Baseline evaluation report: Student survey and multiple-choice test for understanding of clinical pathology

In February 2022, the cLovid project collected baseline data within the context of an elective course in clinical pathology. The participants were 2nd year undergraduate medical students (N=25) who had just completed the mandatory course in pathology. The number of course participants was substantially lower than previously likely due to the fact that the teaching staff in pathology had had few opportunities to teach in person due to the pandemic. A total of 16 students (13 female; 3 male) volunteered to take part in the study.

The basic design of the course was that students elaborated on preparatory questions in small groups before the groups presented their answers to these questions in a final seminar. Due to COVID19 restrictions, the instruction was implemented as distance learning using Zoom. The preparatory questions were related to a patient case.

Table 1: Regular course evaluation items

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
During the preparation for the final seminar... [... learning materials were useful.]	16	3	5	4,37	,719
During the preparation for the final seminar... [...the exercises were useful.]	15	2	5	4,27	1,100
The final seminar--- [... was a good learning experience.]	16	3	5	4,13	,806
The final seminar--- [... was enjoyable.]	16	3	5	4,13	,719
The patient case with the associated tasks... [...did not interest me.]	16	1	2	1,19	,403
The patient case with the associated tasks... [...was clearly structured.]	16	1	5	4,06	,998
The patient case with the associated tasks... [... motivated me to find out and learn new things.]	16	2	5	4,50	,816
The patient case with the associated tasks... [...did not add to what I had previously learned.]	16	1	2	1,56	,512
The patient case with the associated tasks... [...was too complicated.]	16	1	3	1,94	,680
The patient case with the associated tasks... [...was too detailed.]	16	1	4	1,75	,856
Valid N (listwise)	15				

As shown in Table 1, the regular course evaluation items indicate that the elective course in clinical pathology for undergraduate students was rated favorably on average: the ratings for the learning materials, the final seminar and

the patient cases were above 4 on a scale from 1 to 5. The figures also indicate that there was some variance in student perceptions – especially on two items regarding the patient cases (structuring the case and amount of detail).

In addition, student perceptions of the teaching and learning environment were assessed in terms of *perceived teacher suppor*t and *perceived digital support* by means of the following six subconstructs: a) perceived teacher autonomy support, b) perceived teacher relatedness support, c) perceived teacher competence support, d) perceived digital autonomy support, e) perceived digital competence support, and f) perceived digital relatedness. See Table 2. The constructs originate from Chiu (2021); some of the items have been slightly modified or contextualized.

Table 2: An overview of the constructs related to perceived teacher support and perceived digital support

Construct	Item(s)	Cronbach alpha
Teacher autonomy support	• The teacher encouraged me to ask questions. • The teacher answered my questions fully and carefully. • The teacher made sure I really understood the goals of the seminar and what I needed to do.	0.67
Teacher relatedness support	• The teacher supported me. • The teacher was interested in me. • The teacher was friendly toward me.	0.88
Teacher competence support	• I felt that the teacher liked me to do well. • The teacher made me feel that I am able to do the learning activities.	0.72
Digital autonomy support	• I felt I could make a lot of input in deciding how I use the educational technology. • I felt a sense of freedom when using the educational technology in my learning. • I have many opportunities with the educational technology to decide for myself how to learn.	0.86
Digital competence support	• I think I am pretty good at learning with the educational technology. • I have been able to learn interesting new knowledge with the educational technology. • I feel a sense of accomplishment from learning with the educational technology.	0.80
Digital relatedness support	• When I work with the educational technology, I don't feel alone.	--

Finally, *student engagement* was assessed by means of the following five constructs: a) cognitive engagement, b) behavioral engagement, c) emotional engagement, d) social engagement, and e) peer engagement. The four first ones originate in the work by Wang and colleagues (2016), whereas the last one was designed to reflect "engagement with peers" or "effective engagement with teams" from Haidet et al. (2012). Peer engagement was based on items by Hamlyn-Harris (2006). Table 3 provides an overview of the different types of engagement, and Table 4 outlines their descriptive statistics.

Table 3: An overview of the constructs related to student engagement (R=reversed item)

Construct	Item(s):	Cronbach alpha
Cognitive engagement[1]	• I would rather have been told the answers than have had to do the tasks.(R) • I tried to connect what I was learning to things I had learned before. • When the content was difficult, I only studied the easy parts. R	0.75
Behavioral engagement	• I prepared carefully for the final seminar. • During the final seminar, I stayed focused. • During the final seminar, I didn't participate actively. R	0.69
Emotional engagement[2]	• I wanted to understand what is taught in clinical pathology. • I often felt frustrated when learning clinical pathology in this way. R • I am looking forward to future seminars with a similar format. • I found learning in this way boring. R	0.65
Social engagement[3]	• I liked the discussions of the final seminar • During the discussion of the final seminar, I didn't share ideas. R • During the discussion of the final seminar, I built my understanding on ideas expressed by others. • During the discussion of the final seminar, I tried to understand other people's ideas. • During the discussion of the final seminar, I tried to help others to learn.	0.72
Peer engagement	• During the final seminar, the discussions got sidetracked. R • During the discussion of the final seminar, the contributions built on the ideas expressed by others. • During the discussion of the final seminar, individual contributions are somehow lost. R • During the discussion of the final seminar, we were able to express ourselves freely and clearly. • During the discussion of the final seminar, we made good use of our time.	0.74

Table 4: Descriptive statistics for five types of student engagement. Figures leaving definite scope for improvement are marked by a red star. Figures leaving some scope for improvement are marked by a blue star.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Cognitive engagement	16	2,0	5,0	4,0	,74
Behavioral engagement	16	2,0	4,7	3,6	,76 ★
Emotional engagement	16	2,8	5,0	4,1	,55
Social engagement	9	3,0	4,6	3,7	,57 ★
Peer engagement	12	2,8	5,0	3,8	,62 ★

The figures in Table 5 indicate that teacher competence support was rated high (M=4.0; std=0.69) by the students, whereas digital relatedness support was rated lowest (M=2.9; std=0.62). The mean for the other constructs ranged

[1] One student is an outlier. This student is less motivated than the rest.
[2] The Cronbach alpha reliability coefficient would be >0.70 if we discard the item ”I wanted to learn what is taught in clinical pathology”.
[3] Remarkably few (n=9) responded.

between 3.4 (perceived teacher autonomy support) and 3.8 (perceived teacher relatedness support) leaving some scope for improvement.

Table 5: Descriptive statistics for perceived teacher support and perceived digital support variables. Figures leaving definite scope for improvement are marked by a red star. Figures leaving some scope for improvement are marked by a blue star.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Teacher autonomy support	10	2,3	5,0	3,4	,79 ★
Teacher relatedness support	11	2,7	5,0	3,8	,86 ★
Teacher competence support	13	3,0	5,0	4,0	,69
Digital autonomy support	15	2,0	5,0	3,5	,76 ★
Digital competence support	15	2,7	5,0	3,7	,64 ★
...from the educational technology: [When I work with the educational technology, I don't feel alone.]	14	2,0	4,0	2,9	,62 ★

Participating students also answered open-ended questions about what they especially appreciated in the course and what they would change in it. Overall, several students found the tasks (patient cases) motivating, and thought that the tasks were well designed and structured. The final seminar was found to be valuable as a motivator for doing careful studying beforehand, and also a way of summarizing many parts. One student highlighted the value of the annotated histological images. One student praised the Medigi webpages, and one student commented liking "the test of clinical understanding a lot" (i.e. the diagnostic test used in this study). One student felt dispassionate about the technical side of processing samples; another student wished for more technology to be utilized in the course (e.g., in the form of video clips and images). One student suggested clarifying the instructions and another would have hoped for more time. One student brought up that the course design was beneficial for learning about one's own topic, but learning on the other topics was quite superficial. (At least in this student's "team", the work was divided so that each student worked on one preparatory question.)

Performance on the test of clinical understanding

In order to obtain baseline data on students' understanding of the key issues addressed in the seminar, the cLovid team constructed a test based on multiple true-false questions. Multiple-true-false questions was used instead of conventional multiple-choice questions with one correct answer and several distractors for two reasons. First, according to Brassil and Couch (2019) multiple true-false questions have been shown to reveal more thoroughly the nature of students' thinking than conventional multiple-choice tests[1]. Second, the multiple-true false format was more in line with the goals of undergraduate medical education: promotion of differential diagnostic reasoning instead accuracy in deciding upon a specific diagnosis.

The test was scored by giving 1 point for each true positive and true negative. Then a composite index was calculated for each of the 14 questions each containing five true-false items. As can be seen in Table 6, the score of one of the questions had a negative correction with the rest of the composite indices lowering the internal consistency of the scale. Therefore, we decided to remove this set of questions from the scale. The descriptives for each of the 14 set of questions are shown in Table 7.

Finally, the total score on the multiple true-false test (excluding question 2 C) was calculated for each student. (See Table 8.) Descriptive statistics show that the maximum score obtained by any student was 58 and the minimum 43 points out of a total of 65 points (i.e. the theoretical maximum). Importantly, the test was neither too easy (ceiling effect) nor too difficult (floor effect) for the group of students. According to the Shapiro-Wilkes test, the distribution of the test scores conformed to a normal distribution.

[1] For this reason, Brassil and Couch (2019) recommended using multiple-true-false questions instead of conventional multiple choice items in the context of diagnostic assessment.

Table 6: Reliability analysis of 14 sets of true-false questions

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
SUM1A	50,8125	23,096	,101	,647
SUM1B	50,8125	20,029	,439	,594
SUM1C	51,1250	19,317	,544	,575
SUM1D	51,1250	23,850	,048	,648
SUM1E	51,4375	18,929	,473	,583
SUM2A	52,6250	19,583	,506	,582
SUM2B	50,8750	19,983	,451	,592
SUM2C	51,8125	27,629	-,417	,734
SUM2D	51,2500	17,667	,511	,570
SUM3A	51,0000	21,067	,375	,608
SUM3B	51,5625	22,663	,228	,630
SUM3C	51,6250	23,183	,064	,654
SUM3D	51,0625	22,329	,254	,627
SUM3E	51,1250	22,517	,244	,628

Table 7: Descriptive statistics on the students' performance on each of the 14 sets of questions

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SUM1A	16	3	5	4,4	,73
SUM1B	16	3	5	4,4	,89
SUM1C	16	2	5	4,1	,89
SUM1D	16	3	5	4,1	,50
SUM1E	16	2	5	3,8	1,05
SUM2A	16	1	4	2,6	,89
SUM2B	16	3	5	4,4	,89
SUM2C	16	2	5	3,4	,96
SUM2D	16	1	5	4,0	1,21
SUM3A	16	3	5	4,2	,77
SUM3B	16	3	5	3,7	,60
SUM3C	16	1	4	3,6	,81
SUM3D	16	3	5	4,2	,66
SUM3E	16	3	5	4,1	,62
Valid N (listwise)	16				

Table 8: Descriptive statistics for overall performance on the test of clinical understanding (excluding Question 2C)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Total points on MC test sin SUM2C	16	43	58	51,8	5,26
Valid N (listwise)	16				

To summarize, the baseline data indicate that the constructs used in the study have sufficient reliability to be used for research purposes. As for the baseline course, the students rated it favorably even in circumstances of "emergency remote teaching". The results also suggest that there is some scope for improvement in certain aspects of the course. Two things worth highlighting *are the relatively low score (M=2.9) on digital relatedness support and the emergence of a lack of discussion* on preparatory tasks within at least one small group. Promoting deep understanding on all the topics covered is also a clear challenge in group work. *Promoting discussion during the seminar* is also a challenge, because the time was spent on student presentations.

REFERENCES

Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education*, *6*(1), 1-17. https://doi.org/10.1186/s40594-019-0169-0

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.

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

Erasmus+

Partnerships for Digital Education Readiness

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.

medizinische fakultät Westfälische Wilhelms-Universität Münster

UMC Utrecht

TURUN YLIOPISTO