

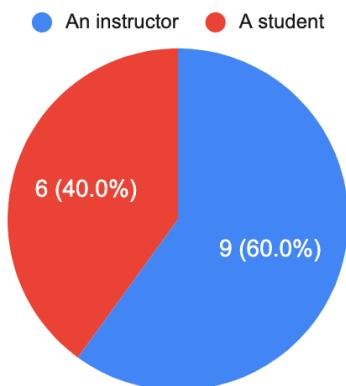
Evaluation Results

This document provides visual representations and analysis of the project's user evaluation.

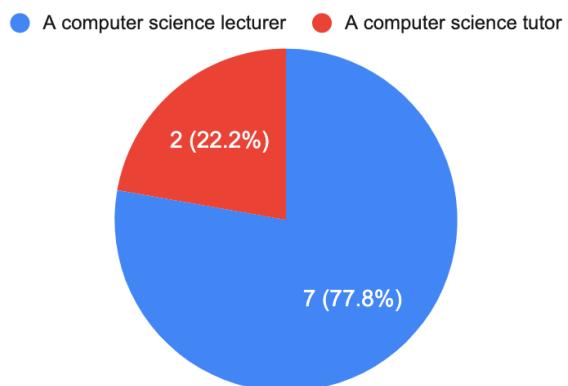
All raw data can be found in the spreadsheet 'Evaluation Raw Data' in the docs folder.

Demographics

Respondent Demographics

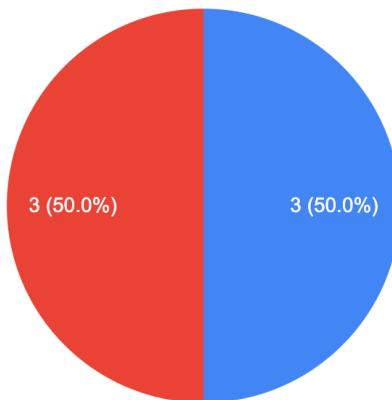


Instructor Demographics



Student Demographics

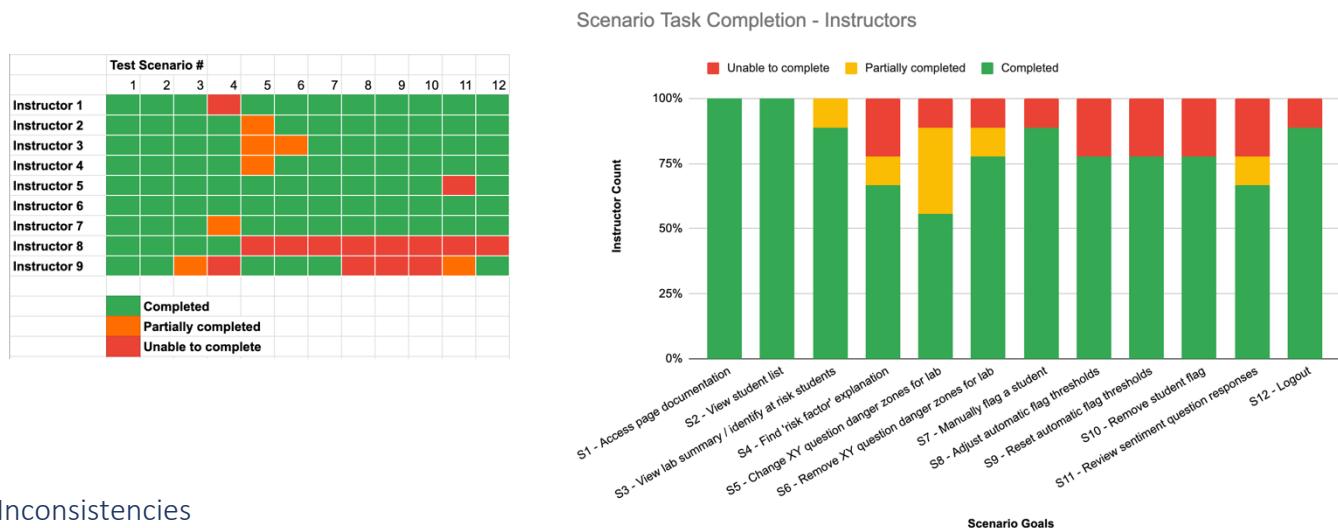
Computer science postgraduate (taught).
Computer science undergraduate (2nd year +).



Instructor Responses

Scenario Testing

The following graph describes the ability of instructors to complete the scenario test tasks.



Inconsistencies

Instructor 8 indicated that the evaluation process was ‘too long’, and this may indicate their ‘unable to complete’ responses were due to time constraints rather than an inability to use the software, particularly as tasks such as ‘logging out’ were straightforward for other respondents (see Figure 1 and Figure 2).

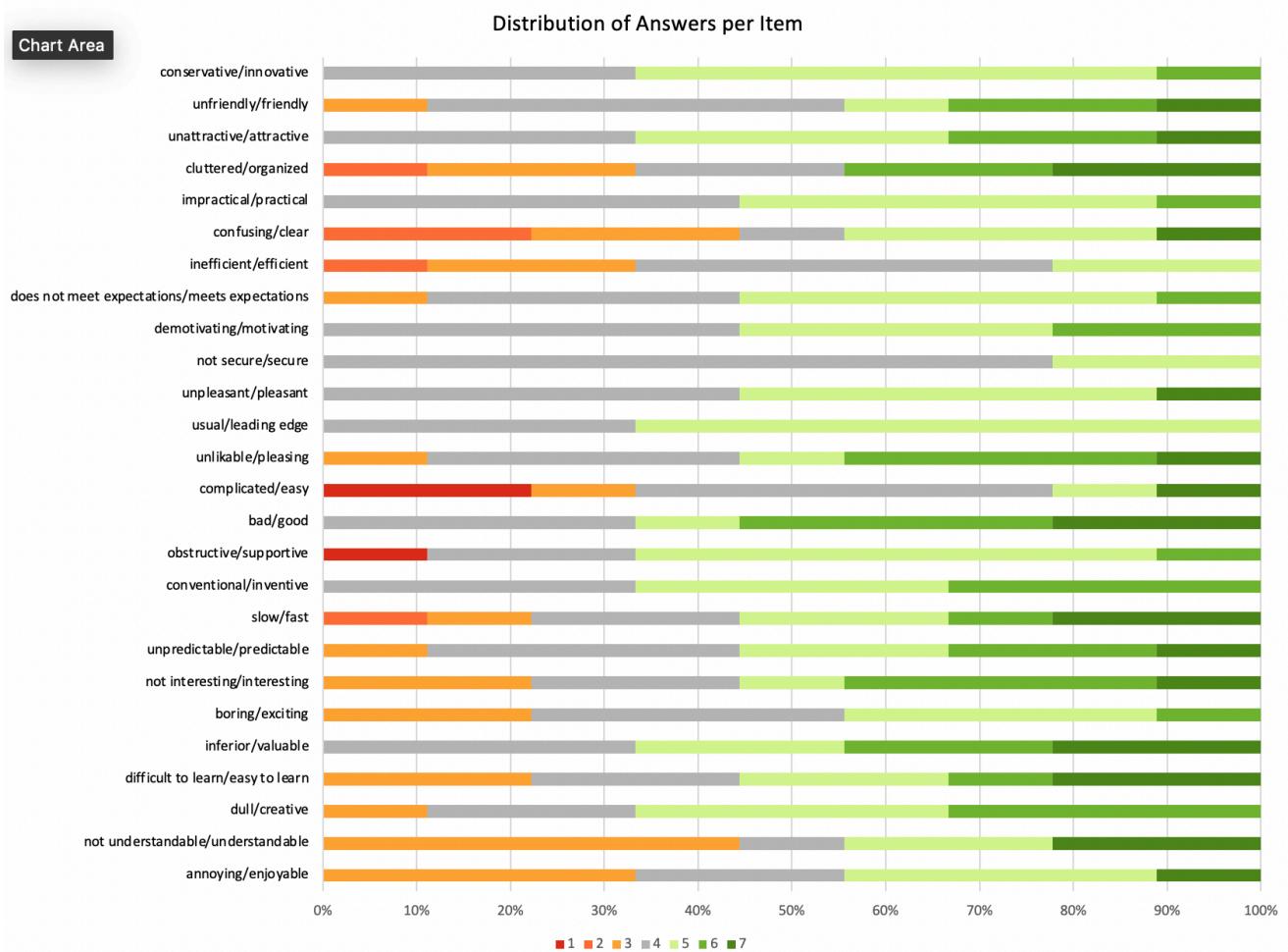
The software is useful but the evaluation far too long to expect large numbers of people to complete

Figure 1 - Instructor 8's comment indicating time constraints

| Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 |
|--------------|-----------|-----------|-----------|--------------------|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| ve successfi | Completed | Completed | Completed | Partially complete | Partially complete | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Partially complete | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Partially compk | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |
| ve successfi | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed | Completed |

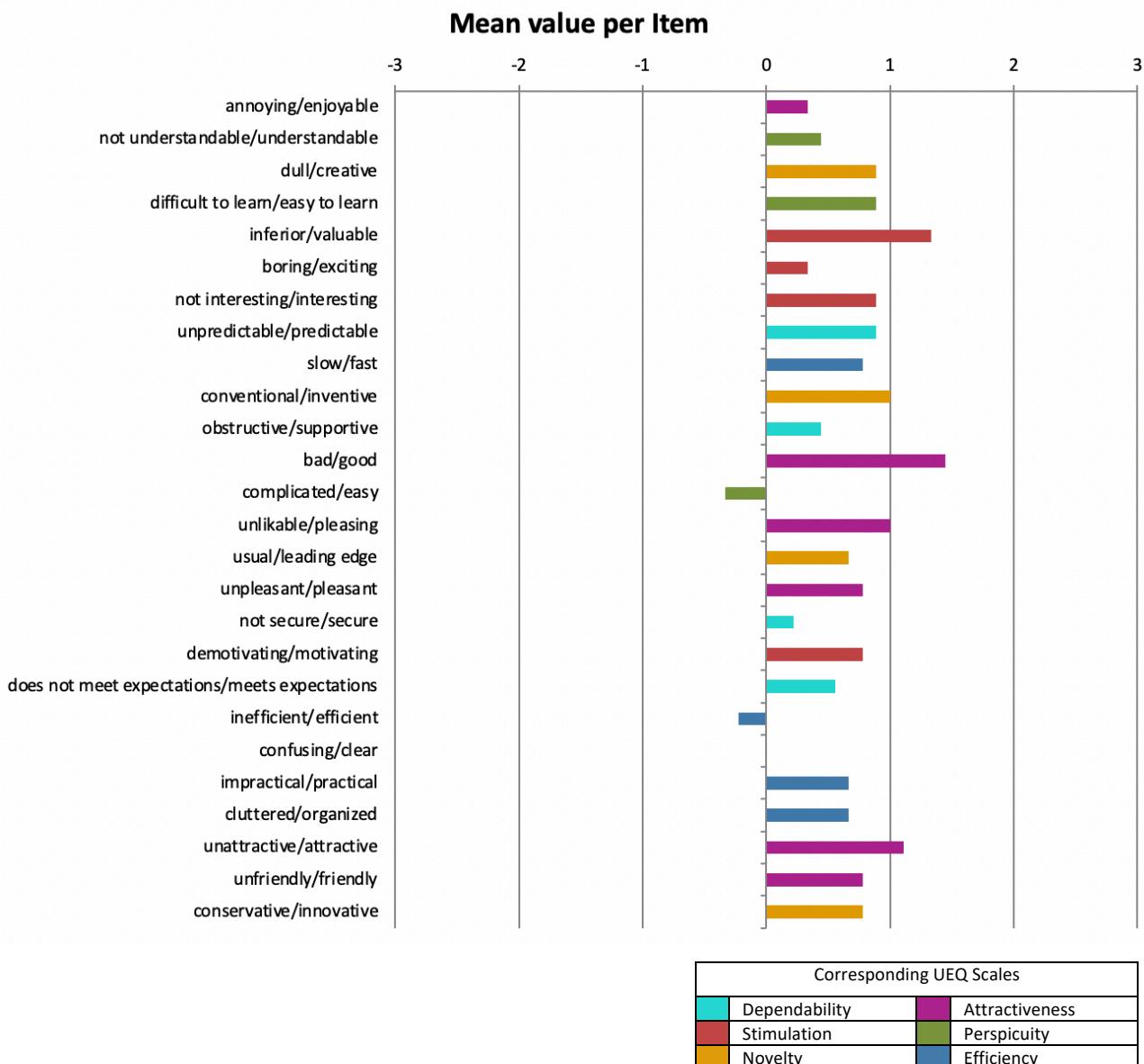
Figure 2 - Instructor 8's scenario responses from the raw data.

User Experience Questionnaire – Likert Item Responses



User Experience Questionnaire – Mean and Variance of Raw Responses

| Item | Mean | Variance | Std. Dev. | No. | Left | Right | Scale |
|-----------|------|----------|-----------|-----|--------------------|----------------------------|----------------|
| 1 ➔ 0.3 | 0.3 | 1.8 | 1.3 | 9 | annoying | enjoyable | Attractiveness |
| 2 ➔ 0.4 | 0.4 | 2.8 | 1.7 | 9 | not understandable | understandable | Perspicuity |
| 3 ⬆ 0.9 | 0.9 | 1.1 | 1.1 | 9 | creative | dull | Novelty |
| 4 ⬆ 0.9 | 0.9 | 2.4 | 1.5 | 9 | easy to learn | difficult to learn | Perspicuity |
| 5 ⬆ 1.3 | 1.3 | 1.5 | 1.2 | 9 | valuable | inferior | Stimulation |
| 6 ➔ 0.3 | 0.3 | 1.0 | 1.0 | 9 | boring | exciting | Stimulation |
| 7 ⬆ 0.9 | 0.9 | 2.1 | 1.5 | 9 | not interesting | interesting | Stimulation |
| 8 ⬆ 0.9 | 0.9 | 1.6 | 1.3 | 9 | unpredictable | predictable | Dependability |
| 9 ➔ 0.8 | 0.8 | 2.9 | 1.7 | 9 | fast | slow | Efficiency |
| 10 ⬆ 1.0 | 1.0 | 0.8 | 0.9 | 9 | inventive | conventional | Novelty |
| 11 ➔ 0.4 | 0.4 | 2.0 | 1.4 | 9 | obstructive | supportive | Dependability |
| 12 ⬆ 1.4 | 1.4 | 1.5 | 1.2 | 9 | good | bad | Attractiveness |
| 13 ➔ -0.3 | -0.3 | 3.5 | 1.9 | 9 | complicated | easy | Perspicuity |
| 14 ⬆ 1.0 | 1.0 | 1.8 | 1.3 | 9 | unlikable | pleasing | Attractiveness |
| 15 ➔ 0.7 | 0.7 | 0.3 | 0.5 | 9 | usual | leading edge | Novelty |
| 16 ➔ 0.8 | 0.8 | 0.9 | 1.0 | 9 | unpleasant | pleasant | Attractiveness |
| 17 ➔ 0.2 | 0.2 | 0.4 | 0.4 | 9 | secure | not secure | Dependability |
| 18 ➔ 0.8 | 0.8 | 0.7 | 0.8 | 9 | motivating | demotivating | Stimulation |
| 19 ➔ 0.6 | 0.6 | 0.8 | 0.9 | 9 | meets expectations | does not meet expectations | Dependability |
| 20 ➔ -0.2 | -0.2 | 0.9 | 1.0 | 9 | inefficient | efficient | Efficiency |
| 21 ➔ 0.0 | 0.0 | 2.8 | 1.7 | 9 | clear | confusing | Perspicuity |
| 22 ➔ 0.7 | 0.7 | 0.5 | 0.7 | 9 | impractical | practical | Efficiency |
| 23 ➔ 0.7 | 0.7 | 3.5 | 1.9 | 9 | organized | cluttered | Efficiency |
| 24 ⬆ 1.1 | 1.1 | 1.1 | 1.1 | 9 | attractive | unattractive | Attractiveness |
| 25 ➔ 0.8 | 0.8 | 1.7 | 1.3 | 9 | friendly | unfriendly | Attractiveness |
| 26 ➔ 0.8 | 0.8 | 0.4 | 0.7 | 9 | conservative | innovative | Novelty |



The above data indicate that many of the responses have significant variance (19/26 answers have a SD > 1.0, 2/26 have a SD <= 0.5), meaning that the results are not definitive and should be assessed with care. A larger sample would be required to reduce such deviations.

User Experience Questionnaire – Scales

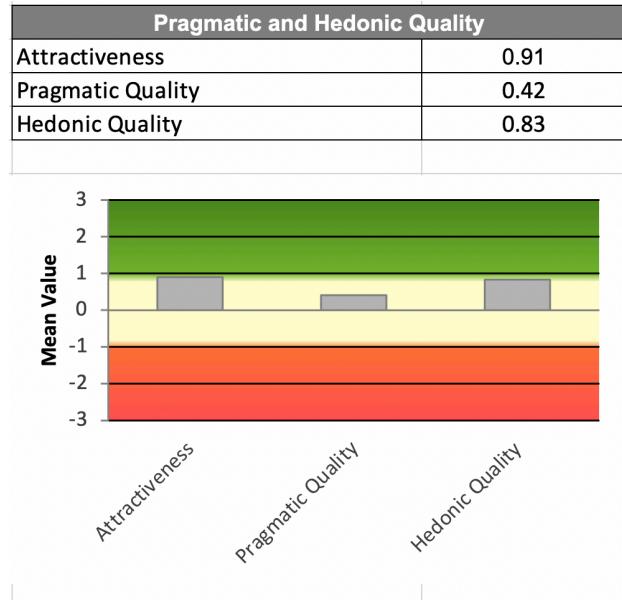
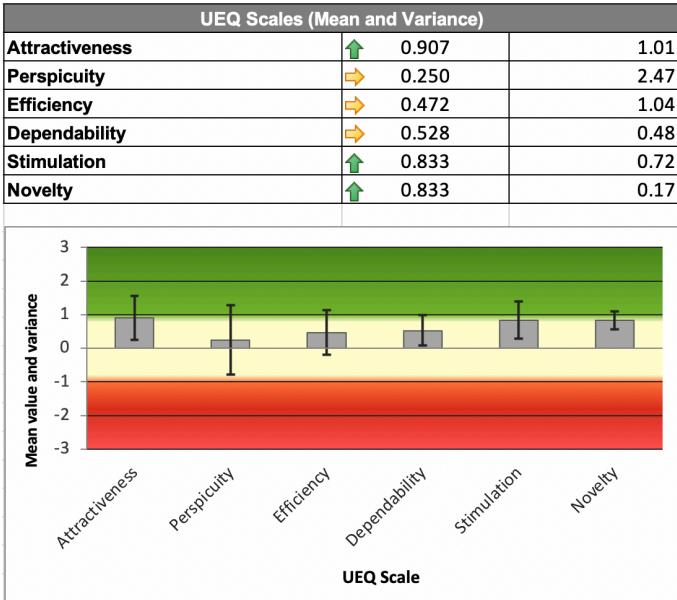
The below text is provided by Schrepp et al for the interpretation of the UEQ [1]:

The UEQ does not produce an overall score for the user experience. Values between -0.8 and 0.8 represent a more or less neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation and values < -0.8 represent a negative evaluation.

The range of the scales is between -3 (horribly bad) and +3 (extremely good). But in real applications, in general, only values in a restricted range will be observed. It is due to the calculation of means over a range of different persons with different opinions and answer tendencies (for example the avoidance of extreme answer categories) extremely unlikely to observe values above +2 or below -2.

Thus, even a quite good value of +1.5 for a scale looks from the purely visual standpoint on a scale range of -3 to +3 not as positive as it really is.

The scales of the UEQ can be grouped into pragmatic quality (Perspicuity, Efficiency, Dependability) and hedonic quality (Stimulation, Originality). Pragmatic quality describes task related quality aspects, hedonic quality the non-task related quality aspects.



| Confidence intervals (p=0.05) per scale | | | | | | |
|---|-------|-----------|---|------------|---------------------|-------|
| Scale | Mean | Std. Dev. | N | Confidence | Confidence interval | |
| Attractiveness | 0.907 | 1.007 | 9 | 0.658 | 0.249 | 1.566 |
| Perspicuity | 0.250 | 1.571 | 9 | 1.027 | -0.777 | 1.277 |
| Efficiency | 0.472 | 1.019 | 9 | 0.666 | -0.193 | 1.138 |
| Dependability | 0.528 | 0.690 | 9 | 0.451 | 0.077 | 0.978 |
| Stimulation | 0.833 | 0.848 | 9 | 0.554 | 0.279 | 1.387 |
| Novelty | 0.833 | 0.415 | 9 | 0.271 | 0.562 | 1.104 |

As previously mentioned, the variance for the UEQ scales seems to be too significant to yield conclusive results at this stage, particularly with respect to perspicuity. However, the above tables are an early indication of the areas for improvement needed.

Item Response Inconsistencies with UEQ Scales

Schrepp et al give the following description for the interpretation of the inconsistency tables below [1]:

All items in a scale should measure a similar UX quality aspect. The idea to detect random or not serious answers is to check how much the best and worst evaluation of an item in a scale differs. If there is a big difference (>3) this is seen as an indicator for a problematic data pattern. Of course such situations can also result from random response errors or a misunderstanding of an item. We suggest to remove answers from the data set that shows a value of 3 or higher in the Critical? column.

| Items | | | | | | | | | | | | | | | | | | | | | | | | | | Scales with inconsistent answers | | | | | | |
|-------|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------------------------|-------------|------------|---------------|-------------|---------|-----------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | Attractiveness | Perspicuity | Efficiency | Dependability | Stimulation | Novelty | Critical? |
| 1 | 3 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | -3 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | | | | | | 1 | | |
| 0 | -1 | 1 | -1 | 1 | -1 | 2 | 3 | 2 | 1 | 1 | -3 | -1 | 1 | 1 | 0 | 0 | 0 | -1 | -1 | 1 | 3 | 2 | 2 | 1 | | | | | | | 1 | |
| 0 | 1 | 1 | 2 | 0 | 1 | 2 | 1 | -1 | 0 | 1 | 2 | 1 | 2 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | -2 | 1 | 1 | 0 | | | | | | 0 | |
| -1 | -1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | -3 | 0 | 1 | 0 | 1 | 1 | 1 | -2 | -2 | 0 | -1 | 1 | 0 | 1 | | | | | | 0 | |
| 1 | 0 | 0 | 1 | 3 | 0 | 2 | 3 | 0 | 1 | 0 | 3 | 0 | 2 | 1 | 1 | 0 | 1 | 1 | 0 | -1 | 2 | 2 | 1 | 0 | 0 | | | | | | 0 | |
| 1 | 3 | 2 | 3 | 1 | 1 | 3 | 0 | 3 | 1 | 1 | 3 | 3 | 3 | 0 | 3 | 0 | 2 | 0 | 1 | 3 | 1 | 3 | 3 | 3 | 2 | | | | | | 0 | |
| 3 | 1 | 2 | 1 | 3 | 2 | 2 | 1 | 2 | 2 | 2 | 0 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | | | | | | 0 | | |
| -1 | -1 | 2 | 0 | 2 | -1 | -1 | -2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | | 0 | |
| -1 | -1 | -1 | 0 | 1 | 1 | 0 | 1 | 2 | 1 | 0 | -1 | 0 | 1 | 0 | 0 | -1 | -1 | 2 | 0 | -1 | 0 | -1 | 1 | 1 | | | | | | 0 | | |

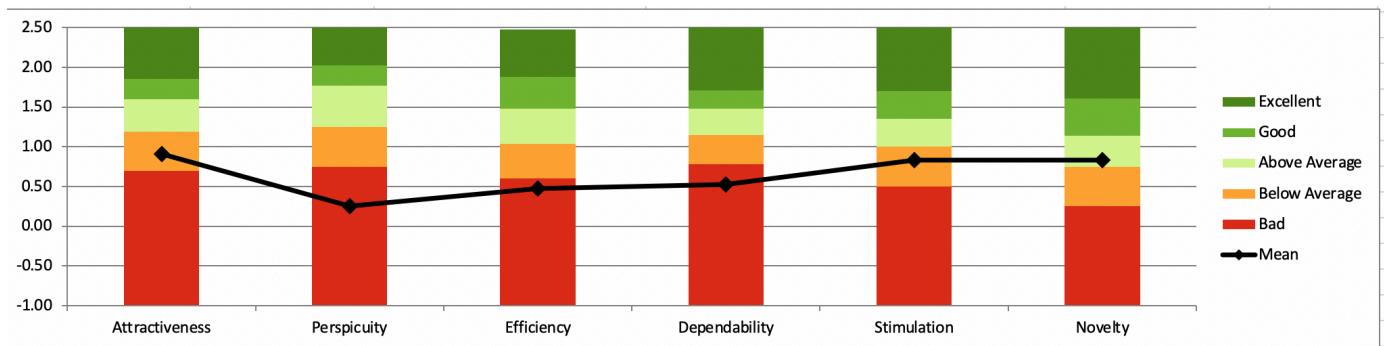
The above tables show no significant inconsistencies in the result set were found.

Benchmark against other software

Schrepp et al give the following description for the interpretation of the benchmark graph below [1]:

The measured scale means are set in relation to existing values from a benchmark data set. This data set contains data from 20190 persons from 452 studies concerning different products (business software, web pages, web shops, social networks).

The comparison of the results for the evaluated product with the data in the benchmark allows conclusions about the relative quality of the evaluated product compared to other products.

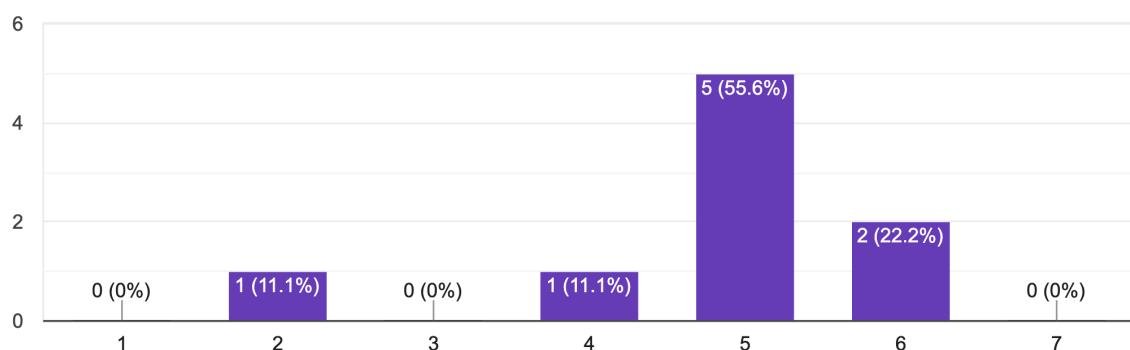


| Scale | Mean | Comparison to benchmark | Interpretation |
|----------------|------|-------------------------|---|
| Attractiveness | 0.91 | Below average | 50% of results better, 25% of results worse |
| Perspicuity | 0.25 | Bad | In the range of the 25% worst results |
| Efficiency | 0.47 | Bad | In the range of the 25% worst results |
| Dependability | 0.53 | Bad | In the range of the 25% worst results |
| Stimulation | 0.83 | Below Average | 50% of results better, 25% of results worse |
| Novelty | 0.83 | Above Average | 25% of results better, 50% of results worse |

Questions relating to pedagogy / retention of CS1 students

This tool would help me detect students at risk quickly (< 2 weeks) and plan interventions.

9 responses



Please elaborate on your answer above.

5 responses

the view that allows you to see red marks next to students' marks is helpful for that.

Having on-hover help might be better, now we have to click through the explanation repeatedly

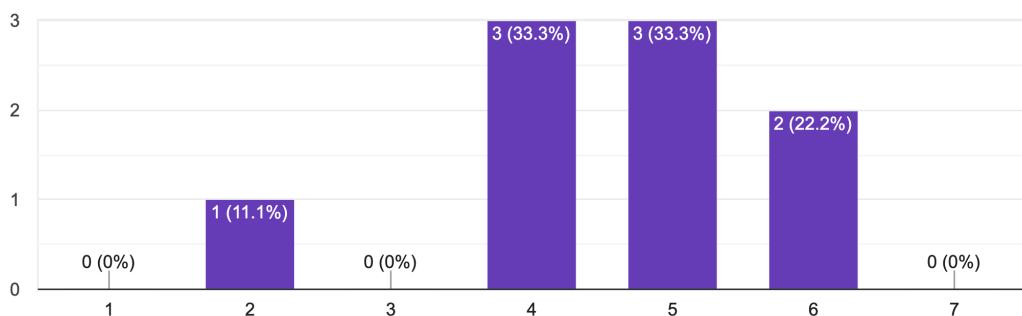
I am not sure how much work is required in preparation of all the questionnaires etc to get to the point where I will get useful insights.

I think the tool is very useful for determining students who are struggling or might struggle in the near future.

I think it captures the state a student is in quite well, both in terms of what they have done tangibly, and also how they are feeling.

I could integrate this tool into my existing pedagogy.

9 responses



Please elaborate on your answer above.

7 responses

Sorry it is a bit too complicated for me.

I wish I'd seen the student side so I could understand what they would need to do.

I think this is a great tool for L1 students and probably also Sem1 MSc students

Again I am not sure how much pre-work is needed

I tend to keep notes of how students are getting on with the material in a course, so having everything in an application might make things easier.

The software is useful but the evaluation far too long to expect large numbers of people to complete

I would like to take feedback from my students more often during a course, and I can see the utility of having a software like this making the most of such feedback.

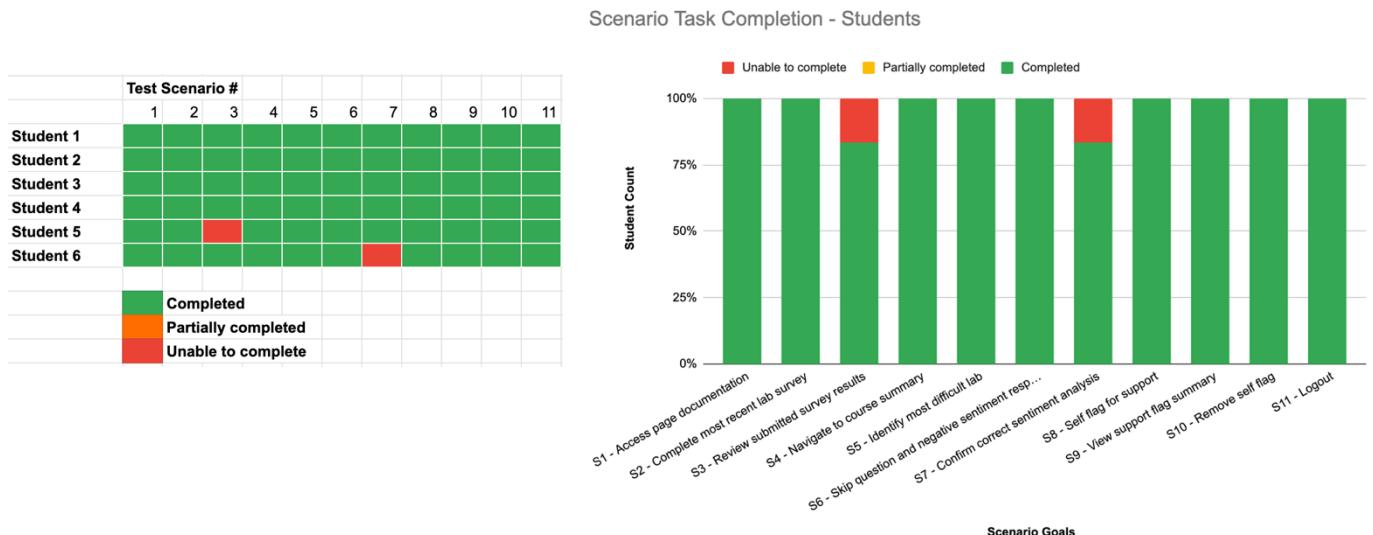
Confidence intervals (p=0.05) per question

| Question | Mean | Std. Dev. | N | Confidence | Confidence interval |
|--------------------------------------|------|-----------|---|------------|---------------------|
| Detect and plan interventions | 4.78 | 1.20 | 9 | 0.785 | 3.993 5.563 |
| Integrate into pedagogy | 4.56 | 1.24 | 9 | 0.808 | 3.748 5.363 |

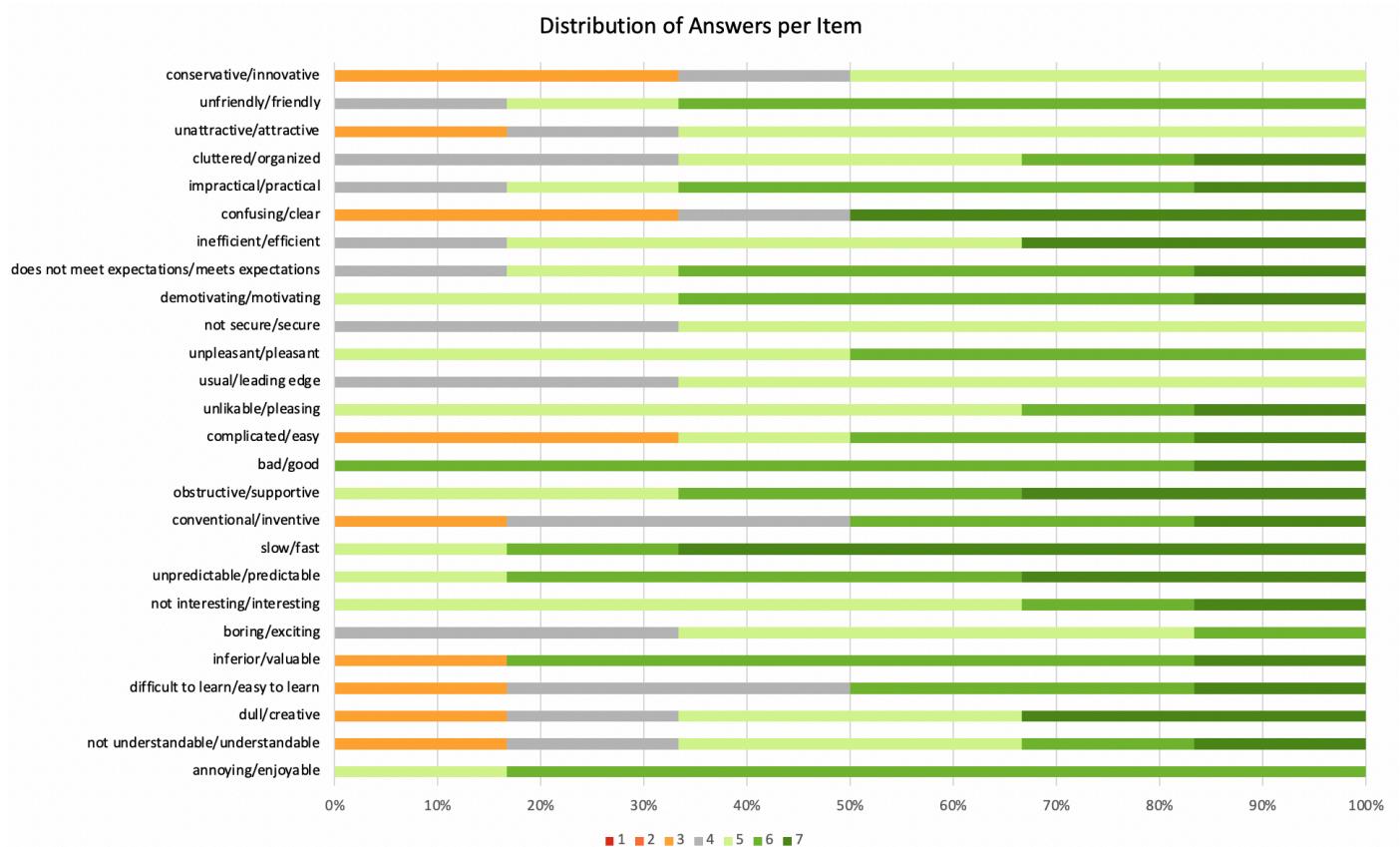
Student Responses

Scenario Testing

The following graph describes the ability of students to complete the scenario test tasks.

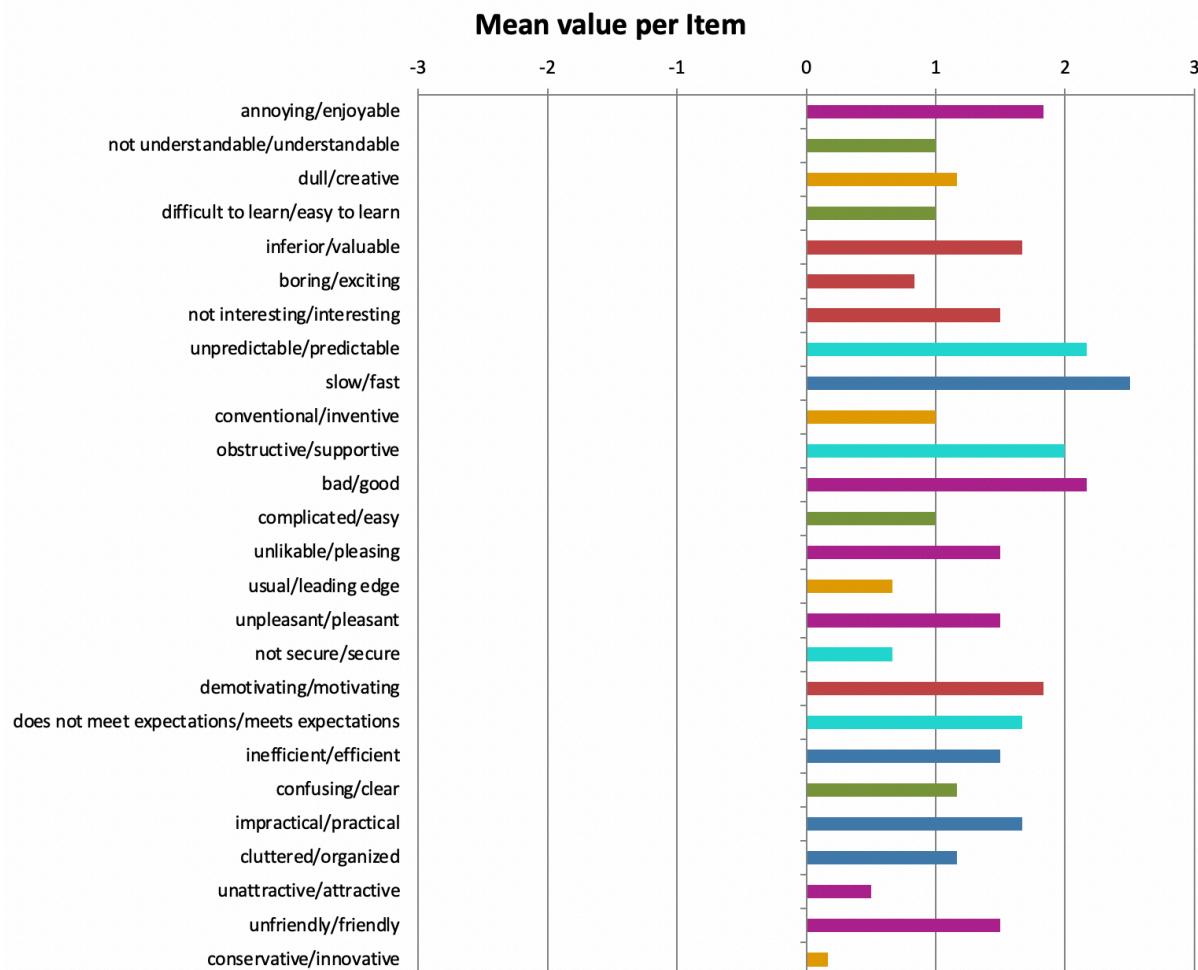


User Experience Questionnaire – Likert Item Responses



User Experience Questionnaire – Mean and Variance of Raw Responses

| Item | Mean | Variance | Std. Dev. | No. | Left | Right | Scale |
|------|-------|----------|-----------|-----|--------------------|----------------------------|----------------|
| 1 | ↑ 1.8 | 0.2 | 0.4 | 6 | annoying | enjoyable | Attractiveness |
| 2 | ↑ 1.0 | 2.0 | 1.4 | 6 | not understandable | understandable | Perspicuity |
| 3 | ↑ 1.2 | 2.6 | 1.6 | 6 | creative | dull | Novelty |
| 4 | ↑ 1.0 | 2.4 | 1.5 | 6 | easy to learn | difficult to learn | Perspicuity |
| 5 | ↑ 1.7 | 1.9 | 1.4 | 6 | valuable | inferior | Stimulation |
| 6 | ↑ 0.8 | 0.6 | 0.8 | 6 | boring | exciting | Stimulation |
| 7 | ↑ 1.5 | 0.7 | 0.8 | 6 | not interesting | interesting | Stimulation |
| 8 | ↑ 2.2 | 0.6 | 0.8 | 6 | unpredictable | predictable | Dependability |
| 9 | ↑ 2.5 | 0.7 | 0.8 | 6 | fast | slow | Efficiency |
| 10 | ↑ 1.0 | 2.4 | 1.5 | 6 | inventive | conventional | Novelty |
| 11 | ↑ 2.0 | 0.8 | 0.9 | 6 | obstructive | supportive | Dependability |
| 12 | ↑ 2.2 | 0.2 | 0.4 | 6 | good | bad | Attractiveness |
| 13 | ↑ 1.0 | 2.8 | 1.7 | 6 | complicated | easy | Perspicuity |
| 14 | ↑ 1.5 | 0.7 | 0.8 | 6 | unlikable | pleasing | Attractiveness |
| 15 | ➡ 0.7 | 0.3 | 0.5 | 6 | usual | leading edge | Novelty |
| 16 | ↑ 1.5 | 0.3 | 0.5 | 6 | unpleasant | pleasant | Attractiveness |
| 17 | ➡ 0.7 | 0.3 | 0.5 | 6 | secure | not secure | Dependability |
| 18 | ↑ 1.8 | 0.6 | 0.8 | 6 | motivating | demotivating | Stimulation |
| 19 | ↑ 1.7 | 1.1 | 1.0 | 6 | meets expectations | does not meet expectations | Dependability |
| 20 | ↑ 1.5 | 1.5 | 1.2 | 6 | inefficient | efficient | Efficiency |
| 21 | ↑ 1.2 | 4.2 | 2.0 | 6 | clear | confusing | Perspicuity |
| 22 | ↑ 1.7 | 1.1 | 1.0 | 6 | impractical | practical | Efficiency |
| 23 | ↑ 1.2 | 1.4 | 1.2 | 6 | organized | cluttered | Efficiency |
| 24 | ➡ 0.5 | 0.7 | 0.8 | 6 | attractive | unattractive | Attractiveness |
| 25 | ↑ 1.5 | 0.7 | 0.8 | 6 | friendly | unfriendly | Attractiveness |
| 26 | ➡ 0.2 | 1.0 | 1.0 | 6 | conservative | innovative | Novelty |



| Corresponding UEQ Scales | | | |
|--|---------------|--|----------------|
| | Dependability | | Attractiveness |
| | Stimulation | | Perspicuity |
| | Novelty | | Efficiency |

The above data indicate that, while less deviant than the instructor responses, several of the responses have significant variance (9/26 options have a SD > 1.0, only 5/26 have a SD <= 0.5), meaning that the results are not definitive and should be assessed with care. A larger sample would be required to reduce such deviations.

User Experience Questionnaire – Scales

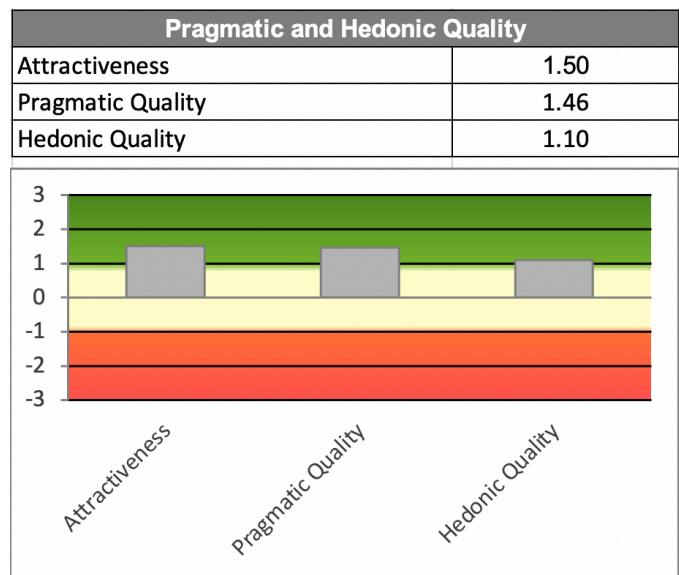
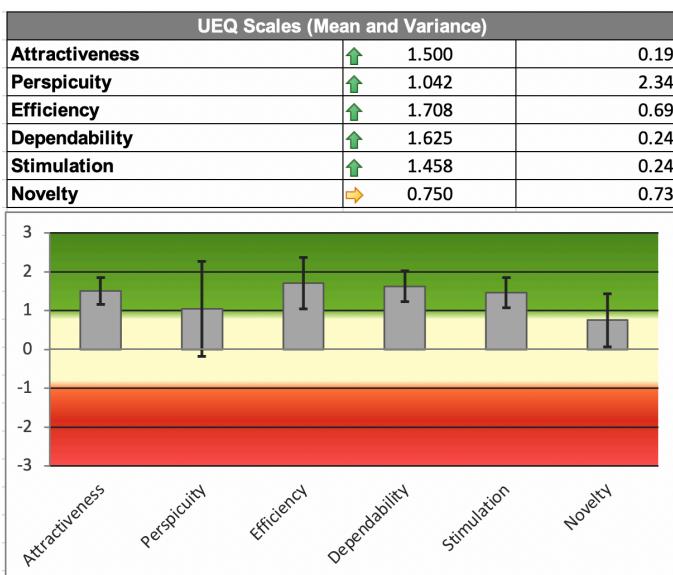
The below text is provided by Schrepp et al for the interpretation of the UEQ [1]:

The UEQ does not produce an overall score for the user experience. Values between -0.8 and 0.8 represent a more or less neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation and values < -0.8 represent a negative evaluation.

The range of the scales is between -3 (horribly bad) and +3 (extremely good). But in real applications, in general, only values in a restricted range will be observed. It is due to the calculation of means over a range of different persons with different opinions and answer tendencies (for example the avoidance of extreme answer categories) extremely unlikely to observe values above +2 or below -2.

Thus, even a quite good value of +1.5 for a scale looks from the purely visual standpoint on a scale range of -3 to +3 not as positive as it really is.

The scales of the UEQ can be grouped into pragmatic quality (Perspicuity, Efficiency, Dependability) and hedonic quality (Stimulation, Originality). Pragmatic quality describes task related quality aspects, hedonic quality the non-task related quality aspects.



| Confidence intervals (p=0.05) per scale | | | | | | |
|---|-------|-----------|---|------------|---------------------|-------|
| Scale | Mean | Std. Dev. | N | Confidence | Confidence interval | |
| Attractiveness | 1.500 | 0.435 | 6 | 0.348 | 1.152 | 1.848 |
| Perspicuity | 1.042 | 1.528 | 6 | 1.223 | -0.181 | 2.264 |
| Efficiency | 1.708 | 0.828 | 6 | 0.662 | 1.046 | 2.371 |
| Dependability | 1.625 | 0.494 | 6 | 0.395 | 1.230 | 2.020 |
| Stimulation | 1.458 | 0.485 | 6 | 0.388 | 1.070 | 1.847 |
| Novelty | 0.750 | 0.851 | 6 | 0.681 | 0.069 | 1.431 |

As previously mentioned, the variance for the UEQ scales seems to be too significant to yield conclusive results at this stage, particularly with respect to perspicuity. However, the above tables are an early indication of the areas for improvement needed.

Item Response Inconsistencies with UEQ Scales

Schrepp et al give the following description for the interpretation of the inconsistency tables below [1]:

All items in a scale should measure a similar UX quality aspect. The idea to detect random or not serious answers is to check how much the best and worst evaluation of an item in a scale differs. If there is a big difference (>3) this is seen as an indicator for a problematic data pattern. Of course such situations can also result from random response errors or a misunderstanding of an item. We suggest to remove answers from the data set that shows a value of 3 or higher in the Critical? column.

| Items | | | | | | | | | | | | | | | | | | | | | | | | | | Scales with inconsistent answers | | | | | |
|-------|----|----|----|----|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------------------------|---|---|--|--|--|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | Critical? | | | | | |
| 2 | 0 | 1 | -1 | 2 | 1 | 1 | 1 | 3 | 0 | 2 | 3 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | -1 | 2 | 2 | -1 | 1 | 1 | 1 | 1 | 1 | | | |
| 2 | 1 | 3 | 2 | 2 | 1 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 3 | 2 | 0 | 1 | 2 | 1 | 0 | 0 | | | |
| 2 | 3 | -1 | 2 | 3 | 1 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 3 | 3 | 2 | 3 | 1 | 2 | -1 | 1 | 1 | 0 | | | |
| 2 | 2 | 1 | 3 | 2 | 0 | 1 | 3 | 3 | -1 | 3 | 2 | 3 | 3 | 0 | 2 | 0 | 3 | 2 | 3 | 3 | 3 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | | | |
| 2 | -1 | 3 | 0 | 2 | 1 | 3 | 2 | 3 | 2 | 1 | 2 | -1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | | | |
| 1 | 1 | 0 | 0 | -1 | 0 | 1 | 2 | 1 | 0 | 1 | 2 | -1 | 1 | 0 | 1 | 0 | 2 | 1 | 0 | -1 | 1 | 1 | 0 | 0 | 0 | -1 | 0 | 0 | | | |

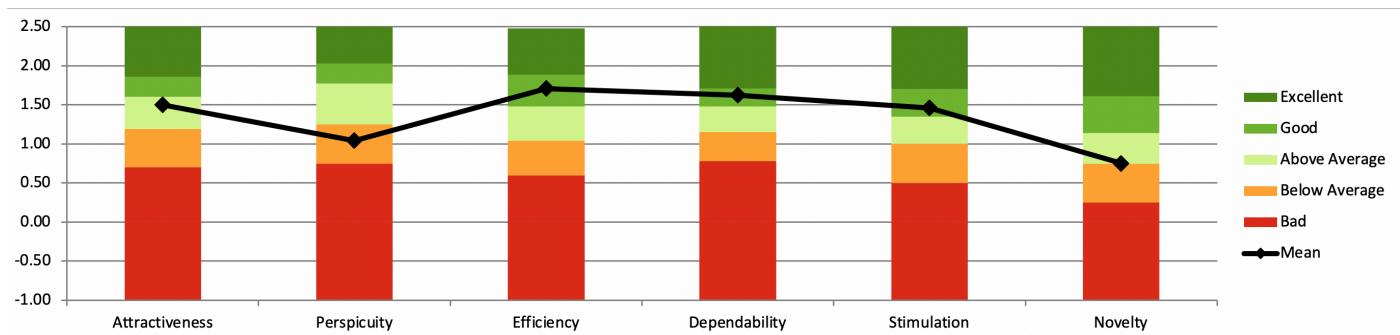
The above tables show no significant inconsistencies in the result set were found.

Benchmark against other software

Schrepp et al give the following description for the interpretation of the benchmark graph below [1]:

The measured scale means are set in relation to existing values from a benchmark data set. This data set contains data from 20190 persons from 452 studies concerning different products (business software, web pages, web shops, social networks).

The comparison of the results for the evaluated product with the data in the benchmark allows conclusions about the relative quality of the evaluated product compared to other products.

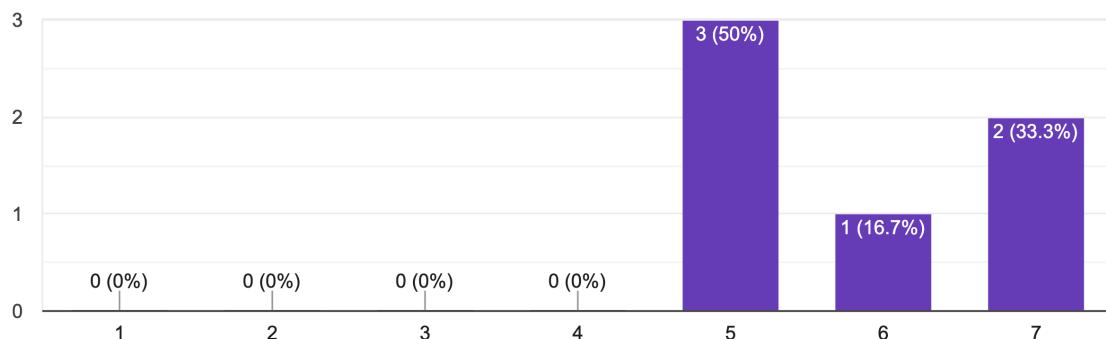


| Scale | Mean | Comparisson to benchmark | Interpretation |
|----------------|------|--------------------------|---|
| Attractiveness | 1.50 | Above average | 25% of results better, 50% of results worse |
| Perspicuity | 1.04 | Below Average | 50% of results better, 25% of results worse |
| Efficiency | 1.71 | Good | 10% of results better, 75% of results worse |
| Dependability | 1.63 | Good | 10% of results better, 75% of results worse |
| Stimulation | 1.46 | Good | 10% of results better, 75% of results worse |
| Novelty | 0.75 | Below Average | 50% of results better, 25% of results worse |

Questions relating to help seeking behaviour / retention of CS1 students

If I were struggling on a course, using this software would encourage me to seek support sooner than if I were not using it.

6 responses



Please elaborate on your answer above.

4 responses

The software is easy to use after the first try, and it is helpful because it makes it easier for a student to ask for help, instead of emailing a member of staff.

It's can be much easier to seek support online rather than having to approach an actual living person

It can be embarrassing to admit you're struggling to tutors/lecturers. A system like this would allow me time to explain how I felt I was struggling clearly and without perceived judgement.

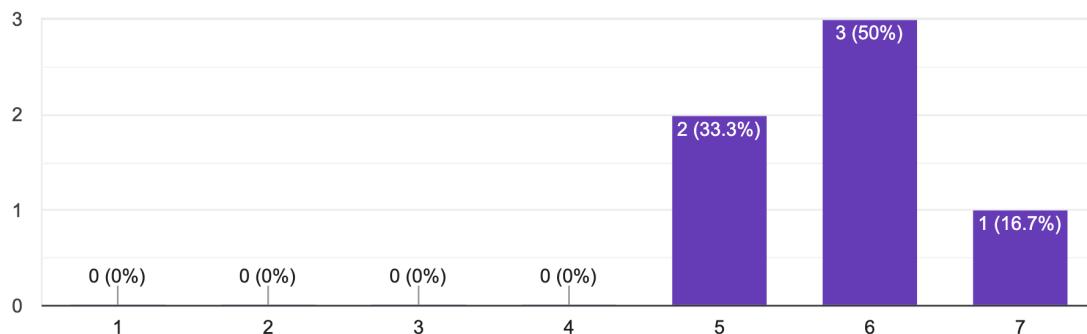
It also makes it easier. I don't need to go and find someone to discuss the problems I'm facing with, I can do it whenever is convenient. Particularly useful in our current online lab/lecture climate.

It would also allow me to provide course feedback earlier in the semester rather than only right at the very end when nothing can change.

nm

Using this software gives me confidence that my instructors will support me.

6 responses



Please elaborate on your answer above.

4 responses

The software itself indicates that a member of staff will assist a student who has difficulties, so it is implied that support will be received.

Knowing that you are flagged in a course gives you some sense of peace of mind as there is a record that you need help and passes the responsibility onto the tutor to ensure you get the support you need.

The software appears very clean, simple and professional. It also worked reliably and quickly in all of the previous tests. This definitely helped build my confidence in it.

However, regardless of how useful the software is for students, to be useful in practice and give me full confidence in the support I would receive, tutors/lectures would need to use this system properly and actually follow up on flags effectively. This is out of control of the software to some extent.

nm

| Confidence intervals (p=0.05) per question | | | | | | |
|--|------|-----------|---|------------|---------------------|-------|
| Question | Mean | Std. Dev. | N | Confidence | Confidence interval | |
| Seek support | 5.83 | 0.98 | 6 | 0.787 | 5.047 | 6.620 |
| Instructor support confidence | 5.83 | 0.75 | 6 | 0.602 | 5.231 | 6.436 |