Action research: Overcoming the Sports Mentality Approach to Assessment/Evaluation

REVIEW

George Bodner, Daniel MacIsaac, and Scott White

Department of Chemistry, Purdue University, West Lafayette, IN 47907-1393
gmbodner@purdue.edu

Introduction

Each time we make significant changes in what we teach or how we teach we are faced with the same question: how can we find out whether the innovation we have brought into our classroom is worthwhile? Chemists, familiar as they are with the criteria for decision-making adopted by physical scientists, find this question so difficult to answer that they often avoid doing the experiment that might provide evidence on which to base an answer. Let us therefore build a metaphor on a recent example from medical research.

In 1997, Bailar and Gornik reported an analysis of age-adjusted mortality rates due to cancer from 1950 to 1994[1]. This paper was picked up by the popular press, who reported that the war on cancer had been a failure[2]. Bailar and Gornik chose to analyze age-adjusted mortality rates because they regarded it as "the most basic measure of progress against cancer" and because it "focuses attention on the outcome that is most reliably reported"[1]. The question before us is simple: would they have reached the same conclusions if they had examined changes in the length of the patient's survival, or changes in the quality of life after cancer had been diagnosed?

Bailar and Gornik's paper provides a metaphor on which discussions of the evaluation of instructional innovation can be based because it illustrates the role that the choice of methodology for evaluation has on the conclusions that are reached. Chemists concerned with improving the way they teach chemistry need to recognize this and act accordingly.

The sports mentality approach to evaluation

Suppose a group of chemistry teachers wanted to evaluate the effectiveness of a set of new curriculum materials or a new method of teaching. What kind of experiment would they be most likely to design? And, what hidden assumptions would underlie their choice of methodology?

History has shown that chemists often base the design of such experiments on the hidden assumption that assessment and evaluation are synonyms. Within the context of the classroom this may be just as incorrect as the assumption that accuracy and precision are synonyms within the context of the chemical laboratory. *Assessment* might best be defined as the process by which the performance of individual students or groups of students is measured[3]. (This use of the term is consistent with the first definition of assessment in the *New Shorter Oxford English Dictionary*: The determination of the amount of a tax, fine, etc.; a scheme of taxation etc.)

Evaluation is the process by which information is collected to make decisions on how instruction can or should be improved[4]. Assessment is therefore a necessary but not sufficient component of evaluation.

Once the fundamental assumption that assessment and evaluation are synonyms is made, chemists often presume that the optimum design for the evaluation experiment is to compare student performance on a common exam for an experimental group using the new curriculum materials (or the new approach to teaching the course) with a control group using the old curriculum[5]. All too often, the result of this experiment is data that are precise, and sometimes statistically significant, but not necessarily useful in answering the original question.

We have described this strategy for the design of evaluation experiments as the sports mentality approach[6]. It provides results that are equivalent to hearing a sports commentator announce the results of a cricket match in terms of one team declaring their innings closed at a score of 318 for 6. For the casual sports fan, this is all the information one needs because it is used to tell us who 'won.' But it is difficult to imagine a coach selecting a team for the next match based only on this information.

There are several potential sources of error in basing evaluations of changes in either course content or the approach to teaching this content on student performance in a summative exam, either during or at the end of the course.

Herron has written about the Principle of Least Cognitive Effort, which presumes that students make the choice that appears to require the least effort[7]. This aspect of human behavior can be a confounding variable in the traditional experiment, which assumes that students will always take advantage of opportunities to do better in a course. What about the students who don't want to do better, who will do whatever is necessary to get a B or even a C? We have found that it is possible to make a change that significantly improves the classroom environment without seeing any effect on exam performance[8]. The traditional experiment is also plagued by the many factors that influence test performance besides the instructional innovation being studied.

A more serious problem with the traditional experiment might best be understood in terms of the metaphor of a drunk searching for a coin beneath the lamp post – not because this is where the coin was dropped, but because it is where the light is. By focusing on how much is learned, the traditional experiment fails to measure differences in *what is learned*, or what knowledge is *retained*, or whether a new instructional technique leads to improvement in students' understanding of knowledge we *value*, rather than knowledge that can be easily tested. In other words, it is not sensible to use the mark scored in a summative exam as a measure of the effect of an educational innovation unless the sole purpose of the innovation was to improve the mark in that exam.

A subtle, but potentially serious, problem with the traditional experiment is its assumption that the change being made is either 'good' or 'bad', i.e. that students, in general, will either benefit from the change or not. What if the change is bipolar? What if some students benefit, while others do not? The traditional experiment does not answer the question: *cui bono*? Who benefits from the intervention? Is it the intended audience? (The traditional experiment presumes that innovations will benefit all students equally, which is seldom if ever true).

Qualitative techniques as an alternative approach to evaluation

Qualitative techniques have been offered by their proponents as a naturalistic alternative to the experimental, quantitative, behaviourist tradition described above[9]. These techniques are built on the social science tradition of ethnography[10,11] and therefore involve extensive interviews[12] that are analyzed in terms of either case studies[13] or cross-case analyses[14].

Qualitative techniques are most often associated with educational research[15]. While qualitative research can inform teachers, it is usually done to inform researchers and has little (if any) effect on classroom practice. As working professionals, teachers are quite aware of what is happening in their classes - they do not believe that they need to be the subjects of anthropological research.As we will see, however, qualitative techniques can be applied to the more pragmatic issue of evaluating the impact of changes in instruction.

Unorthodox methodologies: formative research

In what amounts to a rejection of the 'methodolitry' endemic to both the qualitative and quantitative paradigms, practicing teachers have developed an approach to evaluation known as formative research. Walker has argued that this approach is "usually eclectic in its choice of techniques for eliciting data, including self-reports (in the form of diaries, interviews or questionnaires), observations, tests, and records"[16]. Walker cites the work ofTreisman[17,18] as an example of this approach.

"Treisman... carried out a chain of studies that were by traditional standards methodologically primitive but nevertheless exceptionally productive.... *By any reasonable standards...this [Treisman's study] was an outstanding study....* [Treisman] focused his attention on the crucial practical problem, observed practises closely, kept himself open to a wide variety of evidence at every stage of the inquiry, compared circumstances in which a practice seemed to succeed with circumstances in which it failed, searched for factors in the situation that could be changed, redesigned practises to reflect what he thought he had learned from his observations, and tested the new practises by using the standards of achievement actually employed in the real course. His results have been widely reported and have already begun to influence research and practise in mathematics education... And all this work was accomplished in three years on a modest budget."

Action research as a method for doing formative research

Action research is an approach to formative research that can be traced back to the end of World War II, when the social psychologist Kurt Lewin[19] developed most of its current methodological characteristics. Action research soon fell out of favour among those who pursued "the promise of quantitative methods (uniform regularities, predictability, control, etc.)"[20]. In the '80's and '90's, however, as the positivist foundations of quantitative methods came under increasing attack, action research became increasingly popular, particularly through the work of Kemmis and McTaggert[21-27]. Indeed, a search of the ERIC database brought up 2094 hits under the category 'action research' from 1978 to present.

Kemmis and McTaggert describe action research as a recursive, reflexive, dialectical technique whose goal is to help people investigate reality in order to change it, or to change reality in order to investigate it, by changing their practices in a collaborative, self-reflective spiral of cycles[24]. It is *recursive* because it is a cyclic process in which the product of one step is used as the input for the next. It is *reflexive* because it is characterized by constant reflection on the results of each step in the cycle. It is *dialectical* in the sense of a critical investigation of the truth of people's opinions. Hopkins described action research as an informal, qualitative, formative, subjective, interpretive, reflective, and experiential mode of inquiry in which all individuals are knowing and contributing participants[28].

Our use of action research has been based on a series of assumptions that are so fundamental to this work they might be considered beliefs. We believe that chemists introduce changes in the curriculum or in the way they teach because they have perceived weaknesses in the current situation. Essentially they have formulated an hypothesis (which may or may not be precisely defined) that a particular change will lead to a particular improvement. As concerned scientists they will wish to test or evaluate their hypothesis. This means that a systematic evaluation should be done whenever significant changes are made in an established curriculum or in the way the curriculum is delivered. These evaluations should look behind the facade of answers to the question: "Do the students like it?", toward deeper questions such as "What do students learn that they were not learning before?" and, "If we could provide students with a voice to express their opinions and concerns, what changes would they recommend?"[29].

We believe that any significant intervention into a practicing classroom will have an effect. (If no effect is found, this is more likely to result from poor experimental design than from a flaw in the intervention.) Instead of asking: does the intervention have an effect on the classroom environment, we prefer asking: what is the effect of the intervention, what happens to the teacher; what happens to the students?

We believe that evaluations should assume that any change in instruction will have both positive and negative effects; some students will benefit, others may be harmed. Evaluations should help us to understand what aspects of the intervention are responsible for the positive effects and what facets give rise to the negative effects. One of the goals of evaluation should be modifications of the intervention to increase the positive effects on the target population and minimize any negative effects. We recognise that innovators in education are under severe pressures which prevent them from spending much time evaluating their innovations. Nevertheless, through a knowledge of what is possible, it is possible to select an approach which will generate useful information without spending the amount of time which a committed educational researcher would consider necessary.

How is action research done?

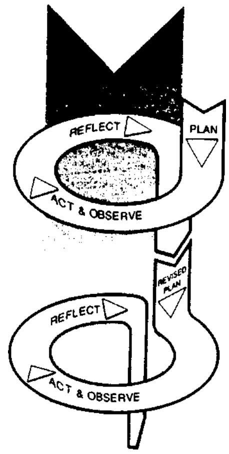
Elliott has described action research in terms of an iterative cycle of four steps or stages[30].

- The Reconnaissance and General Plan: an exploratory stance is adopted, where an understanding of a problem is developed and plans are made for some form of intervention.
- The Action in Action Research: the intervention is then carried out.
- Monitoring the Implementation: during and around the time of the intervention, pertinent observations are collected in various forms.
- The Revised Plan: the data are examined for trends and characteristics, and a new strategy is developed for implementation.

The new intervention strategies are then carried out, and the cyclic process repeats, continuing until a sufficient understanding of (or implementable solution for) the problem is achieved.

Kemmis and McTaggert[24] characterize action research in terms of a spiral of three steps or stages - plan, act and observe, and reflect - as shown in Figure 1.This view of action research has several advantages. By coupling 'act' and 'observe', it emphasizes the formative nature of this methodology. It also emphasizes the cyclic nature of action research as it moves through one iteration after another. In some ways, action research is similar in nature to the numerical technique known as successive approximation - the goal is to achieve a desirable outcome by a process of repeated iterations.

Figure 1: The action research spiral



The role of communication in action research

One of the distinguishing characteristics of action research is the degree of empowerment given to all participants. Whereas educational research has historically been *done on* students or their instructors, action research is *done with* students and their instructors. All participants - students, instructors, and other parties - are knowing, active members of the research project. All participants - including the researchers, the teachers, and the students - contribute to the process by which meaning is extracted from the data and in decisions about modifications that are made in the next cycle or iteration.

Proponents of action research often talk about involving all the major stakeholders in the evaluation process. In the simplest case, this means both the instructors and their students. But it can also involve curriculum developers, researchers, administrators, parents, and so on.

Elliott[30] considers the need for communication between all participants to be of paramount importance: "Since action research looks at a problem from the point of view of those involved it can only be validated in unconstrained dialogue with them." Kemmis[24] argues that action research is a *social* process in which students and teachers work together to improve the processes of teaching and learning. It is *participatory* in the sense that people can only do action research on themselves, either individually or collectively, as a group. It is both *practical* and *collaborative* because it provides those involved with a framework which helps them to avoid making irrational, unproductive and unjust judgements about the topic under consideration.

Every teacher a researcher?

Anyone who has pondered the forces that lead to schism in an established religion should accept the existence of differing opinions on one or more aspects of a methodology, such as action research. Regardless of whether it is applied to curriculum development, professional development, or planning and policy development, there is a consensus that action research is intrinsically *collaborative*. Kemmis and McTaggart[24] argue that it occurs within groups of participants who can be teachers, students, principals, parents, or other community members. What is important is a shared concern among the members of the group. There are proponents of action research whose slogan is 'each teacher a researcher'[31]. Others argue that an outsider should be included in the community being studied, who is neither the instructor nor a student, but who is actively involved with both students and their instructor(s) in the action research cycle and who does not have a vested interest in the success of the change being studied. The latter approach is characteristic of our work using action research.

How do we recognize when an experiment is successful?

In theory, the model of educational research popularized by Campbell and Stanley[5], in which the performance of experimental and control sections is compared, has the advantage that we always know who 'won.' We simply leave the question to the cold, hard, objective test of statistics. In practice, however, as noted in the section on the sports mentality approach to evaluation, we achieve this power at a significant cost. By focusing on measurements that can be subjected to statistical tests we often lose the ability to measure the phenomenon in which we are interested. Or we find the power of our statistical tests diluted by the many confounding variables that influence measurements such as test scores. Or, returning to the metaphor of the cricket match, we find ourselves trading useful descriptive results about individual performance for definitive, but less useful, information about the final score.

This raises an important question: what characteristic of action research plays the role in this methodology that *p* values, *F* values, or tables of two-tailed tests of significance play in more traditional educational research? In particular, how do we ensure that mistakes are not made in deciding which effects of an intervention are 'positive?'

The answer is simple: no research methodology operates in a philosophical vacuum. Quantitative research is based on a philosophical tradition that its proponents describe as scientific and its opponents label behaviourist and positivist[20]. Action research is inexorably coupled to critical theory and often linked most explicitly with the work of the German sociologist-philosopher Jurgen Habermas[32-34]. Rather than delve into a lengthy consideration of critical theory and the implications of Habermas' differentiation among technical, practical, and emancipatory knowledge, we will propose a safeguard against potential abuse of the action research methodology based on the argument of Kemmis and McTaggert that action research is "...a process in which people deliberately set out to contest and reconstitute irrational, unproductive (or inefficient), unjust and/or unsatisfying (alienating) ways of interpreting and describing their world..., ways of working..., and ways of relating to others."[24]. As long as action research is a process done by a group, in which each member of the group is a knowing participant, and decisions or conclusions are agreed to by the group - not just the individual in charge of the course - they are likely to be the correct decisions or conclusions.

How action research can change the questions we ask

Action research has become an increasingly valuable methodology in our research group. It has been used to probe the effect of the implementation of computer simulations in a senior-level chemical engineering laboratory on design[29]; to guide the development and implementation of microcomputer-based laboratories in our introductory physics curriculum[35]; to examine the effect of a novel laboratory course on advanced experiments in chemical engineering[36]; to study the effect on both students and their instructor when an alternative approach is taken to teaching organic chemistry[37,38]; to guide the development of Web-based instruction materials for distance learning in general chemistry[39]; and to bring about significant improvements in student attitude toward a sophomore-level analytical chemistry course for non-majors[40]. To illustrate how the choice of methodology used for evaluation can influence the questions being asked and the conclusions that are reached, let us look at examples of this work.

One of these projects was a response to a request from a colleague in Pharmacy who wanted to change the way he taught his organic chemistry course. The traditional lectures in his section of the course were replaced with a problem-oriented approach in which the instructor presented students with a problem, solicited answers from individual students or groups of students, and then helped the students examine the logical consequences of these answers. To encourage students to work together in groups both in and outside of the classroom, the groups were given time to discuss each hour exam before they split up to write their individual responses to the exam questions.

As one might expect, we analyzed student performance on the exams, which suggested that the individual members of a group often wrote very different answers to the exam questions, but that the class as a whole seemed to understand the questions better than ever before. We also noted that these students did significantly better than students from a traditional lecture section when the two sections of the course were merged for the second-semester of organic chemistry.

Our evaluation of this experiment went far beyond the analysis of exam data, however. We taped and then transcribed each of the 43 classes during the semester and recorded field notes that reported on a day-by-day basis observations about the interactions among the students and between the students and the instructor. We collected attitude data using both Likert scale[41] and open-ended questions. We taped and transcribed extensive interviews with the instructor of the course, his colleagues who taught other sections of the course, and the students taking the course.

It should come as no surprise that the quality of the insight obtained from this information was directly proportional to the effort required in its collection. The results of the Likert scale questionnaires, which took little effort on anyone's part, were pleasing - they suggested that the experiment was worth repeating. The more time-consuming analysis of responses to open-ended questionnaires provided better insight into the aspects of the intervention that needed to be kept and the problems faced by individual students. But it was the transcripts of the lectures and the interviews that provided the information needed to enter the second cycle of the action research iteration.

The action research methodology helped us answer questions that might not otherwise have been asked, such as:

- how do we overcome student resistance to this approach,
- what do we have to do to ensure that groups operate effectively,
- what is the nature of the dissatisfaction that might lead an instructor to change to a problem-oriented approach,
- what factors make it difficult to change the classroom environment,
- what factors interfere with the ease with which this technique can be used by other instructors, or transported to other institutions,
- what effect does this mode of instruction have on the instructor's attitude toward teaching,
- what effect does it have on students' perception of the difficulty of organic chemistry,
- does this approach to instruction produce students who are more likely to think the way an organic chemist thinks?

Another project began with a request for help on the evaluation of computer simulations being used as a substitute for traditional experiments in a capstone chemical engineering laboratory course on design. Our results suggest that it would be a mistake to ask which laboratory format is 'better' for students. They indicate that computer simulations and traditional experiments have different roles in the curriculum because they emphasize different aspects of engineering and require both different levels and types of expertise.

By providing the students with a voice, the action research methodology helped us understand that the environment in which the simulations were implemented had a major effect on students' perceptions of their value and therefore provided useful information on the optimum way in which these simulations could be used. It also clearly showed that computer simulations, by themselves, are not magic bullets that provide instruction and pedagogical benefits for the students in the absence of a human interaction between the students and their instructor. Action research therefore allowed us to provide the authors of the simulations with more information, and more useful information, than they expected.

Our work in analytical chemistry began with classroom observations of students and interviews with students to identify the source of their dissatisfaction with the course, and has extended through three year-long cycles. The work on Web-based learning began with the software developers' efforts to write what they hoped were useful elements of a program. Students were then observed while they used components of this program and revisions were made based on their suggestions. In this case, the action research cycle was significantly shorter, on the order of weeks or at most months.

The key features of our use of the action research methodology could be summarized as follows. Changes are made in what we teach or the way we teach it. Evaluation occurs while the changes are being made. As many sources of information are collected as possible. We never presume that all students will benefit from the change, and are constantly searching for ways to maximize positive effects and minimize negative effects of these changes. And the students are knowing, active participants in the decision-making process about changes that should be made in the next iteration in the innovation cycle.

References

1. Bailar JC III and Gornik HL 1997 Cancer Undefeated *N Eng J Med* **336** 1569-1574
2. Sternberg S 1997 $30 billion 'war on cancer' a bust? *USA Today* 15 29 May 1997 p.1A
3. See for example, Angelo TA and Cross KP 1993 *Classroom Assessment Techniques: A Handbook for College Teachers* (Jossey Bass San Francisco)
4. See for example, Weiss CH 1998 *Evaluation: Methods for Studying Programs and Policies* (Prentice Hall Upper Saddle River NJ)
5. See for example, Campbell DT and Stanley JC 1963 *Experimental and Quasi-Experimental Designs for Research* (McNally Chicago)
6. MacIssac D and Bodner GM 1993 Overcoming the Sports Mentality Approach to Program Evaluation, Action Research as a Metaphor for Curriculum Evaluation paper presented at the National Association for Research in Science Teaching Meeting, Atlanta, GA
7. Herron JD 1996 *The Chemistry Classroom, Formulas for Successful Teaching* (American Chemical Society, Washington DC)
8. Bodner GM, Metz PA and Tobin K 1997 Cooperative Learning: An Alternative to Teaching at a Medieval University *Aust Sci Teach J* **43**(3) 23-2
9. Gage NL 1989 The paradigm wars and their aftermath, A "historical" sketch of research on teaching *Educ Res* 18(7) 4-10
10. Agar MH 1980 *The Professional Stranger, An Informal Introduction to Ethnography* (Academic Press, Orlando)
11. Goetz JP and LeCompte MD 1984 *Ethnography and Qualitative Design in Educational Research* (Academic Press, Orlando)
12. Seidman IE 1991 *Interviewing as Qualitative Research, A Guide for Researchers in Education and the Social Sciences* (Teachers College Press Columbia University, New York)
13. Merriam SB 1998 *Qualitative Research and Case Study Applications in Education* (Jossey-Bass, San Francisco)
14. Patton MQ 1990 *Qualitative Evaluation and Research Methods* 2nd Edition (Sage, Newbury Park CA)
15. Denzin NK and Lincoln YS 1994 *Handbook of Qualitative Research* (Sage, Thousand Oaks CA)
16. Walker DF 1992 Methodological Issues in Curriculum Research, in *Handbook of Research on Curriculum* (ed. PW Jackson) (Macmillan, New York)
17. Henkin L and Treisman U 1984 Final Report, University of California Professional Development Program ERIC ED 2669932
18. Treisman U 1983 Improving the performance of minority students in college-level mathematics *Innovation Abstracts* **5**(17) 1-5

19. Lewin K 1946 Action research and minority problems. *J Soc Issues*, **2**(4), 34-46.
20. Fenstermacher GD, in *Handbook of Research on Teaching* 3rd Ed. (Ed. MC Wittrock) (Macmillan, New York)
21. Carr W and Kemmis S 1986 *Becoming Critical, Education Knowledge and Action Research* (The Falmer Press, London)
22. Grundy S 1987 *Curriculum, Product or Praxis?* (The Falmer Press, East Sussex)
23. Kemmis S and McTaggart R 1988 *The Action Research Planner* 3rd Ed (Deakin University Press: Geelong, Victoria, Australia)
24. Kemmis S and McTaggart R 1990 *The Action Research Reader* (Deakin University Press: Geelong, Victoria, Australia)
25. McTaggart R 1991 *Action Research, a Short Modern History* (Deakin University Press: Geelong, Victoria, Australia)
26. Noffke SE and Stevenson RB 1995 *Educational Action Research, Becoming Practically Critical* (Teachers College Press, New York)
27. Zuber-Skerritt O 1996 *New Directions in Action Research* (The Falmer Press, London)
28. Hopkins D 1985 *A Teacher's Guide to Classroom Research* (Open University Press: Philadelphia)
29. White SR and Bodner GM 1999 Evaluation of Computer Simulation Experiments in a Senior-Level Capstone Course *Chem Eng Educ*, in press
30. Elliott J 1978 What is action research in schools? *J Curric Studies* **10** 355-57
31. Kincheloe JL 1991 *Teachers as Researchers, Qualitative Inquiry as a Path to Empowerment* (The Falmer Press, London)
32. McCarthy TA 1979 *The Critical Theory of Jurgen Habermas* (Massachusetts Institute of Technology Press, Cambridge)
33. Roderick R 1986 *Habermas and the Foundations of Critical Theory* (St. Martin's, New York)
34. Young RE 1990 *a Critical Theory of Education: Habermas and Our Children's Future* (Teachers' College Press, New York)
35. MacIsaac D 1995 Curricular Reformation in Computer-Based Undergraduate Physics Laboratories via Action Research, Ph.D. Dissertation
36. Lauterbach J, White S, Liu Z, Bodner GM and Delgass W N 1997 "A Novel Laboratory Course On Advanced Chemical Experiments" *J Chem Eng Educ* 260-265
37. Lowrey KA 1996 An Alternative Mode of Instruction: A Detailed Analysis of Three Units in Medicinal Chemistry 204, M. S. Thesis
38. Bauer RC II 1998 Teaching Innovation in Organic Chemistry: An Inquiry into What Happens When the Lecturer Stops Lecturing, Ph.D. Dissertation
39. Hunter WJF 1998 Student Participation in World Wide Web-Based Curriculum Development of General Chemistry, Ph.D. Dissertation
40. Holladay, SH, Raftery MD and Bodner GM 1998 Improving Students' Perceptions of Analytical Chemistry Via Action Research, 15th Biennial Conference on Chemical Education, Waterloo, Ontario, Canada
41. Likert scale questions have a single statement to which the respondent provides a graded response, for example from 'strongly agree' to 'strongly disagree'. The questionnaire reported on p3 of this issue used Likert scale questions.