# Collecting actionable and value-focused feedback using multi-criteria evaluation – a procedure and an application

Tuomas J. Lahtinen
tuomas.j.lahtinen@aalto.fi

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis
Aalto University
P.O. Box 11100
00076 Aalto, Finland

**Abstract**

Feedback is essential for improvement. This paper describes how multi-criteria evaluation can be helpful in the collection of feedback. A new procedure is introduced for the collection of course feedback from students. The primary aim is to assist the teacher in her improvement efforts. That is, feedback is collected for a so-called formative purpose. The outcomes of the procedure include a multi-criteria evaluation of the course and a list of improvement ideas. The secondary aim of the procedure is to benefit the students. They get to reflect on a course they have taken and how it might relate to their future paths. In the procedure, multi-criteria evaluation is used by the feedback-givers to form a structured overall picture of the system at hand. Subsequently, the evaluation results inform a creative process where the feedback-givers suggest improvement measures. The creative effort is directed towards overcoming specific weaknesses in light of the criteria used. The procedure is applied in a university course and the results are described. Finally, usefulness and limitations of the procedure are considered, as well as its possible variations and extensions and also questions for future work.

**Keywords**

Multiple criteria analysis; Multi-criteria decision making, Course feedback; Idea generation; Behavioural OR

## 1. Introduction

> "Fundamentally, the most powerful way of thinking about a teacher's role is for teachers to see themselves as evaluators of their effects on students."
> *—John Hattie, Visible Learning for Teachers: Maximizing Impact on Learning*

This quote illustrates a key conclusion based on John Hattie's comprehensive meta-analyses on what works in teaching. The finding is that "almost everything works", which leads to the question of what works the best, and from there, to the notion that teachers should be proactive researchers and developers of their own teaching (Hattie 2012).

Contributing in this direction, the present paper demonstrates a novel application of multi-criteria evaluation to a frequently occurring problem, namely, collecting student feedback on a course. Feedback from students can help a teacher in multiple ways. For example, in finding out the strong and weak parts in a course taught by the teacher, getting ideas on improving the course, and prioritizing improvement efforts (see, e.g., Huxham et al. 2008, Nathenson and Henderson 2018). Alternative methods for collecting feedback are called for to meet the diverse needs and situations arising in practice (Richardson 2010).

To be clear, this paper focuses on the so-called formative purpose of feedback. This means, the aim is to directly help the teacher to improve the effectiveness of teaching and materials. This kind of aim is markedly distinct from a summative purpose of feedback, which refers to using feedback to inform higher level decision making concerned with issues such as rewarding and curriculum planning, for instance (see, e.g. Alderman et al. 2012).

Earlier, when multi-criteria evaluation methods have been used in the collection of feedback from students, the focus has been on the summative purpose. Multi-criteria methods have been used to evaluate different e-learning courses, e.g., regarding course contents and technical implementation (Petasakis et al. 2015), and also in the evaluation of e-learning systems (Zare et al. 2016). Moreover, Hein et al. (2015) describe a case where multi-criteria evaluation of professors, e.g. in terms of interaction with students, was carried out. In contrast to these papers, here our purpose is not primarily in evaluating or ranking distinct courses or other units of analysis, but rather in the generation of measures to help improve a single course.

A new procedure for the collection of student feedback is demonstrated in this paper. In brief, the students first carry out a multi-criteria evaluation of a course they have taken. Subsequently, with the help of the evaluation results, the students generate ideas of measures for improving the course. During the procedure, the students think of the course in terms of criteria, such as "usefulness for future career" and "usefulness for future studies." These criteria reflect the value of the course. Thus, the feedback gathered is likely to be *value-focused* (Keeney 1996). Moreover, when the students generate measures for improving the course, they are instructed to think of measures that help improve specific parts of the course. The improvement ideas generated are thus likely to be *actionable* rather than vague.

The general idea behind the procedure is using multi-criteria evaluation to assess the performance of systems components, which are relatively independent and comparable. And subsequently, using the evaluation results to inform the generation of improvement measures. In earlier literature, this general idea has been applied, e.g., in the context of bank branches (Ferreira et al. 2012) and container terminals (Sharma and Yu 2010). Besides demonstrating the idea in a new context, compared to these earlier studies, the present paper focuses more strongly on the generation of improvement measures rather than on the multi-criteria evaluation per se. Also, in our case, ideas on how to improve the system at hand are generated by the "customers" of the system, i.e., students, instead of managers. Furthermore, it should also be noted that the use of multi-criteria evaluation to support the generation of improvement measures directly links to the decision analysis literature concerned with the generation of alternatives, which is seen as an underdeveloped branch of literature (see, e.g., Keller and Ho 1988, Siebert and Keeney 2015, Colorni and Tsoukias 2020). This linkage has not been explicitly acknowledged in any of the above-mentioned papers.

The new procedure was applied in an introductory operations research course at Aalto University. The author of this paper was the teacher of the course.

The application and its results serve to illustrate the procedure and provides initial evidence on its usefulness.

In the final sections of this paper, a discussion is provided concerning usefulness and limitations of the procedure. Variations and extensions of the procedure are also discussed, as well as the questions left for future work.

## 2. A procedure for collecting feedback

The procedure introduced here is intended to support the identification of promising measures for improving a course. Following the procedure, a group of students, each of them working independently, conducts a multi-criteria evaluation of a course they have taken and then generates measures for improving the course. The procedure is outlined in Figure 1.

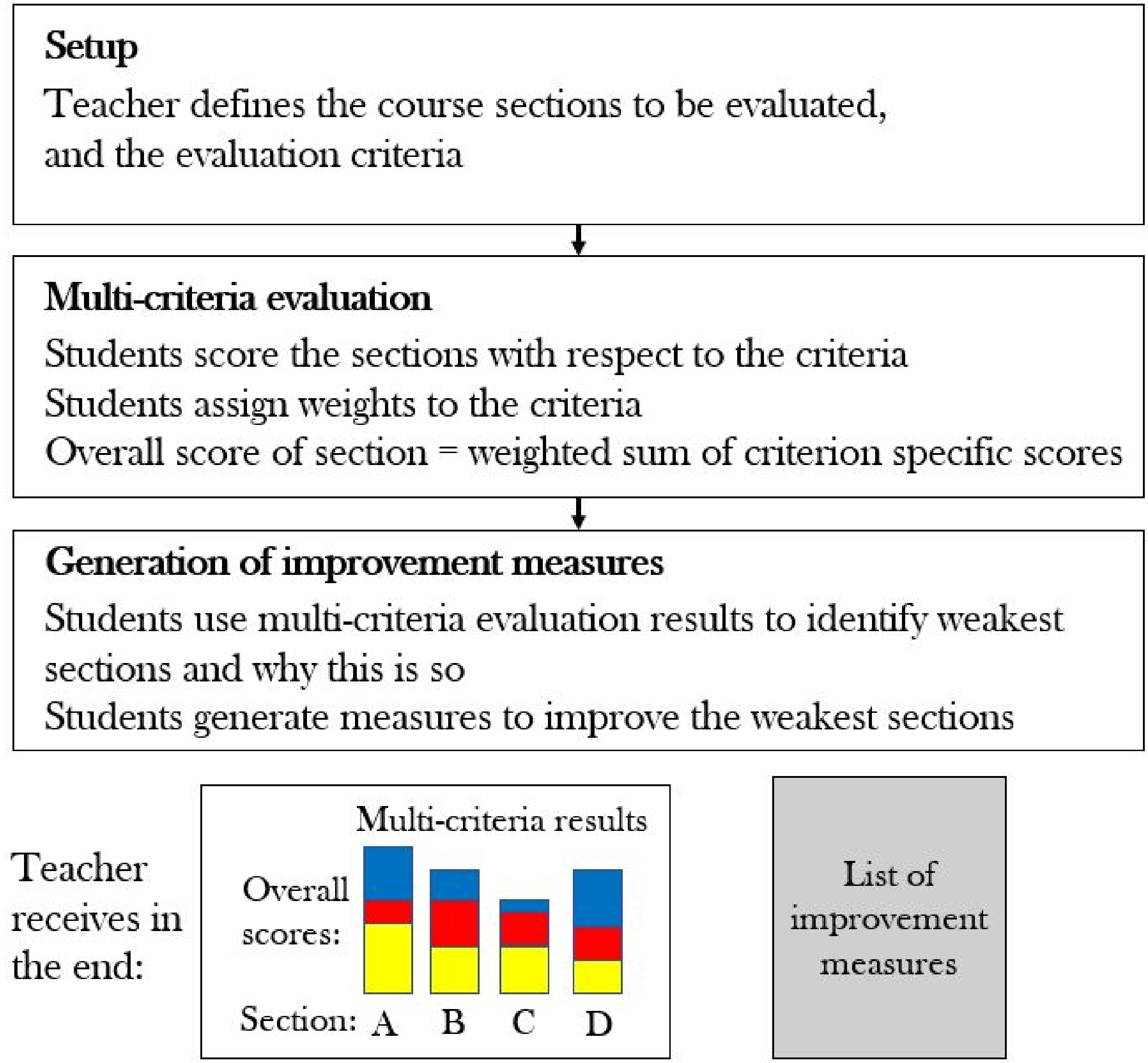
More specifically, from the students perspective, in the procedure,

(i) the course is viewed to consist of multiple sections, which the students score using criteria such as “usefulness for career”, “usefulness for future studies”, and “quality of the teaching materials”, for example. The students also assign weights to the criteria.

(ii) Each student examines her multi-criteria evaluation results to identify the weak parts of the course, and why these parts are weaker than the others. Informed by these results, each student generates ideas of measures to improve the course. For example, a student could find out that a section has received a poor overall score due to its low score on a criterion with a high weight such as perceived usefulness for career. Then, an improvement measure could be to explain better how the section’s contents are relevant to possible future careers.

Furthermore, besides using the list of improvement measures generated, the teacher can also make use of the aggregate multi-criteria evaluation results. If a section consistently receives a poor score in all criteria, the teacher might consider thoroughly overhauling this section.

*Figure 1: Multi-criteria feedback collection procedure*




In the setup phase, preceding the application of the procedure, it is the teacher's task to define the sections and the evaluation criteria. The teacher can define the sections, for example, based on themes covered in the course (theme 1, theme 2, etc.), or on time (e.g., first quarter, second quarter, etc.) Ideally, the sections are non-overlapping and cover all those parts of the course that the teacher is interested in (and capable of) doing something about. The criteria may reflect the course's goals and aims and other concerns of the teacher and the students. The criteria should be understandable for the students, and there should not be too many of them. Ideally, the set of criteria is non-redundant such that there is no double-counting of pros or cons related to the sections. The same set of criteria has to be applicable to the evaluation of each section.

It should be noted that the way multi-criteria evaluation is used in the suggested procedure contrasts with how multi-criteria evaluation is traditionally used. In a typical setting, alternative measures or courses of action already exist, and the problem is to rank the alternative measures or find the most preferred one, for instance. For the interested reader, some basic references in multi-criteria literature are Keeney and Raiffa (1993) and von Winterfeldt and Edwards (1986). A review of applications is presented by Wallenius et al. (2008).

From a pragmatic viewpoint, the procedure described here is primarily a heuristic procedure to stimulate and steer the generation of improvement measures. The list of improvement measures generated is naturally likely to reflect the procedure. The measures are likely to be related to specific parts of the course (rather than to the course as a whole) and reflect the evaluation criteria used. Moreover, in our case, the focus is on overcoming problems and weaknesses. Limitations of the procedure and possible variants of the procedure are discussed in a following section.

## 3. An application of the procedure

Course feedback was collected from an introductory operations research course, Computational Assignments in Applied Mathematics, at Aalto University School of Science in 2014. At the time, the author of the paper was the teacher of the course. The results presented here are based on the feedback collected from 45 students who took the course.

The course consisted of eight thematically distinct sections, each involving a set of computational assignments carried out both in live sessions and as homework assignments. Five criteria were used in the evaluation of the sections. Figure 2 shows the sections and the evaluation criteria in an evaluation matrix, which was presented to the students.

*Figure 2: The evaluation matrix used in the application.*

| Course sections ↓ / Evaluation Criteria → | Usefulness for career | Usefulness for future studies | Quality of materials | Pleasantness of the assignments | Attractiveness of the topic dealt with |
|---|---|---|---|---|---|
| A – Linear optimization with Excel | | | | | |
| B – Statistical analysis with Excel | | | | | |
| C – Differential eqs. with Mathematica | | | | | |
| D – Regression analysis with Matlab | | | | | |
| E – Optimization with Matlab | | | | | |
| F – Monte Carlo simulation with Matlab | | | | | |
| G – Differential equations with Matlab | | | | | |
| H – Dynamic systems with Simulink | | | | | |

The students scored the sections with respect to each criterion on a scale from 0 to 10. They were instructed to give the worst-performing section 0 points and to give 10 points to the best-performing one. The points they give to the rest of the sections represent the performance of these sections in relation to the best and the worst-performing ones. The students gave the weights using the SWING-method (von Winterfeldt and Edwards 1986), i.e., each criterion specific weight represents the perceived subjective value of improvement from the worst to the best performance level in the criterion. The overall score of a section was calculated as the weighted sum of criterion specific scores.

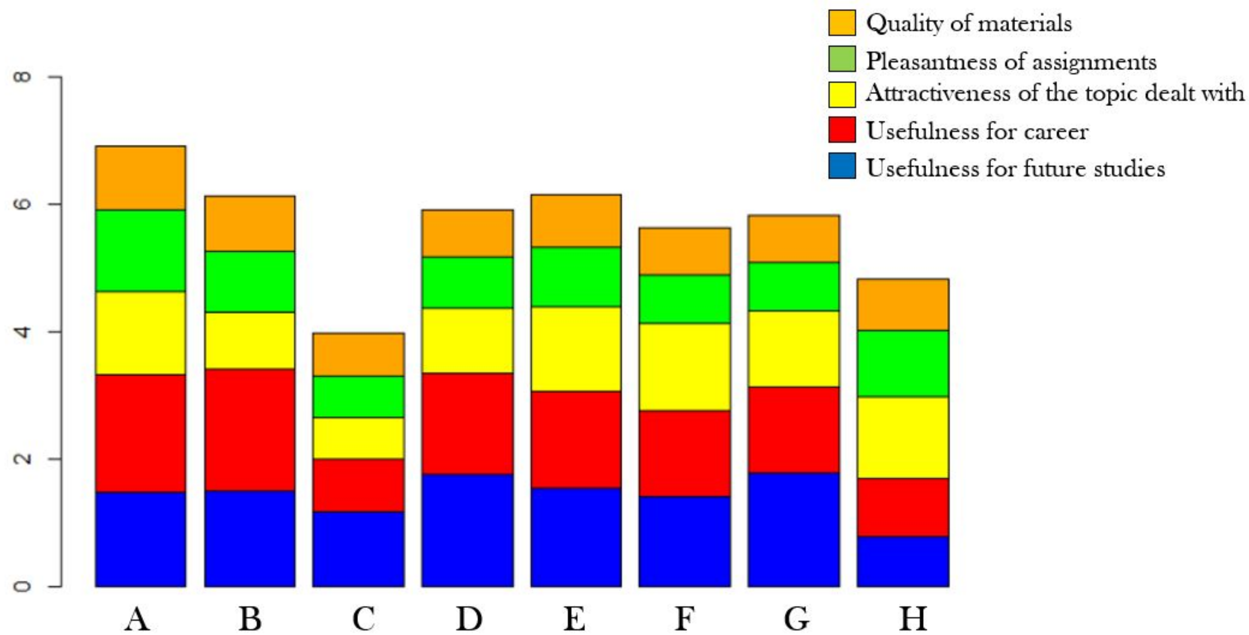
The students examined the multi-criteria evaluation results with an instruction to find out what are the strong and the weak parts of the course, and why this is so. The students were instructed to identify the two lowest scoring sections and to suggest measures for improving these sections.

## 4. Results from the application

The results of the application are two-fold. Firstly, there are the multi-criteria evaluation results, which were used by the students to direct their creative efforts. Secondly, there are improvement measures generated by the students.

Figure 3 shows the overall scores of the course sections based on the multi-criteria evaluations carried out by the students. The heights of the bars represent the overall scores of the course sections. The bars are composed of five shorter bars, which represent criterion specific scores. To be precise, the results presented in this figure are aggregated over individual results by averaging the weighted criterion specific scores given by the individual students.

*Figure 3: Overall scores of the course sections based on the multi-criteria evaluations. (Scores presented in this figure are aggregated over individual results.)*




"Usefulness for career" and "usefulness for future studies" are the criteria that have generally received the most weight. And consequently, the criterion specific scores in these criteria generally contribute most to the overall scores of the sections (i.e., to the total height of the bars in Figure 3).

Sections "C - Differential equations with Mathematica and "H - Dynamical systems with Simulink" have generally received the lowest overall scores. The remarkably poor overall score of section C is due to its consistently low score in all criteria, and particularly due to its low score in the highly weighted criterion "Usefulness for career". The poorer-than-average overall score of section H is explained particularly with its low score in the criteria "Usefulness for career" and "Usefulness for future studies".

A natural follow-up question is, “could the sections C and H be improved in these respects”?

Focusing on these two weakest performing sections, Table 1 shows the measures generated by the students for improving sections C and H. The measures are presented in a condensed form and have been translated from Finnish into English.

*Table 1: The measures generated by the students for improving sections C and H.*

| |
|---|
| C - Differential equations with Mathematica |
| • Change the assignments so that more attention is given to the interpretation of the figures<br>• Increase the variety and attractiveness of the assignments<br>• Give assignments that are more closely related to working life<br>• Give more interesting assignments<br>• Give examples that are more closely related to Industrial Engineering<br>• Add more plotting related tasks to the assignments<br>• Describe the Jacobian matrix more thoroughly<br>• Provide a more thorough introduction to Mathematica syntax<br>• Move the section to the beginning of the course |
| H - Dynamic systems with Simulink |
| • Give assignments that are more closely related to working life<br>• Change the water tank related assignment to something else<br>• Give easier assignments in the beginning of the sessions<br>• Give less arduous assignments<br>• Improve the clarity of the instructions for the assignments<br>• Give several small assignments instead few large ones<br>• Improve the instructions related to how Simulink is used<br>• Increase the number of illustrative examples<br>• Describe the strengths and weaknesses of Simulink<br>• Teach the best practices in naming and positioning the block elements |

- Provide a more thorough description of how blocks work and where they can be found
- Add more simulink related exercise sessions to the course

We can see that many of the suggested measures directly relate to the criteria used in the multi-criteria evaluation. For example, “Give assignments that are more closely related to working life” relates to the criterion “usefulness for career”, “Give examples that are more closely related to industrial engineering” relates to the criterion “usefulness for future studies”. Some of the measures relate to multiple criteria. For example, “Add more plotting related tasks to the assignments” relates to both “usefulness for career” and “usefulness for future studies”. In conclusion, the measures can be characterized as value-focused.

The majority of the measures are specific and concrete, and thus can be called actionable. However, there is variation in this respect. “Provide a more thorough introduction to Mathematica syntax” is a good example of a concrete and specific measure, whereas “Give more interesting assignments” is an example of a more vague improvement measure.

Overall, the results give a clear indication of specific weaknesses in the course and suggest multiple ways in which the weak parts of the course could be improved.

## 5. Discussion

### 5.1. Usefulness of the procedure

Firstly, it is important to note that the procedure is meant to be one of the many ways in which a teacher can obtain student feedback on a course. No single procedure can result in a comprehensive picture of a course or in a comprehensive list of possible improvement measures (see, e.g. Alderman et al. 2012).

The teacher must interpret the results of the procedure, integrating them with feedback obtained with other means and her broader understanding, and then decide what to do, if anything, based on the feedback. For example, after the application described in this paper took place, one of the measures taken was to add an additional assignment in Section C. This section received poor scores particularly with regard to usefulness in career. The assignment was to write a summary of a real industrial application of the software used in Section C. It was possible to implement this measure because reports of such real-life applications were readily available, and based on feedback gathered via other means it was known that the workload of the course could be safely increased a little.

The distinguishing feature of the procedure is the multi-criteria evaluation step. In this step, the students review different parts of the course from multiple perspectives. This is likely to cause the students to reflect on the course more deeply and thoroughly than if such an evaluation was not carried out. At the minimum, this step will help the students refresh their memories about the course. From a cognitive perspective, the multi-criteria evaluation serves as a cue that helps access relevant memories (see, e.g., Perttula and Liikkanen 2005, Johnson and Raab 2003). Moreover, the step helps direct the students' creativity towards the generation of improvement measures that are value-focused and actionable.

However, the structured format of collecting feedback can also create limitations. The course is split into distinct parts, and subsequently the generation of improvement measures is focused on the parts. This characteristic of the procedure is like a double-edged sword; it is likely to help generate specific and concrete improvement measures at the expense of finding more holistic improvement measures. Moreover, a constant set of criteria chosen by the teacher, is used to evaluate all sections of the course. This limits the applicability of the procedure to situations where it is possible to define such criteria in a meaningful way. The fact that the criteria are selected by the teacher may also inhibit creativity of the students.

To sum up the discussion, the procedure is more likely to be useful in situations where the following “conditions” are met:

- The teacher seeks to understand better what are the strong and the weak parts of the course the teacher is giving.
- The teacher feels that the improvement ideas usually received are not sufficiently actionable and value-focused.
- The course can be split, without too much forcing, into distinct sections evaluated with a constant set of criteria.

- Bringing in a structured process is not likely to be a limiting hindrance or strongly opposed by the students. For example, students who are fluently displaying a high degree of creativity, expressivity and thoughtfulness might feel the procedure as rigid.
- The teacher is able to integrate the results with her broader understanding and knowledge. That is, the teacher will not do anything stupid just because it was the result of a procedure.

### 5.2. Pedagogical value of the procedure

The use of the procedure can generate pedagogical value as such. In the context of Operations Research and Management Science (OR/MS) education there is value in that the students experience first-hand an actual application of an OR/MS technique. In particular, the experience is from the perspective of an expert giving subjective inputs to a multi-criteria evaluation. Being really in the shoes of an expert is likely to be a more insightful experience than one arrived at if one were to, e.g., evaluate hypothetical cars described in the course materials.

More generally, the application of the procedure enables students to go through what has been covered in the course, and encourages the students to think of how the contents of the course might relate to their future paths. Furthermore, to enhance pedagogical value, the results could be discussed with the entire group of students. They would see where their views are aligned and where they differ, and they could learn from each other.

### 5.3. Variations and extensions of the procedure

The weighting and scoring methods used are likely to influence, e.g., the ease of using the procedure and also the results obtained (see, e.g. Franco et al. 2020). For example, the likert scale from one to seven, which is familiar to many, could in some contexts be the most suitable scale to be used in the scoring of the course sections. Different weighting methods could be used too, for example, predetermined weights, such as equal weights, could be used. Also, rather than using weights to aggregate the criterion specific scores, multi-dimensional performance profiles could be used as such. Each of these alternative methods exhibit their specific strengths and weaknesses: predetermined weights, for instance, would require less work but could give misleading results. Further variants of the procedure can be created by using different instructions in the generation of improvement measures. Specifically, it is not necessary to focus only on improving the weakest performing sections.

The criteria used do not need to be the same as in the application presented in this paper. For example, one could consider usefulness of the course for self-development and also the broader socio-cultural impacts of the course. To stimulate even greater diversity of perspectives, one could ask the students to role play different perspectives when giving their judgments. For example, they could be asked to evaluate the course from engineer, businessman, and researcher perspectives. In continual use of the procedure, one could ask the students to suggest additional criteria, which could help improve the procedure over time.

Moreover, in the application presented in this paper, the procedure was used at the end of a course. Alternatively, feedback could be collected after each individual section or after every session within the course. It would also be interesting to consider the use of the procedure to monitor the effects of changes over time.

### 5.4. Other questions for future work

The points made above already indicate many interesting avenues for future research. Furthermore, the procedure suggested here, or its variants, need to be tested also in other contexts to explore the broader applicability and scalability of the approach. It would probably be most straightforward to collect experiences from another engineering course. However, more interestingly, experiences could be collected from courses related to different disciplines, and with courses provided in other levels of education, and also in other contexts than university, e.g., in private training programs. Such experiences would support the making of the procedure more widely applicable.

Controlled experiments could help better understand the potential of multi-criteria evaluation in the collection of feedback. One possible experiment would be to carry out a feedback collection procedure with and without a multi-criteria evaluation step. For example, one could compare improvement measures suggested by students when they have first carried out a multi-criteria evaluation compared to a situation where they have not done so. An interesting topic for future would be also to compare the procedure with other approaches intended for similar purposes.

## 6. Conclusion

This paper provides one approach for the proactive teacher (or manager) to consider when looking for ways to collect useful feedback to inform improvement efforts. More specifically, this paper demonstrates how multi-criteria evaluation can be used in the collection of course feedback. In the procedure introduced in this paper, multi-criteria evaluation supports a creative process of developing improvement measures. The application presented in this paper resulted in the generation of value-focused and actionable ideas of measures for improving a university course. The ideas aim to help overcome specific weaknesses in the course in light of the criteria used.

Generally, the use of multi-criteria methods to enhance feedback collection seems like an interesting direction for future work. The collection of feedback is a pervasive issue, and much more could be said and done in this area from the multi-criteria perspective. Hopefully, this paper serves to open doors for valuable cross-fertilization and new insights.

This paper also points out the possibility to generate significant pedagogical value with the integration of multi-criteria evaluation into feedback collection procedures. The procedure suggested in this paper helps students to go through what has been covered in the course, and nudges the students to think of how the contents of the course relate to their future paths. Furthermore, in the context of OR/MS education, the procedure will also result in the students having experienced an actual application of multi-criteria evaluation from the perspective of an expert giving subjective inputs to an evaluation process.

Future research can shed light on the behavioural and cognitive mechanisms related to using multi-criteria methods in feedback collection. Variations of the procedure described here can be considered. And, comparative analyses of different feedback collection methods can be carried out. More evidence needs to be gathered concerning the applicability and scalability of the suggested approach in different contexts.

**References**


Alderman, L., Towers, S., and Bannah, S. (2012). Student feedback systems in higher education: A focused literature review and environmental scan. Quality in Higher education, 18(3), 261-280. https://doi.org/10.1080/13538322.2012.730714

Colorni, A., and Tsoukiàs, A. (2020). Designing alternatives in decision problems. Journal of Multi-Criteria Decision Analysis, 27(3-4), 150-158. https://doi.org/10.1002/mcda.1709

Ferreira, F. A. F., Spahr, R. W., Santos, S. P., and Rodrigues, P. M. (2012). A multiple criteria framework to evaluate bank branch potential attractiveness. International Journal of Strategic Property Management, 16(3), 254-276. https://doi.org/10.3846/1648715X.2012.707629

Franco, L. A., Hämäläinen, R. P., Rouwette, E. A., and Leppänen, I. (2020). Taking stock of Behavioural OR: A review of behavioural studies with an intervention focus. European Journal of Operational Research. In Press. https://doi.org/10.1016/j.ejor.2020.11.031

Hattie, J. (2012). Visible learning for teachers: Maximizing impact on learning. Routledge.

Hein, N., Kroenke, A., and Júnior, M. M. R. (2015). Professor assessment using multi-criteria decision analysis. Procedia Computer Science, 55, 539-548. https://doi.org/10.1016/j.procs.2015.07.034

Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J., and Earl, S. (2008). Collecting student feedback: a comparison of questionnaire and other methods. Assessment & Evaluation in Higher Education, 33(6), 675-686. https://doi.org/10.1080/02602930701773000

Johnson, J. G., and Raab, M. (2003). Take the first: Option-generation and resulting choices. Organizational Behavior and Human Decision Processes, 91(2), 215-229. https://doi.org/10.1016/S0749-5978(03)00027-X

Keller, L. R., and Ho, J. L. (1988). Decision problem structuring: Generating options. IEEE Transactions on Systems, Man, and Cybernetics, 18(5): 715-728. https://doi.org/10.1109/21.21599

Keeney, R. L., Raiffa, H.. (1993). Decisions with multiple objectives: preferences and value trade-offs. Cambridge University Press.

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. European Journal of operational research, 92(3), 537-549. https://doi.org/10.1016/0377-2217(96)00004-5

Nathenson, M. B., and Henderson, E.S. Using student feedback to improve learning materials. Routledge, 2018.

Perttula, M., and Liikkanen, L. A. (2005). Cue-based memory probing in idea generation. In Sixth Roundtable Conference on Computational and Cognitive Models of Creativity, Key Centre of Design Computing and Cognition, University of Sydney, Sydney, 195-210.

Petasakis, I., Theodosiou, T., Kazanidis, I., and Valsamidis, S. (2015). Multicriteria analysis of e–learning courses. International Journal of Data Analysis Techniques and Strategies, 7(2), 203-215. https://doi.org/10.1504/IJDATS.2015.068751

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. Assessment & evaluation in higher education, 30(4), 387-415. https://doi.org/10.1080/02602930500099193

Santos, S. P., Belton, V., Howick, S., and Pilkington, M. (2018). Measuring organisational performance using a mix of OR methods. Technological Forecasting and Social Change, 131, 18-30. https://doi.org/10.1016/j.techfore.2017.07.028

Sharma, M. J., and Yu, S. J. (2010). Benchmark optimization and attribute identification for improvement of container terminals. European Journal of Operational Research, 201(2), 568-580. https://doi.org/10.1016/j.ejor.2009.03.021

Siebert, J., and Keeney, R. L. (2015). Creating more and better alternatives for decisions using objectives. Operations Research, 63(5), 1144-1158. https://doi.org/10.1287/opre.2015.1411

Von Winterfeldt, D., and Edwards, W. (1986) Decision analysis and behavioral research. Cambridge University Press.

Wallenius, J., Dyer, J. S., Fishburn, P. C., Steuer, R. E., Zionts, S., and Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. Management science, 54(7), 1336-1349. https://doi.org/10.1287/mnsc.1070.0838

Zare, M., Pahl, C., Rahnama, H., Nilashi, M., Mardani, A., Ibrahim, O., and Ahmadi, H. (2016). Multi-criteria decision making approach in E-learning: A systematic review and classification. Applied Soft Computing, 45, 108-128. https://doi.org/10.1016/j.asoc.2016.04.020