

✓ How do you evaluate competences?

Leads comparative judgement to better results?

 $\sqrt{}$ How does the comparing tool work?

THE 28 MOST FREQUENTLY ASKED QUESTIONS ABOUT COMPARATIVE JUDGENENT



We need to talk about (comparative) judgement

In the past decades, (higher) education has changed considerably. In the past, education mainly focused on imparting specific knowledge. Nowadays, we want to prepare our students for a rapidly changing world by teaching them complex skills in addition to knowledge. After all, society is waiting for selfdirected, reflective, problem-solving and cooperative professionals.

Teachers show great commitment and creativity in devising didactic approaches to teach such complex skills. They implement practical assignments, reflection exercises and peer feedback and 'flip the classroom' as if they have never done anything else. Only to run into the assessments of all this. Because what teachers feel from their expertise is confirmed by many scientific studies: when it comes to complex skills, there is often something wrong with the reliability and validity of the classic assessment methods.

However, good assessment is important. We want to know what our education brings to students. We also want assessments that bring to the surface where our education needs to be adjusted and, if possible, create learning opportunities for students.

And so we must talk about assessment. After more than 7 years of research, the pitfalls of conventional assessment have been exposed.

Judging methods no longer hold any secrets for us, the Comproved team. We would like to share this knowledge with you through this e-book. We want to provide alternatives and tools with which teachers can better find their way in the more complex educational landscape. Comparative judgement is one of those alternatives. This method uses the way teachers usually assess spontaneously: by comparing students' work with each other. Research has shown that this leads to very reliable and valid results when assessing complex competences. Moreover, the method creates many learning opportunities, both for assessors and students. Comparative judgement therefore deserves a place in every evaluation policy.

In this e-book, we have combined insights from research, case histories and testimonies into 28 questions about (comparative) assessment. If you want more in-depth information after reading, you can always contact me or my colleagues.

Roos Van Gasse

Comproved

P4 PART 1: CONTEXT

- 1 What is the common way of assessing in education?
- 2 How do you evaluate competences?
- 3 What makes assessment so difficult?
- 4 Is there an alyernative?
- 5 What is comparative judgement?
- 6 How to implement comparative judgement?
- 7 What are the main advantages of comparative judgement?
- 8 When does it make sense?
- 9 Is comparative judgement efficient?

PART 2: COMPARATIVE JUDGEMENT IN DETAIL

- P16 10 For What kind of tasks is this method particularly suitable?
 - 11 What's the amount of product needed?
 - 12 How many assessors do you need?
 - 13 Who can be an assessor?
 - **14** Does comparative judgement take longer?
 - **15** Does comparative judgement lead to better results?
 - 16 Does comparative judgement lend itself to giving (peer) feedback?
 - **17** Does comparative judgement lead to additional learning opportunities?

P23 PART 3: COMPARATIVEJUDGEMENT

WITH COMPROVED

- **18** What is Comproved?
- **19** How does the comparing tool work?
- 20 How to convert ranking to grades?
- 21 What are the main advantages of the comparing tool?
- 22 What kind of assessments can the comparing tool be used for?
- 23 Who can work with the comparing tool?
- 24 How do you 'steer' assessors?

P32

PART 4: SPECIFIC APPLICATIONS

- **25** Can comparative judgement be used in peer assessments?
- 26 Can you also evaluate live with comparative judgement?
- 27 Can you also have group works assessed comparatively?
- 28 How does comparative judgement works with externals?

P37 ABOUT US

PART 1 CONTEXT

What is the common way of assessment in education?

In the past, education focused on knowledge transfer. Assessing whether a student had studied well and had the necessary knowledge was done relatively simply by means of multiple-choice questions, closed questions and fill-in-the-blanks questions.

Since the end of the 1970s, education has focused mainly on competences: the integration of knowledge, skills and attitudes and the ability to apply them in other contexts. Competencies are much more difficult to assess, especially when it comes to complex skills. Does a presentation show originality, creativity and critical thinking? Is a writing assignment strong in content and grammatically correct and good structured? That is much harder to measure.. In an attempt to assess complex skills in the best and most streamlined way possible, teachers and lecturers often use rubrics or criterion lists. In these, the assignment is divided into sub-competencies, each of which is scrutinised separately, and these are then scored.

are added up. Working with such lists of criteria can sometimes be useful, for example in making feedback tangible. But it is also a time-consuming task that often gives a false sense of reliability: there is a good chance that the criteria overlap and are open to different interpretations. Moreover, you push the bigger picture out of the picture by merely zooming in on the parts..

2 How do you evaluate competences?

TIP 1: Start from your learning goals

Evaluation starts with clear learning goals. Were you aiming to teach the students knowledge, or were you aiming more at skills and competences? Think about what you want to measure exactly. Choose tests or tasks that match your learning goals.

TIP 2: Rely on your expertise, not on selection criteria

Suppose you give a separate mark to criteria such as spelling, sentence structure, word usage, layout, content and structure of an essay and then add up the marks. This gives you a 'logical' final score - but one in which you often have the feeling that you are wrong. Recognisable? The quality of a piece of work is indeed not the sum of its parts, but is primarily determined by the interaction between the parts. So look atthe final product and rely on your expertise to make a judgement .

TIP 3: Compare tasks

First go through all the assignments or watch all the presentations. Write down a series of pluses and minuses, but don't grade them yet. Finally, put all tasks in order, from lower to higher quality, and distribute your final scores at the same time..

TIP 4: Work together

The "four eyes principle", where complex tasks are looked at by two assessors, was created to avoid too large individual differences. So ask a departmental or parallel colleague to go through your tasks and make their own ranking.

How does he or she rank the same tasks? Are you on the same page? Discuss what you liked and disliked and come to a consensus. Such discussions are often very inspiring. Exchanging how you deal with certain topics and what you expect from the students is conducive to a better coordination between subjects and teachers.

3 What makes assessment so difficult?

The most common way of assessment is absolute assessment: you examine each piece of work individually, with or without a list of criteria. This method of assessment inevitably brings up a number of bottlenecks that are not only inconvenient but also prevent reliability.

What aspects do you look for?

The first text is smoothly written, the second is full of language errors but better argued and the third has a tight structure. What gives you the edge? What about your own hobbyhorses and blind spots as an assessor? And are you perhaps influenced by aspects that have nothing to do with the skill being assessed, such as handwriting?

What judgment do you make?

You know the whole range of scores, for example from 0 to 20 points, but while you are assessing you do not yet have an overview of all possible variations in quality. You are therefore in the dark about which score corresponds to which quality and, as a result, do not usually use the full assessment scale.

At what point do you judge?

In the morning, you may judge the first writing assignment much more leniently than an equally strong piece of work after a long tiring day. And if you are starving, you are guaranteed to judge a presentation differently than after a tasty snack.

In what order do you proceed??

If the first task is excellent, the tasks that follow may not live up to your expectations and will therefore be judged more harshly. And after a series of poor works, you might be cheered by a task of average level. The order of your evaluations definitely has an impact. Also notorious is the so-called raters' drift: over time, evaluators tend to pay attention to other things.

Who is judging?

When different teachers assess the same piece of work, the conclusions are often very different. One teacher is stricter or milder than the other, or attaches importance to other aspects. Research shows this clearly: if teachers have to give a score out of 20 for the same task, it can differ by up to 8 points from one teacher to another.



Caren Legius Teacher and principal examination commission, Avans Hogeschool (Breda) "As a coordinator of vocational tasks, I supervise a ten-week programme in which 200 first-year students of the teacher training college for primary education do a little research in a school and then pour their findings into a report. In order to assess these reports, we have been using Comproved's comparing tool for a number of years now. The reports are graded and provided with feedback by the entire teaching team. It gives me a really good feeling that the ranking we obtain in this way always does justice to what the student has submitted. The result is unanimous, well-considered and supported by every assessor."

Comproved =

 \vee Reliable

- $\sqrt{\text{Solid}}$
- √ Scientifically proven



4 Is there an alternative?

The good news: there are alternative assessment methods that are more in line with the workings of the human brain. How do you naturally make decisions? Indeed: by making comparisons. If you are going to buy shoes, for example, you are unlikely to go into the shoe shop with a typed-out list of criteria on which you are going to judge all the shoes. What you are more likely to do, and without thinking about it, is to compare pairs of shoes: pairs of shoes in pairs or pairs of shoes with an ideal image in your head. That is comparative judgement. When teachers and lecturers assess, they do the same. Give them one task, and it is difficult to make a quality assessment. Offer them tasks in pairs, ask them to compare.

And yes: they effortlessly answer the question of which paper is the better of the two. Because of their expertise, they quickly see which opinion piece is better argued, which source research is more thorough or which visual work shows more originality and creativity.

5 What is comparative judgement?

Comparative judgement = making comparisons. And you do this almost without thinking. For example, it is impossible to estimate the height of an overflying aircraft. But if two planes fly over, you can quickly see which one is highest in the sky.

The principle of comparative judgement is thus: you

do not assess pieces of work as separate products but in relation to each other. From two tasks you always choose the best. By comparing all products in pairs in this way, you arrive at a ranking from 'least good' to 'top'.

Comparative judgement is the opposite of absolute assessment. In absolute assessment you look at one task, in comparative judgement you situate the task in relation to other tasks. **Comparative judgement takes a holistic approach:** the task is assessed as a whole. This is in contrast to analytical methods that usually work with criterion lists: a piece of work is analysed in detail by looking at partial aspects and partial competences.

Comparative judgement reflects the consensus

between assessors. And that without lengthy discussions, but simply by the statistical model underlying the method (Bradley-Terry-Luce). It calculates a quality scale from lesser to better quality. It is also quickly clear which assessors deviate from the consensus, so who often chooses differently, and about which products opinions are strongly divided. This provides useful information that can be further explored.



Ranking

If you have fewer than 15 or 20 products, you can simply construct the quality scale yourself. You do this by comparing the physical workpieces and ranking them from lower to higher quality. Because it is a number of products, you can work with the whole group and you do not have to work in pairs..

Working with benchmarks

Sometimes you have anchor tasks (also known as bench marks): tasks that have already received a value judgement or mark, for example during a previous school year. You can then compare new tasks with those benchmarks. This can be done manually or through a tool that supports comparative judgement. Then you determine the ranking of the task by giving it a place among the benchmarks.

Random comparative

No benchmarks available? Then the comparisons are simply compiled randomly: you compare each task with a random other task. For a large number of tasks, this can be done digitally using a comparing tool.

What are the main advantages of comparative judgement?

These are the reliability and validity of the method and the results.

This is why:

- Those who assess comparatively make more consistent decisions: no matter what time of day or what tasks you have seen before, the same piece of work will always stand out in a comparison. As an assessor you are fairly certain of your judgement; that gives you confidence.
- Four eyes see more than two, and eight eyes see even more than four. By evaluating together with colleagues, you can more quickly achieve a higher degree of reliability and validity. After all, there is a greater chance that several assessors will choose the same product as the best in a comparison. And because more perspectives are included in the assessment, you assess the competence in its entirety.

No colleagues who can co-assess? Then ask others, such as professionals from the field. Or call in the students. They find it difficult to use criteria lists to assess competences, but they are usually very good at comparing products. Did five students give a presentation in class? Their classmates can perfectly rank it from 'least strong' to 'top'. When using peers, you have to take into account that they may come to a very reliable ranking (because they agree with each other), but that this may deviate from the teacher's assessment.

See also question 15: LDoes comparative judgement lead to better results?



See aso question 16: Does comparative judgement lend itself to give (peer) feedbackk?

When does it make sense?

Comparative judgement is especially suitable for evaluating:

- v complex competences: Think of presentations, portfolios, reflections or visual works.
- open tasks and questions (not suitable for multiple choice or closed knowledge questions),
- V different formats (not only texts, but also films, music, drawings or images).

tasks to be assessed by **multiple assessors. peer assessments**, because it helps induc- tive learning of students: they get a better idea of the quality criteria, which benefits their own work.

> See also question 10: For what kind of tasks is comparative judgement particularly suitable?

Lore Staes

Teacher and Coordinator Thomass More Hogeschool "I coordinate the internships of our students. Each student has a tutor who accompanies him or her throughout the internship process and proposes a grading afterwards which we discuss with the tutor team. These discussions usually take a long time, which made it impossible to organise them in an online meeting. So we came up with the idea to have each lecturer make a one pager of his student's internship trajectory and submit it anonymously, after which the lecturers could assess the trajectory comparatively. During an online consultation, we briefly checked the ranking that came

out of the Comproved tool. In this way, we were able to avoid a lot of discussions and long meetings."

Comproved =

 $\sqrt{\text{Clear criteria and expectations}}$

- $\sqrt{}$ Widely supported assessment by all assessors
- \vee Asource of new insights for all parties

Is comparative judgement efficient?

Do you assume that comparative judgement is rather cumbersome, especially if you have to involve colleagues? Then be sure to read these (scientifically supported) insights.

Judging always takes time.

This is certain: assessment is a time-consuming task in any case, whichever way you go about it. But.....

Comparative judgement is easier.

Do you think you are doing double work because you are judging in pairs? Not true. In the case of pairs with a clear difference in quality, it is quickly clear which product is the best. If the products are of similar quality, then it is logically slower. But the total assessment time will never be longer than with an assessment with a criteria list.

Comparative judgement achieves higher reliability more quickly.

Reliable results require a certain investment of time, independent of the method. But if you want to be sure that your judgement is not too dependent on you or on the moment, comparative judgement is - whether or not together with an extra assessor- faster and with a reliable result.

No more endless lists of criteria.

The biggest time-saver in comparative judgement? No need to develop and validate criteria lists. Comparative judgement relies on the expertise of assessors, which has proven to be very reliable. Moreover, it works intuitively. Assessors do not need to be trained to look at (the same) aspects. ►

► The biggest gain

Because comparative judgement is so intuitive, it is much less perceived as a chore. In fact, most people find it fun to do and they also feel more confident about their assessments. Comparative judgement gives more peace of mind for another reason: others see the work too; it is a shared responsibility.



Veerle Meuleman

Head of Department for Educational, Karel de Grote Hogeschool "We often work with large-scale projects at the college that are worth a lot of credits. When assessing them, we like to work with Comproved's comparing tool. Each assessor must weigh up whether a work scores satisfactorily or unsatisfactorily. Because the assessors must also note why their assessment is positive or negative, they are challenged to put their assessment into words and make it explicit. Combined with the principle that several assessors look at the same work, these opinions - if they differ - can also be expressed effectively. In doing so, we create much more language for the underlying frame of reference that teachers usually have and that is often a blind spot. We give words to something for which there were no words before. To me, that is the biggest plus of comparative judgement."

Comproved=

 $\sqrt{\text{Reliable due to the 'multiple eyes' principle}}$

 $\sqrt{\text{Giving language to aspects that often remain unspoken}}$

 \sqrt{A} substantiated judgement and a rich assessment

For what kind of tasks is is this method particularly suitable?

Comparative judgement is not useful when it comes to closed and straightforward tasks. Think of closed questions, fill-in-the-blank questions or multiplechoice questions.

This way of assessment is especially interesting when the competence is too complex to be captured in one or a few aspects. Think of reflective strokes, action plans, solution strategies, training videos, visual works, music pieces, presentations, theses or portfolios.

See also question 8: When does it make sense?



It is possible - logically - from as few as two products. Up to 15 to 20 tasks can be perfectly arranged manually from least good to best. For more than 20 products, this becomes more difficult. A tool that supports comparative judgement can then be a solution.



How many assessors do you need?

Two assessors are sufficient to rank tasks with some reliability. But assessors can differ greatly in what exactly they value in a piece of work. To ensure that all aspects of compe- tition are taken into account (and that validity is assured), it is advisable to work with at least four assessors.



Laura Dewinter Teacher training primary and nursery school and internship supervisor, Thomas More Hogeschool "Practice committees used to take a long time. Thanks to Comproved they are much more efficient, because it is no longer necessary to discuss each student individually. That saves an enormous amount of time. In addition, it is also enriching to be able to compare your opinion with that of a fellow lecturer. It gives the feeling that determining the grade is not an individual task but in the first place a team effort. Moreover, the result often gives a feeling of relief and confirmation to a teacher: you know the student in question, you have had an internship interview and read his internship report, and based on that you have a judgment ready in your head.

If you then see the same verdict come out of the Comproved ranking, it indicates that you were right in your initial opinion."

Comproved =

 \lor Peace of mind and a shared vision \lor An invitation to self-reflection \lor Time saving



5 Who can be an assessor?

Assessors do not need specific training. Everyone who is considered capable of assessing the competence can become an assessor. This can be experienced teachers as well as starting teachers,

See also question 23: Who can work with the comparing tool? but also (fellow) students or external people from the field, such as for example supervisors. Of course this all depends on the competence to be assessed. Suppose the students were assigned to design an advertisement.

Anyone can judge whether the advertisement in question appeals or not. But to judge whether the advertisement took certain marketing principles into account, the reviewer must have knowledge of marketing..

1 Does comparative judgement take longer?

See also question 9: Is comparative judgement efficiently?

The assessment time is certainly not longer. In the case of pairs with a clear difference in quality, it is quickly clear which product is best. If the products are of similar quality, then it is logically slower. But the total time is shorter than with a assessment with a criteria list

Because criteria lists need to be developed and validated, and people need to be trained to use them properly, comparative judgement actually saves a lot of time.

15 Does comparative judgement lead to better results?

Research shows that comparative judgement effectively leads to high reliability and validity. For example, assessors make more consistent decisions: in a pairwise comparison, they will choose the same product as the best every time. And this regardless of the time of day, mood and order of work. If more assessors are used, the number of assessors increases. the validity and quality of the judgements.

And when the assessors have strong differences of opinion and make divergent judgements, this is only valuable: by exchanging what they think is good and bad about a task, they are challenged to name qualities, something that with other methods is often a matter of chance. In this way, they can also give richer and more in-depth feedback to the students.

See also question 7: Wat zijn debelangrijkste voordelen vancomparat What are the main advantages of comparative judgement?



Inge Van Genechten lecturer in Construction, Thomas More Hogeschool "A great deal of creativity is required of our students, and comparative judgement is, in my opinion, ideally suited to creative contexts. It also provides an enormous and extremely valuable stream of feedback."

Comproved =

 \bigvee Smooth and easy

 \sqrt{An} enormous return for the students

 \sqrt{A} lot of formative feedback in a short period of time

16 Does comparative judgement lend itself to giving (peer) feedback?

Comparative judgement is a powerful way of giving feedback. By having several assessors look at the works, students not only get more feedback than just from their teacher, they also get it from multiple perspectives. This makes feedback from comparative judgement very valuable. Research has shown that students from this feedback provides clear learning points to hone their competences (feed forward). Above all, the method works intuitively, which means that students can also (anonymously) view the work of fellow students and provide them with feedback. This leads to more learning opportunities for students and less time pressure for teachers.

17

Does comparative judgement lead to additional learning opportunities?

No matter how you use the method, it provides great learning opportunities for students.

Do you evaluate with fellow teachers? Then, first of all, you produce learning opportunities through the rich feed-back. Moreover, because the students can see the ranking of the works, they can better assess where they stand themselves. They get the chance to look at better and less good examples of the task and discover why these differ from their own work in terms of competence level. These are two clear learning opportunities.. Do you use the method for peer assessment? Then there are learning opportunities. The (anonymous) assessment of fellow students' work, for example, is a learning opportunity in itself. After all, by weighing the works against each other in a comparison, students learn bottom-up to recognise the most important aspects in quality tasks. By explicitly mentioning them in the feedback they give their fellow students, they also activate this knowledge within themselves, which (hopefully) benefits their follow-up assignments.

Ö

What is Comproved?

Comproved is a spin-off that grew out of a project by researchers from the University of Antwerp, Ghent University and imec. The central question of this research project was: what is the added value of comparative assessment for the evaluation of complex skills? Comproved grew out of this project. Comproved wants to help lecturers and assessors to carry out honest, highquality assessments by giving them knowledge and tooling that are both useful and necessary for the future.

For example, Comproved developed a digital tool that supports comparative judgement. If you have to rank, say, 50 or 100 tasks from 'least good' to 'top', it is not practically feasible to put together random pairs yourself.

The comparing tool automates this process and makes it possible to quickly and reliably compare products in pairs in an online environment.

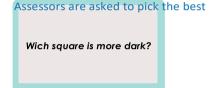




How does the comparing-tool work?

- The students upload their own work into the digital tool, completely anonymously.
- The evaluator is then presented with a series of pairs composed at random and chooses the best of the two each time. Each product is compared equally: the algorithm always selects for a new pair the product that has been compared the least.
- ✓ For even more certainty about which workpiece is the best, the tool uses multiple assessors. Each assessor sees the same products, but in different combinations.
- ✓ Ultimately, the tool brings all this input together.
 The result: a quality scale that ranks the products. ■

STEP 1



H

STEP 2 **STEP 3** Multple assessors make multiple comparisons





FROM LESS DARK TO MORE DARKR



ZU

How to convert ranking to grades?

Have all tasks been assessed and is the ranking established? That is the first step. The next step is to award scores. But how do you do that? The comparing tool does not automatically convert the ranking into points. The evaluators have to do that themselves. There are several possibilities.

- Two tasks one of the best and one of the worst - are given a grade. Then the grades of the other tasks are given via the values in the ranking.
- Two or more random tasks are graded before they are included in the comparisons; these are the benchmarks or anchor points. This way, in the final ranking, it quickly becomes clear which products are situated around certain grades. You can also use tasks from a previous school year as benchmarks. In this way, you immediately build up a stable competence scale that transcends the year.

Who has passed and who has not? The assessors themselves determine where the caesura or pass mark lies, i.e. the turning point between pass and fail. A precise method to determine the cut-off point is to load the word "pass" as one of the tasks to be assessed. The word thus regularly appears together with a piece of work, with which the de facto question is: is this work better or less good than "sufficient"? In this way, the "sufficient" becomes one of the tasks in the ranking: here lies the caesura, or an indication of it. The team of assessors checks whether the position of the caesura is correct and does justice to what the students have learned and could expect. Next, the team assigns a grade to a work at about 2/3 of the grade. Now that two grades are fixed, the other grades can be derived from them.

21

What are the main advantages of the comparing tool?

- The tool is user-friendly. The students upload their own work, after which the tool constructs comparisons and sends them to the assessors. The tool calculates the ranking and provides the feedback to the students..
- As a teacher, you can use the comparing tool to set up and manage assessments quickly and easily. There are also many options for customising the assessment: How many comparisons do you want? What kind of feedback do you ask for (with or without criteria)?
- The tool gives teachers more guidance by basing scores on multiple assessments and combining the expertise of assessors.

- There is also support in the form of a team of experts. The Comproved-Academy shares its knowledge, insights and know-how through publications, presentations and customised advice. Teachers and educational teams who want to use the tool can count on training and guidance. The Comproved-Academy gives webinars and workshops, both knowledge- and practice-oriented.
 - Handy: the comparing tool is available 'stand alone' in the cloud environment, but can also be perfectly integrated in Learning Management Systems (LMS), such as Blackboard, Canvas or Brightspace.

22 What kind of assessments can the comparing tool be used for?

- In summative assessment, where you determine whether a student achieves the expected level of performance, the tool helps to make honest, objective and qualitative judgements. The results are valid and more reliable than conventional methods, because multiple assessors provide multiple perspectives and each product is compared to other products several times. The students can look at their own work, the feedback and the other works afterwards.
- Formative assessment focuses on the learning process and on strengths and areas for improvement. Development-oriented feedback is particularly important here. The comparing tool has a function that this makes extensive feedback possible.

Students also learn a lot from giving feedback on each other's work. The tool is therefore particularly suitable for peer assessments (set up in a jiffy!), where students assess each other's products and provide feedback.

Live judging is done, for example, with physical products that cannot be digitised, such as an installation or a scale model. Each workpiece is assigned a code. Using Comproved, the jury is led through the products 'in real life' and must each time compare two products, indicate the best one and give feedback. The results reflect the consensus of the jury members. ►



- The tool can be used in any context in which complex assessments must be made and a selection made. Think for example of the evaluation of subsidy and project proposals. In personnel recruitment, the tool can help to quickly and efficiently select the best candidates from a large number of CVs. Even when thinking about the vision or mission of a company or department, it can be a useful tool: which themes do we as a group consider most important?
- ✓ With the comparing tool there is no need for the assessors to be physically together or to be assessing at the same time. They can assess perfectly remotely: where and when it suits each of them, completely independent of time and space. This way, there is no need to consider distances or matching agendas.. ■

See also question 25: Can comparative judgement be used in peer assessments?



See also question 26: Can you also evaluate live with the comparative judgement?



Gijs Leenders Teacher Dutch, Amadeus Lyceum (Utrecht) "Comproved makes it easier to check in a short time how your students score in general and in relation to other classes. It is also an extra check to see to what extent teachers agree with each other."

Comproved = \lor Intuitive \lor Fast and efficient \lor Practical and clear



23 Who can work with the comparing tool?

The tool can be used by anyone who can judge the piece of work and its competence.

- Students who assess each other's work and give feedback on it (peer assessments) undeniably reap the benefits. They learn to evaluate the 'overall' quality of a product, develop a clear idea of the expected level of performance and learn to give and receive feedback.
- ✓ Teachers are encouraged to make maximum use of their expertise in assessment, which contributes to the validity and reliability of the results.
- External experts from the field or sector can be brought in to help assess tasks that are closely related to what the students will have to do in practice later on, for example.



24 How do you 'steer' assessors?

Or better: how do you make sure that assessors look at the right things? The comparing tool can be used in different ways to give guidance and points of attention.

V In the description of the task objectives, for example, you can choose to highlight certain subcompetencies from the rubrics.

 You can also give feedback in a structured way.
 For example, with each piece of work you can have them write down plus and minus points or only give feedback on the most important criteria.



25 Can comparative judgement be used in peer assessments?

Comproved's comparing tool offers the possibility of having students assess each other's products and provide feedback..

How does it work?

The teacher sets up the assessment (it takes a maximum of 10 minutes), after which the students receive an e-mail with a link to the platform. They must submit their own work and assess each other work comparatively: they see two works side by side, indicate which is the best and formulate feedback. The tool ranks the works from least good to best based on the comparisons. The students see where their work stands in the ranking and can view each other's work and feedback.

Advantages?

- V Students receive a lot of feedback, and quickly. They also learn to give and receive feedback.
- V By comparing the products of peers, They have a good idea of the expected level of performance and can construct a mental scale of "quality" for themselves
- Seeing different products from peers also makes them better able to reflect on their own work.



26 Can you also evaluate live with comparative judgement?

Comproved can be used when non-digitisable products (such as visual works, models or installations) have to be assessed by a jury.

How does it work?

The products are 'exhibited'. The teacher sets up an assessment and gives a code to each product. If necessary, the students are given the opportunity to explain their work. The jury is guided through the products via Comproved and has to compare two products each time, indicate the best piece of work and give feedback. The results reflect the consensus of the jury members, without any group discussion (and its disadvantages).

Advantages?

- V The intuitive and holistic way of assessment does more justice to the often creative products.
- V Several assessors give their opinion (the so-called multi-rater perspective).
- V Each work receives equal attention.
- Each jury member has an equal impact on the final result. And in the determination of the points, each vote counts equally. There is less opportunity for discussion.

See also question 22: For which assessments can the comparing tool be used?

27 Can you also have group works assessed comparatively?

That is perfectly possible. It suffices that each group uploads its work into the digital tool. The process then proceeds in the same way as for individual pieces of work. The group works are presented in pairs to the assessors, after which the tool puts the input together and ranks the group works. The assessors can be peers or teachers. In the case of peer assessment, individual students thus give feedback to groups of students. The advantage is that fewer comparisons have to be made to generate a lot of feedback, and many learning opportunities for students.



Tim Leeijen-Van Geelen

Applied Sciences (Breda)

Lecturer 'Programme Design' and chairman of

the board of examiners, Avans University of

"In our training, complex skills are key. In large projects, where students have to demonstrate a variety of skills, we can use the comparing tool to give an unambiguous and holistic overall assessment."

Comproved=

 \vee Innovative and convenient \vee A well-founded compromise \vee A wider supporting base

comprov/ec

28 How does comparative judgement work with externals?

If you want to involve external assessors, the comparing tool makes it possible to do so independent of time and space, as the whole process takes place via the cloud. They do this where and when it suits each of them and take the time they need. The big advantage: no distances or scheduling problems need to be taken into account, as it is not necessary to physically bring the judges or experts together at the same time.

So it is not necessary to bring the judges or experts physically together at the same time.

SPECIFIC

APPI ICATIONS

comproved Experts in education, testing and assessment

In 2014, the seed was planted from which Comproved has grown. At that time, University of Antwerp, imec and University of Ghent jointly started a research project with the central question: what is the added value of comparative judgement for the evaluation of complex skills? The project was named D-PAC (Development of a Platform for the Assessment of Competences). In the years that followed we, the researchers, became experts in education, testing and (comparative) assessment. The research project yielded a good number of doctorates and publications. Because we needed real-life assessment situations, we worked a lot with teachers. They increasingly asked: "Will we still be able to use the tool if the research project stops?

This is how Comproved was born. Comproved helps teachers and assessors to assess fairly and with quality. We do this with knowledge and with practical tooling.



Maarten Goossens

Former teacher and an expert in translating research results into practice. His analytical view helps educational institutions to sharpen their assessment practices. maarten@comproved.com



Dr. Marije Lesterhuis obtained her PhD on the quality of assessments with comparative judgement and can teach you everything you need to know about reliability, efficiency and validity. *marije@comproved.com*



Dr. Roos Van Gasse

All-round assessment expert. Her PhD research sheds light on how (test) data can improve the educational practice of schools and teachers.. <u>roos@comproved.com</u>



Prof. dr. Sven De Maeyer - advisor

Full-time professor at the Department of Training and Educational Sciences at the University of Antwerp. Examines everything related to measurement and evaluation. *sven.demaeyer@uantwerpen.be*

in

in

Dr. Renske Bouwer - advisor

Assistant Professor of Language and Education. Her research focuses on the improvement of writing education, of which assessment of writing skills is an essential part. *r.bouwer@uu.nl*

Shareworks Solutions

> Shareworks - Software development Develops apps to make learning experiences more motivating and effective. Does this in cooperation with the education sector itself, because userfriendliness is the. info@shareworks.nl

> > comproved





