Measuring modelling

Oscar Patrick René van Vliet[1]

Department of Environmental Systems Science, ETH Zurich
8092 Zurich, Switzerland

Urs Brändle[2]

Department of Environmental Systems Science, ETH Zurich
8092 Zurich, Switzerland

Abstract

Modelling is an essential skill in many scientific fields, including environmental science. We designed a Modelling Competence Inventory (MCI) to measure the progress of students in acquiring competence in modelling during the bachelor curriculum. As models in environmental science borrow from many disciplines, and modelling is by nature an abstract activity that requires critical thinking, we find that designing an MCI is difficult compared to competence inventories for more physical subjects. We discuss the design process, two iterations of our MCI, and the results of testing these on a group of students before and after a modelling course. Results suggest that students understanding of the learning goals taught in the course improved somewhat, but their score on other learning goals decreased. Overall, we find that bachelor students need more supervised independent practice with modelling and building of confidence in their modelling abilities. The MCI needs further development and differentiated questions specific to the course in which the MCI is administered. The process of searching for competencies to track and developing the MCI, in cooperation with lecturers in the environmental science bachelor, by itself helped build a community of practice and led to steps to better align courses in our curriculum.

1 Introduction

Monitoring students' progress is an essential feedback loop in continuously improving education, as it allows educators to target additional efforts towards aspects of teaching that need improvement.

Several top universities have adopted a backward design approach to their courses that lends itself well to monitoring progress (e.g. Stanford University 2019, Volk 2019). Their courses are designed to fulfil learning goals, or specific elements of knowledge, skills, and attitudes that institutions want their students to acquire within a course, to a minimum level of competence (see Bloom et al. 1956, Anderson et al. 2001). These learning goals are to be phrased S.M.A.R.T.[3], for example as a measurable goal in an object-verb structure. This approach to curriculum design is being adopted more widely, although implementation depends on available time, mentoring and other resources for lecturers (e.g. Bosma et al. 2016).

Measuring and monitoring student progress towards these learning goals can be done on two levels: per course, and over the entire curriculum. On a course level, it allows lecturers to clarify specific classes and assignments, shift emphasis where needed, and determine where their

[1] o.p.r.vanvliet@uu.nl
[2] urs.braendle@env.ethz.ch
[3] Acronym for Specific, Measurable, Achievable, Realistic, and Time-bound (see Doran, 1981 and many variations since).

own teaching skills may need improving. On a curriculum level, it can be used to identify weak courses, but also to identify gaps and synergies between courses, and evaluate if the curriculum as a whole teaches what it advertises to. This in turn enables an evidence-based approach to planning of future courses and curriculum revisions.

One possible measurement tool for monitoring student progress is a competence inventory. These are multiple-choice questionnaires that test students' understanding by giving at least one correct answers and multiple *distractors*, answers that seem reasonable and correct, but are based on commonly held misconceptions (e.g. Hestenes et al. 1992; Champagne Queloz et al. 2017). The logic is that students that harbour specific misconceptions will choose the incorrect answers that reflect these misconceptions. Lecturers can use this information to pinpoint where in their course the students fail to absorb the knowledge or skills the lecturer is trying to teach.

In our study, we wanted to measure how student competencies evolve in computational modelling. Modelling is an essential skill in many scientific fields, including environmental science. To measure the progress of students in acquiring competence in modelling during the bachelor curriculum, we developed a Modelling Competence Inventory (MCI).

Our MCI development took place as part of the larger project, Model 4 Modelling (M4M), that aims to improve modelling in the curriculum of the 3-year bachelor program in environmental science (UMNW) at ETH Zürich.

2 Teaching concept

We developed an MCI to monitor students' progress in modelling, to collect input for an evidence-based approach to improve both individual courses and the overall curriculum. At the course level, having students take the MCI some time at the start and end of a course provides a measure of learning goal attainment, and allows the lecturers to focus their efforts on improving the teaching of learning goals that are not sufficiently attained.

At a curriculum level, this entire process is intended to support the following analysis:

1. Identify competencies that are required for advanced courses and compare these to competencies taught basic courses to identify learning gaps.
2. Identify competencies that are taught repeatedly to identify overlapping courses and then determine how these courses' learning goals overlap.
3. Use the MCI to see how the overlaps work out in practice. For example, if there is overlap between two courses:
    a. If students understand after the first course, then the topic can be removed from the second course.
    b. If students do not understand after the first course, then the first course needs to be improved. However, it could be that the learning goal is simply hard, and students require multiple courses, with teaching from multiple angles, to get their heads around a specific competency. This is especially true for levels of competency that allow for independent reflection and creation of new knowledge. In that case:
    c. If students do not understand after the first and second courses, then both courses need to be improved.

Another reason why overlap in learning goals between courses can be necessary and even desirable is if students do not take the same consecutive courses, e.g. the first and/or second courses are electives. However, the result of the MCI should always be considered together with qualitative assessment of the courses. An MCI is therefore best used as a signalling system, as part of a larger process of evaluation.

To separate the process of conceptualizing, developing and using a model into manageable learning goals, we split the process of computational modelling into five successive steps, with feedbacks to earlier steps (visualised in Figure 1):

1. Acquire system knowledge, i.e. learn about the natural or social system to be represented by the model
2. Design the model, i.e. decide on its structure and mathematical foundations
3. Implement the model, i.e. code it, debug it, improve its efficiency, and/or expand it.
4. Run & evaluate model, i.e. iterate to achieve results that make sense.
5. Interpret results, i.e. help the rest of the world make sense of the outputs.

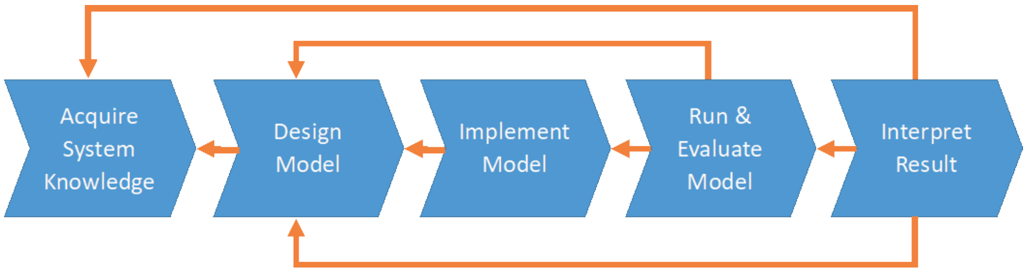

Figure 1: Five steps in modelling, with feedback loops in arrows.

However, these five steps do not distinguish in depth of competence. For this, we can refer to Henning & Keune (2007), who describe three levels of competency in modelling. However, we embed these three levels into the more widely used Bloom taxonomy (Bloom et al. 1956, Anderson et al. 2001), as shown in Table 1.

Henning & Keune	Bloom taxonomy
1. Recognition and Understanding	1. Reproduce 2. Understand
2. Independent modelling	3. Apply 4. Analyse
3. Meta-reflection on modelling	5. Evaluate 6. Create

Table 1: Levels of competency

Combining these steps and levels together yields a matrix of 30 categories for competencies and questions to measures these competencies. We can use this matrix to check our coverage of different aspects of modelling in the MCI. In interviews and discussions with lecturers, we can also use this matrix to ensure that we cover the full span of competencies in their courses, and help them make explicit the way their courses build competencies on top of each other (also see Wijngaards-de Meij & Veenhoven 2016).

3 Analysis of student learning

One of the core elements of a competence inventory is to test for misconceptions, phrased as distractors. We crowd-sourced an initial set of misconceptions by sending an open question to a group of lecturers who teach some aspect of modelling within the department of Environmental System Science (D-USYS) at ETH Zürich, and who had recently been in contact with the authors. This yielded two immediate results:

- 13 misconceptions (see Appendix A), submitted by 8 lecturers (not including the authors), that we used as source material for the first version of our MCI that contained 13 questions.
- Expressions of interest from a number of staff members, mostly the same group who also replied with misconceptions. We gathered these in an M4M special interest group.

The next step was to interview 7 other lecturers in the M4M special interest group about their courses. The authors distilled and combined these interviews into an initial list of competencies.

This list was discussed in a workshop with 10 of the M4M special interest group, where we validated the list of competencies and examined which competencies were taught in the respective lecturers' courses, which were assumed or required for these courses, and which were seen as missing or underdeveloped.
Based on the list of learning goals and the interviews, we compiled a second version of the MCI, with a total of 17 questions (see appendix B, 10 of which were new and 7 of which were adapted from the previous version. Each of these questions has an introduction, describing a model and situation, a question, and several possible answers. We also gave students an opportunity to comment on their answers, if they wanted. This gave us feedback on the questions, as well as an opportunity to identify more misconceptions from the students' point of view rather than the lecturers.

We tested both versions of the MCI on a group of 3rd year bachelor students in environmental science, majoring in human-environment systems. The group was taking a mandatory practical course in the spring semester that included two full days of energy system modelling for three consecutive weeks, and repeated use of an energy system model for another three weeks. MCI v1 was filled out by 21 students before their modelling course segment, and 19 students afterwards. MCI v2 was filled out the next year by 22 students before, and 16 afterwards. This resulted in a small dataset of 14 questionnaires that were filled out by students both before and after the modelling course segment.
The first attempt at our MCI did not work as intended. We did not see much change in the answers before and after the modelling practical course. Comments suggested that many of the questions were poorly understood. This made sense in hindsight given that questions were based on misconceptions suggested by veteran modellers and lecturers, rather than focusing on the specific competencies the students need to complete a course. The misconceptions we used reflected lecturers' experience with their research and public outreach, rather than bachelor-level modelling competencies.

To gather better raw material to mine for questions, we collected learning goals in our modelling courses from interviews and validated these in a workshop. This led to a table of 75 separate learning goals for modelling, spread over the 5 steps and 6 Bloom levels (see Table 2).

	1. Reproduce	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
Acquire system knowledge	Recall the formula for a previously discussed system interaction	Describe the components of a system using modelling terms	Identify interacting system components using a known method.	Describe interactions in a given system qualitatively.	Evaluate if a system model contains the necessary components to answer a given research question.	Formulate a research question about a known system that can be approached with modeling tools.
Acquire syste		Explain the mathematical formulation of a real-world system using simple terminology/everyday language	Rank processes according to their potential importance for system behaviour	Identify key processes (to be included in the model)	Understand/rank parameter uncertainty based on experimental data	Design an experiment to capture the data needed to answer a given research question with modelling tools.
Acquire syste			Rank processes according to level of empirical knowledge		Evaluate completeness of process understanding for a given system	
Acquire syste					Point out where to complement models with other methods	
Design model	Write down definitions of given modelling terms	Describe the individual activities of research using modelling methods based on a simple example	Convert equations and parameters into formal model description	Select the most appropriate modelling method based on a given dataset and research question	Discuss advantages and disadvantages of different modelling methods applicable to a given dataset and research question	Suggest modelling methods appropriate to analyse a research question and dataset
Design mode	Write down definitions of different types of models.	Classify different models based on a verbal descriptions.	Mathematical foundation to quantitatively describe a relationship between system components	Argue if a specific hypothesis can be tested with a specific model	Describe to what extent an existing model differs from the system to be modelled	Decide upon relevant model complexity in terms of: a) processes to be included to reflect behaviour of the system, b) physical-temporal resolution required to resolve relevant processes
Design mode		Explain procedures for chosing modelling methods based on given datasets and research questions	Identify sources of data used in an existing model	Design the data flow diagram for a given model	Deduce necessary changes in an existing model design based on given changes in a system model	
Design mode		Explain the mathematical formulation of a model (component) using simple terminology/everyday language		Draw the data flow diagram for a given model based on the code		
Implement model	List modelling environments or software packages available to use with a method	Identify code that corresponds to a given modelling component	Write code to automate model runs (eg. Periodicity)	Suggest what sections of code a new model component should hook into	Chose a modelling library to implement a given model	Write model code based on a given dataset, method and research question
Implement n	Describe software packages available for modelling and their advertised range of application to models	Structured code development (hierarchy of subroutines)	Adapt exisiting modelling code or data to changes in the model design	Troubleshoot broken code for a given model design for minor errors		
Implement n			Expand modelling code with existing components from libraries	Choose appropriate software / programming language		
Implement n			Systematic debugging			
Implement n			Testing of every implementation step (e.g. mass conservation)			
Implement n			Add new data to an existing dataset			
Run & evaluate model	#N/A	Explain the function of a parameter in a model implementation	Describe the qualitative and quantitative effects of changes to a parameter	Estimate necessary changes in input parameters to get a given output value	Judge if a (statistical) analysis was carried out correctly	Suggest appropriate methods to test (statistical) significance of a model output
Run & evalua			Test various input values to see if an output stays with a given range (robustness)	Test model results (processes, stability)	Conduct a sensitivity analysis on a given model implementation	
Run & evalua			Decide on a schedule and strategy for the intended model runs		Model-data intercomparison	
Run & evalua			Name and verify units of input and output parameters in a model implementation			
Run & evalua			Development of analysis code (e.g. for statistical tests)			
Run & evalua			Model tuning / process optimisation			
Interpret results	#N/A	Explain what hypotheses a familiar model has been used to test	Write up the results of this model run in modelling jargon	Test if qualitative conclusion can be drawn based on quantitative outputs	Evaluate whether the result of a model is realistic and meaningful to stakeholders	Design visualisations for model outputs and inputs
Interpret resu		Compare model input parameters and output values to historical ranges	Draw simple graphs that show model results	Translate model outputs in to human-readable conclusions	Synthesise model results	Develop interpretations of model for different stakeholder groups
Interpret resu				Compare results against null hypothesis (e.g. is the model better than mean of data?)	Inform on uncertainties in result	
Interpret resu				Identify results that support a particular stakeholder's point of view	Inform on model limitations	
Interpret resu					Provide outlook how to improve model in the future (e.g. with better data)	

Table 2: Learning goals in modelling courses in the UNW curriculum, as supplied by the M4M special interest group.

During the workshop with the M4M special interest group, the lecturers suggested that the learning goals in our bachelor courses mostly address the lower levels of the Bloom taxonomy: reproducing, understand, applying, and sometimes analysing (levels 1-4, see Table 2). This matches the learning goals for a bachelor curriculum set out in the Dublin level descriptors (see Framework for Qualifications in the First Cycle, European Consortium for Accreditation, 2014); Independent work is specified at the Master level and beyond, though some level of critical thinking is specified for the bachelor level, and ETH Zürich promotes critical thinking as an important skill for everyone (CTETH 2019).

Our overall set of competencies is very modelling-specific, with only a few competencies in critical assessment of methods and results, as well as effective communication of results and uncertainties. This leaves out many other cross-disciplinary competencies like meta-learning, developing character, and skillsets like creativity and effective collaboration (c.f. Center for Curriculum Redesign, 2019). While some of these are taught in the modelling courses (e.g. effective collaboration), they were not framed specifically as modelling competencies in the M4M special interest group discussions.

We found that the goals that the lecturers defined were mostly set in Bloom levels 2-5 (see Table 3), and the questions in the second version of the MCI reflect this distribution (see Table 4).

	1. Reproduce	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
Acquire system knowledge	1 (3)	2 (9)	1 (3)	1 (3)	1 (5)	2 (5)
Design model	2 (5)	3 (10)	1 (2)	3 (5)	3 (6)	1 (1)
Implement model	0 (0)	1 (2)	3 (6)	2 (3)	0 (0)	1 (2)
Run & evaluate model	0 (0)	1 (4)	3 (11)	1 (3)	2 (4)	0 (0)
Interpret results	0 (0)	2 (7)	2 (6)	2 (7)	2 (5)	1 (2)

Table 3: Sum of learning goals in a particular step and Bloom level applied in all modelling courses taught by M4MSIG workshop participants. Bloom levels from left to right, modelling steps from top to bottom. Some goals are taught more than once; number of times a goal in the category was taught by participants is in parentheses).

Reproduction (level 1) was difficult to put into generic questions, i.e. applicable to multiple courses. This is because the reproduction questions that we found or invented were very specific to the disciplinary case/scientific context that informs most modelling cases in environmental science, or else very abstract in terms of math or programming. Considering the Dublin level descriptors for a bachelor's degree, we avoid questions to test Create (level 6).

We tested the learning goals with questions in MCI v2. As testing all of the Bloom level 2-5 learning goals would lead to an unacceptably long questionnaire, and devising questions to test some competencies was difficult, we ended up testing 27 different learning goals in 17 questions.

In MCI	1. Reproduce	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
Acquire system knowledge	1	1	1	0	0	0
Design model	0	5	3	2	2	0
Implement model	0	0	2	1	0	0
Run & evaluate model	0	1	2	0	2	0
Interpret results	0	0	1	1	2	0

Table 4: Sum of learning goals in a particular step and Bloom level tested in the MCI v2 questions. Bloom levels from left to right, modelling steps from top to bottom.

The results, as seen in Figure 2, were varied. Individual scores rose and fell, but the net change in aggregate score of all questions and students fell well within the margin of error. The correlation coefficient between before and after scores was also very low, at just 0.45. Scores on the 'before' MCI were no clear predictor of scores on the 'after' MCI. Fortunately, the worst-scoring student 'before' improved more than any other. In total, standard deviation in scores on a 0-10 scale dropped from 3.80 to 3.68, indicating a slight convergence. This suggests that as many misconceptions evolved during the course as were cleared up, a sobering thought for any educator. Students comments at least suggested that they better understood the questions in MCI v2.

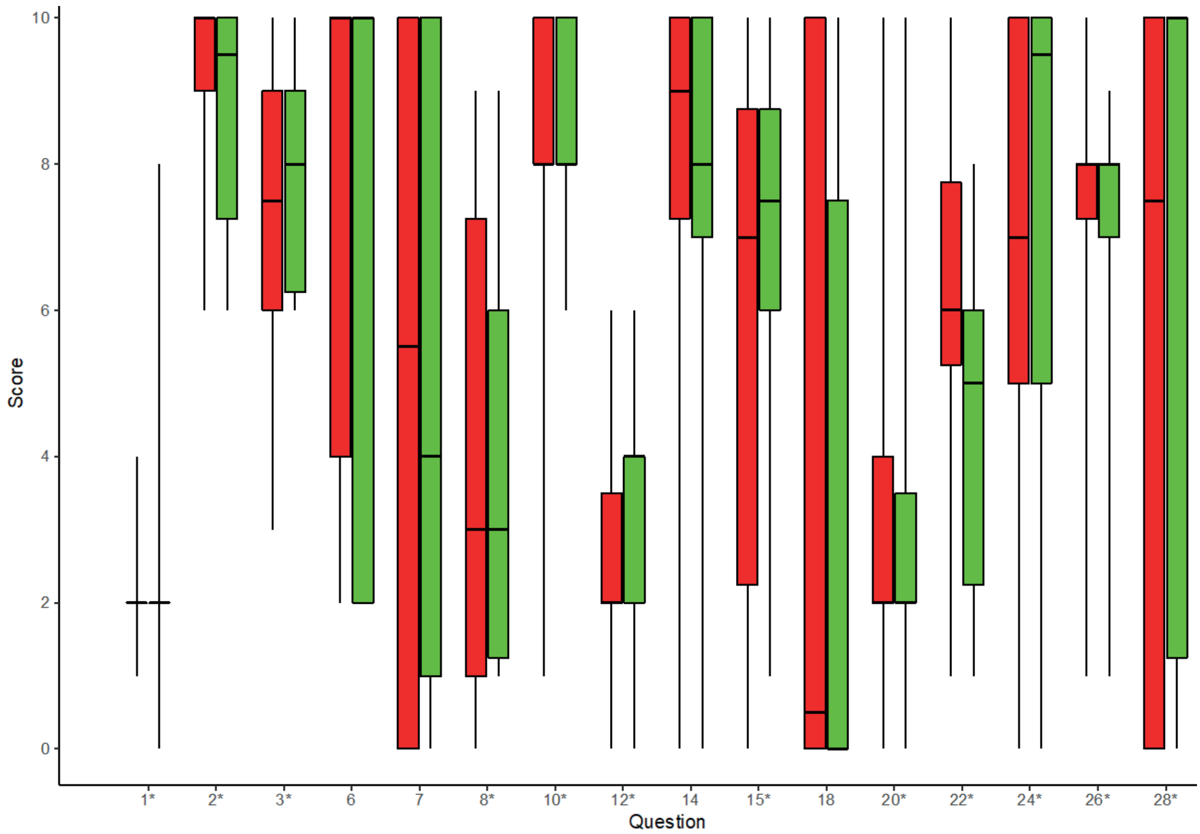

*Figure 2: Boxplot of MCI v2 scores per question before (left, in red) and after (right, in green) the course segment (n = 14). Question marked with * were addressed in the practical modelling course segment. The coloured boxes show results from 25% to the 75% quartile, and the black line in the middle shows the median. Thin lines show the full range (minimum to maximum) among all students.*

Looking at only the questions that tested learning goals that were addressed in the practical modelling course segment that the students took (1, 2, 3, 8, 10, 12, 15, 20, 24, 26, and 28), we see an overall increase in score, with a still very low correlation of 0.48. The rise is due to overall higher scores in just two of the questions (see Appendix B for full descriptions):

12. Which two tasks do you expect to be most time-consuming of the following [activities in the modelling process] (choose two answers)?
28. Can a model result show you something that you didn't expect?

These questions were both directly addressed repeatedly in the course segment, 12 by having the students work and expand on a small energy systems model, and 28 by having the students repeatedly draw conclusions from their own modelling and present their findings. This suggests that repeated practical training of learning goals is more effective at improving students' understanding than lectures.

We also observed a reduction in scores after the course segment when specifically asking about a type of model that was not used in class:

22. What limitation(s) do you think most restrict(s) the usefulness to policymakers of computable general equilibrium (CGE) model projections the most?

This was surprising, as the learning goal tested ('Discuss advantages and disadvantages of different modelling methods applicable to a given dataset and research question') was the same as one of those tested in question 28. We can suggest that the environmental science students were more familiar with climate models than CGE, and therefore felt that they could extrapolate from their experience more easily even though both classes of models have similar properties at the level of abstraction in the two questions, but no student commented on their answers this this question.

4 Lessons learnt

Finding the learning goals in our curriculum and developing the MCI has increased our understanding of modelling in the curriculum.

In curriculum reform in general, we found that the fact-finding process alone was very helpful for curriculum development, as talking to lecturers and having a workshop got lecturers started on aligning courses and gaps. Similarly, knowing where the learning goals are in our conceptual framework (see Table 3) has revealed gaps and overlaps between courses without even testing the students. More generally, having a curriculum-wide set of learning goals could support reflection and course planning by both students and lecturers (Wijngaards-de Meij & Veenhoven 2016). We conclude that the preparatory work for our MCI, particularly the interviews and a workshop with lecturers, is a good basis for a community of practice, and already provides sufficient input to begin aligning courses. The learning goals we found are ambitious when compared to the Dublin level descriptors for a bachelor's degree, though critical thinking is also an explicit overall learning objective at ETH.

Only for step 1 (Acquire System Knowledge) do lecturers consistently rely on students having already acquired competencies, for the rest they mostly start from scratch in each course. In the authors limited and anecdotal experience, this is because some of the course work involves following 'recipes', which allow students to construct a model without truly understanding what they are doing. While this is a form of learning by doing, we conclude that a recipe needs to be accompanied by reflection and supervised independent practice for a student to reach Bloom levels 3-4. Ideally, this is followed in the master program by combining the experience from several modelling courses to facilitate reflection on modelling in general (Bloom level 5, as opposed to remembering the drawbacks of specific models, which falls in Bloom level 1).

For the MCI, we found in testing that the phrasing of the questions is very difficult; the abstract and varied nature of modelling does not lend itself to making a generic test on modelling competencies. This seems especially pressing in environmental sciences, as it uses models from many different disciplines (e.g. biology, chemistry, physics, economics, psychology). Despite this diversity, we conclude from the general agreement between lecturers about our 75 learning goals, that there are many more similarities in the practice and even methods of computational modelling than differences between models of different disciplines. Unsurprisingly, the learning goals also strongly overlap with the core skills identified for students in disciplines that contribute to environmental science (see the Biology, Economics, and Sociology sections developed by the Measuring College Learning panels, 2016).

A complicating factor for interpreting the MCI results is that students may be more hesitant than professionals to extrapolate their experiences from one genre of modelling to another. This may be a question of confidence as well as experience. We also find this lack of confidence in students' limited ability to apply statistical methods to problems sets that are slightly different than a previous example. To test student populations that are not confident in their modelling competencies, we therefore recommend developing an MCI that tests the same competencies to facilitate monitoring at a curriculum level, but with differentiated questions. That is, the learning goals tested stay the same, but the phrasing of the questions is adjusted to examples from the course during which the MCI is administered. For example, the questions may use aquatic biology models instead of energy systems models. This seems similar to the context-specific performance described in Musekamp et al. (2014).

In addition to customising MCI questions, improving students' confidence in their modelling abilities seems possible. For example, in statistics, providing a quiz session in which students have to practice and explain their method choices repeatedly in quick succession can build insight and confidence. We suggest a similar approach could be taken to application of specific model types.

Students also disliked the length of the MCI questionnaire, most of them taking over half an hour to fill it out. The 17 substantive questions in various forms (e.g. multiple choice, single choice, ranking) were seen as a burden, especially with the added text fields in our development version. Having to fill out the same questionnaire several weeks later was not welcomed, even though we explained to the students that we needed before and after samples. Keeping the MCI as short as possible is recommended. We suggest choosing some 10 competencies among those taught that are important for the interconnections between courses, and focussing questions on those few competencies.

As our curriculum-wide project is still ongoing, we envision the following next steps:

- Use the learning objectives to begin aligning courses, with a focus on gradually moving up in Bloom levels, and ensure that basic courses provide all the competencies needed for advanced courses, or else help these advanced courses integrate the teaching of the missing competencies.
- Iterate on the MCI, as we expect that several more iterations and testing on different groups of students will be needed to develop an MCI that can be applied across an entire (modelling) curriculum including an differentiation to connect to recent courses.

Acknowledgements

The authors thank the modelling lecturers in the Department of Environmental Systems Science at ETH who offered their time and enthusiasm: Nele Schuwirth, Meike Vogt, Monika Niederhuber, Claude Garcia, Christoph Baumberger, Reto Knutti, David Bresch, Christian Pohl, Tony Patt, and Bernhard Wehrli. We also thank the students in the Praktikum Anthroposphäre in spring of 2018 and 2019 who filled out the MCI.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R. et al. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom’s Taxonomy of Educational Objectives. New York: Addison Wesley Longman.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. New York: David McKay Company.

Bosma, R. H., Phung, L. D., Le, A. V., Ngo, A. T., Tran, H. M. T., Pham, S. H. et al. (2016). Making Agriculture Curricula Competence-Oriented While Inserting Issues on Gender and Climate Change at Vietnamese Universities. *Tropicultura* 34; 2016, p. 3-19. https://library.wur.nl/WebQuery/wurpubs/fulltext/405294

Center for Curriculum Redesign (2019). Competencies/Subcompetencies framework. https://curriculumredesign.org/framework/

Champagne Queloz, A., Klymkowsky, M. W., Stern, E., Hafen, E., Köhler, K. (2017). Diagnostic of students' misconceptions using the Biological Concepts Instrument (BCI): A method for conducting an educational needs assessment. *PLOS ONE* 12(5); 2017, p. e0176906. https://doi.org/10.1371/journal.pone.0176906

CTETH (2019). Critical Thinking. ETH Zürich Website accessed on 28.08.2019: https://ethz.ch/en/the-eth-zurich/organisation/executive-board/president/critical-thinking.html

Doran, G. T. (1981). There's a S.M.A.R.T. way to write managements' goals and objectives. Management Review, 70, 35–36. Dörnyei, Z. (2001). Motivational strategies in the language classroom. Cambridge: Cambridge Language Teaching Library.

European Consortium for Accreditation. Dublin Descriptors. Website accessed on 26.02.2020: http://www.ecahe.eu/w/index.php/Dublin_Descriptors

Henning, H. & Keune, M. (2007). Levels of Modelling Competencies. In: Blum, W., Galbraith, P. L., Henn, H. W., Niss, M. (eds). Modelling and Applications in Mathematics Education. Boston: Springer.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30. Online: https://doi.org/10.1119/1.2343497

Measuring College Learning panels (2016). Essential Concepts and Competencies: Disciplinary Frameworks for Teaching, Learning, & Assessment. The Social Science Research Council. http://highered.ssrc.org/wp-content/uploads/MCL-Frameworks_Compiled.pdf

Musekamp, F., Spöttl, G., Mehrafza, M., Heine, J.-H., Heene, M. (2014). Modeling of Competences for Students of Engineering Mechanics. *International Journal of Engineering Pedagogy (iJEP)*, 4(1). https://online-journals.org/index.php/i-jep/article/view/2917

Stanford University. Learning Goals. Stanford University Website accessed on 2.7.2019: https://teachingcommons.stanford.edu/resources/course-preparation/creating-syllabus/learning-goals

Volk, B. (2019). Backward Course Design. ETH Zürich Website accessed on 2.7.2019: https://ethz.ch/en/the-eth-zurich/education/educational-development/unterrichtsgestaltung/backward-course-design.html

Wijngaards-de Meij, L. & Veenhoven, J. Improving the Coherence in the Psychology Curriculum: a Pilot Study. *The Future of Education*. Conference Proceedings: libreriauniversitaria.it

Appendix A: Crowdsourced misconceptions

The rest of this section are direct quotes from the replies we received to our call for students' misconceptions in modelling among lecturers:

'- Being not aware enough that what we learn about the model behavior, e.g. with sensitivity analyses, might not necessarily apply to the modelled system, because models are usually a very simplified description of the very complex nature.
- Mechanistic ecosystem models usually rely on an empirical descriptions of input-output relationships, e.g. for the description of temperature dependence in process rates. Even though we mention that there are different possibilities for describing such relationships and teach several alternatives, students often just remember the formulation we used in our examples and then think that this is the way in which it should be done, instead of being aware that it is just one option out of several.

Other misconceptions I often encounter when talking with practitioners (not students) is that people either overestimate the predictive capacity of models or they totally mistrust any modelling result. Both reactions are not adequate but show the difficulty to communicate the value of models.'

'the most important misconception that I have met in my Biogeochemical Modeling Class is that once you've physically coded your model, your work is done - whereas clearly, this is where your work actually starts. During the past few years, I have had comments in our course evaluation where students said that they were surprised to learn that there is something called "model evaluation" and "model-data inter-comparison", and that we put so much emphasis on the analysis and interpretation of the model results, because clearly, the aim of the modeling course was to learn how to program, or wasn't it?

Since then we explicitly highlight that the analysis, validation and evaluation phase is the aim we work towards, and that the programming serves as a tool to answer our research question, and to test the hypotheses that we have posited at the start of the project. This is tough for students to understand at fist while they are still struggling with the elementary programming skills. So: Programming is not modeling.

Another common misconception (that, sadly, certain publications show that many scientists seem to share) is that one model can be made to fit all questions, whereas we now emphasize in our course that the model complexity and set-up has to fit the purpose of the work, and needs to include process formulations, and a well set-up spatio-temporal grid, resolution, or certain considerations about the uncertainty in your parameters, etc. that allow you to actually test your hypothesis against some sort of null hypothesis.

All in all, what students seem to struggle with is to link a new methodology "modeling" with the scientific method, which of course should not differ between research done using model simulations, and that conducted in a laboratory.'

'Students are often not very well aware of the fact that different types of purposes (e.g., prediction/projection vs. attribution/explanation, global vs. regional, long term vs. short term) may require different types of models. The reason is that they do not see that models are tools that are more or less adequate for specific purposes. Moreover, students sometimes take as simply given the data that are used to calibrate and to evaluate models and do not acknowledge that the data themselves are typically the result of complex modelling processes. Another thing that comes to my mind is that students do not always clearly see how modelling, simulating and experimenting are related. As mentioned by [other lecturer], we discuss such questions in our seminar “Philosophical Issues in Understanding Global Change”.’

‘One thing I observe a lot (and I had this myself when I started twenty years ago with modeling) is that students a) use a model without asking or understanding why they are using that particular model, and not some other model, and that b) they quickly think the model is reality. So the whole thing of the model being a tool to learn, to explore the consequences of specific assumptions, the model being an idealization of reality, with the scientist making choices about what can be neglected, what is parameterized, what is resolved, and then the actual uncertainty not being in the model itself, but in the inference step of transferring the model results to reality, they never or rarely think about those questions. We address some of those questions in our Philosophical Issues Seminar (with [other lecturer]) but it’s not something that is covered a lot in other courses I think.’

‘As for (mis)conceptions, its very diverse, ranging from students with a belief that they would not need solid programming skills in their career, some hesitation to get one’s hands dirty to some students fully embracing and throwing themselves into… [other lecturer] and I teach modeling very in an embedded fashion, making students aware not only of model limitations, but also to the (social) context in which they are (co-)developed and applied (as tools to offer facts to underpin decision-making, not more but also no less than that).’

‘More an open question than a misconception: Does a model represent real reality or a constructed reality?’

‘I would say that they don’t “get” the idea of starting really simple, with what they know, and gradually increasing complexity from there.’

‘Misconception: A model is something like a mathematical problem – you solve it and then the job is done. The model “works”. The iterative and explorative aspects of modeling are not in the DNA of my students in semester 5.’

Appendix B: MCI version 2

Each MCI question has the following:
Learning goal(s): What are we trying to test? These goals are drawn from the list in Table 2.
Situation: Description of what the students should evaluate
Question(s): What the students are asked to consider
Options: Multiple options that the students can choose, usually including a ‘none of the above’.
Answer: How we should rate the answers.

Most questions were followed by a text field with the heading: ***Why did you choose this answer? Feel free to explain your reasoning.***
Scoring the answers was done so that the minimum points for each answer was 0, and the maximum was 10. For multiple choice questions, points were distributed so that the total reaches 10. For questions with a maximum number of answers (e.g. ‘choose two’), points were subtracted for exceeding the number of answers. Selecting ‘I don’t know’ yielded 1 point. For question that required ranking answers, the correct order(s) received 10 points, with 1 point deducted every step an answer was in the wrong place.

Learning goal: 133

Situation: You have to write an energy-economics model that finds the cheapest scenarios for electricity production in your country.

Question 1: In what order would you rank the kinds of input parameters that will go into this energy-economics model? (order from most common to least common)

Options:
1. Indisputable facts
2. Opinions (relevant and with justification)
3. Guesses
4. Scientific results from previous work
5. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: In order of common to least: 2, 4, 3, and 1.

Learning goals: 221, 222 (for both question 2 and 3)

Situation: There are generally two types of computer models. The first type is called optimisation models, where you set a number of conditions and limits, and the model optimises on a single variable. For example, you give it the list of ingredients for cookies and the prices of these ingredients at different shops, and the model tells you which brands to buy in order to make the cheapest cookies. The second type is called simulation models, where you set behaviours and initial conditions, and then let it run on its own and see how the system develops. For example, you give a robot the recipe for cookies and put it in a well-stocked kitchen, and see how the cookies turn out.

Question 2: Which of these are optimisation models:
Question 3: Which of these are simulation models:

Options:
1. A route planner like in Google Maps or Apple Maps on your phone.
2. An algorithm that decides on what computer characters in games like The Sims, Sim City, or Civilisation do.
3. An energy systems model of all possible renewable sources and grid lines, that calculates the least-cost way to build to carbon-free electricity system.
4. An atmospheric model with all the chemistry and physics of rain clouds, that calculates the likelihood of rain tomorrow.
5. A financial derivatives trading model that suggests to a trader whether the price of a stock will go up or down.
6. An actuarial model for that calculates the health care insurance premium that you should pay based on your age and lifestyle.
7. None of the above are correct.
8. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 and 3 are optimization, 2 and 4 and 5 are simulation, 6 is too simple to be classified really, but could be argued either way.

Question 5: Please add another example of an optimisation or simulation model (and describe why it is one or the other).

Learning goals: 111, 122

Situation: You use Newton's Second Law of Motion in a model that calculates speed or force over time.

Question 6: What is the unit of F in this related formula?
$F = m \cdot a$

Options:
1. J · s
2. Kg · m^2 / s^2
3. N
4. J
5. None of the above are correct.
6. I do not remember enough high school physics and math to answer this question.

Answer: N is most commonly used answer, J · s is equivalent, Kg · m^2/s^2 is incorrect and should be Kg · m / s^2, and J is incorrect.

Learning goal: 434

Situation: You use Newton's Second Law of Motion in a model that calculates speed or force over time (same as question 6).

Question 7: The following formula is incorrect; You can make it correct by removing one of the elements or adding one in the square box at the end. What would you remove or take away to make the equation correct?
$\int_0^t F \cdot a_t \cdot dt = m \cdot \ v_t \cdot \square$

Options:
1. F = force
2. a_t = acceleration at different times
3. ∫dt = integral with respect to time
4. m = mass
5. v_t= velocity at time t
6. None of the above are correct.
7. I do not remember enough high school physics and math to answer this question.

Answer: a_t should be removed, or added on the other side.

Learning goal: 233

Situation: Integrated Assessment Models have often evolved over decades and have been used to analyse impacts of, for example, our choice of energy sources on climate, land use, water availability, air pollution and food scarcity.

Question 8: Who selected most of the natural and economic relations contained in these models? (order from most inputs selected to fewest inputs selected)

Options:
1. Policymakers and other clients.
2. Government panels of experts.
3. Lone post-doctoral researchers, sitting at their desks.

4. Groups of modellers, at their institute meetings.
5. Someone else entirely.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: nearly entirely 3); 4) and 2) sometimes happen, 1) is very rare but would be better. 5 is this question's version of 'none of the above'.

Learning goals: 222, 242

Situation: Imagine you have an optimisation model for the Swiss electricity system that will calculate the cheapest technically possible electricity system based on data about electricity demand, costs, land availability, CO_2 emissions, and transmission line bottlenecks.

Question 10: What can you do with such an optimisation model?

Options:
1. Generate scenarios for future power plant construction.
2. Calculate the electricity mix with the lowest costs.
3. Evaluate different combinations of electricity sources.
4. Change public attitudes towards renewable electricity.
5. None of the above are correct, but you can do something else with it.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 2 is correct; 3 needs more info than you can out in a model; 1 is wrong as the scenario are input, and 4 is way out there because it would need some really awesome visualisations and a message that non-technocrats care about.

Learning goals: 231, 334, 336, 342, 532 (these cover the tasks described in the answers)

Situation: Imagine you are going to build a reasonably complicated model in your chosen field (meteorology, ecology, energy systems, economics, or something else) and use it to answer a question from a policy maker.

Question 12: Which two tasks do you expect to be most time-consuming of the following (choose two answers)?

Options:
1. Gathering input data, such as national energy statistics, consumption elasticities, chemical reaction coefficients, or species population data
2. Programming the model code, including parsing your data and interfacing with different code libraries
3. Calibrating the model, i.e. adjust inputs so that the outputs match the real world
4. Analysing and visualising the outputs, to produce human-understandable tables and graphs
5. Adapting your research questions to what you find and running your model again with adapted inputs to get new results.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 and 5 are nearly always the worst; 3 can be awful, depending on the model, 2 and 4 are usually the lightest tasks unless you want cinematic outputs.

Learning goals: 452 (question 14), 431 (question 15)

Situation: Suppose you have model that projects climate change until the end of the century. This model calculates the climate from physical and chemical equations and parameters for those equations, as well as assumptions on the volume of different greenhouse gasses that are emitted from burning fuels, agriculture, and other human activities.

Question 14: What do you think is/are the most common way(s) of analysing the likelihood of specific climate outcomes in this climate model?
Question 15: What do you think is/are the most informative way(s) of analysing the likelihood of specific climate outcomes in this climate model?

Options:
1. Vary input parameters into chemical and physics equations that describe the natural mechanisms represented in the model.
2. Leave out some of the chemical and physics equations that describe the natural mechanisms represented in the model.
3. Vary assumptions about volumes of greenhouse gas emissions.
4. Vary the way outputs are described and visualised, i.e. change graphs and phrases.
5. None of the above are correct.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 is common and useful, 2 is rarely done because you'd want all of the them turned on, 3 is very common and probably has the widest influence, 4 is only done implicit in the framing of results but interesting.

Learning goal: 451

Situation: Suppose you have model that projects climate change until the end of the century. This model calculates the climate from physical and chemical equations and parameters for those equations, as well as assumptions on the volume of different greenhouse gasses that are emitted from burning fuels, agriculture, and other human activities (same as questions 14 and 15).

Question 17: Match the options 1-4 in question 15 to one of the following categories of uncertainty:

Options (repeat per Q1 line):
A. Uncertainty from imperfect internal/endogenous parameter values
B. Uncertainty from choices in external/exogenous parameter values (e.g. scenarios)
C. Uncertainty from suitability of the model design
D. Uncertainty from imperfect interpretation of results
E. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1A; 2C; 3B; 4D

Learning goal: 451

Situation: Suppose you held a survey among 10000 people and asked two questions, 1) how worried they were about environmental degradation and 2) how they see themselves in the political spectrum (left or right).

Question 18: If the result is that being worried about the environment on 1) is associated with leaning politically left in 2) with a high significance ($p = 0.004$), what does that mean?

1. People who lean left politically also tend to worry about environmental degradation.

2. Being worried about environmental degradation often leads people to lean left.
3. Politically leaning left often leads people to worry about environmental degradation.
4. Something causes people to both lean left and worry about environmental degradation.
5. Leaning left and environmental degradation have nothing to do with each other.
6. None of the above are correct.
7. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 is correct, 2, 3, and 4 assume different causal relationships that the questionnaire cannot prove, 5 is unlikely because of the high significance and chance that the null hypothesis is true.

Learning goal: 551

Situation: You colleague made a novel economic forecasting model.

Question 20: What do you think are sufficient reasons for to consider this model suitable for making economic forecasts (choose as many as you think are needed):

Options:
1. Model is mathematically correct.
2. Model reproduces (historical) real world dynamics.
3. Model gives the same results as other models in the field.
4. Model gives a result that no other model has given before.
5. None of the above are sufficient.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: None are objectively sufficient; 1 is almost orthogonal, 2 is most often cited but historical accuracy could be an accident, and the future has different dynamics than the past (see: 2008 financial crisis); 3 and 4 are both good reasons if you want to publish a paper that uses it. Note that 3 and 4 contradict each other.

Learning goal: 251

Situation: There is a class of economic models, called computable equilibrium models, which can calculate how money, jobs, and/or resources flow between different sectors of the economy. These models use large datasets of economic statistics. Changing some of the equations that describe how an individual sector uses resources and produces goods and services allows the user to analyse questions like 'What would happen in the rest of the economy if this industry changes because of X?' The results would include rates of growth and decline of industries, change in GDP, and unemployment. Such models have been used to project the future state of the economy in many countries.

Question 22: What limitation(s) do you think most restrict(s) the usefulness to policymakers of these projections the most?

Options:
1. It can project in the short-term only, due to unknown trends that are not captured in the model ('exogenous' trends).
2. The change that we test in the model has to not also change everything else in the economy.
3. The quality of the economic statistics used as a basis for the model.
4. An inability to calculate anything about who gets profits, i.e. not answer to distributional / fairness questions.
5. None of the above are correct.

6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 is probably the largest issue, 2 is also an issue, 3 more of a reason why economists would not make that model rather than a problem with outputs, and 4 is true but not the epistemological point of CGE models.

Learning goals: 232, 243

Situation: In a classic model of predator-prey dynamics, we have rabbit and foxes that live together. Rabbits and foxes breed. Rabbits have enough food, foxes survive on eating rabbits, and the death rate of rabbits depends on the number of foxes.

Question 24: What can you say about this model and the equations in it?

Options:
1. Birth rate of rabbits depends on the number of foxes.
2. Births of foxes rise together with the birth of rabbits.
3. Deaths of foxes depend on the number of rabbits.
4. The long-term average of foxes and rabbits remain stable.
5. None of the above are correct.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 1 is wrong, depends on the number of rabbits, which in turn and depends on the number of foxes and other things; 2 is wrong because there is a lag; 3 is true because foxes need to eat; 4 is true because the model oscillates.

Learning goal: 551

Situation: Models are often used to influence the way people talk and think (i.e. social discourse). For example, models are used to produce economic forecasts or inform warnings about climate change.

Question 26: What roles do such models play in the way people talk and think about the economy, the environment, society, and their futures (choose as many as you think is correct)

Options
1. Models provide great detail on how the future will play out
2. Models are used to support arguments
3. Models are wrong, so they often create mistrust
4. Models help reduce uncertainty and confusion
5. None of the above are correct.
6. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 2 is correct; 4 is mostly wrong but what we hope for; 3 is said by critics; 1 is how non-modellers often seem to see it.

Learning goals: 251, 541

Question: Can a model result show you something that you didn't expect (choose one answer)?

Options:
1. Only if you don't know the model you work with

2. Yes, for some models but not for others
3. Always, that is why we use them in the first place
4. None of the above are correct.
5. I do not understand the question and its concepts well enough to give a sensible answer.

Answer: 2 is correct, and it applies to simulation models; 1 is always true for linear optimisation from models but not necessarily for others; 3 is untrue, just a version of 1 for someone very clueless