RCSI UNIVERSITY OF MEDICINE AND HEALTH SCIENCES

Royal College of Surgeons in Ireland

repository@rcsi.com

Evaluation Research in Education

AUTHOR(S)

Pauline Joyce

CITATION

Joyce, Pauline (2010): Evaluation Research in Education. Royal College of Surgeons in Ireland. Preprint. https://doi.org/10.25419/rcsi.10780373.v1

DOI

10.25419/rcsi.10780373.v1

LICENCE

CC BY-NC-SA 4.0

This work is made available under the above open licence by RCSI and has been printed from https://repository.rcsi.com. For more information please contact repository@rcsi.com

URL

https://repository.rcsi.com/articles/preprint/Evaluation_Research_in_Education/10780373/1

Evaluation Research in Education

Dr. Pauline Joyce EdD
Institute of Leadership
RCSI
pjoyce@rcsi.ie

Introduction

This paper gives an overview of evaluation and evaluation research, particularly how it fits with education. Reference to some evaluation debates over the years is presented as well as some seminal works in the topic area. A brief synopsis of evaluation in the education setting is then presented before outlining various approaches to evaluation. Fourth generation evaluation is presented as a suitable approach to consider and one, which has been used by the writer for her doctoral studies in learning approaches of postgraduate healthcare professionals (Joyce, 2010).

Evaluation Debates

Evaluation, as a form of systematic inquiry, occupies an increasingly major place in making decisions about public policies (Virtanen and Uusikylä, 2004). Calls for accountability through evaluation research, particularly in the USA have increased the demand for measurements of performance (Cousins and Aubry, 2006). The demand for an appropriately skilled workforce in an evolving global economy makes evaluation of higher education a high priority. Effective evaluation can be a significant contributor to quality but does not necessarily guarantee that those in authority will heed the outcomes of evaluation and take needed corrective action. The term quality assessment has been used synonymously with evaluation in the context of regulating higher education. According to Kells (1992) institutions and programmes can be strengthened substantially through effective evaluation.

Some writers place evaluation as a distinct research school with its own identity (House 1993; Scriven 2005) while others consider it a specialism within social science, placing emphasis on meeting information needs of decision makers and policy makers (Patton, 1997, 2002). Many authors have highlighted debates about various approaches within evaluation (Shadish *et al* 1991; Chen 1996; Ong 1996;

Pawson and Tilly 1997; Shaw 1999; Tones and Tilford 2001; Robson 2002) while others have focused on analysing the contributions of influential evaluation theorists and the congruence of their theoretical positions (Shadish *et al* 1991; Clarke and Dawson 1999; Shaw 1999). The developments which have taken place in evaluation over the last 40 years or so reflect broader movements which have taken place around research paradigms and methods in the social sciences. There has been a tendency to break away from the classical, objectivist, outcome-based and performance orientated evaluation or traditional evaluation towards a multiplicity of models. Among these models or alternatives to traditional evaluation are responsive evaluation as illumination (Stake, 1983), utilisation-focused evaluation (Patton, 1997), fourth-generation evaluation (Guba and Lincoln, 1989), empowerment and self-evaluation evaluation (Fetterman, 1996) and others. In the recent past there has been a move from debates between positivists and post-positivists to a dialogue between paradigms.

Education Evaluation

The term education evaluation can be related back to the seminal work of Ralph Tyler in the early 1930s (Tyler, 1930). His approach was distinguished by its concentration on clearly stated objectives, as discussed in chapter two in the context of a product curriculum. The resulting behavioural objectives movement influenced curriculum design away from the content to be taught towards the student behaviours to be developed. Tyler's approach concentrates on direct measures of achievement, as opposed to indirect approaches that measure such inputs as quality of teaching or community involvement. This approach set the stage for how educators and other programme evaluators viewed evaluation for the next twenty-five years. During the 'Tylerian Age' in the US and subsequently in many other countries standardised tests were developed to reflect the objectives and content of the curricula. However, the influence of Tyler began to wane. Cronbach (1963) sharply criticised these approaches for their lack of relevance and utility and argued that the purpose of evaluation differentiates it from scientific research. Calling for a reformation in evaluation years later he recommended that its mission should be to 'facilitate a democratic, pluralistic process by enlightening all the participants' (Cronbach *et al*, 1986:1)

Evaluation Approaches

More recently Stufflebeam and Shinkfield (2007) assessed evaluation approaches by classifying them on the basis of their level of conformity to the definition of evaluation given by the Joint Committee of Congress' standards which focuses on the systematic assessment of the worth or merit of an object. Twenty six evaluation approaches were analysed under five categories: pseudoevaluations; questions- and methods-oriented evaluation or quasi-evaluation studies; improvement- and accountability-oriented evaluations; social agenda and advocacy approaches; and eclectic evaluations.

Pseudoevaluations categorise those evaluations which fail to produce and report valid assessments of worth or merit and are often motivated by political objectives (Stufflebeam and Shinkfield, 2007). The questions- and methods-oriented evaluation or quasi-evaluation studies group evaluations tend to narrow the evaluation's scope, often delivering, according to Stufflebeam and Shinkfield (2007), less than a full assessment of merit or worth. An example of this approach is the objectives-based evaluation and theory-based evaluation. These approaches list the programme's activities and desired end results with the main strength of such an approach lying in its causal inferences (Weiss, 1998). Improvement- and accountability-oriented evaluations summarise approaches that stress the need to fully assess a programme's value. The central thrusts of these approaches are to foster improvement and accountability through informing and assessing programme decisions, assist consumers to make wise choices among optional programmes and services and to help accrediting associations certify meritorious institutions and programmes for use by consumers. This approach is represented by Stufflebeam and Shinkfield (2007) as the Context, Input, Process and Product (CIPP) model. Context evaluations assess pertinent needs, assets, opportunities and problems to assist in formulating or judging goals. Input evaluations identify and assess competing programme strategies for meeting beneficiaries' assessed needs. Process evaluations assess the implementation of a selected programme strategy. Product evaluations search out, analyse and judge programme results.

Challenging the privileged status of traditional evaluation Lincoln and Guba (1985) encouraged a ‘paradigm wars’ (Caracelli, 2000:99) type of debate in the field of evaluation. From a constructivist viewpoint Lincoln and Guba argue that each truth is socially constructed. The following approaches developed from such debates. The social agenda and advocacy approaches are aimed at increasing social justice through programme evaluation. These approaches seek to ensure that all segments of society have equal access to educational and social opportunities and services. They favour a constructivist orientation and the use of qualitative methods. They provide for democratic engagement of stakeholders in obtaining and interpreting findings. The classic responsive evaluation approach by Stake (2003) is included in this category, which emphasises the evaluator’s role in interacting continuously with, and responding to, the needs of clients and stakeholders. This approach contrasts with Scriven’s (2005) objectivist orientation in that the client must be willing to endorse a quite open, flexible evaluation plan as opposed to a well-developed, detailed one. Clients must also be receptive to ambiguous findings and multiple interpretations. They must be sufficiently patient to allow the programme evaluation to unfold and find its direction based on ongoing interactions between the evaluator and stakeholders. Stake’s approach calls attention to the complexity and the uncertainty of the programme, the difficulty in measuring outcomes and the importance of descriptive and judgemental data (Viser, 2009). Again the evaluators and programme stakeholders are placed at the centre of the inquiry process.

Fourth-generation evaluation (FGE) follows three earlier generations of constructivist approaches to evaluation by Guba and Lincoln (1989). They suggest that the first three generations were focused on measurement, description and judgement. Guba and Lincoln (1989) identify inherent flaws in these evaluation methods as a tendency towards managerialism, a failure to accommodate value-pluralism and overcommitment to the scientific paradigm. I would generally agree with these criticisms as in many cases managers have had the ultimate power in determining what questions the evaluations pursued and how the data was collected and interpreted. Evaluations have not always acknowledged differences in values of the stakeholders involved in the evaluation. Finally, the overuse of the scientific method has ignored alternative ways to think about evaluation. Presenting quantifiable data as hard facts does not always encourage a responsibility in following up on findings.

FGE was designed to counteract problems with classical experimental or quasi-experimental designs in evaluation. Some of Guba and Lincoln's harsh attacks on quantitative evaluation methods were viewed as one-sided interpretations (Virtanen and Uusikylä, 2004). FGE was introduced as a participatory pluralistic process that provides a framework through which the interests of stakeholder groups and individuals can be put onto the agenda and renegotiated. It was thus presented as a responsive evaluation methodology.

Personalising evaluation (Kushner, 2000), in the tradition of responsive and democratic evaluation, grew out of concerns about the distortions generated when a programme is seen as the principal or exclusive context within which to attribute significance to people's lives and work. It proposes instead, the portrayal of people's lives and work as contexts within which to read the significance of the programme. Personalised evaluation promotes the view that evaluators must be their own methodologists and seek personal voice and personal meaning in their evaluations.

The newest addition to programme evaluation under the constructivist paradigm is the deliberative democratic approach advanced by House and Howe (2003). It envisions programme evaluation as a principled, influential societal institution, contributing to democratisation through the issuing of reliable and valid claims. Equity of all interested stakeholders is stressed and power imbalances are not tolerated. Methods employed include discussions with stakeholders, surveys and debates.

Eclectic evaluations include those approaches which draw on a broad and diverse range of sources. The most widely used of these approaches is Patton's (1997) utilisation-focused evaluation. The approach is geared towards maximising evaluation impacts and fits well with the key principle of change. It engages stakeholders to determine the evaluation's purposes and procedures and uses their involvement to promote the use of findings. Rather than trying to reach all stakeholders a select, representative group is chosen. A limitation may include the possibility of its vulnerability to corruption by user groups, since they are given much control over what will be examined, the questions asked, methods employed and questions to be asked. Stakeholders with conflicts of interest may influence the evaluation inappropriately.

Many of these models represent a form of evaluation which involves judgements made through the eyes of the external evaluator and the connotation persists of evaluation as an external monitoring of professional practice. In contrast, fourth generation evaluation takes a constructivist position, allowing access to participants' interpretations of their world, because they can construct and interpret realities, which are shaped and perceived by cultural and linguistic meanings. Evaluation within a naturalistic stance requires the analysis and description of participants' meanings and interpretations of the social world examined within the world settings they occupy (Brewer, 2003). A key assumption underpinning this type of evaluation is that evaluators' interactions with their participants is itself part of the evaluation exercise (Galvin, 2005). Critiquing evaluative techniques, which have been used over the years Guba and Lincoln (1989) draw attention to a tendency towards managerialism, where researchers determine what is to be evaluated and what will happen to the findings, so disempowering other stakeholders.

Fourth Generation Evaluation

Fourth-generation evaluation (FGE) is presented by Virtanen and Uusikylä (2004) as a goal-free evaluation (portraying a holistic portrayal of the programme) rather than a goal bound approach (where the evaluator maps causal links between objectives, inputs and outputs). The role of the evaluator is to provide a methodology through which different concerns and constructions of stakeholders can be understood and critiqued. Stakeholders are asked to provide their own (emic) constructions and evaluators include their own (etic) constructions as well as constructions from other sources e.g. relevant documentation and academic literature (Lay and Papadopoulos, 2007). The use of a hermeneutic dialectic circle (Guba and Lincoln, 1989) is proposed in FGE. Being *hermeneutic* means it is interpretative and being *dialectic* means it represents a comparison and contrast of divergent views. It therefore allows a cross fertilisation of data with a connection between them that allows for mutual exploration by all stakeholders (Guba and Lincoln, 1989). Participation between the stakeholders and the evaluator is recommended in following up on the findings, indicating a responsive evaluation. Continuing a search for illuminating constructions the evaluator posits that there can be no definitive conclusions (Stufflebeam, 1999).

FGE has been criticised for representing an 'over-socialized' interpretation of programme reality, in neglecting the programme goals in favour of attention to negotiations between stakeholders and consensus building (Virtanen and Uusikylä, 2004:83). Having used the methodology for a study (Joyce 2010) I refute this claim as the interviews connected back to the overall programme aims and the learning outcomes. FGE begins with a philosophical base in constructivism, where the evaluator shares constructions of other stakeholders in order to form a joint construction around which some consensus can be built. It is therefore a democratic methodology where as many people as possible can agree on the outcome (Heap, 1995). The study by Joyce (2010) did not fulfil such a description preferring instead to present the findings that can be judged by the reader on the holistic viewpoints as part of a pluralistic evaluation.

Conclusion

Learning from and about evaluation often requires us to change our mental models – to rethink our assumptions and beliefs and to develop new understandings about our programmes and evaluation processes (McNamara *et al*, 2010). This logically should lead on to an organisational learning approach to evaluation. Such an approach to evaluation would be context- sensitive, ongoing, support dialogue, reflection, and decision-making at department and organisation-wide levels, and contain strong commitments to self-evaluation and practitioner empowerment. A vision of evaluation for the twenty-first century may be one that is made honest, accurate, and useful by engaging in a partnership with practitioners, people, and programmes. Evaluators will be held to a higher standard and will be expected to do good through evaluation. Most experts agree that there is no one best model for evaluation. It is up to practitioners to agree their own philosophy about evaluation and to choose an approach which best matches their context and student needs.

References

Brewer J (2003) Naturalism. In Miller RL and Brewer JD (Eds.) *The A-Z of social research*. Sage, London. (210-212)

Caracelli VJ (2000) Evaluation use at the threshold of the twenty-first century. In Caracelli VJ and Preskill H (Eds.) The expanding scope of evaluation use. *New Directions for Evaluation*, No. 88. Jossey-Bass, San Francisco. (99-111)

Chen HT (1996) A comprehensive typology for program evaluation. *Evaluation Practice* 17(2): 121-130.

Clarke A and Dawson R (1999) *Evaluation research: an introduction to principles, methods and practice*. Sage, London.

Cousins JB and Aubry T (2006) *Roles for government in evaluation quality assurance*. Discussion paper. www.tbs-sct.gc.ca (Accessed 13/3/09).

Cronbach LJ (1963) Course improvement through evaluation. *Teachers College record* 64, 672-683.

Cronbach LJ, Robinson Ambron S, Dornbusch SM, Hess RD, Hornik RC, Phillips DC, Walker DF and Weiner SS (1986) *Toward reform of program evaluation*. Jossey-Bass Publishers, London.

Fetterman DM (1996) Empowerment evaluation: introduction to theory and practice. In Fetterman DM, Kafitarian SJ and Wandersman A (Eds.) *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Sage, Thousand Oaks. CA. (3-48).

Galvin K (2005) Navigating a qualitative course in programme evaluation. In Holloway I (Ed.) *Qualitative research in health care*. Open University Press, Berkshire (229-249).

Guba EG and Lincoln YS (1989) *Fourth generation evaluation*. Sage, Thousand Oaks.

House E (1993) *Professional evaluation: Social impact and political consequences*. Thousand Oaks, Sage CA.

House ER and Howe KR (2003) Deliberative democratic evaluation. In Kellaghan T and Stufflebeam DL (Eds.) *International handbook of educational evaluation*. Kluwer, Norwell MA (79-102).

Joyce P (2010) Approaches to learning of postgraduate healthcare professionals in an outcomes-based curriculum. Unpublished doctorate thesis, Dublin City University, Ireland.

Kells HR (1992) *Self-regulation in higher education. A multi-national perspective on collaborative systems of quality assurance and control*. Jessica Kingsley Publishers, London.

Kushner S (2000) *Personalising* evaluation. Sage, London.

Lay M and Papadopoulos I (2007) An exploration of fourth generation evaluation in practice. *Evaluation* 13(4):495-504.

Lincoln YS and Guba EG (1985) *Naturalistic Inquiry*. Sage, Thousand Oaks.

Lomas L (2007) Are students customers? Perceptions of academic staff. *Quality in Higher Education* 13(1): 31-44.

McNamara, G, Joyce, P and O'Hara, J (2010) The Evaluation of Adult Education and Training Programmes. *The International Encyclopedia of Education,* 3: 548-554.

Ong BN (1996) *Rapid appraisal and health policy*. Nelson Thornes, Cheltenham.

Organisation for Economic Corporation and Development (2008) *Thematic review of tertiary education.*

Patton MQ (1997) *Utilization-focused evaluation: The new century text*. 3rd Edition. Sage, Thousand Oaks.

Patton M. (2002) *Qualitative research and evaluation methods,* 3rd Edition. Sage Publications, Newbury Park, CA.

Pawson R and Tilley N (1997) *Realistic evaluation*. Sage, London.

Perkins D (1999) The many faces of constructivism. *Educational Leadership* 57(3): 6-11.

Robson C (2002) *Real world research: A resource for social scientists and practitioner researchers*. 2nd Edition. Blackwell, Oxford.

Scriven M (2005) Review of DM Waterman and A Wanderman (Eds.) Evaluation principles in practice. *American Journal of Evaluation* 26(3): 415-147.

Shadish W, Cook T and Leviton L (1991) *Foundations of program evaluation: Theories of practice*. Sage, Newbury Park, CA.

Shadish W, Cook T and Leviton L (1991) *Foundations of program evaluation: Theories of practice*. Sage, Newbury Park, CA.

Shaw I (1999) *Qualitative evaluation.* Sage, Thousand Oaks, CA.

Tones K and Tilford S (2001) *Health education: Effectiveness, efficiency and equity*. 2nd Edition. Nelson Thornes, London.

Stake RE (2003) Responsive evaluation. In Kellaghan T and Stufflembeam DL (Eds.) *International handbook of educational evaluation*. Kluwer. Norwell. MA (63-68).

Stufflebeam DL (1999) *Foundational models for 21st century program evaluation*. The Evaluation Centre Occasional Papers Series. 1-96. http://www.unssc.org/web/programmes/LS/unep-unssc-precourse-material/.pdf (Accessed 23/3/09).

Stufflebeam DL and Shinkfield AJ (2007) *Evaluation theory, models, and application*. John Wiley and Sons, Inc., San Francisco.

Tyler RW (1930) Measuring the Ability to Infer. *Educational Research Bulletin* 9(17): 475-480.

Virtanen P and Uusikylä P (2004) Exploring the missing links between cause and effect: a conceptual framework for understanding micro-macro conversations in programme evaluation. *Evaluation* 10(1):77-91.

Viser RME (2009) *Trends in program evaluation literature: the emergence of pragmatism*. TCALL Occasional research Paper No. 5. http://www-tcall.tamu.edu/orp/orp5.htm (Accessed 5/6/09).

Weiss CH (1998) *Evaluation: Methods for studying programs and policies*. 2nd Edition. Upper Saddle River, NJ, Prentice Hall.