

The Feedback Loop: How Student Evaluations Shape Teaching at AELO Swiss Academy

On a Tuesday evening in week five of the autumn term, a statistics instructor at AELO Swiss Academy opened her mid-course survey results and felt her stomach drop. Roughly six in ten students said they had been lost since week three. Not bored, not unmotivated, lost. The free-text comments told the same story from a dozen angles: the pace had picked up around hypothesis testing, and nobody had wanted to be the first to say so out loud.

She had two options. She could treat it as a bad week and push on, trusting the exam revision sessions to sort things out later. Or she could rebuild the next two lessons around worked examples and a gentler runway into the hard material. She chose the second path, and by the final evaluation the same cohort rated clarity of instruction among the strongest parts of the course.

That, in miniature, is the feedback loop we try to run at AELO Swiss Academy. It is not glamorous. It is a survey, a conversation, a change, and a follow-up, repeated until it becomes a habit. Run consistently, though, it is the most practical tool we have for keeping teaching honest, and the one that has changed more classrooms here than any training day or curriculum rewrite.

Why we ask twice: the mid-course pulse

Most schools gather course evaluations at the end of the term, after the grades are in, when the feedback can only help future students. We do that too, and we value it. But an end-of-term survey always arrives too late to rescue the people who wrote it. So a few years ago we added a second touchpoint: a short mid-course pulse, sent in week four or five, that takes about five minutes to complete and is anonymous by default.

The pulse asks four questions:

- How does the pace feel so far: too fast, about right, or too slow?
- Which topic or activity has helped your learning the most?
- Which part is giving you the most trouble right now?
- If we could change one thing for the second half of the course, what should it be?

Four questions, deliberately blunt. Anything longer and completion rates sag; anything vaguer and the answers become too soft to act on. The timing is where the value lives. Week five is early enough that a course can still change direction, but late enough that students know what they have signed up for. Response rates for the pulse usually land somewhere between seventy and ninety percent, comfortably above what we see at term's end, partly because it is short and partly because students have learned that answering it actually changes something.

Designing questions that get honest answers

The wording of an evaluation question does more work than most people expect. Early in our process we leaned on the classic items: overall satisfaction, a rating for the instructor, a box for comments. The numbers came back flattering and nearly useless. A satisfaction score of 4.2 out of 5 tells you nothing you can teach differently on Monday.



So we rebuilt the question set around behavior rather than sentiment. Instead of asking whether students like the instructor, we ask whether each session connects clearly to the previous one. Instead of asking whether the materials are good, we ask whether students could complete the weekly tasks using the materials provided. These are things a student can verify from their own week, which makes the answers concrete and, crucially, comparable across courses.

Anonymity is the other half of honesty. We only process results for groups of five or more students, because in a class of three, every comment is guessable and students know it. For smaller cohorts, the instructor runs a structured live feedback conversation instead, with the same questions asked aloud and the teacher stepping out of the room for part of it. It feels old-fashioned, and it works.

One more detail took us years to appreciate: we keep the instructor questions and the course design questions in separate blocks, and we tell students why. When people believe a survey exists to help their teacher rather than to judge them, the tone of the comments

changes. People stop performing and start troubleshooting. A student who would never write "the teacher goes too fast" when they suspect the survey feeds a personnel file will happily write "the Thursday examples lost half the room after minute forty" when they believe the survey exists to help them learn. That shift, from verdict to suggestion, is the whole game.

From spreadsheet to Monday morning

Every pulse lands in a shared dashboard that the instructor and the program lead look at together. Within a week of the survey closing, the two of them sit down for twenty minutes with the results on screen. The discipline in that meeting is simple: read the pattern, not the loudest comment. If eight students say the Thursday workshop keeps running out of time, that is a design problem with a design fix. If one student writes a paragraph about the seating, it goes into the notes for the facilities team and nowhere near the teaching plan.

The instructor leaves the meeting with at most two changes to make. Two is a deliberate ceiling. In our early years we sometimes handed teachers a full page of suggested improvements, and the predictable result was that nothing changed, because a teacher mid-term can absorb one or two meaningful adjustments, not ten. Better to fix the weekly recap handout properly than to half-fix five things at once.

The statistics course from the opening story is a good example of the ceiling doing its job. That meeting produced exactly two decisions: rebuild the next two sessions around worked examples with a visible [pilot demand forecast](#) solution path, and open each class with a ten-minute structured review for a fortnight. Nothing

else. The instructor told me later that the rule of two was the only reason the changes actually happened. With permission to fix everything, she would have fixed nothing.

You said, we did

The step most institutions skip, and the one that makes our loop work, is telling students what changed. Within a week or so of the pulse, every instructor opens a lesson with a short "you said, we did" moment. It takes three minutes. "You told us the pace jumped at hypothesis testing. Here is what the next two weeks look like instead, and here is a recap sheet you can use in the meantime."

This small ritual does three jobs at once. It validates the students who filled in the survey, which keeps response rates high next time around. It shows the skeptics that the survey is not a filing cabinet. And it quietly teaches people how to give useful feedback, because they watch which comments produced action and which did not. Over a term, the free-text answers get noticeably more specific. By the final evaluation, comments read like notes from a careful observer rather than steam let off at the door.

The awkward cases, and how we handle them

No honest account of student feedback is complete without the messy parts. Some evaluations are unfair. Some are careless. Occasionally a comment has nothing to do with teaching at all and everything to do with a grade a student did not like, or worse, with bias about an instructor's accent, age, or background. Pretending this away would be dishonest, so we have a few working principles, learned the hard way:

- Read patterns, not outliers. One harsh comment is information about one student; the same comment from several is information about the course.
- Separate delivery from message. A rude comment can still contain an accurate observation worth acting on.
- Act on behavior and design, never on identity. "The lectures run ninety minutes with no break" is actionable. Comments about how someone looks or sounds are not, and they get removed from any material the instructor sees.
- Check context before concluding. When one cohort's numbers look strange, we compare them against the same students' other courses before anyone draws a lesson from it.

The program lead plays an important filtering role here. Instructors see the results after this pass, not the raw feed, which protects them from the two or three lines in every survey season that exist only to wound. Some colleagues initially felt this was paternalistic. Most changed their minds after their first encounter with a genuinely spiteful comment, and after seeing how much easier it is to discuss real criticism when it is not buried in noise.

What feedback can and cannot fix

Some of what the pulse surfaces cannot be solved by the instructor at all. Room temperature, timetable collisions, a workload pile-up caused by three courses assigning essays in the same week. Those items go into a separate program-level review that the academic team at AELO Swiss Academy runs at the end of each term, and the fixes there tend to be slower and less visible. Being straight with students about this boundary matters. If twelve people ask for smaller classes and smaller classes are not on the table this year, we say so in the "you said, we did" moment rather than letting the comment vanish into silence.

There are also judgment calls about whether to act at all. One evening cohort asked us to drop a group project entirely, because coordinating schedules around jobs was proving painful. The instructor and the program lead talked it through and decided the project served a learning purpose worth protecting, so they changed its shape

instead: pairs rather than groups of four, with dedicated working time inside the scheduled session. The follow-up pulse showed satisfaction with the project rising sharply. The lesson we took from that term is one we repeat often internally: feedback tells you where the friction is, not always what the fix should be. The loop informs professional judgment. It does not replace it.

The long arc: term-end results and instructor growth

At the end of each term we run a fuller evaluation, longer than the pulse and with items that can be compared across the program. Those results feed three things: the design of next term's courses, the program-level review, and each instructor's development conversation. Instructors see their own trend lines across terms, and that is where the real value shows up. A score that wobbles once means little. A score that drifts downward for three terms starts a genuine conversation, usually a constructive one.

A colleague teaching economics gives me my favorite example. His clarity scores sat roughly half a point below the program average for two consecutive terms, on otherwise strong results. Rather than sending him to a generic workshop, we paired him with a fellow instructor whose scores were high on exactly that dimension, and he sat in on two of her sessions with one narrow question: how does she signal transitions between ideas? He noticed she previewed the structure of each session in the first two minutes and returned to it at each handover. He borrowed the technique, adapted it to his own style, and his scores moved up the following term. Small, targeted, peer-based. A one-off training day could never have produced that, because a training day does not know what your specific number is doing.

A habit, not a project

If I had to name the single reason the loop works here, it would be this: we never treated it as a project with an end date. There was no launch event, no rebrand, no poster campaign. It is a survey, a twenty-minute meeting, a three-minute message to students, and a trend line, repeated term after term until nobody can remember teaching any other way.



The payoff is cultural more than technical. Students learn that their voice has weight, which makes them more careful and more generous with it. Instructors learn that feedback is not a verdict on their worth but a running commentary on their course, which makes it easier to hear. And the school learns, term by term, where its real weaknesses live, which is usually somewhere more mundane than anyone expected. Ask our statistics instructor.

The fix that saved her course was not a new teaching philosophy. It was two weeks of worked examples and a ten-minute review at the top of each class, suggested by the very people the course was built for.