, Self-Evaluation and Program Review*. Utrecht: Lema
- van Vught, F. and Westerheijden, D. (1993). *Quality Management and Quality Assurance in European Higher Education: Methods and mechanisms*. Brussels: Commission of the European Communities.
- Vessuri, H. (ed.) (1993). *La Evaluación Académica*. Paris: CRE-UNESCO.
- Vroeijenstijn, A.I. (1995). *Improvement and Accountability: Navigating Between Scylla and Charybdis: A Guide for External Quality Assessment*. London: Jessica Kingsley.
- Westerheijden, D., Frederiks, M. and Weustof, P. (1994). 'Effects of quality assessment in Dutch higher education', *European Journal of Higher Education* (29), 181–200.
- Yorke, M. (1996). 'Shouldn't quality be enhanced', *Tertiary Education and Management* 2(1), 86–94.

Address for correspondence: H.R. Kells, Knabrostraede 28, Copenhage 1210, Denmark

Fax: +4533 138828