Evaluation report: Pilot implementations

In 2020, the European Union launched a special call (Erasmus+ strategic partnership for digital readiness in higher education) to address challenges posed by the COVID crisis. In response to this, the cLovid project gathered set out to build, test, and evaluate an integrated system to promote meaningful learning in a similar way in the domain of microscopic pathology. More specifically, the three partners (UMC, Univ. of Muenster & UTU) set out to build an integrated online learning environment featuring:

- an open-source webmicroscope (an extension to the OMERO-viewer) with enhanced features for annotating Whole Slide Images, allowing integration with assessment and feedback software
- a student assessment system (e.g., VQuest, in our design) for designing and constructing assignments for individual and collaborative activities
- an open-source software application/dashboard (PRISMA) for visualizing students' responses in tasks using various types of responses (e.g., marker questions that are ideal for visual domains) in order to provide collective feedback to a group of students

The cLovid project organized two pilots in order to try developed tools in practice. In Pilot I, the learning environment was applied at a national level, for the purpose of preparing Finnish residents for their final examination in pathology. In Pilot II, the learning environment was applied at the universities of Maastricht and University of Turku, in an international seminar that was part of undergraduate pathology courses in each university. Table 1 summarizes key details of each pilot. In what follows, we outline and reflect upon these pilots, also providing an early look at selected data regarding both pilots.

Pilot I: Seminar for Finnish pathology residents

In collaboration with pathology teachers Vesa-Matti Pohjanen (University of Oulu) and Otto Jokelainen (University of Eastern Finland), the cLovid project arranged a pilot for the integrated learning environment targeting pathology residents at a national level in Finland. The pilot was arranged online, and it took place during December 2022 and January 2023. During December 2022, residents spent approximately three hours of individual work on tasks developed by the pathology teachers Pohjanen and Jokelainen together with cLovid team. The tasks consisted of eight patient cases with whole slide images, and were built on VQuest, and applied different functionalities and question formats. After this individual work, the teachers had two to three weeks for familiarizing themselves with the residents' output and prepare a feedback discussion based on the given responses. A total of 16 pathology residents participated only in the first part involving individual assignments, while eight also participated in an online debriefing session (described below). These latter eight were from four different Finnish universities, and ranged from first to third year in their residency training.

medizinische fakultät Westfälische Wilhelms-Universität Münster · UMC Utrecht · TURUN YLIOPISTO · Co-funded by the European Union

Table 1: Core details about two teaching pilots (from Kainulainen et al., accepted)

	Pilot I	Pilot II
Participants	8 pathology residents from different universities within one country	70 undergraduate medical students from two universities in different countries
Time for teacher preparation for feedback session	~2 weeks, well-prepared feedback	None, spontaneous feedback
Timing and length	Two sessions: first residents' individual work (1–2 h) with VQuest, and a plenary feedback session (2h) ~2 weeks later	One session
Purpose of the teaching	To serve as training opportunity for the final examination in pathology	Capstone online seminar in an elective course in clinical pathology
Pretraining	No pretraining	Pretraining with online learning materials and two assignments
Interaction during session	There was interaction between the two teachers, but despite teacher prompts to ask questions, there were no questions from the participants. . (Overt teacher-student interaction was low; no interaction among the learners)	Students were presented with a new clinical case. They explored the case in two parts in small groups in break-out sessions (Zoom) and documented their responses in VQuest. The online interaction (teacher-student; student-student) was substantially higher than in the past.
Main teaching activities during the session	Interestingly, one teacher only made use of open questions, whereas the other teacher made more use of the affordances of the integrated system (e.g. marker questions). The teachers elaborated on the responses (incorrect responses and correct ones) of the participants.	Visiting the break-outs rooms (monitoring and social presence); asking the small groups justifications during the subsequent plenary session; providing a histological summary of the case

The debriefing session, which took place on January 13th, 2023, lasted for nearly 2 hours, during which the two pathology teachers discussed the cases one by one. Teachers provided further context for the cases, and lead the discussion through the PRISMA dashboard, where the residents' responses were represented. Having the responses well before the session allowed the teachers to prepare a discussion focusing on possible misconceptions, difficult details, or on elaborating the reasoning behind certain diagnoses or justifications for specific decisions regarding additional procedures such as staining the samples.

Evaluating the pilot

The cLovid team collected a set of data during this pilot through video recording (for illustration: https://vimeo.com/medicampusvhp/preparation-medical-pathology-exam) and questionnaires, and is in process of analysing the data and material in more detail. Below, we report on preliminary findings. Our observation of the debriefing session showed that the teachers managed to use PRISMA without significant problems. They were also able to utilize many of its functionalities (e.g. hiding/revealing resident responses) in meaningful ways. One aspect that became apparent already during the planning of the pilot, was that the two teachers had differing approaches to the functionalities of VQuest and PRISMA. One of the teachers preferred using only text-based (multiple choice, long-list menu, and open-ended type) items, while the other teacher preferred a combination of both text-based items and marker items (i.e. items in which responses are placed in locations inside the OMERO webmicroscope). Another aspect we noticed during the debriefing session was a complete lack of resident-resident and resident-teacher interaction. In other words, the session was mainly one-directional communication from the teachers to the residents. This was not something that was planned by neither the cLovid team nor the teachers. And in fact, the teachers did ask the residents several times if they had questions or comments. However, the only few residents' comments during the session were in the Zoom chat. While we are not sure of the reasons for the lack of vocal activity during the session, there are some possible explanations (e.g., labelling the session as "debriefing", emphasizing the anonymity of responses, or the recording of the session for research purposes).

Feedback from teacher and residents

Feedback regarding the Pilot I was collected from both the residents and the teachers. Feedback was received from one of the two participating teachers. The teacher was largely satisfied with how the session turned out and the quality of the cases (rating 5/5). However, he was not fully satisfied with the interaction during the session. The teacher noticed some difficulties with residents being able to understand and use some question formats (ones with several microscope samples or ones with long-list menu). The teacher considered both PRISMA (rating 5/5) and VQuest (rating 4/5) as very useful for pathology teaching. In both software packages, the teacher especially appreciated the marker questions. The teacher saw high educational value in the scenario (rating 5/5) and was interested in pursuing similar sessions in the future. He also noticed how it was a relatively heavy effort to plan the session and the tasks. But considering future sessions, the materials might be—at least partially—re-usable, thus saving time from planning in the future.

Feedback from residents

The residents were asked to reflect on the session through a questionnaire with 7-point likert-scale and open-ended questions. The residents considered both the individual work and debriefing session as valuable to contributing to their pathological knowledge and for preparing them for their national specialist examination. From these, especially the individual work was considered valuable for preparing for the national specialist examination (avg. rating: 6.1/7), and the debriefing session for contributing to one's knowledge on microscopic pathology (avg. rating: 6.3/7). The questionnaire also showed that residents did not contribute much during the session and did not explain concepts to peers. Yet, many residents tried to think about ways of connecting their own ideas to ones expressed during the online session (avg. rating: 4.5/ 7). Despite that

lack of visible and audible student activity during the session, residents were still relatively interested in the interaction (avg rating: 5.4/7) and did not feel very anxious during the session (avg. rating: 2.3/7). Residents indicated mildly positive ratings of involvement in a learning community – something that could surely be improved by more learner involvement during the session. In open-ended questions, one resident also recommended combining small group discussion for similar future sessions—a solution that could also likely help in forming a sense of belonging to a learning community. In other responses to the open-ended questions, residents found the functions of the learning dashboard useful. Several residents especially mentioned finding the possibility to show the range of (more or less correct) responses together as valuable to their learning. Finally, residents considered it highly likely that they participate in similar offers for preparing the national specialist exam in the future (avg. rating: 6.6/7).

Pilot II: International seminar for undergraduate level

On January 27th, 2023, the cLovid team implemented another pilot, this time for undergraduate level medical education. This pilot took place online, and was co-arranged by the University of Turku (n=37) and University of Maastricht (n=33), with students from both universities. Researchers from the University of Twente also participated in arranging this seminar. The students were required to prepare for the seminar by studying a set of self-study material prepared by the cLovid team and by completing two case assignments in VQuest containing multiple-choice questions. The seminar began with a brief introduction in Zoom, after which students were split into smaller groups break-out rooms. In the breakout groups, students worked collaboratively in order to solve one more patient case in VQuest. In between, and after the collaborative small group work, all participants joined in for collective feedback discussions, led by pathologists Pauliina Kronqvist (University of Turku) and Myrurgia Abdul Hamid (Maastricht UMC+).

Evaluating Pilot II

Comparison to findings from previous year's course

In 2022, we collected a set of baseline data regarding the final seminar of the elective course in clinical pathology that year (i.e. a year before our pilot implementation). A year after, this allows us to make some comparisons across the years. Table 2 outlines items related to student experiences of the final seminar and its tasks. Regarding general course evaluation items, all the item comparisons with statistical significance marked negative developments brought by the pilot course. Students participating the pre-pilot implementation found the preparatory learning materials more useful. They also found the final seminar more enjoyable and a better learning experience, and motivating for learning. They also found the patient case and associated tasks more interesting and more contributing to learning new information.

We also assessed several constructs related to student engagement (Haidet et al., 2012; Hamlyn-Harris, 2006; Wang et al., 2016) in both years. This assessment was done using five constructs: cognitive engagement, behavioral engagement, emotional engagement, social engagement, and peer engagement. The composite indices for these five constructs are presented in Table 3. Four of the five aspects of engagement remained at the same level as in 2022, but emotional engagement was at a lower level in 2023 leaving scope for improvement.

A similar pattern could be observed in the ratings of aspects of teacher support (teacher autonomy support, teacher relatedness support and competence support, Chiu, 2021): there were no statistically significant differences between the groups of -22 and -23. As for digital support, there was no statistically significant difference between the two years, whereas the year of -22 scored higher on the score for digital agency support probably due to the fact that for the students, the technology was new and required some learning.

c-vid-project

medizinische fakultät
Westfälische
Wilhelms-Universität Münster

UMC Utrecht

TURUN
YLIOPISTO

Co-funded by
the European Union

Table 2. Comparison of course evaluation items between the teaching pilot of -23 and the previous course of -22. Statistically significant differences are circled red.

	Year	N	Mean	Std. Deviation
During the preparation for the final seminar... [... learning materials were useful.]	22	16	4,38	,719
	23	37	3,59	1,013
During the preparation for the final seminar... [...the exercises were useful.]	22	15	4,27	1,100
	23	37	3,86	1,058
The final seminar--- [... was a good learning experience.]	22	16	4,13	,806
	23	37	3,08	,795
The final seminar--- [... was enjoyable.]	22	16	4,13	,719
	23	36	3,28	,974
The patient case with the associated tasks... [...did not interest me.]	22	16	1,19	,403
	23	37	1,97	,928
The patient case with the associated tasks... [...was clearly structured.]	22	16	4,06	,998
	23	37	3,51	1,096
The patient case with the associated tasks... [... motivated me to find out and learn new things.]	22	16	4,50	,816
	23	37	3,30	,878
The patient case with the associated tasks... [...did not add to what I had previously learned.]	22	16	1,56	,512
	23	37	2,84	1,167
The patient case with the associated tasks... [...was too complicated.]	22	16	1,94	,680
	23	37	1,78	,821
The patient case with the associated tasks... [...was too detailed.]	22	16	1,75	,856
	23	37	1,86	,918

As for the Test of Clinical Understanding, there was no statistically significant difference in mean between the groups of 2022 and 2023. Interestingly, the minimum score was 4 points higher (47 points out of 65 compared to 43 out of 65) in 2023 suggesting that the pilot course may have promoted learning at the lower end of achievement.

The results were also compared by splitting the participants of -23 into two groups: those scoring higher than the mean in the Test of Clinical Understanding and those scoring lower than the mean (see, Table 4 in Appendix). Although any of the differences are not statistically significant (small groups), it may be worth noting the group scoring lower, rated the preparatory exercises more useful (M=4.2) than the group scoring higher (M=3.8), but the group scoring higher appeared to enjoy the seminar more (M=3.4) than the group scoring lower (M=3.1).

Table 3. Comparison of aspects of student engagement between the teaching pilot of -23 and the previous course of -22. Statistically significant differences in means are highlighted with a red circle

	Year	N	Mean	Std. Deviation
Cognitive engagement	22	16	4,0	,74
	23	37	4,0	,53
Behavioral engagement	22	16	3,6	,76
	23	37	3,5	,73
Emotional engagement	22	16	4,1	,55
	23	37	3,7	,65
Social engagement	22	9	3,7	,57
	23	37	3,7	,62
Peer engagement	22	12	3,8	,62
	23	37	3,9	,53

To conclude, although the international undergraduate pilot left scope for improvement, we feel that the results are encouraging. The teachers and the students clearly appreciated the opportunity to discuss pathology in an international setting. The course format and tools were clearly accepted by the students, and they were very constructing in their feedback – despite the fact that time management failed: the seminar ended 30 minutes late. Once the constraints of project (e.g. necessity to record the session) are removed, the teachers assume full ownership of their teaching once again, and the feedback provided by the students is considered, we believe that the new form of teaching will reach its full potential. Many of the suggestions made by the students are easy to implement, e.g, requiring the use of a web-camera, using ice-breaking techniques in the group, attending to time management, attending to clear instructions, and rethinking the role of the moderator. Some suggestions are harder to implement, e.g., designing case which are sufficiently (but not overwhelmingly) challenging and ”less straightforward”, i.e., requiring serious discussion. We believe that the solution is to build a new, more challenging case for the seminar requiring differential diagnostic reasoning.

Feedback from teacher

Feedback was received from one of two teachers. The teacher appreciated both VQuest and PRISMA software. VQuest was appreciated especially due to its possibility of engaging “students in really studying the slides by making annotations. There was also the possibility to use several interesting and different kinds of

tasks (not only the traditional correct/incorrect -alternatives).” PRISMA also worked well for the teacher, but she also found it “quite complicated for a single teacher” as there is “so much data and so many items to follow simultaneously. I couldn't have managed alone, by myself, and in normal teaching circumstances we do not have a possibility for an ‘assistant teacher.’” Regarding PRISMA, the teacher also considered whether “it could be possible to group student responses in some way. Now you could only see an exhaustive list of responses and it was difficult to find any general trends among them, at least in the course of the seminar.” The teacher found the overall teaching scenario engaging and activating for students (rating 4/5), but was doubtful or in-between opinions about the originality and alignment with course objectives (rating both 3/5). The teacher did not find that the time in the teaching scenario was well spent (rating 2/5).[1] Regarding the possibilities of applying similar teaching scenarios to different levels of medical education, the teacher had doubts regarding undergraduate education at large:

> *At this phase, I am quite doubtful about the advantages of the used digital teaching tools for undergraduate education. The best part was, namely, the possibility to annotate yourself and that is actually in controversy with the learning goals we use in my university …: undergraduate students are not required to actively identify histological criteria of diseases. I think that the annotation feature is definitely advantageous for post-graduate pathology education. But for undergraduate education it is suitable only for students with special interest and ambitions in the field.*

Overall, the teacher saw value in all of the components of the learning system (microscope, VQuest, PRISMA), but also found it necessary to carefully consider how to spend the time applying them, and for which students to introduce such software packages.

Feedback from students

Qualitative feedback and evaluation data were collected from participating students on several occasions. First, during the seminar, students from both Turku and Maastricht were asked if they had comments or suggestions that might help further development and application of this form of education. In this regard, responses can be divided into five broad themes. First, some students considered the tasks too easy and/or found them to provide no significant new opportunities for learning. Second, some students found they had too little time to complete the tasks. In some cases, students even had to return to the plenary session before finalizing their work with the case. Third, some students commented on difficulties related to getting a discussion going in the small groups. One key factor contributing to this, according to the feedback, may have been that many participating students did not want to turn their cameras on,[2] and thus several students did not feel like they could get a connection with the other group member. This difficulty was especially present in mixed groups (with students from both universities) who did not already know each other. Some students suggested modifying and structuring the role of the “operator” in ways that would better facilitate conversation. Fourth, several students were not happy with the arrangement in which only one student from the group had access to VQuest, and through it, to the microscope and the sample. Since the sharing of the microscope activity (zooming, moving around, pointing) was not always fluent with a single shared screen, some students felt like they could have made much more progress with individually working with VQuest before the plenary session. Finally, some students felt a bit lost in the process, and hoped for clearer instructions.

[1] One main issue regarding time management related to spending too much time in the beginning discussing technical instructions.

[2] This may have been due to the fact that the session was recorded for research purposes.

In a later questionnaire with open-ended items to the Turku students, most expressed being satisfied with the self-study materials before the actual seminar. In those tasks, students largely felt they had clear instructions, and they also felt that the tasks were well suited for their level of expertise. The Turku students found working with VQuest very fulfilling when they could place markers inside the integrated microscope. In the seminar part, the Turku students appreciated the possibility to work together with students from another country and to get a change discuss pathology in English. Similar to the above mentioned feedback from all students, many of the Turku students also mentioned not being fully happy with the time management in the seminar. However, the opinions were not all similar, as some students hoped for more time, whereas others felt they needed much less time. Overall, students commented upon the tasks being both too easy and too difficult. Therefore, there could be some need for differentiation in similar future seminars.

General discussion

The evaluation results indicated that the teaching pilot for the pathology residents was very well received by the residents and could be implemented as such into the training of residents. The teacher taking part in the study agreed and pointed out that he greatly valued the annotation function in the integrated learning environment. Finnish residents receive some formal training locally on a weekly basis. In addition, the International Academy of Pathology Finland organizes training on a national basis twice a year. This training requires preparatory work with virtual samples. The only aspect that obviously needs further analysis and attention is the promotion learner-teacher interaction during the online seminar.

As for the international online seminar for undergraduate students taking an elective course in pathology, the results indicated that the flipped classroom learning scenario and the tools developed in the project were accepted by the students although the implementation left some room for improvement. Also, in this teaching pilot, one key challenge was to stimulate interaction. One obvious solution is to build cases which are less straightforward to solve such as ones requiring differential diagnostic reasoning. The plan is to continue developing the course in this direction.

It is important to note that at the undergraduate level, the learning objective in microscopic pathology is that students are able to make distinctions and identify features differing from normal histology, whereas in residency training the learning objective is to make diagnostic interpretations based on large amounts of visual material. For this reason, it may be more straightforward to apply the tools developed in the cLovid project to residency training.

References

Chiu, T. K. (2021). Digital support for student engagement in blended learning based on self-determination theory. *Computers in Human Behavior*, *124*, 106909. https://doi.org/10.1016/j.chb.2021.106909

Haidet, P., Levine, R. E., Parmelee, D. X., Crow, S., Kennedy, F., Kelly, P. A., ... & Richards, B. F. (2012). Perspective: guidelines for reporting team-based learning activities in the medical and health sciences education literature. *Academic Medicine*, *87*(3), 292-299. https://doi.org/10.1097/ACM.0b013e318244759e

Hamlyn-Harris, J. H., Hurst, B. J., Von Baggo, K., & Bayley, A. J. (2006). Predictors of team work satisfaction. *Journal of Information Technology Education: Research*, *5*(1), 299-315.

Kainulainen, M., Helle, L., Kronqvist, P., Vincken, K. L., Pawelka, F., Korpinen, K., & de Leng, B. (accepted). Towards an integrated online learning system for microscopic pathology: Two teaching examples.

c-vid-project

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The math and science engagement scales: Scale development, validation, and psychometric properties. *Learning and Instruction, 43*, 16-26. https://doi.org/10.1016/j.learninstruc.2016.01.008

The creation of these resources has been (partially) funded by the ERASMUS+ grant program of the European Union under grant no. 2020-1-DE01-KA226-HE-005813. Neither the European Commission nor the project's national funding agency DAAD are responsible for the content or liable for any losses or damage resulting of the use of these resources.

APPENDIX.

Table 4. Students course evaluation with comparison of students split into two groups: those scoring higher than the mean in the Test of Clinical Understanding and those scoring lower than the mean.

	LowVSHigh	N	Mean	Std. Deviation	Std. Error Mean
During the preparation for the final seminar... [...learning materials were useful.]	Low	15	3,80	1,014	,262
	High	16	3,69	,946	,237
During the preparation for the final seminar... [...the exercises were useful.]	Low	15	4,20	1,014	,262
	High	16	3,81	1,047	,262
The final seminar--- [...was a good learning experience.]	Low	15	3,07	,961	,248
	High	16	3,19	,655	,164
The final seminar--- [...was enjoyable.]	Low	15	3,07	1,163	,300
	High	15	3,40	,910	,235
The patient case with the associated tasks... [...did not interest me.]	Low	15	2,33	,900	,232
	High	16	1,56	,814	,203
The patient case with the associated tasks... [...was clearly structured.]	Low	15	3,53	,915	,236
	High	16	3,31	1,302	,326
The patient case with the associated tasks... [...motivated me to find out and learn new things.]	Low	15	3,33	,976	,252
	High	16	3,31	,704	,176
The patient case with the associated tasks... [...did not add to what I had previously learned.]	Low	15	3,07	1,033	,267
	High	16	2,31	1,138	,285
The patient case with the associated tasks... [...was too complicated.]	Low	15	2,07	,884	,228
	High	16	1,69	,704	,176
The patient case with the associated tasks... [...was too detailed.]	Low	15	2,20	1,014	,262
	High	16	1,50	,632	,158