ENETOSH
European Network Education and Training in Occupational Safety and Health
FACTS 04 2018

Evaluation of quality of academic safety, health and environment education

Paul Swuste, Safety Science Group, Delft University of Technology, The Netherlands
Frank van Dijk, Learning and Developing Occupational Health Foundation (LDOH), ICOH Scientific Committee Education and Training in Occupational Health (SCETOH)

Quality of education and training in safety and health is a difficult topic. In practice, we are used to performing some evaluations. In contrast, there are hardly any scientific evaluation studies, presumably due to a lack of tradition and financial constraints. Another reason may be a lack of consensus on what to evaluate (Alliger & Janak, 1989; Dijk et al., 2015; Hale, 1984; Heath, 1982; Kennedy, Chyung, Winiecki & Brinkerhoff, 2013; Mann, 1996). Quality of education in occupational safety and health can be viewed from different perspectives – from the view of participants, from the perspective of managing the course, from companies where participants are working, from the government, etc. Quality is a relative concept, its operationalisation is dependent on the interested party.

Quality

A possible definition of quality of education can be derived from a definition on quality of healthcare: 'Quality of safety and health education is the degree to which organisations providing these trainings and educational courses will increase the likelihood that desired educational goals are reached, and are consistent with current professional and academic knowledge' (IOM, 2001). This definition implies that educational goals, 'learning objectives' or 'learning outcomes' should be set beforehand.

Quality assessment

Almost sixty years ago, Donald Kirkpatrick from the Wisconsin University, US and co-workers published a paper on assessing the quality of training courses (Catalanello & Kirkpatrick, 1968; Kirkpatrick, 1959a, 1959b, 1960a, 1960b). Generally, training courses cover a timespan of days or weeks. His assessment is also applied to education, covering a much longer period. In this factsheet, the term education will be used, which also includes training. The literature still refers to Kirkpatrick's levels, because they are simple and easy to understand (Liebermann & Hoffmann, 2008):

Level 1 Reaction: Do trainees like the programme? The trainees' evaluation is based on the assumption that a satisfied student will learn more and better than one who is not satisfied. Most educational programmes use this perspective for their course evaluation (Bollmann, Gründler & Holder, 2018). A 2011 survey on post academic education in safety and health in Europe supports this conclusion (Arezes & Swuste, 2012). The limitations of this tool are clear: students lack the necessary understanding and will primarily judge the form of the programme, rather than its content. Measuring the reaction of trainees does not evaluate learning (Heinrich, 1956; Kirkpatrick, 1959). Some teachers communicate very well, without offering much content or even teaching unreliable content.

Level 2 Learning: Do trainees understand the facts, principles, theories, models and approaches presented? Classroom activities as individual performance, quizzes, discussions and written tests are evaluation techniques to assess actual learning. Many programmes have some sort of examination, either at the end of the programme or several times during the programme. Most examinations test knowledge. In some examinations or evaluations, skills and attitudes are assessed. In occupational medicine, skills in medical interviewing (e.g. occupational history) and in physical examinations can be tested. The attitudes of medical students towards the specialist field of occupational medicine can be evaluated before and after education with a validated questionnaire. A complication is that evaluation tools are, in practice, mostly restricted to 'internal tools' which only monitor the reactions of trainees and individual teachers.

Level 3 Behaviour: Do students apply models, tools and approaches of the programme in their jobs? An evaluation may include a pre and post education survey, preferably some time after finishing education, say six months. Not many organisations organise such an evaluation.

Level 4 Results or impact: Are workers, companies or organisations safer or healthier as a result of activities of the students or postgraduates who successfully finished their education? Such an evaluation implies one or more measurements of

safety or health. Accident, and incident frequencies are used as indicators for safety. Good examples for education at the level of workers are the studies from Yu et al. (2017) and of Chatterjee & Agrawal (2017). Of course, studies should be aware of biased safety outcomes (Kirkpatrick, 1977). Only using accidents as an indicator can be unreliable as this indicator is subject to all sorts of variations. Accident processes, or more specifically accident scenarios, and the quality of measures to prevent accidents might be better indicators. Another example is the incidence of occupational and work-related diseases before and after education, e.g. skin diseases, musculoskeletal diseases or burnout. It is also the case here that an evaluation which takes into account processes, systems and culture has to be considered for a more sustainable solution.

Kirkpatricks' levels are mainly output and outcome oriented. They lack a quality evaluation of the content and processes offered in the course. These aspects are highlighted more by Donabedian (1966) from the University of Michigan, US, with his input-process-output-outcome scheme. Here, the input or infrastructure refers to the state-of-the-art of the knowledge provided by the course and the quality of the teachers. A course organiser should have an overview of the subject areas that are taught in order to adequately select teachers who have up-to-date knowledge of their subject areas and can provide them with feedback on their teaching. Donabedian also addresses the process and the immediate outputs of the process: (a) the relevance and quality of the selected educational activities and learning materials: do these conform with the learning objectives, are they complete and valid, and (b) the quality of the teaching itself, such as interactive learning and learning by doing. Are all participants involved in active learning?

Transfer of education

The logic of Kirkpatrick's levels has been questioned. A positive reaction of trainees does not include an evaluation of learning in the sense that the trainees have understood the principles, models, essential facts, theories and techniques which have been taught (Kirkpatrick, 1959b; Mann, 1996). Therefore, the second and third flow charts in Figure 1 do not have an arrow between reaction and learning.

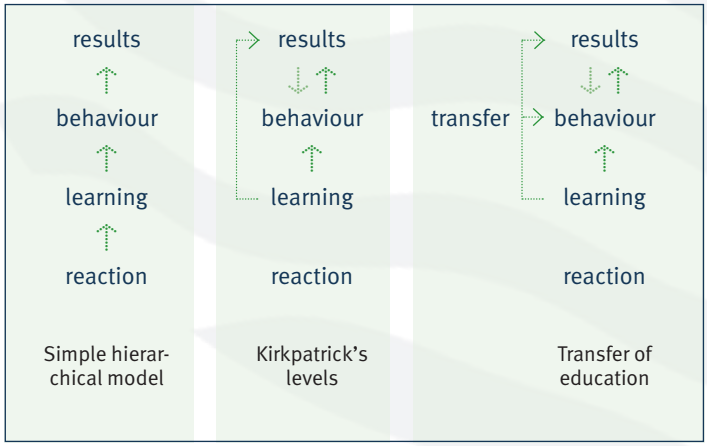

Figure 1 Educational models (Alliger & Janak, 1989; Mann, 1996)

However, the relationship between learning and behaviour on the job is not obvious. Therefore, the literature from the 1990s onwards has placed more emphasis on the transfer of education. Transfer of education can be evaluated as the degree to which trainees effectively apply the knowledge, skills and attitudes gained in education to their jobs. In order for this transfer to happen, the trainee has to feel a need to improve, as well as recognise his or her weaknesses. Supporting factors at the workplace of the trainees include working in an encouraging climate, receiving help from someone interested and skilled, and the opportunity to try out new ideas (Kirkpatrick, 1960a).

Transfer at different levels of education

Learned behaviour should be in accordance with actual job conditions of the participant. Therefore, education should be connected to the practical settings of the trainees, including teaching awareness about the conditions needed for acceptance of interventions, and accounting for possible resistance to change. Incorporating the working environment into education, or vice versa, has proven to be effective (Swuste & Arnoldy, 2003). Unfortunately, much of the education fails to transfer to job settings (Liebermann & Hoffmann, 2008; Mann, 1996). Transfer of learning must be evaluated by assessing if the educational goals or learning objectives have been met or not.

A well-known problem in academic undergraduate student education is the lack of actual job experience. Various techniques are used to compensate for this such as role playing, using representative situations in the daily life of the students, virtual reality, site visits and internships. For secondary and higher education in safety and health, but also in academic education, the focus of the transfer process can be more on practical aspects.

Transfer in (post) academic education differs from non-academic higher education in safety and health, due to its goal to teach trainees not only 'facts' but also critical reflection. An example is the Dutch post-graduate course 'Management of Safety, Health and Environment' (MoSHE). The vision is that a postgraduate safety and health expert is a direct advisor of the chief executive officer (CEO) of a company or organisation. He or she should provide functional leadership to risk management of SHE (safety, health and environment) processes, implement proactive SHE management with colleagues, and be responsible for the quality of SHE advice, having access to relevant reliable SHE expertise and sources. He or she should be independent, understand cross-border influences, and be able to analyse problems and provide solutions to new situations. Critical reflection implies a willingness to discuss divergent points of view on the topic concerned. This requires an overview of models, metaphors and theories of safety science in order to be able to analyse problems encountered at a meta-level. The overview of safety science is tested during homework assignments and in the final examination (Swuste & Sillem, 2018).

Evaluation of outcomes

Behaviour and results (impact) are interdependent since people tend to continue behaviours that are perceived to be effective even when this is not the case (Alliger & Janak, 1989). Evaluating the impact is difficult, and sometimes even impossible. For example, a pre-post study design such as a comparison of safety records one year before and one year after the education may show a decrease in figures. A causal relationship between the education and accident figures remains questionable, due to statistical variability and different forms of bias. In an interrupted time series design, a series of measurements is performed before and after, followed by a trend analysis (Schelvis et al., 2015). Another possibility to evaluate results is to focus on working relationships between middle managers and front line workers (Kirkpatrick, 1960b, 1977, 1978). There are also some comments on the levels proposed. Many evaluation studies that have evaluated education using Kirkpatrick's levels have reported a different effect on different levels. Because of difficulties in assessing levels three and four, often due to organisational disinterest of the organisation in which course participants work, evaluation of education remains mostly limited to the first two levels of Kirkpatrick's levels (Kennedy et al., 2013). On the other hand, Kirkpatrick's model may never have been meant to be more than a first, global heuristic for education evaluation. As such, it has served this purpose well (Alliger & Janak, 1989).

Evaluation in daily practice

The evaluation of a course in daily practice and a scientific evaluation have much in common but are not the same. Evaluation in daily practice aims to evaluate one particular course and to improve that. The aim of a scientific evaluation is often to evaluate only a few aspects of education, but for a wider audience and application. Conceptually, however, there are not many differences. In educational practice and in scientific studies, clear and realistic aims and learning objectives are always needed for an evaluation. Kirkpatrick's levels are a valid point of reference in every evaluation showing what can be evaluated[1]. Donabedian offers complementary topics such as the quality of the teachers and the learning infrastructure, the quality of the content and of the teaching and assessment methods. We recommend course organisers widen their scope in evaluation including, where possible, the evaluation of behaviour and final impact. Due attention is needed to evaluate the quality of input or infrastructure and to evaluate educational processes.

Scientific evaluation

In addition to these considerations, a scientific evaluation has to meet quite a number of other crucial requirements.[2] Scientific evaluation studies require a great deal of effort but have two major advantages: (1) in general, the reliability of the results and conclusions is much better than in-practice evaluations, and (2) it is possible to disseminate the results. When a study is indexed in a common literature database, e.g. PubMed, experts all over the world can easily find and use the results. Studies with a low number of participants can still be included in a systematic review or meta-analysis.

We recommend both: better evaluations in practice and more and better scientific evaluation studies.

1 the students' experience, (2) knowledge, skills and attitude, (3) behaviour in practice and (4) the final impact such as for workers, a company or society.

2 A scientific publication is mostly more focused on one aspect of education and starts with a background description showing publications regarding the topic studied. Next, the aim of the study, and the study objectives or hypotheses have to be clarified. A study design, population size (power of the study) and methods have to be selected. Subsequently, results are presented, often tested on significance, and discussed in comparison with literature findings. Strengths and weaknesses of the study are considered. Finally conclusions and recommendations are formulated, and a list of references is added.

References

Alliger, G., & Janak, E. (1989). Kirkpatricks levels of training criteria. *Personnel Psychology 42(4)*, 331-342.

Arezes, P., & Swuste, P. (2012). Occupational Health and Safety post-graduation courses in Europe: A general overview. *Safety Science 50*, 433-442.

Bollmann, U., Gründler, R., & Holder, M. (2018). *Integrating of safety and health into education. An empirical study of good-practice examples on www.enetosh.net.* IAG Report 1/2018e. Berlin: German Social Accident Insurance (DGUV).

Catalanello, R.F., & Kirkpatrick, D.L. (1968). Evaluating training programs, the state of the art. *Training and Development Journal 22(5)*, 2-9.

Chatterjee, S., & Agrawal, D. (2017). Primary prevention of ocular injury in agricultural workers with safety eyewear. *Indian Journal of Ophthalmology 65(9)*, 859-864.

Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly. 44*, 166–206.

Dijk, F. van, Bubas, M., & Smits, P. (2015). Evaluation Studies on Education in Occupational Safety and Health: Inspiration for Developing Economies. *Annals of Global Health, 81,* 548-560.

Hale, A. (1984). Is safety training worthwhile? *Journal of Occupational Accidents 6*, 17-33.

Heath, E. (1982). Workers training and education in occupational safety and health: a report on practice in six industrialised western nations. Part III. *Journal of safety research 13*, 121-131.

Heinrich, H. (1956). Recognition of safety as a profession, a challenge to colleges and universities. National Safety Council Transactions. *Proceedings of the 44th National Safety Congress, October 22-26, Chicago, Ill*, 37-40.

IOM (2001). Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington DC: National Academies Press.

Kennedy, P.E., Chyung, S.Y., Winiecki, D.J., & Brinkerhoff, R.O. (2013). Training professionals' usage and understanding of Kirkpatricks' level 3 and level 4 evaluations. *International Journal of Training and Development 18(1)*,1-21.

Kirkpatrick, D. (1959a-1978) techniques for evaluating training programs *Journal of American Society for Training and Development 13(11)*, 3-9; (1959b). Techniques for evaluating training programs part 2 – learning. *Journal of American Society for Training and Development (12),* 21-26; (1960a). Techniques for evaluating training programs part 3 - behaviour. *Journal of American Society for Training and Development 14(1)*, 13-18; (1960b). Techniques for evaluating training programs part 4 – results. *Journal of American Society for Training and Development 14(2),* 28-32; (1977). Evaluating training programs: evidence vs proof. *Training & Development Journal 31(11),* 9-12; (1978). Evaluating in-house training programs. *Training and Development Journal 32(9),* 6-9.

Liebermann, S., & Hoffmann, S. (2008). The impact of practical relevance on training transfer: evidence from a service quality training program for German bank clerks. *International Journal of Training and Development 12(2)*, 74-86.

Mann, S. (1996). What should training evaluations evaluate? *Journal of European Industrial training 20(9)*, 14-20.

Schelvis, R.M., Oude Hengel, K.M., Burdorf, A., Blatter, B.M., Strijk, J.E., & van der Beek, A.J. (2015). Evaluation of occupational health interventions using a randomized controlled trial: challenges and alternative research designs. Scand J Work Environ Health;41(5):491-503.

Swuste, P., & Arnoldy, F. (2003). The safety advisor/manager as agent of organisational change: a new challenge to expert training. *Safety Science 41,* 15-27.

Swuste, P., & Sillem, S. (2018). The quality of the post academic course Management of Safety Health and Environment (MoSHE) of the Delft University of Technology. *Safety Science 102*, 26-37.

Yu, I.T., Yu, W., Li, Z., Qiu, H., Wan, S., Xie, S., & Wang, X. (2017). Effectiveness of participatory training in preventing accidental occupational injuries: a randomized-controlled trial in China. *Scandinavian Journal of Work Environment and Health 43(3),* 226-233.

Education and Culture
Leonardo da Vinci