

# Journal of Judicial Administration

## FEATURING

The Changing Face of Case Management: The New South Wales Experience  
*The Hon Justice J R T Wood*

Evaluating the Impact of Judicial Education  
*Livingston Armytage*

The Criminal Process in Transition:  
Balancing Principle and Pragmatism—Part 1  
*Justice J Badgery-Parker*

## EDITOR

Peter A Sallmann



The Law Book Company Limited

## Evaluating the Impact of Judicial Education\*

Livingston Armytage

Judicial Commission of New South Wales

*Evaluation is essential to judicial education. Most directly, evaluation measures the quality of the learning process for the individual judge. Of equal significance, evaluation measures the impact of continuing education on judicial performance, and provides the means to demonstrate the judiciary's concern for the development of competence. Thus, educational evaluation integrates the pursuit of competence and, in the evidence of success in that endeavour, a means of social accountability.*

*This article critiques the practice of evaluating judicial education, which it finds to be generally inadequate, inappropriate and of limited utility. It is argued that there has been a failure to develop an appropriate evaluation model which can provide meaningful measurements of value within the constraints imposed by the doctrine of judicial independence.*

*To rectify these deficiencies, the article postulates a Judicial Systemic Performance model, which uses systemic rather than personal indicators of judicial performance, including trial disposal, appellate disposal and complaints. This model supplements existing practice to provide the means to assess and demonstrate the impact of judicial education without infringing judicial independence.*

*"The evidence is impressive that continuing education programs can have a major impact on performance. However, such benefits are not automatic and they are not easily assessed."<sup>1</sup>*

*"Evaluation is a vital component of almost any program that wishes to remain competitive. Evaluation becomes the watch-dog of efficiency in the planning and implementation phases, and the guarantor of effectiveness in the end results."<sup>2</sup>*

*"Attesting to the need for evaluation is somewhat akin to deciding to take exercise more regularly. Both are resolutions that are deemed important and necessary, but both are, for whatever reasons, rarely implemented."<sup>3</sup>*

\* This article is based on a paper presented at the National Association of States' Judicial Educators (NASJE) Annual Conference, San Francisco, 10-13 October 1993. The views expressed in the article are purely those of the author, and do not necessarily represent any official views of the Judicial Commission of New South Wales.

## Evaluation Techniques

### *What is Educational Evaluation?*

Evaluation is the process of assessing the value or worth of any educational endeavour, in terms of its effectiveness in accomplishing its goals or results.

In common parlance, *evaluation* is the process of judging or assessing worth,<sup>4</sup> and is commonly referred to as the means of assessing success or failure in attaining planned objectives. In educational terms, this assessment involves obtaining information for use in judging the worth of an instructional programme, product, procedure or objective.<sup>5</sup> More specifically, *educational evaluation* can be described as making informed judgements on the overall value of the learning programme and whether or not the programme accomplished what it set out to do. This usually involves making judgments concerning the quality of the impact of training, by reference to an assessment of goals and objectives with their outcomes.<sup>6</sup> On the basis of these judgments, decisions can then be taken about the feasibility, effectiveness and educational value of the programme.<sup>7</sup>

At a more technical level, Cook and Reichardt define evaluation as the systematic endeavour to describe, understand and judge the worth of the activities and experience which actually occur within the instructional events of a programme (process), and of the outcomes (impact) on participants delineated in cognitive, affective and/or psychomotor outputs, usually measured within three environments (social, environmental and organisational).<sup>8</sup>

Chinapah defines educational evaluation in the generic terms of social research, as the systematic application of procedures to assess the conceptualisation, design, implementation and utility of a learning project or programme:

“As a process, evaluation can be seen as an attempt to assess the relevance, effectiveness (significance) and impact of an intervention, usually in the light of its objectives.”<sup>9</sup>

Others see the definition of evaluation depending on the broader, underlying philosophy of education, which affects how one intends to use the acquired evaluation information. Grotelueschen documents a range of different applications which include evaluation as the process of correlating the congruence of learner outcomes and programme objectives; comparing performance data with commonly accepted standards; comparing actual effects of a programme with a variety of demonstrated needs; and judging a programme critically using expert knowledge.<sup>10</sup> Gardner provides five alternative definitions depending on application, which classify evaluation as educational measurement, using a quantitative index of performance; a process of professional judgment, with experts giving considered assessment of quality; a comparison of performance data with clearly specified objectives, which identifies discrepancies by comparing actual performance to standards or benchmarks of performance; a process of identifying and collecting information to assist decision makers; and as a goal-free process, noting actual programme effects.<sup>11</sup>

Ultimately, it is argued that the most useful definition of educational evaluation for general discussion is provided by Scriven who describes evaluation as the systematic collection and use of information to make informed decisions about an education programme.<sup>12</sup> This definition possesses the dual qualities of being generally descriptive but permitting the context of the ensuing argument to define more precisely the meaning which is most appropriate. Accordingly, the term “evaluation” will be defined in this general educational sense for the purposes of on-going discussion throughout this article.

### *Evaluation versus Research*

It is argued in this article that evaluation embodies a distinctive research methodology.

Evaluation and research both share the process towards systematic collection and analysis of information. Research, on the one hand, is defined (at least in social science terms) as an activity aimed at obtaining generalised knowledge by contriving and testing claims about relationships among variables or phenomena. This knowledge results in theoretical models, functional relationships or descriptions which may be obtained by empirical or other systematic methods, and may or may not have immediate application.<sup>13</sup> Evaluation, on the other hand, is an applied form of research and is directed towards practical applications of knowledge, with immediate utility. The goals of this inquiry include answering questions pertaining to the worth of educational materials and activities. Evaluation is concerned not just with knowledge but with knowledge for action, with description, judgment and the facilitation of understanding and decision-making.<sup>14</sup>

Thus, educational evaluation is an applied means of research with recognisable method, which is defined by its purposes and goals, and is orientated towards utility by fulfilling an inquiry for its client, being specifically an assessment of value. It follows that the question which must ultimately be addressed is what evaluation method is both useful and appropriate in any assessment of the value of judicial education. Finding an answer to this question is the objective of this study.

### *Fact versus Value*

It is also argued that any notion of evaluation is dependent on the declared or latent framework of value within which it is conducted and is inescapably normative. In effect, it is not possible to conduct an evaluation without having first defined the purpose of that evaluation.

Just as the notion of assessment hinges on the concept of value, so any notion of evaluation is dependent on normative criteria applied in making judgments about educational programmes, and determining their value or merit. Brookfield argues that evaluation becomes pragmatic in a normative vacuum. Any approach to evaluation requires criteria and procedural features with which to operate. Preferably, these should emerge from the nature of the adult learning process itself.<sup>15</sup> Thus evaluation is “inescapably” a value-judgmental concept,<sup>16</sup> and requires a normative structure which consists of measurement against value scales, making value statements and the justifying of those measurements.<sup>17</sup>

In this sense, evaluation is clearly more than just gathering data.<sup>18</sup> To operate, the evaluation process fits within the context of a model that derives criteria relating to quality, value, worth, effectiveness and procedural features which arise from an organisational context and are channelled by the adult learning process.

Within this context, there is a debate among educational researchers relating to *Fact* and *Value*. Purists argue that evaluators ascertain facts and clients/sponsors place values on these facts—that is, that it is a value-neutral process; pragmatists assert this dichotomy is a theoretician's fiction, and claim that all social science research is value-based. This debate, therefore, is at what point does the process become normative and judgmental—finally in the hands of the decision-maker, or from the outset in the hands of the evaluator?

Brookfield argues that scientifically “pure” data is of little validity unless it can be applied in settings where individuals are able to relate it to value components. He declares that the adult educator cannot escape valuing by burrowing into data.<sup>19</sup>

House goes further to argue that evaluation can variously be seen, as a process of illumination, advocacy or judgment. For example, evaluation of teaching will differ depending upon whether one sees teaching as labour, craft, profession or art. If one sees teaching as labour, as a set of standard operating procedures planned and programmed by administrators then, House argues, evaluation becomes direct monitoring of *teacher* performance according to set standards. If however it is seen as art, then an altogether more subjective and qualitative process must be engaged upon.<sup>20</sup> Using this illustration, House concludes that the traditional social science position of the separation of fact and value is fallacious.

The implications of this discussion are hardly polemic. Ultimately, this argument may lead researchers to drop the burden of ostensible objectivity to become advocates of information in collaboration with programme developers rather than a judge. Few issues are as fundamental to evaluation as this debate in the process of educational evaluation continuing to evolve an adequate model for adult and professional learners.

It follows that if we recognise that evaluation invokes a methodological process of assessing value, and if we also recognise the inescapable non-neutrality of that process, then it becomes of critical importance that we can explicitly define the norms upon which any assessment of judicial education rests: what are the criteria of value which are to be applied? In effect, we must state precisely what we want to measure, before it is possible to make any rational assessment. These criteria of value define the framework of educational objectives, standards or benchmarks of competence, and indicators of performance essential to provide any process of measurement or assessment. It will be argued that the practice in judicial education is generally deficient in failing to provide any clear statements of criteria. The nature of these criteria is largely determined by the purpose of the evaluation.

#### *Purpose of Evaluation*

The purpose of evaluation is generally to assess the quality of training intervention in terms of resultant behavioural change, through enhanced

professional competence, and its systemic effects on organisational performance.<sup>21</sup>

It is now argued that evaluation in fact serves two fundamentally different purposes.<sup>22</sup> Historically the first is to provide institutional accountability, or justification, usually to external sponsoring bodies. More recently, evaluation has also become important in providing a learner-based method of assessing the personal worth of the individualised learning process. The selection of any evaluation methodology ultimately depends on which purpose is being served by the evaluation.

In the domain of judicial education, it is argued that providing judicial learners with the means to assess the value of their own learning is the more appropriate endeavour of any educator who is sensitive to the implications of judicial independence. In judicial education, no less than elsewhere, however, it will be argued that evaluation must also provide the means to assess the value of continuing education on the systemic performance of the justice system.

Thus, the nature of the evaluation process varies depending on which purpose is being met: external accountability generally requires greater reliance on objective outcomes, while internal accountability is more concerned with the qualitative learning process. It will be argued that the evaluation of judicial education should combine both, but in practice supplies neither satisfactorily.

#### *Institutional Accountability or Individual Learning?*

The historic origins of modern educational evaluation have had a significant influence on its character and development in the public sector. Evaluation became popular in the United States, as a systematised procedure in education programming during the mid-1960s, and since then grant funding has become closely linked with evaluation.

This genesis has important implications in terms of accountability. Accountability has played an influential role in the continuing rationale of evaluation, which has been intimately tied to funding bodies which are usually governmental.<sup>23</sup> As a corollary, the development of educational evaluation has traditionally been determined by these funding entities which are invariably external to the organisational environment within which the education is undertaken. It follows that significant changes in evaluation practices have, on occasion, resulted from shifts in government policies and requirements. This alignment of educational accountability to external interests is a cause for concern to some educationalists. Some studies have demonstrated that these government accountabilities may be inflexible and inappropriate, and may have diverted attention towards artificial directions in the educational process.<sup>24</sup>

These different purposes underpin distinctively different rationales for educational evaluation. Each purpose has profound implications on the selection of appropriate methodology, and gives rise to the need to clarify the purposes for the evaluation of judicial education from the outset. In effect, is evaluation for the purpose of providing external accountability, or is it for the discrete purpose of equipping individual judicial learners with tools with which to monitor and critique their own progress? This writer

argues that the latter is the only means by which judicial learning will ultimately provide value; notwithstanding the reality of practice that most educational evaluation endeavour is traditionally undertaken for the former purpose.

#### *Formative or Summative Evaluation*

There are in fact numerous purposes for the evaluation of judicial or any other form of education. These can be categorised in a variety of different ways.<sup>25</sup> Grotelueschen, for example, identifies three categories of purposes for evaluation which he describes in terms of time: past, present and future.<sup>26</sup> For Grotelueschen, the central question is “what is the purpose of the evaluation—is it intended for justification, improvement or planning?”<sup>27</sup> Evaluation of past activities, or summative evaluation, is for justification; evaluation of current programme activities, or formative evaluation, is for improvement; and evaluation of future activities is for planning.

The question whether an evaluation is formative or summative determines the time at which the evaluation is undertaken which, in turn, is relevant to the purpose of the exercise. *Formative* evaluations are undertaken during the course of the education process usually for the purpose of modifying current proceedings, as required. *Summative* evaluations are undertaken at the end of the process to draw conclusions on the programme at large, usually for the purpose of refining or developing future programmes. In practical terms, the relevance of this distinction is often masked within the more important and adjacent issue of what is to be evaluated: the education process or its outcome?

Houle analyses the purposes of evaluation from an environmental perspective and argues that the appraisal of the quality of continuing education can be conducted at three levels. These levels consist of evaluation of the activity itself (*activity*), the extent and quality of the individual’s developed ability (*learning*), and the profession’s general performance (*outcome*).<sup>28</sup> It is argued that this “environmental” approach is of considerable practical utility: in the experience of the writer, it assists in providing a focus to any evaluation. It requires the process to select a specific perspective and to consistently address each of the three levels of application in turn, inevitably culminating, as it should, in the outcome.

Hudzik and Wakeley provide another important means of categorising evaluation, and its purposes. They argue that evaluation can be classified into two approaches which determine the object and focus of any study. The first is the programme monitoring approach, also known as process evaluation. This approach assesses effort, efficiency, effect, changes in behaviour, and process.<sup>29</sup> The second is the training or learning approach, also known as impact or outcome evaluation. This approach assesses reactions, learning, job behaviour change, and impact or results.<sup>30</sup>

The dichotomy leads to one of the central debates in the literature of educational evaluation, which is polarised around exercising the choice between assessing the process and assessing the outcome of an educational endeavour. The resolution of this debate is of considerable importance as

it affects the fundamental character of any evaluation exercise and is highly influential in determining what is actually done and why. This debate is postulated within two separate but related choices confronting the evaluator: *should the evaluation be formative or summative?* and, *should the evaluation be of process or outcome?* The significance of these choices to the judicial educator are that they succinctly identify the key issues which determine the selection of appropriate evaluation methodology.

#### *The Process or Outcome Debate*

This debate addresses whether the educator must demonstrate the provision of value directly through an assessment of the results of education in terms of enhanced performance, or inferentially through an assessment of the quality of the educational process.

Outcome or impact evaluation, on the one hand, is concerned with measuring the outcome of a programme and the extent to which it produces desired change. Ideally, an impact evaluation should be able to attribute change to the implementation of the programme, while ruling out other possible factors which might have affected the change. This should involve considering the many factors external to the actual programme that can account for some or all of any noted unforeseen outcomes. In practice, however, many impact evaluation designs fail to consider how to distinguish these external influences from the actual programme.

Process evaluation, on the other hand, is concerned with the manner in which a project or programme is implemented, especially as regards the stated guide-lines and design. It is primarily intended to pin-point problems or hindrances which can interfere with implementation of the programme.

The judicial educator must distinguish between the two different purposes for evaluation, and provide means of measuring and demonstrating value for both learner and stakeholder. Clearly, this is done most directly by assessing impact or outcome on performance.

#### *Review of Evaluation Practice*

A review of the broad literature on educational evaluation reveals that practice usually dwells on formative, process evaluation in an effort to encourage use of findings for programme improvement.

For example, Knox argues that formative, process evaluation is the most common form of educational evaluation because convincing evidence about programme impact (such as change in performance) is difficult to obtain, and it is usually assumed to be associated with evidence of satisfactoriness (such as participant satisfaction or knowledge acquisition). Formative evaluation assumes that satisfactory process will lead to impact, while summative evaluation emphasises documentation of the extent and type of impact that results.<sup>31</sup> Other commentators see the assessment of outcome as the *raison d’être* of evaluation. Cervero sees impact evaluation—assessing the application of learning through the quality of a participant’s resulting performance—as potentially the most important means of evaluation because it often deals with the long-term goals of a continuing education programme. However this form of evaluation can also be the most difficult.<sup>32</sup> For Chinapah, the goal of the evaluation endeavour—the

essential challenge for the evaluator—is to overcome the difficulties of impact evaluation.<sup>33</sup> Green and Walsh indorse this view. They castigate the practice of continuing professional education at large for a “glaring weakness” in its failure to document the nature of its effects on professional services.<sup>34</sup> Indeed, they see impact evaluation as the imperative for continuing professional educators, and describe evidence of the impact of continuing education activities on proficiency, professional performance and service outcome as the “missing link” in continuing [medical] education.<sup>35</sup>

Certainly, there is at least a measure of consensus on the difficulties of impact evaluation.<sup>36</sup> As Knox propounds: “The scarcity of excellent impact evaluation reports for continuing education testifies to how difficult they are to conduct.”<sup>37</sup>

There are a number of reasons for these difficulties, not least being the causal difficulty of isolating extraneous influences and persuasively linking cause and effect.<sup>38</sup> There is also the difficulty of obtaining consensus on a few major desirable outcomes that can be readily assessed and whose results can be mainly attributed to educational activities.<sup>39</sup>

#### Judicial Education

In the specific domain of judicial education, these difficulties are no less acute. Hudzik agrees that while this form of evaluation is “often difficult if not impossible” to undertake:

“Ultimately, evaluation ought to concern itself with the question of outcome and impact: have conditions changed, and does the change represent an improvement or a deterioration of performance when set against our objectives?”<sup>40</sup>

These difficulties oblige the judicial educator to fall back to process-based evaluation techniques or some combination of process and formative techniques in order to attain any measurement of the value of the educational endeavour. Pearson confirms these difficulties:

“It is extremely difficult to show a direct cause-and-effect relationship between the stimulus of a single judicial education activity and the response of a specific change in behavior [sic] attributable solely to that activity.”<sup>41</sup>

Hudzik notes that goal divergence is a particular problem for public sector entities:

“In the public sector evaluations typically produce *mixed evidence* [sic] of impact against which the professional *judgment* and experienced decision makers will be set in determining ultimately whether the programming effort had value, and if that value sufficiently offsets its costs. The answer depends on which conflicting values or goals are given by the decision maker.”<sup>42</sup>

It follows that there may be markedly different goals in the arena of judicial education between the perspectives of the state legislature and, for example, the chief justice. This divergence is perhaps best typified by competing priorities for speedy trial versus fair trial or, in effect, quantitative compared to qualitative goals. For the trial judge, sitting in the

middle, the goal is more likely to be seen in terms of participation to the extent that any educational activity can provide immediate benefit in solving actual problems.<sup>43</sup> Meanwhile, other more disparate goals will be held by litigants, the consumers of the justice service, or other stakeholders.

To compound the challenge of educational evaluation, it can be argued that attainment of educational objectives relating to enhancing judicial competence may be very difficult to discern, owing to the difficulty of selecting appropriate performance indicators for an essentially intellectual quality, and the reluctance of judges to subscribe to the process of being assessed through any means other than the formal appeals process.

In relation to the difficulty which may exist to identify and link cause (programme implementation) and effect (performance outcomes), Cervero argues that:

“While this may be true in a strictly scientific sense, the alternative of not collecting any data will not improve program development efforts at all. While this type of evaluation is not commonly done, it should be used when the situation permits it. This assertion is based on the assumption that some information is better than none, as long as the limitation of the data is recognised.”<sup>44</sup>

Knox sees this difficulty of assessing the relationship between education and action as providing “unique opportunities” for educational evaluators.<sup>45</sup>

In the domain of judicial education, no less than elsewhere, it is concluded that the ultimate purpose of any evaluation is to assess the value of continuing education on the professional performance of judges and systemic performance of the justice system. In practice, however, practical as well as doctrinal difficulties frequently lead to an expedient reliance on inferential measurements of the quality of the education process rather than its outcomes, with the result that qualitative assessments are frequently used to provide quantitative measurements.

The continuing challenge for judicial educators remains the development of a model of impact evaluation which selects appropriate performance indicators and viable means to measure their enhancement. How, then, should this be done?

#### Evaluation Methodology

##### Criteria for Assessment

The classic goal of professional education is to ensure the development of competencies or proficiencies that can be translated into professional performance. Evaluation can be seen in Tylerian terms as the process of measuring the congruence between learning objectives and outcomes.<sup>46</sup> The manner in which this measurement is carried out depends on what is being measured, the prior selection and definition of criteria for key performance goals and standards and appropriate performance indicators. It follows that if the educational objectives can be measured directly against outcomes, rather than indirectly or inferentially, then there is a much greater likelihood that the evaluation will be deemed competent.<sup>47</sup>

### Various Formulations

There is a general consensus in the literature on the criteria for evaluation, although these criteria are formulated differently by various commentators. The classic formulation of evaluation criteria was made by Kirkpatrick, who organised four foci for evaluation—reaction, learning, behaviour and results.<sup>48</sup>

These criteria have been refined by Houle who identified the extent of participation, extent of learner satisfaction, accomplishment of a learning plan, and measurement of performance as the four major criteria for evaluation.<sup>49</sup> Cervero indorses this general approach and recasts Houle's assessment criteria into three frames of reference, being the results of formal education activities (*programme*); the impact of learning on quality of practice (*learning*); and the level of performance of the entire profession (*standards*).<sup>50</sup> Within this framework, Cervero developed a number of categories of evaluation questions organised around programme design and implementation, learner participation, learner satisfaction, learner knowledge, skills and attitudes, application of learning from the programme, and the impact of application of learning, which he described as "the holy grail" of evaluators.<sup>51</sup> This formulation is similar to Hudzik's programme evaluation approach which comprises an amalgam of six criteria: effort, efficiency, reactions, learning, job-behaviour change and results.<sup>52</sup>

Although it has been argued that there is a consensus in the literature that the evaluation exercise is ultimately directed towards attainment of Cervero's "holy grail"—that is, the assessment of results or impact of learning on behaviour and performance—observation of practice frequently reveals that these other subordinate criteria dominate the assessment process owing to the significant practical and doctrinal difficulties of assessing impact on judicial performance through direct means.

### Performance Indicators—The Need for Benchmarks of Proficiency

It has already been argued that evaluation embodies the assessment of outcomes against objectives. This assessment involves measurements within a framework of standards defined in terms of behavioural proficiencies.

In order to study the nature and extent of any educational impact, educators need to specify benchmarks against which this impact can be assessed: most frequently, educators appear to rely on expressions of participant satisfaction as a benchmark of impact. This practice is problematic, however, as there is mixed evidence of the association between satisfaction with the programme and change in performance.<sup>53</sup> Because of the equivocal nature of the relationship between satisfaction and enhanced performance, Knox argues that the selection of performance indicators should

"deal directly with specific and achievable changes in performance that are important to the adult learner, are amenable to educational influence, and that can be readily documented."<sup>54</sup>

Some performance indicators are readily used for educational or proficiency evaluations, such as personnel and production records. However clear-cut measures of effective performance are less available for some professional, managerial and highly technical positions, and are unsuitable to measure a range of learning gains or objectives where visible, quantitative outputs may not be readily apparent or available. Indeed, it becomes successively more difficult to link education and performance where the measurements relate to changes in knowledge, skills and attitudes respectively. Knox observes that the complexity of professional performance and the lack of quantifiable standards and measures of excellent performance make impact evaluation more difficult than in other occupational training.<sup>55</sup>

These difficulties are no less acute in the judicial arena than in any other professional domain. Indeed, they are compounded by impediments associated with the doctrine of independence which render a range of conventional measuring techniques inappropriate with judges. Brookfield notes that these difficulties of assessment are exacerbated and become "somewhat tortuous" when dealing with a host of highly interpretive, essentially intellectual, frequently discretionary political, social, moral and ethical questions which frequently arise.<sup>56</sup> He argues that such difficulties—which are endemic in the judicial work-place—may be overcome by relying on indirect and inferential performance indicators, and by triangulating<sup>57</sup> a number of different methods and perspectives in the assessment exercise. However, the position remains that it is not possible to evaluate in the absence of a framework of values, as has been previously argued.

Notwithstanding these significant doctrinal and practical difficulties, it remains essential to design an evaluation framework for judicial education which incorporates an appropriate framework of behavioural benchmarks by which to assess the proficiency of performance.

### Methodological Issues

The design of any evaluation of judicial education is determined by the nature of the information required, its feasibility, and the resources available. Commentators agree that there can be

"quite obviously, no single technique or method that can serve all of the varying roles and types of evaluation. . . . [and evaluation design] must not always be confined to a single applied method".<sup>58</sup>

A range of methodologies or techniques are available for possible application in judicial education. These include experiment, correlation, surveys, client assessment, systematic expert judgment, clinical case-studies, and observation.<sup>59</sup> Data is usually gathered from a variety of sources including tests, interviews, logs, observations, ratings, records, clinical exams, expert opinion and hearsay. Houle argues that the selection of supplementary methodologies can be utilised to validate findings (at possibly considerable expense). These methodologies can include peer appraisal, complex self-assessment programmes (such as professional specialisation programmes assessed by examination), or formal staff appraisal.<sup>60</sup> In practice, it is observed that reliance on any single methodology is frequently inappropriate and may provide unreliable findings.

The validity and reliability of findings is usually best assured through the triangulation of methodologies: for example, the combination of interviews, client surveys, observation and expert judgment is generally likely to test hypotheses and tentative conclusions to reach a consensus on the efficacy of any assessment findings. Consequently, this is the course adopted by the writer in assessing need and evaluating judicial education which is outlined below.

#### *Qualitative versus Quantitative*

The design of any evaluation is affected by the methodological debate surrounding whether quantitative or qualitative techniques and instruments should be used to collect data. While it may be acknowledged that this is an issue which, at least in the opinion of Grotelueschen, usually "generates more heat than light",<sup>61</sup> the outcome of this debate is influential in determining the design of the evaluation.

It is argued here that selection of quantitative or qualitative methodology is affected by the purpose of the evaluation.<sup>62</sup> Process analysis, for example, normally involves qualitative assessments while impact analysis ideally requires quantitative measurements. Although trite, it can be observed that qualitative data is best for in-depth evaluation requirements, while quantitative data provides breadth to the issues under review.<sup>63</sup> On the one hand, quantitative methodology is more appropriate for evaluations which provide accountability (to stakeholders and funding agents), where a need to demonstrate value to external parties may rely more heavily on objective measurements. On the other hand, qualitative methodology is more appropriate to measure the value of the learning experience relying on the learner's own perceptual assessment. It follows that where educators are presented with the need to evaluate for a variety of purposes, and audiences, it will frequently become necessary to combine methodologies in the evaluation design.

Chinapah observes that the real concerns of educational evaluation can become obliterated in this debate:

"Much of the methodological debate about educational evaluation centres on the importance given to 'technical rationality' in the selection of evaluation approaches, methods and techniques . . . There has been a tendency to polarize evaluation approaches, methods and techniques into quantitative and qualitative ones. This polarization has greatly contributed to a narrow perception of the success or failure of given educational programs in that evaluation became either 'output-orientated' or 'process-orientated'."<sup>64</sup>

Most recently, educators have increasingly recognised the complementary value of qualitative methodologies in evaluation research. It is argued that this has led to a shift away from total reliance on the traditional "scientific" quantitative approach, towards an integration of both approaches as an improved means of validated findings.<sup>65</sup> For this reason, it is the writer's practice to design evaluation programmes and survey instruments, in particular, to combine both quantitative and qualitative methodology. This serves a dual research purpose: first, to cover the field of data to be collected, providing alternative means for respondents to supply that data

in the most meaningful way; and, second, to provide a means to cross-validate and to interpret data supplied. A simple example of this is where respondents have been asked to rate responses on a quantitative scale, and they are then asked to describe and comment in their own words on their reasons for these preferences. This qualitative data can be useful in explaining, amplifying and qualifying the quantitative data, and vice versa.

Ultimately, it is the task of the educational designer to devise an evaluation methodology which is appropriate to each situation, in terms of providing valid and reliable data upon which well-informed assessments can be made. For these reasons, the requirements of validity and reliability generally militate in favour of an integration of both qualitative and quantitative techniques into any evaluation strategy. In practice, however, the viability and range of available methodologies will be affected by a variety of constraints within the educational environment, some of which are unique to particular professionals, as is the case in judicial education.

#### *Constraints*

There are a number of constraints that affect the nature of the evaluation process. These constraints may be both obvious and concealed. The most obvious include limitations on resources such as cost, time and expertise. However, constraints will also frequently include qualified institutional support, goal ambiguity, and fear of results. Technical constraints such as complexity, research design problems, lack of data and limited experience and expertise also play influential roles in confining the scope of the evaluation and the methodologies employed.<sup>66</sup> These constraints require educational designers to ensure the viability of any evaluation methodology in terms of both available resources and shared objectives.

It is argued that the most significant constraint facing evaluators of judicial education is that of the doctrine of judicial independence. This constraint is also the least frequently recognised and accommodated. The existence of this doctrine tends to constrain and deter methodological rigour in the evaluation of judicial education. Put simply, it can be argued that it is improper to measure judicial performance in any way other than through the existing, formalised appellate process. Judges generally have an aversion to any scrutiny other than by the formal appeals process; observation also suggests that they are uneasy with quantitative, statistical measurements of any kind.<sup>67</sup> These constraints combine to obstruct the application of conventional educational evaluation procedures. As a consequence, judicial educators have tended to abdicate impact assessment altogether, to rely on qualitative assessments of the education process, while retaining the hope that some improvement must flow through to the level of performance.<sup>68</sup>

However, this response is inadequate and inappropriate. It is argued that rather than abdicating the endeavour, educators should be confronting the challenge of finding an appropriate means to make these measurements within these doctrinal constraints, through the development of a distinctive evaluation model appropriate for judicial education.

## Evaluation Models

It is argued that there is a marked discrepancy between theoretical models of educational evaluation and practice. This proposition is supported by numerous commentators. Brookfield, for example, observes that across the expanse of educational evaluation "The need for evaluation . . . [is] deemed important and necessary but [is] for whatever reasons, rarely implemented".<sup>69</sup>

Chinapah agrees with Brookfield, and argues that although there are increasingly concerted efforts among policy-makers, programme administrators, front-line implementors and target beneficiaries to institutionalise educational evaluation, "the support is minimal at present".<sup>70</sup> Chinapah remarks that there is a need for theorists, methodologists and empiricists to establish closer linkage between the theory and practices of educational evaluation, "there is a challenge ahead, namely that of facing the practitioners in their day-to-day 'real world' before advocating any recipe for educational evaluation".<sup>71</sup>

There are several reasons for this inconsistency. Partly, this is due to the need for trade-offs in real life between data collection research requirements and practical constraints: between quality and utility on the one hand, and cost and feasibility on the other.<sup>72</sup> Partly, however, the reality is that some forms of evaluation are undertaken without methodological rigour, and rely entirely on the subjective, unverifiable opinions of one or a very few individuals. All evaluations are inescapably subjective in part because values, preconceived notions and existing preferences cannot be screened out entirely.<sup>73</sup> In addition, evaluators are likely to err on the side of fulsome rather than understatement, to avoid offending paying clients and get further work. As Brookfield points out: "Evaluators therefore are under many pressures to suspend rigorous, clinical scrutiny when they examine program accomplishments."<sup>74</sup>

To test the appropriateness of this critique of educational practice, it is necessary to survey the range of evaluation models available in order to assess the adequacy of prevailing evaluation practice, within which context the evaluation practice of judicial education can then in turn be assessed.

### Classic Models

Commentators have classified a number of educational evaluation models.<sup>75</sup> Each model reveals features which render it more or less suitable for selection for particular purposes. The distinctive features and characteristics of each of these models should be specifically considered when planning an evaluation strategy; they exemplify Deshler's observation that:

"It is now widely recognised that appropriate selection of models to match requirements of particular situations produces the evaluation results that are most likely to be useful for specific purposes. It is no longer acceptable for practitioners or evaluators to apply to everything the one model with which they are familiar."<sup>76</sup>

There can be no single evaluation model which may be ideal for continuing judicial education; however, the appropriate model should be

selected deliberately and with informed method in any situation. It is also observed that the reality of practice is frequently at odds with the rhetoric of academe: practical difficulties such as pressures of time, money and expediency militate against conducting systematic evaluations.

### Adult Education Models

Despite the multifarious nature of educational evaluation, particular practices and conventions have emerged within the domain of adult education, which are distinctive.

Traditionally, adult education programmes have been evaluated according to criteria chosen by those funding the evaluations, rather than by those participating. An analysis of the literature of adult education has led Brookfield, for example, to endorse the proposition that if one had to assign adult educators to some school of evaluation thought, then it would have to be one identified with goal attainment.<sup>77</sup> Brookfield argues that adult educators have adopted the Tylerian school-based model of evaluation that operates on the assessment of previously specified performance behaviours as criteria and indicators, using tests, grades, measurements and judgments of achievement.<sup>78</sup> This model has been inherited from a view of education based on schooling and relies on the pedagogic methodology of teaching.<sup>79</sup> Brookfield argues that this is inappropriate, and postulates two possible alternatives:

- *Participatory Evaluation Model*—This participatory model is suited to the needs and dynamics of the adult education process, and provides an evaluative framework that is grounded in and derived from some central features of adult learning. Brookfield describes this approach as compelling and attractive.<sup>80</sup> He argues that participation should be recognised as the key to accuracy of data, prompt learning, improved communication, and increased commitment and support for programmes.<sup>81</sup> Such a model would allow adult learners to assume control of evaluation of their learning. It would be based on premises of freedom and democracy (with accountability vesting in participants), individualised and varied criteria, and subjective data collection.
- *Androgogy and Collaborative Model*—An alternative to the participatory approach to evaluation has been devised by the Nottingham Androgogy Group. This model assesses effective facilitation—as the rationale of adult education—as the extent to which it enables participants to evolve from passive recipients of transmitted knowledge to take responsibility for controlling their own learning and extending their boundaries of knowledge and experience.<sup>82</sup>

The approaches of both Brookfield and the Nottingham Androgogy Group are forceful in recasting the purpose of adult educational evaluation, and are radical in their implications. Most simply, these "radical" approaches to educational evaluation postulate that participants must prescribe their own criteria for evaluation if learning is to be meaningful. Adoption of either radical model, which refocuses criteria on the participant's notions of relevance, will require a corresponding shift of methodology from the traditional quantitative emphasis to a more qualitative approach.



While untested, these radical models appear logical and humanistic, and consistently embrace the distinctive imperative for self direction which characterises the androgical approach to continuing education. The participatory approach to educational evaluation is radical because it restores control of—and responsibility for—the formalised learning process on to the adult learner. It is also compelling because it realigns the rationale of educational evaluation from a judgmental metaphor, concerned predominantly with providing external justification, to a private self-critique where the participants become the actors rather than the targets in taking responsibility for any learning outcome. The approach transfers responsibility for assessing learning to the shoulders of participants, who should also determine or at least participate in discerning and subscribing to defining the objectives and standards, selecting indicators and exercising judgments about quality and value. For this transfer of responsibility to be effective, however, it is essential that any such self-critique is facilitated by experts to ensure that it is a rigorous process embodying meaningful criteria.

It is concluded that these radical models are better aligned with the attainment of the ultimate objective of continuing education, that is, the qualitative learning of professionals. Despite being denuded of the classic, scientific respectability of the quantitative approach, it is appropriate to develop an evaluation process for judicial education which embodies these participatory qualitative elements when evaluating for the purpose of the learner. As other purposes, however, usually coexist to provide accountability to external stakeholders, an extended quantitative methodology is also usually required in the evaluation process. Accordingly, an indorsement of this radical learner-based approach to evaluation alone is not sufficient.

### The Practice of Evaluating Judicial Education

It has already been argued that the reality of educational evaluation practice is often far from the theoretical. It is now argued that this discrepancy is also a characteristic of practice in judicial education. Hudzik, for example, observes from an analysis of judicial education that “the usual level of evaluation systematically undertaken is minimal”.<sup>83</sup> This assessment is validated from a study of prevailing practice which also postulates that doctrinal constraints relating to the non-measurement of judicial performance frequently combine with the other general constraints to render the evaluation of judicial education to be inadequate, lacking in methodological rigour, inappropriate and of limited utility.

#### *The United States “Empirical” Participant Reaction Model*

In the arena of continuing judicial education there is effectively one prevailing approach to evaluation, which has been described by Hudzik as the “empirical” model. This model of practice consistently relies on participant reaction combined with some notional cost/efficiency assessment. No attempt is made to measure either learning or impact on performance.

In the decade between 1981 and 1991, Hudzik undertook and documented the only research available on the practices of evaluation

undertaken by continuing judicial educators throughout the United States. He found that all respondents measured the reactions of participants in education programmes. Some educators asked participants whether they thought various presentations added to their knowledge, and others asked “what is the likelihood of you implementing changes in your court as a result of this programme?” A few used a technique of action planning, which requested participants to describe any programme-suggested actions that they planned on return to court, with follow-up on how they had done. However, he found that none took any steps to measure the results of those programmes or, more directly, learning or job behaviour change.<sup>84</sup> His conclusions were disparaging: “In sum, the usual level of evaluation undertaken systematically by the various organizations is minimal.”<sup>85</sup>

In other work, he observed that most courts do not usually even *attempt* to measure learning or knowledge gain, to determine if participants change their behaviour on the job, and if the changes improve their performance.<sup>86</sup> He also observed:

“[It is] to no great surprise [that] the vast majority of both state and national respondents report either never or only occasionally employing the other evaluation methods (ie impact on participants and impact on courts). We suspect that respondents who indicate that they occasionally use these methods are, for the most part, collecting anecdotal feedback data as opposed to rigorous and systematic evaluative data.”<sup>87</sup>

The nature and characteristics of this empiric model, as documented by Hudzik, accord with the writer’s observations of evaluation practice in the United States, and in some measure, also in Australia.

#### *The British Process-based Approach*

British practice appears at first glance to adopt a similar reaction-based approach to the United States. However on closer appraisal, it is observed that the principal educative agency, the Judicial Studies Board, adopts a markedly more restricted and less formal approach to the evaluation process. The Board reports that “the efficacy of judicial studies cannot be measured directly”,<sup>88</sup> relies primarily on monitoring the process of education through participant feedback, and offers no apology for avoiding any assessment of outcome whatsoever:

“The Board cannot take hold of a judge and make him better. It would be unrealistic and impertinent to try. This is a task for the judge himself, with such help as we can give . . . the quality of the hearing depends in the last resort upon the personal qualities and attainments of the judge himself. But we believe that judicial training can help in this respect as well, elusive and delicate as the problem undoubtedly is.”<sup>89</sup>

The British position may seem at first to be bizarre: formally abdicating any suggestion of impact evaluation. Clearly, the Board places considerable importance on avoiding any appearance of assessing judicial competence; indeed, it goes to considerable lengths to avoid any appearance of assessing judicial performance. This position becomes easier to understand within the doctrinal context of judicial independence, and distils the fundamental

importance of avoiding any confusion between assessing education and assessing the results of education through performance within the judicial domain, however academic this distinction may appear in any practical context. In effect, the Board prefers to avoid assessment, rather than risk infringing independence.

The Board claims to meet the requirements of evaluation—measuring the value of its educational endeavour—through indirect and inferential means:

“Putting the task of the Board at its lowest, and stating the equation crudely in terms of money alone, the Board would not have to show more than the prevention of a modest handful of mistakes amongst the many thousands of hearings in the Crown Court every year, to demonstrate that the Board has more than paid for itself.”<sup>90</sup>

It is concluded that the issue of educational evaluation is evidently at least as problematic for the British judiciary as it is for their American colleagues. Notwithstanding the circumlocution of the Board’s formulation, it does not offer any measure of prevention, however circumstantial; nor is any evidence offered to demonstrate that such prevention has occurred.

#### *The New South Wales’ Participant Reaction/Client Appraisal Approach*

The approach of the Judicial Commission of New South Wales is an adaptation of the classic Tylerian and Kirkpatrick models. All educational programme activities are subject to formalised needs assessment and evaluation processes. Needs are identified using compound methodologies including interviews (both within and beyond the judiciary), observation, surveys and analysis of courts’ management data. Educational objectives are then defined to meet these needs. All programmes are subjected to both formative and summative assessment focusing primarily on participant reaction, together with appraisal by instructors, education staff and education committee. Participant reactions are scored on a numerical index which creates relativities on an aggregated scale of participant satisfaction. Scoring participant reaction provides a consistent—albeit non-objective—measure of satisfaction rather than learning. Because it is recognised that satisfaction need not correlate with effective learning, this measure is frequently overridden by the expert appraisal of the education committee in final assessment for purposes of future planning.

The evaluation effort does not at the present time extend to make any direct assessment of the impact of education on judicial performance.<sup>91</sup> Efforts are, however, made to triangulate indications of positive impact through regular consulting with professional and client representatives in a continual process of assessing educational needs and endeavours as they pertain to judicial performance and the justice service.

#### *Michigan’s Action-Planning Approach*

An alternate, more elaborate approach to assessing the effects of judicial education was developed by the Michigan Judicial Institute specifically to improve the measurement of learning and behavioural change. The approach creates an index of intended behavioural change to measure actual change, using a self-reporting procedure.

The model operates at three levels by measuring reactions (immediately and three months later), learning (by examination before and three months later) and behaviour change (participants state behaviour change goals immediately on completion which, three months later, they are asked to reconcile with actual behaviour).<sup>92</sup>

Proponents of this approach claim that it provides useful additional data on the effectiveness of programme activities. They also claim that it is a compatible technique which retains focus on the learning experience and provides on-going reinforcement.<sup>93</sup>

#### *The JERITT Model of Intended Behavioural Change<sup>94</sup>*

Hudzik argues that the prevailing practice of evaluating judicial education outlined above is inadequate. A number of specific difficulties and constraints impede the application of conventional evaluation methodology to judicial education. These relate to fundamental philosophical barriers pertaining to the doctrine of judicial independence, resource constraints and the practical difficulties of defining goals in meaningful, concrete terms and selecting performance indicators on which there is a consensus.

“The ultimate value of training in criminal justice is measurable only against a very elusive and not concretely defined set of objectives—‘securing law and order’ [sic], ‘providing justice that is swift and sure’, and ‘ensuring that justice is indeed just’.”<sup>95</sup>

To cure these deficiencies, Hudzik postulates the JERITT model which is designed to take account of the peculiar difficulties and constraints which beset the judicial evaluator in practice. This model combines approaches to obtain as much objective data as possible within the constraints that operate within the judicial context. He argues that of the two broad approaches to evaluation—programme monitoring and programme-training—the latter is better suited to substantive evaluation of judicial education. The JERITT model modifies the Kirkpatrick approach for the judicial setting into three tiers of assessment: *participant feedback*, *impact on participants* and *impact on courts*.<sup>96</sup> While evaluation ought to concern itself ultimately with the question of outcome and impact—and the assessment of the impact of educational programmes on overall court performance is “laudable”—Hudzik argues that impact evaluation can be very difficult and very expensive:

“[It is] wholly unrealistic to expect that judicial education organizations will be able to conduct such evaluation systematically and directly . . . about the closest most judicial educators will come to measuring impact on court performance is through self assessment of intended and actual job behaviour change, which can be supplemented with anecdotal opinion data.”<sup>97</sup>

Additionally, Hudzik argues that it is difficult to establish consensus about the specific goals of judicial education, and it is almost impossible to measure behavioural achievement of those goals:

“About the only way to do a results evaluation is to obtain the opinions of a powerful person or group of powerful people who have management or control responsibility for the organization.”<sup>98</sup>

Because of the technical difficulties associated with objectively and systematically evaluating learning, behavioural change and results, Hudzik proposes “surrogate” measures for these which, although weakening validity, do give some means for measuring programme-related effects. He argues that at a minimum all programmes should be assessed for *effort*, *efficiency* and *intentions* to change job behaviour. The approach provides a means of measuring intentions to change job behaviours:

“About the closest most, if not all, judicial educators will come to measuring impact on court performance is through some of the information provided by participants in a questionnaire [(self reports of individual behavioural changes (if any)) . . . [which] could be supplemented with anecdotal opinion data solicited from key judicial system personnel as to whether they *think* [sic] court-system performance has improved, and whether they *believe* [sic] the educational program had anything to do with the improvement.”<sup>99</sup>

It has been previously argued that the ultimate purpose of evaluation for accountability is to assess and demonstrate the value of judicial education in terms of its effectiveness in enhancing performance; it is concluded from this review that the actual practice of evaluation in judicial education discloses marked discrepancies with theoretical models through the universal reliance on subjective, usually formative, participant-reaction technique, and an almost total abdication of any meaningful attempt at assessing outcomes in terms of changes in judicial performance (with the exceptions of the JERITT model of intentions, and the Judicial Commission’s integration of client appraisal). As such, it is concluded that prevailing practice is deficient in concentrating on participant-reaction process evaluation at the expense of providing direct assessment of outcomes for the purposes of external accountability.<sup>100</sup>

## Conclusion

### *Critique of Practice*

In this article, it has been argued that the existing practice of evaluation is inadequate, inappropriate and has limited utility. These discrepancies are largely caused by the practical difficulties of selecting assessment criteria and data collection, shortages of resources, measurement impediments and a lack of methodological rigour. Most significantly, however, the evaluation process fails to accommodate the distinctive requirements of the doctrine of judicial independence and a variety of related features of the judicial environment within which this education is provided. Consequently, it is argued that it is necessary to develop a distinctive evaluation model which can provide an appropriate assessment of the impact of judicial education.

Any critique of practice must acknowledge that the JERITT model is unique in recognising the practical constraints which limit the application of theoretical models to continuing judicial education and is, consequently, the most appropriate and useful model available. The JERITT model recognises, for example, that external assessment is anathema to the judiciary and is, for fundamental doctrinal reasons, inappropriate in judicial

education. Similarly, it recognises the unsuitability of otherwise conventional evaluation mechanisms such as pre/post testing and the use of control groups. The importance of these features had not previously been recognised. In addition, it incorporates a participatory methodology which relies on self-reporting by participants and is compatible with the principles of adult education and professional development, and it restores responsibility and control of the measurement of the learning process to the participants, thereby encouraging self-direction and providing reinforcement and a means of action planning. On this basis, it is a realistic and progressive approach to evaluating judicial education.

The model does, however, have certain drawbacks:

- The model does not offer any means of objectively assessing impact and, to this extent, is less than ideal. The discussion advocates the theoretical importance of impact assessment—in this instance, judges’ learning and its outcomes—but does not formulate any means of doing so in view of the practical difficulties. Instead, it adopts a subjective mechanism of measuring intentions which hinges on participant perception—on this occasion exchanging perception of benefit for perception of application.
- The model does not set target standards for evaluation: it constructs an index of intention, but does not provide any criteria with which to monitor that index or to rate the nature or quality of the intentions being measured.
- Validation of assessment procedure is difficult. The assessment procedure selected in the model reverts from visible external behaviour or performance to internal intention. This renders any rigorous validation of assessment impossible externally, and thereby leaves the process dependent on participants’ judgment rather than any professional or community appraisal. Ultimately, it may be naive to expect such a procedure to sustain credibility in the face of consistent scepticism.
- The model relies on the mechanism of self-reporting which, in practice, is vulnerable to the continuing commitment of participants to remain involved and motivated for protracted periods following the programme.
- It demands significantly greater resources to service the monitoring procedures of any assessment.

On balance, the JERITT model incorporates important new features which warrant trial and evaluation in turn. However, it is concluded that the model does not go far enough in providing meaningful assessments of outcome for the stakeholders who ultimately fund the programme—be they within the community, the justice system itself, or the executive branch of government.

The need to demonstrate value, and thereby to provide accountability, is an imperative of evaluation.<sup>101</sup> Ultimately, the education programme which can demonstrate value survives; the more discernible that value, the less intervention is likely. Thus objective outcomes from the education process are the most persuasive in providing accountability. To this extent, the JERITT model remains incomplete.

In conclusion, this study supports Hudzik in his assessment that current practices in evaluating judicial education are minimal. Indeed, these practices are inadequate and inappropriate. This study recognizes the need to develop a range of evaluation models to accommodate the various types of evaluation which may be required, within a framework which incorporates the principles of adult education and professional development, generally, and the organisational and situational constraints of the judiciary, specifically. Additionally, this study concludes that the JERITT approach is the most responsive and sophisticated attempt yet to develop an evaluation model for continuing judicial education. However, this model remains incomplete to the extent that it requires the addition of stronger mechanisms to visibly demonstrate impact and value to external funding bodies.

The challenge which remains for judicial educators is, therefore, to develop a distinctive evaluation model which can provide both value to judicial learners and accountability to external stakeholders without infringing judicial independence.

#### *A Judicial Systemic Performance Model Postulated*

There is a need to develop a consummate, distinctive model of evaluation for judicial education which can overcome the deficiencies of existing practice. This model should assess both process and outcome in order to meet the two respective purposes of evaluation, that is, for the learner and for those other parties with a stake in the process.

Any proposed model must overcome the difficulty of evaluating judicial education, being the problem of identifying appropriate criteria and indicators to measure judicial performance in a meaningful way without infringing judicial independence. The doctrinal significance of judicial independence has a fundamental impact on the selection and modification of any model of education evaluation for judges. The constraints which this doctrine imposes should not be under-estimated, as is evidenced by the lengths to which the British approach circumlocutes the problem. However, it is argued that the British approach of abrogating any assessment of results is equally problematic for educators concerned with ensuring educational effectiveness.

It has already been argued that judicial evaluation should ideally measure results. Specific difficulties, however, militate against measuring results in terms of judicial performance directly. Several models attempt to overcome these difficulties by measuring results through utilisation of surrogate criteria or indicators of performance.<sup>102</sup> These indicators reflect the intention to improve performance, and the clients' perceptions of improved performance. Although inferential, these indicators are capable of contributing some potentially useful measurements for purposes of triangulation. These measurements are however limited in their validity and utility.

It is now postulated that a distinctive evaluation model for judicial education should integrate methodologies which can assess appropriate criteria of impact. Such criteria should be objectively recognisable and accordingly quantitative (since qualitative indicators are difficult to

measure precisely or with any replicability); they should measure outcome in terms of judicial service; they should be recognisable and credible both to the judiciary and its stakeholders; and critically, they should operate within the constraints of the doctrine of judicial independence by measuring systemic rather than individual performance.

Criteria conforming with these characteristics is already available within the judicial management and administration system. This criteria includes, among other possible indicators, the following:

- *trial disposal rates* and through-put times (which measure the passage of time for particular proceedings and, thereby, the efficiency of judicial service);
- *appeal rates* and disposal outcomes (which measure an incident of judicial competence and error at the level of the trial bench); and
- *complaints rates* (which measure an incident of perceived satisfaction by clients).

Using these criteria, it is possible to construct a valid and meaningful evaluation mechanism to measure the value of judicial education in terms of its costs and its benefits. This mechanism would integrate objective, quantifiable measurements of outcome in terms of enhanced judicial performance (for purposes of accountability), with the prevailing practice of qualitatively assessing the education process (for purposes of learning).

It is the goal of evaluation to demonstrate value through identifying causal relationships between educational intervention and enhanced performance outcomes. In practice, however, it is often difficult, if not impossible, to eliminate extraneous influences on these relationships; since outcomes may have a number of causes. Even so, it is argued that this does not impair their validity as evaluative indicators. A lack of exclusive causality does not vitiate causality, although it may qualify it. Thus it is argued that evaluation need only demonstrate that judicial education has contributed to desired performance outcomes, in order to provide the accountability required.

#### *Summing-up*

It is essential to assess the value of judicial education. Evaluation measures the quality of the learning process for the individual judge. Evaluation also provides the means to demonstrate the worth of the educational endeavour for the judiciary as a profession. The existing practice of evaluating judicial education, however, is generally inadequate and inappropriate. There remains a need for educators to formalise techniques of evaluating the impact of judicial education, both for the purpose of promoting effective learning and as a means of demonstrating the development of competence flowing from the education endeavour. To rectify these deficiencies, this study postulates a *Judicial Systemic Performance* model. This model provides educators with the means to assess and demonstrate the attainment of professional competence without infringing judicial independence.

It is still too early to assess the value of continuing judicial education in Australia. If evidence from overseas can serve as any guide, the very

existence of a formalised process of continuing education provides a means of accountability to the community and raises the standing of the judiciary.

The challenge of evaluation remains to ensure that continuing education does promote learning which contributes to enhancing the quality of justice.

### Endnotes

<sup>1</sup> Knox, "Assessing the Impact of Continuing Education" *New Directions for Continuing Education (No 7)* (Jossey-Bass, San Francisco, 1979), p 117.

<sup>2</sup> Chinapah and Miron, "Evaluating Educational Programs and Projects" *Socio Economic Studies No 15* (UNESCO, Belgium, 1990), p 25.

<sup>3</sup> Brookfield, *Understanding and Facilitating Adult Learning* (Jossey-Bass, San Francisco, 1986), p 261.

<sup>4</sup> *The New Collins Dictionary* (Collins, London, 1986), p 340.

<sup>5</sup> Worthen and Sanders, *Educational Evaluation: Theory and Practice* (Wadsworth, Belmont, Calif, 1973), pp 20-23.

<sup>6</sup> Pearson, "Evaluation" in *Adult Education Perspectives for Judicial Education* (Judicial Education Adult Education Project (JEAEP), University of Georgia, Georgia, 1992), pp 8.1-8.13 and 8.7.

<sup>7</sup> House, *New Directions in Educational Evaluation* (Falmer Press, London, 1986), p 87.

<sup>8</sup> Cook and Reichardt, *Qualitative and Quantitative Methods in Evaluation Research* (Sage, Beverly Hills, Calif, 1979), p 27.

<sup>9</sup> Chinapah and Miron, op cit n 2, p 25.

<sup>10</sup> Grotelueschen, "Program Evaluation", in Knox (ed), *Developing, Administering and Evaluating Adult Education* (Jossey-Bass, San Francisco, 1989), pp 75-123. The author reviews the diverging approaches of Tyler (1950), Rivlin (1971), Scriven (1972), and Eisner (1976), respectively, p 76.

<sup>11</sup> Gardner, "Five Evaluation Frameworks: Implications for Decision Making in Higher Education" (1977) 48 *Journal of Higher Education*, 571.

<sup>12</sup> Scriven, "Curriculum Evaluation—Definitions and Boundaries" in Tawney (ed), *Curriculum Evaluation Today: Trends and Implications* (Macmillan, London, 1991), 40. See also Scriven, *Evaluation Thesaurus* (Sage, Newbury Park, Calif, 1991).

<sup>13</sup> Worthen and Sanders, op cit n 5, p 19.

<sup>14</sup> *Ibid*, pp 20 and 24-25.

<sup>15</sup> Brookfield, op cit n 3, p 262.

<sup>16</sup> *Ibid*, p 264.

<sup>17</sup> Worthen and Sanders, op cit n 5, p 25.

<sup>18</sup> Grotelueschen, op cit n 10, p 77.

<sup>19</sup> See, eg, Steele, *Cost Benefit Analysis and the Adult Educator: A Literature Review* (ERIC Clearinghouse on Adult Education, Syracuse, NY, 1971) cited in Brookfield, op cit n 3, p 264.

<sup>20</sup> House, op cit n 7, p 7.

<sup>21</sup> Houle, *Continuing Learning in the Professions* (Jossey-Bass, San Francisco, 1980), p 237.

<sup>22</sup> This proposition is advanced by a number of commentators; see, eg, Chinapah and Miron, op cit n 2, pp 26-27; see also Grotelueschen, op cit n 10, pp 78-79.

<sup>23</sup> House, op cit n 7, p 6.

<sup>24</sup> See, Abramson, Tittle and Cohen (eds), *Handbook of Vocational Education Evaluation* (Sage, Beverly Hills, 1979), pp 573-574.

<sup>25</sup> Anderson and Ball, *The Profession and Practice of Program Evaluation* (Jossey-Bass, San Francisco, 1978).

<sup>26</sup> Grotelueschen, op cit n 10, p 79.

<sup>27</sup> *Ibid*, p 88.

<sup>28</sup> Houle, op cit n 21, p 237.

<sup>29</sup> Hudzik and Wakeley, "Evaluating Court Training Programs" (1981) 64(8) *Judicature* 369. Evaluation can be utilised to improve or adapt an on-going programme, or it can be used for accountability, certification or selection. The former is often referred to as a *formative* function, and the latter as the *summative* function. Examples of summative accountability

categories include the cost effectiveness of a programme, refers to its efficiency, or more specifically, its beneficial impact in relation to its costs; Grotelueschen, op cit n 10, pp 79-80. See also Chinapah and Miron, op cit n 2, pp 26-27.

<sup>30</sup> Hudzik, *Judicial Education Needs Assessment and Program Evaluation* (Judicial Education Reference, Information and Technical Transfer Project (JERITT), 1991), p 44. A number of commentators argue this is the ultimate value goal of evaluation. However, difficulties exist in establishing causal linkage, data access/collection, and need for pre/post measurement. This is also known as the "Kirkpatrick" approach; see Kirkpatrick (ed), *Evaluating Training Programs* (American Society of Training and Development, Madison, Wis, 1975).

<sup>31</sup> Knox, op cit n 1, p 2.

<sup>32</sup> Cervero, *Effective Continuing Education for Professionals* (Jossey-Bass, San Francisco, 1988), p 144.

<sup>33</sup> Chinapah and Miron, op cit n 2, discuss the importance and difficulties of impact assessment (p 29) and its attainment through the prior definition of tangible educational objectives, p 33.

<sup>34</sup> Green and Walsh, "Impact Evaluation in Continuing Medical Education—The Missing Link" in Knox, op cit n 1, pp 81-87, 82.

<sup>35</sup> *Ibid*, p 86.

<sup>36</sup> Knox op cit n 1, p 118.

<sup>37</sup> *Ibid*, p 6.

<sup>38</sup> Chinapah and Miron, op cit n 2, p 29. Knox reports on growing evidence that continuing education can have an impact on practice. He cites Guire and others (1964) who monitored a two day workshop on diagnosing heart complaints for practising physicians. Using data on gain and retention with pre/post and follow-up testing and control groups, this study found that learning gain and retention depended on opportunity to practice the studied procedure, and established the importance of periodic instruction, assessment and reinforcement to produce a lasting change in performance. He also cites Parnes (1976) who demonstrated abundant evidence that adults who participated in vocational continuing education experienced greater career success than adults with similar levels of education and age who did not; Knox, *ibid*, pp 13-16.

<sup>39</sup> Knox, *ibid*, p 4.

<sup>40</sup> Hudzik, op cit n 30, p 37.

<sup>41</sup> Pearson, *Adult Education Perspectives for Judicial Educators* (JEAEP, State Justice Institute, 1992), p 8.8. Evaluation can measure cumulative effect, and claim a contributory role. This is described as a "good faith" effort to improve the judicial education programme.

<sup>42</sup> Hudzik, op cit n 30, p 38.

<sup>43</sup> For empirical validation of this observation, see Catlin, "An Empiric Study of Judges' Reasons for Participation in Continuing Professional Education" (1982) 7 *The Justice System Journal* 236.

<sup>44</sup> Cervero, op cit n 33, p 145-6.

<sup>45</sup> Knox, op cit n 1, p 119. Continuing education programmes typically have multiple benefits, several of which can usually be assessed in a specific impact study: included are a personal sense of understanding or mastery, adoption of changed practices, economic benefits such as increased productivity, organisational survival, and various benefits to the larger society such as creativity and support for sound policies. The assessment of such changes and benefits can be based on self-reports, records, ratings by peers and supervisors. Evidence of impact can be collected at one particular time, but tends to be most convincing through longitudinal or series of studies.

<sup>46</sup> Tyler, *Basic Principles of Curriculum and Instruction* (University of Chicago Press, Chicago, 1949), p 106. The predetermined objectives approach was conceived by Tyler in 1932, and has been successively re-indorsed in the literature. The Tylerian approach sees evaluation as "the process of determining to what extent the educational objectives are actually realised . . . since educational objectives are essentially changes in human beings, . . . then evaluation is the process for determining the degree to which these changes in behaviour are actually taking place" (p 106). Brookfield, however, argues that this view of evaluation is "flawed" in its application to adult learning; Brookfield, op cit n 3, pp 267-268.

<sup>47</sup> Chinapah and Miron, op cit n 2, p 33.

<sup>48</sup> Kirkpatrick, *Evaluating Training Programs*, (American Society of Training and Development, Madison, Wis 1975). *Kirkpatrick's hierarchy*, which is widely adopted as a balanced and practical approach—emphasises four levels of evaluation: reaction, learning,

performance (transference of behaviour), and organisational/community impact (results). This model is useful in redirecting attention from participant and organisational perceptions to focus on outcome and results.

<sup>49</sup> See Houle, op cit n 21, pp 237-265, for a description of these criteria: extent of participation, extent of learner satisfaction, accomplishment of a learning plan and measurement of performance. Recent studies disclaim the value of satisfaction measures, identifying a negative correlation between satisfaction and cognitive gain: experiments describing the "Dr Fox Effect" found that students positively correlated the cognitive scores with the level of content presented, also found that high-seduction presentations produced higher scores than low-seduction; and the cognitive scores were higher for students watching the high seduction low content film than for the low seduction high content film; Williams and Ware, "Validity of Student Ratings Under Different Incentive Conditions: A Further Study of the Dr Fox Effect" (1976) 14 *Journal of Educational Psychology* 449, indorsed by Houle, *ibid*, pp 245-246.

<sup>50</sup> Cervero, op cit n 33, pp 131-146.

<sup>51</sup> *Ibid*, p 143.

<sup>52</sup> Hudzik, op cit n 30, p 49.

<sup>53</sup> Knox, op cit n 1, p 121. Note particularly findings of the "Dr Fox Effect" experiments, n 51, above.

<sup>54</sup> *Ibid*, p 12.

<sup>55</sup> *Ibid*, p 12.

<sup>56</sup> Brookfield, op cit n 3, p 275.

<sup>57</sup> Triangulation involves the use of multiple methods of data collection in the same aspect of human behaviour, frequently including the integration of qualitative and quantitative techniques.

<sup>58</sup> Chinapah and Miron, op cit n 2, p 41.

<sup>59</sup> See eg. Anderson and Ball, op cit n 26.

<sup>60</sup> Houle, op cit n 21, pp 252-265.

<sup>61</sup> Grotelueschen, op cit n 10, p 122. Grotelueschen explains the difference in indicators on the basis that qualitative indicators emphasise holistic and tacit understanding demonstrated by illustrations and case-studies; while quantitative indicators emphasise the rigorous analysis of key indicators of worth.

<sup>62</sup> The quantitative approach to educational evaluation is generally still seen as the dominant paradigm in the theory of educational evaluation. It is concerned with reliability, objectivity of data, focused on impact and outcomes, operates on large samples and in generalisations. It is the classic "scientific" methodology. Quantitative methods can include random experiments and probability, and the instruments used increasingly include surveys with closed questionnaires, and cognitive and affective tests. Qualitative evaluation is regarded as an alternative paradigm, and in practice is frequently relied upon considerably more than may be formally acknowledged. It is concerned with validity, utilises subjective data, focuses on process, is interested in uniqueness, tends toward the anecdotal and operates on case studies. It is "practitioner-friendly". Qualitative methods usually rely on detailed descriptions of people, events, and situations, direct quotations from people on their experiences, attitudes, beliefs, thoughts and needs, and observable behaviours using a range of interview and observation techniques.

<sup>63</sup> The quantitative paradigm is characterised in the literature as having "a positivistic, hypothetico-deductive, particularistic, objective, outcome-orientated, and natural science world view"; and the qualitative paradigm as subscribing to a "phenomenological, inductive, holistic, subjective, process-orientated and social anthropological world view". See Cook and Reichardt, op cit n 8, pp 10-11.

<sup>64</sup> Chinapah and Miron, op cit n 2, p 23.

<sup>65</sup> The triangulation effect on data also reduces bias.

<sup>66</sup> Hudzik, op cit n 30, pp 44-47.

<sup>67</sup> The writer relies upon numerous informal discussions with senior members of the Australian judiciary to support this proposition. It is observed that these views are both strongly and broadly held within the judiciary. The lack of literary authority on this point highlights an existing paucity of expressed views or researched scholarship in judicial education.

<sup>68</sup> See Hudzik's analysis of the practice of evaluation in judicial education, above; and Hudzik, *Issues and Trends in Judicial Education* (JERITT, Michigan State University, 1991), p 135.

<sup>69</sup> Brookfield, op cit n 3.

<sup>70</sup> Chinapah and Miron, op cit n 2, p 17; and Brookfield, *ibid*, p 261.

<sup>71</sup> Chinapah and Miron, *ibid*, p 18.

<sup>72</sup> Hudzik, op cit n 30, p 39.

<sup>73</sup> *Ibid*, p 37.

<sup>74</sup> Brookfield, op cit n 3, p 265.

<sup>75</sup> Ulschak identifies a number of evaluation methods and techniques which include nominal group, delphi, critical incident, competency models and exit interviews among others; Ulschak tabulates the distinctive features of a range of methodologies: Ulschak, *Human Resource Development: The Theory and Practice of Need Assessment* (Virginia, Reston, 1983), pp 96-98.

House classifies eight models by reference to their major audiences, assumptions, methodologies and outcomes. These are classified as the systems analysis approach, behavioural objectivity, decision-making, goal-free, art criticism, professional review and legal, and the case-study approaches. House, *Evaluating with Validity* (Sage, Beverly Hills, Calif, 1980) and House, "Assumptions Underlying Evaluation Models" (1978) 7 *Educational Researcher* 4.

<sup>76</sup> Deshler, "Evaluation in Program Development", *New Directions in Continuing Education* (Jossey-Bass, San Francisco, 1984), p 12.

<sup>77</sup> Brookfield, op cit n 3, p 262.

<sup>78</sup> *Ibid*.

<sup>79</sup> *Ibid*, p 267.

<sup>80</sup> *Ibid*, p 277.

<sup>81</sup> *Ibid*, pp 276-277. He also alludes to an alternative model: Perspective Discrepancy Assessment (Mezirow 1978: "the education process can be best understood by examining how those involved perceive and understand the process and themselves in relation to it" (p 52)). It identifies discrepancies in the perceptions of key personnel regarding current and future practice within the programme. For a critique of evaluation typologies, see Brookfield's discussion, pp 266-282.

<sup>82</sup> Nottingham Androgogy Group, *Towards a Development Theory of Androgogy*, (University of Nottingham, Department of Adult Education, Nottingham, 1983); indorsed in Brookfield, op cit n 3, p 280.

<sup>83</sup> Hudzik, op cit n 67, pp 132 and 135.

<sup>84</sup> Hudzik, *The Continuing Education of Judges and Court Personnel* (Judicial Education Network, Lansing, Michigan, 1989), pp 12-13.

<sup>85</sup> *Ibid*, p 13.

<sup>86</sup> Hudzik and Wakeley, op cit n 29, at 371.

<sup>87</sup> Hudzik, op cit n 67, p 135.

<sup>88</sup> *Report of the Judicial Studies Board 1983-87* (HMSO, London, 1988), p 4.

<sup>89</sup> *Ibid*, p 21.

<sup>90</sup> *Ibid*, pp 20-24. This inferential method of cost/benefit analysis to justify judicial education is also adopted by Riches, "Judicial Education—A Look at the Overseas Experience" (1990) 64 ALJ 189 at 192; and Wood, "The Prospects For a National Judicial Orientation Programme In Australia" (1993) 3 JJA 75.

<sup>91</sup> Impact evaluation is not presently undertaken for a number of reasons which are doctrinal, educational and practical. The doctrinal reasons relate to constraints imposed by the doctrine of judicial independence, and the considerable sensitivities relating to any measurement of judicial performance other than through the formal appeal process: see related discussion of "constraints", above. Educationally, evaluation can arguably be limited to assessing the efficiency of training and not its effectiveness. In practice, impact assessments are difficult to make and consume substantial resources: shortages of resources available tend to focus educational resources "at the sharp end" in delivery.

<sup>92</sup> Hudzik and Wakeley, op cit n 29 at 374.

<sup>93</sup> Hudzik, op cit n 30, pp 56-57.

<sup>94</sup> JERITT is the acronym for "Judicial Education Research and Information Technology Tranfer" project which sponsored Hudzik's research, and acts as a publishing network within the judicial education community in the United States.

<sup>95</sup> Hudzik and Wakeley, op cit n 29 at 371.

<sup>96</sup> Hudzik, op cit n 30, p 48; see also Hudzik, op cit n 67, p 133. The JERITT approach encompasses assessment of reactions, learning, behaviour change and results. Reactions are essentially subjective and comparatively easily measured. Learning and behaviour change are more difficult and costly to measure. Results are very difficult to measure because assessment

must be undertaken not just within the participant's work environment but also of the court as an organisation, whether training is responsible for any changes.

<sup>97</sup> Hudzik, op cit n 30, p 59.

<sup>98</sup> Hudzik and Wakeley, op cit n 29 at 373.

<sup>99</sup> Hudzik, op cit n 30, p 59.

<sup>100</sup> A detailed discussion of the doctrine of judicial independence has been provided earlier. The "Kilmuir Rules" have operated in Britain since 1955 with the effect that judges are constrained by convention from public comment. Notwithstanding, the judiciary is keenly sensitive to criticisms of failure to be accountable, as remedial steps following extensive public criticism of a South Australian judge, Justice Bollen, in 1992 illustrates. These steps include publicising the participation of a number of judges from around Australia in a special "train-the-trainer" workshop of gender awareness in Canada.

See also Phillips J, unpublished address, Monash Law School Foundation, 2 December 1993 reported in "Judges, Too, Have The Right To A Hearing", *The Australian*, 3 December 1993. In this edited address, the Chief Justice of Victoria, Sir John Phillips, is reported as indicating that he will depart from the Kilmuir Rules to publicly defend the judiciary against unfair public criticism where necessary.

<sup>101</sup> See, eg, Jones (ed) *Evaluating Training* (Lakewood Publications, Minneapolis, MN, 1989). This monograph extracts selected articles from *Training* magazine, many of which emphasise the need to demonstrate worth for survival, and outline a variety of management and accounting techniques for doing so. This approach to the training function is typical of the private sector as distinct from government or academe.

<sup>102</sup> Notably, the JERITT and the Judicial Commission models referred to in preceding discussion.