Social Work Education

ISSN: 0261-5479 (Print) 1470-1227 (Online) Journal homepage: https://www.tandfonline.com/loi/cswe20

Evaluating Social Work Education: A Review of Outcomes, Measures, Research Designs and Practicalities

John Carpenter

To cite this article: John Carpenter (2011) Evaluating Social Work Education: A Review of Outcomes, Measures, Research Designs and Practicalities, Social Work Education, 30:02, 122-140, DOI: 10.1080/02615479.2011.540375

To link to this article: https://doi.org/10.1080/02615479.2011.540375

Published online: 06 Feb 2011.

Submit your article to this journal

Article views: 1348

View related articles

Citing articles: 53 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cswe20

Social Work Education
Vol. 30, No. 2, March 2011, pp. 122–140

Routledge
Taylor & Francis Group

Evaluating Social Work Education: A Review of Outcomes, Measures, Research Designs and Practicalities

John Carpenter

This review considers how the outcomes of social work education can be identified with reference to Kirkpatrick's framework of levels of outcomes and Kraiger et al.*'s general model of learning outcomes. It presents examples of different approaches to the measurement of outcomes and to the use of research designs in social work education and interprofessional education involving social workers and social work students. These examples are drawn from a search of the English language literature (1997–2010). Three-quarters of the studies (23 out of 31) found measured changes in attitudes and/or knowledge and skills only. Seven measured changes in social workers' behaviour and only one measured the impact on service users' lives. Three-quarters employed pre–post one group designs. The advantages and disadvantages of alternative research designs are discussed. Finally some practicalities regarding the development of an outcomes-based culture in university-based education are considered.*

Keywords: Evaluation Social Work Education; Outcomes Social Work Education; Outcome Measures; Research Designs; Learning Outcomes; Social Work Education

Introduction

This article is a revised, updated and condensed version of a discussion paper written for the Social Care Institute for Excellence (SCIE) and the former Scottish Institute for Excellence in Social Work Education (SIESWE)[1] (Carpenter, 2005). Over the last five years that paper achieved its aim of stimulating discussion amongst social work educators in the UK on outcomes and how they can be evaluated, specifically considering/focusing on:

1. What we mean by the 'outcomes' of social work education.
2. How these outcomes might be measured.
3. The advantages and disadvantages of different research designs for the evaluation of outcomes in social work education.

Correspondence to: Professor John Carpenter, School for Policy Studies, University of Bristol, 8 Priory Road, Bristol BS8 1TZ, UK. Email: j.s.w.carpenter@bristol.ac.uk

ISSN 0261-5479 print/1470-1227 online © 2011 Taylor & Francis
DOI: 10.1080/02615479.2011.540375

4. Some of the methods and measures which have been used to evaluate outcomes. (Carpenter, 2005, p. 4)

The paper has been cited extensively in the UK literature and internationally, although citation tracking revealed that most references have been used to note, and lament, the paucity of outcome studies. The paper did, however, lead directly to an action learning project, supported by SCIE, SIESWE and the Higher Education Academy Subject Centre for Social Policy and Social Work (SWAP), involving nine English and Scottish universities (Burgess and Carpenter, 2008) and subsequently to the publication of a monograph comprising reports of the outcome studies undertaken (Burgess and Carpenter, 2010). Around the same time, the Department of Health commissioned a major evaluation of the new social work degree in England that employed mixed research methods, including a focus on outcomes (Orme *et al.*, 2009). Consequently, in preparation for this paper, an updating literature search was undertaken.

Literature Search

The inclusion criteria for the updating literature review were studies in the English language published between 2004 and 2010 on the outcomes of social work education, and of interprofessional education (IPE) involving social workers. The outcomes specified were of student learning: outcomes had to be measured and the paper had to report differences between two or more time points. Studies using only qualitative methods, where data were not reported, and those which measured student satisfaction alone were excluded. The search terms were: 'education' AND 'outcomes' AND 'evaluation' AND 'measures' AND 'social work'.

The following electronic archives were searched: Web of Science, ASSIA, CINAHL, Social Care online and Embase plus Google Scholar. Hand-searching was also carried out of the two most recent issues of journals and online advanced access papers in the most frequently cited journals: *Social Work Education*, *Journal of Social Work Education*, *British Journal of Social Work*, *Child and Youth Services Review*, *Research on Social Work Practice*, *Journal of Interprofessional Care*, and *Health and Social Care in the Community*. 'Snowballing' and citation tracking were employed to identify further publications.

Abstract were screened and full texts of papers which appeared to meet the inclusion criteria were obtained and assessed for quality and for relevance to this paper. A summary of the number of studies identified in the original review combined with the update review is presented in Table 2 below. Studies were categorised in terms of their research design and the learning outcomes evaluated. However first, it is necessary to discuss learning outcomes.

Learning Outcomes

What do we mean by outcomes? The best known and most widely used classification of educational outcomes was devised by Kirkpatrick (1967). This model defined four levels of outcomes: learners' reactions to the educational experience; learning,

Table 1 Levels of Outcomes of Educational Programmes [Carpenter (2005) after Kirkpatrick (1967) and Barr *et al.* (2000)]

Level 1	*Learners' Reactions*	Participants' views of their learning experience and satisfaction with the training
Level 2a	*Modification in Attitudes and Perceptions*	Changes in attitudes or perceptions towards other professionals, service users and carers, their problems and needs, circumstances and care
Level 2b	*Acquisition of Knowledge and Skills*	The concepts, procedures and principles of working with service users and carers, and the acquisition of thinking/problem solving, assessment and intervention skills
Level 3a	*Changes in Behaviour*	The implementation of learning from an educational programme into practice, prompted by modifications in attitudes or the application of newly acquired knowledge and skills
Level 3b	*Changes in Organisational Practice*	Wider changes in the organisation/delivery of care, attributable to an education programme
Level 4	*Benefits to Users and Carers*	Any improvements in the well-being and quality of life of people who are using services, and their carers, which may be attributed to an education programme

conceptualised mainly as the acquisition of knowledge and skills; behaviour change, including the application of learning to the work setting; and results, assessed in relation to intended outcomes. This model was elaborated by Barr *et al.* (2000) for a review of interprofessional education in order to include the modification of attitudes as a learning outcome and to divide 'results' into change in organisational practice and benefits to patients/clients. A generalised version is shown in Table 1.

Table 2 Published Studies of the Outcomes of Social Work Education 1997–2010 ($N = 31$)

Participants	Social work only	18
	Interprofessional	13
Outcome levels (ref. Table 1)	2a: Attitudes and perceptions (only)	4
	2b: Knowledge and skills (only)	6
	2a + 2b: Attitudes + knowledge and skills	13
	Subtotal	*23*
	2b + 3a: Knowledge and skills + behaviour	5
	2a + 2b + 3a: Attitudes + knowledge and skills + behaviour	2
	2a + 2b + 4: Attitudes + knowledge and skills + outcomes for users	1
	Total	*31*
Reliable and validated measures	Yes	20
	No	11
Research design (all pre–post)	One group only	23
	Comparison group	5
	Control group	3

As the framework suggests, learning is conceptualised both as a response to positive reactions to training and as a causal determinant of changes in the trainee's behaviour. The advantages of Kirkpatrick's model are that it focuses attention on possible different levels of evaluation and implies that a comprehensive approach should be concerned with all these levels. Thus, it is insufficient to evaluate training according to whether or not the students enjoyed the presentations and found them informative, or to assume that it is adequate to establish that students acquired particular skills, in communication, for example, without investigating whether or not they were able to transfer those skills to practice. Further, since the ultimate purpose is to benefit service users and/or carers, a comprehensive evaluation should ideally ask whether training has made any difference to their lives.

Findings From the Literature Search

Thirty-one studies met the inclusion criteria, of which 18 had been published in the six years since the previous review. (Multiple reports from the same study were counted as one study.) Seventeen studies derived from the UK, 11 from the US, two from Hong Kong and one from Norway. Over half the studies concerned social workers or students only, the rest being of IPE initiatives involving social workers (Table 2).

Three-quarters (23) of the studies found measured changes in attitudes (2a) and/or knowledge and skills (2b) only. Seven measured changes in behaviour, all using simulations rather than actual practice. Only one measured changes in service users' lives, employing standardised measures. Overall, two-thirds employed standardised measures or at least measures whose reliability was reported. Three-quarters were simple pre–post studies with one group of students. Research designs are considered later.

Specifying and Measuring Learning Outcomes

Kraiger *et al.* (1993) presented a theoretically based general model of learning outcomes. They elaborated Kirkpatrick's level 2, distinguishing cognitive, skill-based and affective outcomes. For each of these they classified key variables and suggested how they could be measured. These range from basic skills to higher level compilation skills. In Carpenter (2005), Kraiger *et al.*'s model was applied to social work education, making reference to empirical studies of outcomes of social work education. This review is updated to include more recent studies (Table 3).

1. Cognitive Skills

Kraiger *et al.* (1993) classified *cognitive skills* as verbal (declarative) knowledge, knowledge organisation and cognitive strategies. For example, a student on an interviewing skills course with declarative knowledge should be able to define a concept such as 'active listening'. This outcome can be measured in written or multiple choice tests (e.g. Freeman and Morris, 1999).

Table 3 Knowledge, Skills, Attitudes and Behaviour: Measuring Learning Outcomes [from Carpenter (2005)]

Dimension		Measurement
Cognitive	Declarative (verbal knowledge)	MCQs; short exam questions
	Procedural (knowledge organisation)	Concept mapping; vignettes
	Strategic (planning, task judgement)	Reflections on simulated interviews with 'standardised clients'
Skills	Initial skill	(Self-ratings); observer ratings (scales)
	Compilation of skills	Observer ratings of DVDs of communication skills; self-rating of competences
	Advanced skills (automaticity)	Observation (e.g. of assessment interviews)
Affective	Attitudes to users; values	Attitude scales
	Motivational outcomes, self-efficacy	Self-ratings; confidence ratings
Behaviour	Implementation of learning (and barriers)	Self-report; practice teacher/manager report; rating scales
Impact	Outcomes for users and carers	User-defined scales; self-esteem & empowerment; measures of social functioning, mental health, quality of life, child behaviour etc.

The next level is '*procedural' knowledge*—that used in the performance of a task, as organised into a mental map; the more developed the knowledge, the more complex (interrelated) the mental map. This 'internalisation' of knowledge is conventionally assessed by academic essays, although this procedure is probably not reliable and is unrealistic as an outcome measure; students will not want to write the same assignment at the start and finish of a module.

'Concept mapping' in which students are asked to link a series of concepts in relation to a particular topic, shows promise and has recently been adapted/applied to the arena of social work education (Anghel *et al.*, 2010; Webber *et al.*, 2010). Students are first trained in the method and then, before the teaching and without the help of books or papers, are asked individually to draw a map of their existing knowledge. Visually, differences between maps drawn at T1 and T2 can be impressive. The maps can be scored in terms of both their structural and relational qualities (Anghel *et al.*, 2010).

Procedural knowledge may also be tested using written vignettes, specially designed outline case studies to which students are asked to respond. Trained raters can score their written answers, e.g. in relation to 'criticality' (MacIntyre *et al.*, 2011). Similarly, Lefevre (2010) rated responses using a taxonomy of core capabilities for effective communication with children and young people which had been derived from a systematic review.

Once knowledge has been internalised, we can think strategically about its use, a process known as 'metacognition'. *Strategic skills* include planning, monitoring and revising. An example of high level skills would be reflecting on the process of an interview with a family group so as to modify the worker's alliances with different family members and also think about the overall direction of the interview, while at the same time engaging (cognitively) in active listening with the person who is talking.

These are not easy to measure in direct practice, but Bogo *et al.* (2011) report that students' reflections on their practice in an objective structured clinical examination (OSCE) adapted for social work can be assessed reliably using structured probes and rating scales. This procedure has not yet been used to measure outcomes, but it does distinguish between students and experienced social workers. The potential is there and Lozano *et al.* (2010) have recently reported a small RCT of brief training in motivational interviewing for paediatricians.

2. Skills

Kraiger *et al.* (1993) posited three levels of skills: initial skill acquisition; skill compilation, or the grouping of skills into fluid behaviour; and, through practice, 'automaticity'. Automaticity enables you to accomplish a task without having to think about it consciously and to complete another task at the same time. A familiar example is learning to drive a car; at the third level you are able to talk to passengers while monitoring road conditions, changing gears and reacting to sudden hazards.

Nerdrum (1997) measured *initial skill acquisition*. Student social workers were invited to suggest helpful answers to 10 videotaped statements from 'simulated clients'. The students' written responses were then rated by researchers using a standardised measure of 'empathic understanding'.

More directly, observer rating of students' communication skills in simulated interviews offers a method for measuring initial and compilation skills. Cheung (1997) had both trainers and trainees assess the content of videotapes used in an interview protocol for social workers and police in child sex abuse investigations. Similarly, Freeman and Morris (1999) measured higher level compilation skills used by child protection workers in simulated interviews. They employed a coding system to assess the support and information provided by the trainee, as well as the more basic questioning skills. The measure used samples of interactions between interviewer and interviewee, although in this case, only the interviewer's behaviour was rated. Independent raters achieved substantial agreement (90%). Interestingly, although there were improvements in results on a knowledge questionnaire, there was little evidence of improvement in trainees' skills. The authors suggested this might have been a consequence of the artificiality of the simulated interviews as well as deficiencies in the training programme.

Evaluators have also measured *compilation skills* by asking students to self-rate their competence. For example, Wong and Lam (2005) developed the Competence and Aptitude in Social Work Scale (CASWS) to assess self-perception of skills and also knowledge and values. CASWS has good internal reliability and the skills and values subscale scores correlate significantly with course marks, demonstrating concurrent reliability. Similarly, Vitali (2011) has used the National Occupational Standards for social workers in England as the basis for a rating scale of competences which can be employed by both students and fieldwork supervisors.

Not surprisingly, the measurement of the highest level of skill development, *automaticity*, poses significant problems, even when attempting to assess apparently straightforward tasks such as computer programming. Possibly the best indication of automaticity in social work is when students appear, to a trained observer, to have stopped monitoring their own behaviour in the accomplishment of a high level task, or report less conscious awareness of their own actions. There were no examples of this in the literature searched.

3. Affective (Attitudinal) Outcomes

Affectively-based outcomes (Kraiger *et al.*, 1993) include attitudes (level 2a in Barr *et al.*'s framework) and values and commitment to organisational goals. These are conventionally measured by means of standardised self-rating scales. For example, Forrester *et al.* (2008) used the Alcohol and Alcohol Problems Perceptions Questionnaire to measure level 2a outcomes of a course in motivational interviewing. Similar Likert-type scales have been used to measure 2a and 2b outcomes of interprofessional education for social work and health care students (Pollard *et al.*, 2006).

Kraiger and colleagues also propose *motivational outcomes*, an example of which might be a greater determination to change one's own behaviour in response to learning about racism. Related to this is the idea of 'self-efficacy'; that is the (realistic) feeling of confidence that you have the ability to carry out a particular task in particular circumstances. This is particularly important in relation to difficult and/or complicated tasks, such as carrying out a complex child care assessment. Holden and colleagues (2002) have pioneered self-efficacy as an outcome measure in social work, developing self-rating scales measuring self-efficacy in three areas: hospital social work, research and evaluation. Holden *et al.* (2007) have reviewed recent studies and investigated 'response shift bias' by inviting students to make retrospective assessments of their baseline confidence, 'If I knew then what I know now'. The effect was not statistically significant in their 2007 study of evaluation self-efficacy, but it seems well worth incorporating this measure.

Self-efficacy scales are becoming quite popular because they can quite easily be adapted to the measurement of confidence in other tasks such as communication (Lefevre, 2010; Koprowska, 2010) and aspects of placement learning (Parker, 2006). However, further work is required to show test–retest reliability and the concept validity of the scales should be explored. Greater confidence in the use of this method would be provided by demonstrating that self-ratings were associated with independent ratings by supervisors. This cannot be assumed. For example, Fortune *et al.* (2005) found that students' 'achievement motivation', which included self-efficacy, was unrelated to field instructors' (practice educators') evaluation of students' skills and Vitali (2011) found no significant correlation between students' self-ratings and practice educators' ratings of performance in relation to social work competences. Further, Koprowska's (2010) comparison for two students of self-efficacy, user-rating and staff assessment indicated that relationships between these variables may be complex.

4. Changes in Behaviour

How can we know whether learning has been implemented? Most studies have relied on follow-up surveys using postal questionnaires or interviews, and in some cases both. Wehrmann *et al.* (2002) asked trainees to rate their acquisition and use of skills (as defined in the course learning outcomes) at the end of a course and again six months later. They reported a moderate reduction in mean ratings at follow-up, but without a baseline (pre-course) measure it is not possible to know whether the trainees had gained in skills overall.

The evaluation of changes in behaviour is most straightforward when there is clear evidence as to whether the trainee carried out the learned behaviour or not. For example, Bailey (2002) used a pre–post design to monitor changes in assessment for people with interrelated mental health and substance misuse needs. At the start of the course, trainees were asked to complete a proforma on the care they were providing to service users with whom they were currently working. They were subsequently asked to complete a similar proforma for the same clients a month after the training. Of interest was whether or not they had undertaken an assessment in the manner taught on the course; the existence of written assessments could therefore provide evidence of the effectiveness of the course in this respect. Platt (2011) provides a similar approach.

This approach would, however, be more difficult to apply with social work students. First, unlike practitioners, it may not be possible to measure a behavioural baseline if the programme design involves students being taught a skill and then going into practice placements. Second, it might be difficult to implement the method of working because of organisational constraints. For example, it would be possible to collect evidence that students were using task-centred casework as taught on the programme (e.g. written, signed contracts with service users' goals and tasks, tape recordings). However, a particular student's failure to implement the method may have more to do with the practice agency's function or management than any lack of learning on the part of the student (Luger, 2011).

5. Impact: Outcomes for Service Users and Carers

Barnes *et al.* (2000) reported that when mental health service users were asked to define the desired outcomes of training, they focused on level 2 outcomes, changes in attitudes, knowledge and skills. The researchers described the development of a questionnaire to determine user-defined outcomes of post-qualifying education in mental health. The questionnaire may be used in confidential postal surveys or structured interviews with an independent researcher. Barnes *et al.* (2006) used it to compare users' ratings of trainees with a comparison group of professionals who had not received training.

From a professional perspective, outcomes for service users and carers are generally considered in terms of changes in such factors as the quality of life, skills and behaviour, self-esteem and levels of stress (level 4). Standardised instruments have

been developed to assess these factors and may be used in assessing the outcomes of education. However the study mentioned above is the only one identified in the literature search to have done so, along with level 2a and 2b outcomes. It used measures of mental health, life skills and social functioning and psychiatric symptoms of users, demonstrating positive changes overall (Carpenter *et al.*, 2006).

Research Designs

Having identified outcomes and measures, the next challenge is to develop experimental or quasi-experimental designs which are feasible to employ in the evaluation of social work education. Potential designs are summarised in Table 4.

1. Post-test Only

The most common form of evaluation in social work education (and across most of higher education in the UK at least) is almost certainly the completion of feedback questionnaires at the end of the course or programme. Such a 'post-only' design is useful as formative feedback to the educators. However, in the absence of baseline data it cannot tell us about the *outcomes*. The post-only design is therefore inadequate for outcomes evaluation and arguably should be discouraged because it is not that much more difficult to obtain more informative data.

2. Pre-test, Post-test

This is the 'entry level' design and the one used in three-quarters of the studies found. In practice this means asking students to complete the same questionnaire, respond to the same vignette or take part in a similar interview, on two occasions. This is not particularly difficult once a suitable measure of outcome has been chosen.

Table 4 Possible Research Designs for Assessing Outcomes [from Carpenter (2005)]

	Design	Comment
1	Follow up (post-test): single group	Useful as formative feedback to the trainers, but cannot inform outcomes
2	'Before and after': single group	Quite commonly used, but outcomes cannot be ascribed exclusively to the training intervention
3	Post-test: two groups	Requires random allocation of students to different conditions
4	Pre-test, post-test: two groups	Natural comparison groups can be used. Random allocation preferable
5	Repeated measures, two groups	Students randomly assigned to two groups, both of which get the intervention at different times, but requires three measurements
6	Times series: one group	Requires multiple, unobtrusive observations rather than formal tests
7	Single-subject experiments	Requires repeated measures of the same person before, during and after the intervention. Small samples

Persuading students to complete measures on two occasions does not seem to be difficult if the baseline measure (T1) is presented as part of the introduction to the module. Knowledge tests and attitude questionnaires are simple to administer in this way. Students are generally willing to complete the measure again at T2 as part of a course review, so long as there is time to do this in the final session itself and students hand in their forms before leaving the session. For example, Carpenter and Hewstone (1996) measured changes in social work and medical students' interprofessional stereotypes over a short IPE programme. This latter study also asked students to rate how interesting and useful they had found the course and, because it had also asked about the students' expectations at T1, the evaluators were able to put the feedback into perspective. For instance, they concluded that some participants' experiences were quite positive given their low expectations; further, because there were two distinct groups of participants, two-factor analysis of variance could be employed to distinguish the differential effects of the programme on the medical and social work students.

The limitation of simple pre–post designs, however, is that any changes observed cannot be ascribed exclusively to the training intervention. For example, improvements may in part be attributed to 'maturational effects', i.e. a general increase in knowledge, skills and improvement in attitudes as a consequence of being engaged in professional training. Second, changes may also be attributed to 'Hawthorne' effects, that is improvements resulting from involvement in an intervention in its own right, irrespective of its content—a simple response to 'something happening'. Third, they may be an artefact of practice in completing the measures; for example thinking more about the answers to questions about knowledge or appreciating the socially desirable responses in attitude scales. Consequently, it is preferable to introduce a control group, as will be discussed later; but first we should consider one design which eliminates the effects of maturation or practice.

3. Post-test, Two Groups Design

Here, participants are randomly allocated to one of two groups and assessed only at the end of the course. This design may be used to compare two different approaches to learning and teaching using any of the outcome measures outlined above. The assumption is that, provided the cohort size is large enough, initial randomisation evens out differences (for example in initial knowledge, learning ability and motivation) between participants in the two groups. Consequently the difference in the average T2 scores on whatever measure is used gives an indication of which of the two methods is superior. Because the measure is used only once, there is no opportunity for contamination by practice or maturation. Of course, because there is no baseline, what this method is unable to tell us is how much the participants have actually learned. Although this is actually the simplest of all experimental designs, it did not appear in the literature reviewed for this paper. In practice, it would mean persuading a cohort of students to agree to being randomly allocated to one of two methods of learning a topic or skill. This would be easier to achieve if the topic was not

formally assessed, which would avoid students' anxiety about being at a disadvantage in relation to their colleagues.

4. Pre-test, Post-test: Two Groups

To be confident that differences between T1 and T2 scores may be attributed to the educational intervention, we need a no-intervention control group. This is equivalent in design terms to making a comparison between two different methods of intervention. Ideally, the participants are selected randomly for one of the two interventions, as in the post-test, two group design described above; this is a 'true' experiment. However, 'quasi-experimental' designs are generally easier to accomplish, especially in the context of higher education. Here, seemingly equivalent groups which experience different interventions are compared in terms of outcomes.

Nerdrum (1997) compared approaches to training in empathy skills by examining outcomes for students on different social work degree programmes in Norway. This approach is potentially very useful for two main reasons. First, it eliminates the problem of some students on a particular programme feeling that they are getting a 'worse' educational intervention than colleagues on the same programme; all get the same intervention. Second, the sample sizes generated for statistical analysis will be much larger if two, or more, whole programmes participate than if one programme's students are divided into two groups. This increases the possibility that the analysis will have sufficient statistical power to detect differences between the two interventions.

There are a number of aspects of comparative, between-programme, evaluations which must be taken into account. The first, and most obvious, is that the student groups must be as closely matched in terms of prior education and social work experience. It is probably safest to consider comparison groups as being 'non-equivalent' and to test statistically for socio-demographic differences such as age. There may be differences in the baseline scores on the outcome variables (test scores), i.e. one group performing better on average than the other at T1. In this case, it is possible to adjust statistically for these differences by using analysis of covariance (ANCOVA).

Another factor in cross-programme comparisons is the timing of the interventions; comparing first and second year students would not be useful. A factor that is more difficult to standardise for is the teacher: some educators are more inspirational or technically proficient than others. It would therefore be relevant to ask students to rate the trainer's competence; if there was an apparent difference in the ratings made by the students on different programmes, the potential impact on learning outcomes could be checked.

On a degree programme it would not be possible to have a 'non-intervention' group. However, there are examples of comparisons of outcomes between people who received the training and an equivalent group who did not. Thus, Sharples *et al.* (2003) reported the evaluation of an 'introduction to management' course which employed a control group of staff from the same local authority who were matched for job

responsibility and gender. Unfortunately, from a research perspective, the two groups were found to differ in some important respects because newly appointed managers had apparently been given priority in selection for the course.

There were no examples of randomised controlled trials of educational interventions in social work education and there are very few examples in health and medical education. Again, difficulty with the assignment of students to groups is probably an important factor.

However, an instructive example of an approach which could be used in social work education was published by Cook *et al.* (1995). They evaluated a two-day programme on community development for mental health professionals in the US. They used a variation on the pre-test, post-test two groups design to assess the trainees' attitudes towards people with mental illness in the roles of service recipient, service deliverer and trainer. Trainees received the same training on the first day, delivered by someone who was not a user of mental health services. On the second day the trainees were randomly assigned to receive training from either a service user or a trainer who did not have direct experience of using mental health services. There is no mention of the trainees' response to being assigned in this way.

Trainees completed two attitude questionnaires before the first day of training and again at the end of the programme. The authors reported that compared to those who had been trained by a non-user, those who were trained by the user trainer expressed more positive attitudes towards people with mental illness overall. However, this study, although strong in terms of experimental design, illustrates some of the problems of interpretation of findings. The positive change in attitudes reported could be due to the trainees having a different trainer on the second day of the programme. Alternatively, it could be due to some other personal characteristic of the trainer, as opposed to their status simply as a user of mental health services; thus, the generalisations that can be drawn from the study are limited.

Waiting list controls with repeated measures

If it is unreasonable or impossible to deny training to a control group, a 'waiting list' control may be acceptable. Here, all participants are given a baseline assessment (T1) and then divided at random into two groups (Table 5). Group 1 receives the intervention, for example a three-week intensive module in communication and interviewing skills, and all members are reassessed at the end (T2). Group 2, the (waiting list) controls then start their course and are assessed at the beginning (T2) and the end (T3). Group 1 students are also re-assessed at T3. If the training was

Table 5 Example of a 'Waiting List' Controlled Design for the Evaluation of a Training Intervention [from Carpenter (2005)]

	T1 (baseline)	3 weeks	T2	3 weeks	T3
Group 1	Assessment	Training intervention	Assessment	Other studies	Assessment
Group 2	Assessment	Other studies	Assessment	Training intervention	Assessment

successful, we would expect a greater improvement in mean scores between T1 and T2 for Group 1 than for Group 2 (we might anticipate some improvement in Group 2 because of practice effects and other generalised learning on the programme). However, we would expect a greater increase in mean scores in Group 2 between T2 and T3, i.e. while they were receiving the training, than for Group 1 (although once again we might anticipate a further small improvement in this group on account of continued non-specific learning). If these assumptions proved correct, we could reasonably conclude that there was consistent evidence of improvement associated with the training.

A more manageable approach in terms of measures would employ 'counter balancing'. Here the participants would be randomised into two groups. Group 1 would be trained in Method A, e.g. brief solution-focused interventions while at the same time Group 2 would learn Method B, e.g. groupwork. Then, the groups would cross over, with Group 2 learning Method A and Group 1 learning Method B. Both groups would be assessed at the end of each module. This approach aims, through randomisation, to deal with 'order effects', associated with practice and generalisation of learning on the one hand and fatigue on the other. Despite the advantages of this design, it does not appear in the literature, possibly because it appears complex and statistical analysis demands slightly more sophisticated analysis of variance techniques. However, training small groups of students in intervention methods seems to be quite common on social work programmes and there may be potential for the use of this design.

5. Time Series Designs

Time series designs do not require a control group. Instead they need multiple measures of the group members who receive the educational intervention, including multiple baseline measures. Conclusions about the effects of the intervention are based on an analysis of trends before, during and after the intervention. In more sophisticated designs, the intervention is withdrawn and subsequently reintroduced and the effects noted. They are often employed in the assessment of interventions for people with severe learning disabilities, but there is no reason why they should not also be used in the evaluation of students trained to use these methods. The focus in this case would be on the students' behaviour rather than that of the service user. Note that in this approach, the numbers of participants need only be small. The argument about the effect of a training intervention relies on the repeated demonstration of the same pattern in a small number of individual case studies rather than the aggregated measure of a group. There were no examples of these designs in the literature.

A variant on this approach is the double baseline with follow up group design (e.g. Carpenter *et al.*, 2011). Here participants complete the measures on four occasions: T0, at course registration (some weeks in advance); the start (T1) and end (T2) of the course; and some weeks later (T3). This design tests the hypothesis that, in the absence of any training, there would be no change in mean total ratings between T0 and T1; at

the end of the course (T2) any change in mean scores compared to T1 could be attributed to the training; finally, T3 ratings would indicate whether or not learning had been sustained.

Practical Considerations

This section reviews some practical considerations, including how to engage students and service users, how evaluation might be linked to assessment and the potential for collaboration between programmes. Some of these issues are elaborated in Burgess and Carpenter (2010).

1. Engaging Students

It is both desirable in practice and essential for ethical reasons to engage students in the systematic evaluation of their own learning. The first step can be to involve student representatives with staff in an advisory group, where they can be engaged in discussions about the desirability and feasibility of the various approaches under consideration. They will want to consider the implications for their time and their learning and to be confident that the findings will be used appropriately. Thus, they might be engaged if they considered that the findings would be reported carefully to them as individuals as well as to the group and that they could use the information to monitor their own performance and learn how to improve. Similarly, students could be interested in the systematic collection of self-report data for use in the portfolios which are required on many programmes. These could include ratings of self-efficacy in a range of relevant practice skills.

Test results, e.g. MCQs, concept mapping scores and observer ratings of communication skills DVDs, could all be used in this way. In some cases students could score these test themselves, or if they completed a test (e.g. a case study simulation or knowledge test) on a computer, the programme could generate an automatic individual score while feeding an anonymous score to a database for group analysis. If the results are to be used in summative as well as formative assessment, the measures could be included in the range of assessment methods which feed into the overall score for a course module.

The experience from the OSWE project was that response rates were high when students were *required* to participate in the measurement of their performance for formative and/or summative assessment purposes and given the opportunity of *opting out* of the evaluation (see Vitali, 2011).

2. Involving Service Users and Carers

Many of the same arguments may be applied to the involvement of service users and carers in the design of systematic evaluations (Fox and Ockwell, 2010). Discussion can take place with users and carer consultants to individual programmes (Dowson *et al.*, 2010); this may come about through the planned review of a course module and/or practice placements. They can advise on the appropriateness and feasibility of

proposed measures and research designs. In addition, their contribution to the interpretation of findings, particularly from practice, can be invaluable.

3. Engaging Staff

Much course evaluation already happens in universities; the problem is that it is too restricted in scope and design so that the information it provides is limited. The increasing prevalence in the UK in particular of quality assurance templates which collect end of course satisfaction ratings only, and fail to measure outcomes, is an impediment. Faculty committees need to be persuaded that more effective and useful evaluations can be employed.

Evaluation should be built into the design and re-design of modules and programmes and seen as part of students' learning. The systematic collection of test data becomes part of assessment as well as providing evidence for the evaluation of the module for the university. The assessment of learning outcomes can provide not only case studies for research methods training but also, fundamentally, an enthusiasm for the collection of data and its analysis within the general framework of evidence-based practice.

To put this another way, the collection and analysis of course evaluation data could be more interesting, and less of a chore, if programme staff: (1) used a wider range of measures which could provide information about the learning outcomes in which they were interested; and (2) attempted some of the research designs suggested here, so that they could have evidence to see whether the learning outcomes had been achieved.

As a first step the learning outcomes for course modules may be reviewed and measurement discussed. Since these are often poorly specified, this exercise would be valuable in itself: if they are to be measured they will have to be revised in order to be observable. This process should improve the alignment of learning and teaching methods with learning outcomes. The next step would be to assess these in a pre-test, post-test single group design. This has limitations, as noted above, but it gives an indication of whether the outcomes are being achieved. The third step would be to use some of the stronger designs discussed above.

4. Collaboration Between Programmes

As suggested there are significant advantages in programmes collaborating in evaluation studies. This allows the comparison of different methods of teaching and learning and the creation of experimental and control/comparison groups which would be large enough to generate sufficient statistical power to detect significant differences between conditions. However, there is a further reason for collaboration: since we know so little about *how* these methods of evaluation might be put into practice, it is only through enthusiasts sharing ideas and experiences that we could hope to make progress.

Conclusion

The literature search showed that rigorous evaluation of the outcomes of social work education is still at an early stage of development; social work in certainly not unique in this respect. Few evaluations have been published and most of these have focused on changes in attitudes and/or knowledge and skills only. There is, as yet, little evidence of changes in behaviour as a result of social work education and only one study which can claim to show a positive impact on the lives of service users. There has been progress in developing outcome measures and some of these, particularly self-efficacy ratings, are being used with increasing sophistication. Nevertheless, there is much still to be done to establish the reliability and concurrent validity of self-report measures.

Published research is heavily dependent on pre–post designs, which are limited because observed outcomes cannot be ascribed exclusively to the education intervention. There are very few comparative studies and none involving controls. This is an essential area for development, which could be achieved by collaboration between universities and evaluators.

Finally, one of the lessons from the OSWE project (Burgess and Carpenter, 2008) was that this kind of research is challenging, even when there was the expressed support of senior staff and some funding. The lack of time and resources was a significant obstacle but one which could be overcome with support from colleagues within and without the institution. Nevertheless, if this line of research is to develop—as it should, given the substantial investment made in education and training in social work—the commitment of research funders is required. At a time of financial constraint, it is all the more important that evidence about the outcomes of social work education is accumulated so that the public and policy makers can be confident that their investment is producing high quality social workers.

Acknowledgement

Thanks to Verity Rees, MSc Social Work student at the University of Bristol for her help with the literature search.

Note

[1] Now Institute for Research and Innovation in Social Services (IRISS).

References

Anghel, R., Fox, J. & Warnes, M. (2010) 'An exploration of concept mapping as a method of evaluating student learning in social work', in *The Outcomes of Social Work Education: Developing Evaluation Methods*, eds H. Burgess & J. Carpenter, Higher Education Academy SWAP, Southampton, pp. 34–50.

Bailey, D. (2002) 'Training together—part two: the evaluation of a shared learning programme on dual diagnosis for specialist drugs workers and Approved Social Workers (ASWs)', *Social Work Education*, vol. 21, pp. 685–699.

Barnes, D., Carpenter, J. & Bailey, D. (2000) 'Partnerships with service users in interprofessional education for community mental health: a case study', *Journal of Interprofessional Care*, vol. 14, pp. 189–200.

Barnes, D., Carpenter, J. & Dickinson, C. (2006) 'The outcomes of partnerships with mental health service users in interprofessional education: a case study', *Health and Social Care in the Community*, vol. 14, pp. 426–435.

Barr, H., Freeth, D., Hammick, M., Koppel, I. & Reeves, S. (2000) *Evaluating Interprofessional Education: A United Kingdom Review for Health and Social Care*, BERA/CAIPE.

Bogo, M., Regehr, C., Katz, E., Logie, C. & Mylopoulos, M. (2011) 'Developing a tool for assessing students' reflections on their practice', *Social Work Education*, vol. 30, no. 2, pp. 186–194.

Burgess, H. & Carpenter, J. (2008) 'Building capacity and capability for evaluating the Outcomes of Social Work Education (the OSWE project): creating a culture change', *Social Work Education*, vol. 27, no. 8, pp. 898–912.

Burgess, H. & Carpenter, J. (eds) (2010) *The Outcomes of Social Work Education: Developing Evaluation Methods*, Higher Education Academy SWAP, Southampton.

Carpenter, J. (2005) *Evaluating Outcomes in Social Work Education,* Scottish Institute for Excellence in Social Work Education (SIESWE) and the Social Care Institute for Excellence (SCIE), Dundee [online]. Available at: http://www.scie.org.uk/publications/misc/evalreport.pdf, accessed 11 January 2011.

Carpenter, J. & Hewstone, M. (1996) 'Shared learning for doctors and social workers. Evaluation of a programme', *British Journal of Social Work*, vol. 26, pp. 239–257.

Carpenter, J., Barnes, D. & Dickinson, C. (2006) 'Outcomes of interprofessional education for Community Mental Health Services in England: the longitudinal evaluation of a postgraduate programme', *Journal of Interprofessional Care*, vol. 20, pp. 145–161.

Carpenter, J., Patsios, D., Szilassy, E. & Hackett, S. (2011) 'Outcomes of short course interprofessional education in parental mental illness and child protection: self-efficacy, attitudes and knowledge', *Social Work Education*, vol. 30, no. 2, pp. 195–206.

Cheung, K. M. (1997) 'Developing the interview protocol for video-recorded child sexual abuse investigations: a training experience with police officers, social workers and clinical psychologists in Hong Kong', *Child Abuse and Neglect*, vol. 21, pp. 273–284.

Cook, J., Jonnikas, J. & Razzano, L. (1995) 'A randomized evaluation of consumer versus nonconsumer training of state mental health service providers', *Community Mental Health Journal*, vol. 31, no. 3, pp. 229–238.

Dowson, E., Gee, M., Ingram, R., Leeson, B. & Mackenzie, G. (2010) 'Service user, carer, practice educator and academic partnership in evaluating decision making by social work students', in *The Outcomes of Social Work Education: Developing Evaluation Methods*, eds H. Burgess & J. Carpenter, Higher Education Academy SWAP, Southampton, pp. 88–95.

Forrester, D., McCambridge, J., Waissbein, C., Emlyn-Jones, R. & Rollnick, S. (2008) 'Child risk and parental resistance: can motivational interviewing improve the practice of child and family social workers in working with parental alcohol misuse?', *British Journal of Social Work*, vol. 38, pp. 1302–1319.

Fortune, A., Lee, M. & Cavasoz, A. (2005) 'Achievement motivation and outcome in social work field education', *Journal of Social Work Education*, vol. 41, pp. 115–129.

Fox, J. & Ockwell, C. (2010) 'Service user and carer involvement in the OSWE Project in England: the experiences of two service user researchers', in *The Outcomes of Social Work Education: Developing Evaluation Methods*, eds H. Burgess & J. Carpenter, Higher Education Academy SWAP, Southampton, pp. 118–126.

Freeman, K. A. & Morris, T. L. (1999) 'Investigative interviewing with children: evaluation of the effectiveness of a training program for child protective service workers', *Child Abuse and Neglect*, vol. 23, pp. 701–713.

Holden, G., Barker, K., Rosenberg, G. & Onghena, P. (2007) 'Assessing progress towards accreditation-related objectives: self-efficacy as an outcome', *Research on Social Work Practice*, vol. 17, pp. 456–465.

Holden, G., Meenaghan, T., Anastas, J. & Metray, G. (2002) 'Outcomes of social work education: the case for social work self-efficacy', *Journal of Social Work Education*, vol. 28, pp. 115–133.

Kirkpatrick, D. L. (1967) 'Evaluation of training', in *Training and Development Handbook*, eds R. L. Craig & L. R. Bittel, McGraw-Hill, New York, pp. 87–112.

Koprowska, J. (2010) 'Are student social workers' communication skills improved by university-based learning?', in *The Outcomes of Social Work Education: Developing Evaluation Methods*, eds H. Burgess & J. Carpenter, Higher Education Academy SWAP, Southampton, pp. 73–87.

Kraiger, K., Ford, K. & Salas, E. (1993) 'Application of cognitive, skill-based and affective theories of learning outcomes to new methods of training evaluation', *Journal of Applied Psychology*, vol. 78, pp. 311–328.

Lefevre, M. (2010) 'Evaluating the teaching and learning of communication skills for use with children and young people', in *The Outcomes of Social Work Education: Developing Evaluation Methods*, eds H. Burgess & J. Carpenter, Higher Education Academy SWAP, Southampton, pp. 96–110.

Lozano, P., McPhillips, H., Hartzler, B., Robertson, A., Runkle, C., Scholz, K. A., Stout, J. W. & Kieckhefer, G. M. (2010) 'Randomized trial of teaching brief motivational interviewing to pediatric trainees to promote healthy behaviors in families', *Archives of Pediatrics & Adolescent Medicine*, vol. 164, no. 6, pp. 561–566.

Luger, L. (2011) 'Enhancing cultural competence in staff working with people with drug and alcohol problems—a multidimensional approach to evaluating the impact of education', *Social Work Education*, vol. 30, no. 2, pp. 223–235.

MacIntyre, G., Green Lister, P., Orme, J., Crisp, B.R., Manthorpe, J., Hussein, S., Moriarty, J., Stevens, M. & Sharpe, E. (2011) 'Using vignettes to evaluate the outcomes of student learning: data from the evaluation of the new social work degree in England', *Social Work Education*, vol. 30, no. 2, pp. 207–222.

Nerdrum, P. (1997) 'Maintenance of the effect of training in communication skills: a controlled follow-up study of level of communicated empathy', *British Journal of Social Work*, vol. 27, pp. 705–722.

Orme, J., MacIntyre, G., Green Lister, P. *et al.* (2009) 'What (a) difference a degree makes: the evaluation of the new social work degree in England', *British Journal of Social Work*, vol. 39, pp. 161–178.

Parker, J. (2006) 'Developing perceptions of competence in practice learning', *British Journal of Social Work*, vol. 36, pp. 1017–1036.

Platt, D. (2011) 'Assessments of children and families: learning and teaching the skills of analysis', *Social Work Education*, vol. 30, no. 2, pp. 157–169.

Pollard, K., Miers, M., Gilchrist, M. & Sayers, A. (2006) 'A comparison of interprofessional perceptions and working relationships among health and social care students: the results of a 3-year intervention', *Health and Social Care in the Community*, vol. 14, pp. 541–552.

Sharples, A., Galvin, K., Holloway, I. & Brown, K. (2003) 'The impact of training: an evaluation of an "introduction to management" course for social services staff', *Learning in Health and Social Care*, vol. 2, pp. 37–50.

Vitali, S. (2011) 'The acquisition of professional social work competencies', *Social Work Education*, vol. 30, no. 2, pp. 236–246.

Webber, M., Currin, L., Groves, N., Hay, D. & Fernando, N. (2010) 'Social workers *can* e-learn: evaluation of a pilot post-qualifying e-learning course in research methods and critical appraisal skills for social workers', *Social Work Education*, vol. 29, pp. 48–56.

Wehrmann, K., Shin, H. & Poertner, J. (2002) 'Transfer of training: an evaluation study', *Journal of Health and Social Policy*, vol. 15, pp. 23–37.

Wong, D. & Lam, D. (2005) 'Problem-based learning in the BSW programme; a study of learning outcomes', in *Problem-based Learning in a Social Work Context*, ed. V. Pearson, Monograph No. 53, Department of Social Work and Social Administration, University of Hong Kong.