

Article

Using an Evidence-Informed Approach to Improve Students' Higher Order Thinking Skills

Mireille Desirée Hubers

Department Educational Science, University of Twente, 7500 AE Enschede, The Netherlands;
m.d.hubers@utwente.nl

Abstract: This study aimed to determine why university students performed relatively low on a writing assignment. In phase 1, questionnaire and performance data were collected and analyzed. This evidence-informed approach revealed that students' relatively low performance was caused by their relatively underdeveloped higher order thinking skills. A six-step research-informed procedure was designed to assist students' development of these skills. In turn, questionnaire data, performance data and observational data were collected in phase 2 to determine the effectiveness of the design and the quality of the implementation. The results indicated that students improved their higher order thinking skills. Alternative explanations did not play a role; the designed procedure was implemented with high fidelity and positively rated by the students. This study can aid university teachers in designing course materials for the development of students' higher order thinking skills.

Keywords: action research; evidence-informed teacher improvement; higher order thinking skills; lower order thinking skills



Citation: Hubers, M.D. Using an Evidence-Informed Approach to Improve Students' Higher Order Thinking Skills. *Educ. Sci.* **2022**, *12*, 834. <https://doi.org/10.3390/educsci12110834>

Received: 11 October 2022

Accepted: 14 November 2022

Published: 19 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In their Special Issue about failure in academic development Bolander Laksov and McGrath [1] made a compelling case for learning more about failure and failed initiatives. One way to do so is via evidence-informed school and teacher improvement [2]. This approach builds on the work around data-based decision making [3,4] and research-informed practice (e.g., [5]), and is composed of eight steps. The approach starts by defining the current educational situation and setting a concrete goal for educational improvement one wants to achieve. Then, possible causes are identified and data about those potential causes are collected and analyzed. After conclusions are drawn based on the data, research evidence is used to inform the potential solutions. Subsequently, an action plan is made, implemented and evaluated (see [2], for an elaborated description of each step).

The use of an evidence-informed approach is important because decisions informed by data and research evidence are more likely to be effective than decisions based on intuition, personal judgement, and experience [4,6]. Furthermore, data can support teacher learning by stimulating reflective processes and providing insight into the teachers' strengths and weaknesses. As a result, teachers may change their teaching behavior, which can result in improved teacher performance and subsequently, improved student achievement [7,8].

One area where such an evidence-informed approach might be beneficial is in teaching students higher order thinking skills in academic education. These higher order thinking skills reflect a complex mode of thinking that often generates multiple solutions [9]. According to Lewis and Smith [10], higher order thinking "occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations" (p. 451). Such thinking skills include, amongst other things, learning experiences that require students to analyze, evaluate, and synthesize information [9]. Given that our societal problems are complex and do not have straightforward solutions, teaching our students these higher order thinking skills is essential.

Higher order thinking skills are an expected outcome of higher education, but they are often not explicitly taught [11]. Usually, higher order thinking skills are taught only when lower order thinking skills have been mastered. Shepard [12] argued that the most serious consequence of such an approach is that higher order skills are introduced quite late, which means that students often never get to have the opportunity to engage in higher order thinking skills. The goal of the current study is to determine whether students' relatively poor performance on a written assignment was indeed rooted in their underdeveloped higher order thinking skills. The evidence-informed school and teacher improvement approach [2] is used to study this case and to develop, test and evaluate a concrete educational intervention. The next section provides more information about the case.

2. The Study's Context

The current study focuses on a course about leadership and organizational change that is taught at the master's level at a university in the Netherlands. The course is assessed with both an exam and a final paper that is written in dyads; the former targets mainly the level of reproduction and understanding and the latter targets the level of understanding and more complex forms of application. Each assessment type counts towards 50% of students' final grades. For their final paper, students have to analyze a hyped leadership model or organizational change strategy and determine whether this hype makes sense from a scientific point of view.

In past years, the quality of students' final papers was not at the desired level. A descriptive analysis of their grades showed that although students scored a passing mark on average, there was substantial variation in students' performance; see Tables 1 and 2. Moreover, dyads scored above or at the mid-point for the writing style/grammar, and around the mid-point or higher for APA style/layout, meaning that the problem was mainly in the content of the assignment. The aim of the current evidence-informed study is to improve students' performances on this assignment.

Table 1. Number of Points Student Dyads (n = 17) Achieved per Element of the Assignment in Year 1 (all transformed to a 10-point scale).

	Overall Grade	Description of Vision	Description of Learning	Argument Vision	Argument Learning	Style & Layout
M	6.08	6.29	5.26	5.76	4.96	6.59
St. dev.	1.42	1.90	1.64	2.05	1.60	1.14
Min–Max	4.33–9.19	3.5–9.5	3–9	2.8–10	5–9	5–9

Table 2. Number of Points Student Dyads (n = 8) Achieved per Element of the Assignment in Year 2 (all transformed to a 10-point scale).

	Overall Grade	Description of Vision 1	Description of Vision/Actions 2	Argument 1	Argument 2	Conclusion	Writing Style & Grammar	APA Style & Layout
M	5.92	5.70	5.11	5.83	5.06	4.44	6.78	6.33
St. dev.	1.85	2.08	2.03	2.42	2.57	2.47	1.37	1.87
Min–Max	3.52–8.83	3.33–8.67	2.67–8.67	2–9	2–8.5	1.33–10	5–8.5	4–9

3. Hypotheses

It was hypothesized that three causes could play a role in explaining students' performance. All three causes are examined in this study. The first potential cause is the clarity of the assignment. If an assignment is unclear or even confusing, it can result, amongst other things, in less confidence in one's own reasoning abilities and self-efficacy [13], although this is only weakly positively related to students' final grades. In addition, when students are not satisfied with the assignment, this can be an important source of demotivation for the entire course [14].

The second potential cause for students' relatively poor performance is students' expectancy. Students' expectancy is a component of student motivation that refers to

students' beliefs about their ability to perform the task [15]. It reflects whether they believe they can do a task and are responsible for success in this task. Research has shown that students with higher levels of expectancy are more likely to show active cognitive engagement in learning through the use of strategies such as rehearsal, elaboration, and organization of the learning materials [15]. Some studies have also linked this higher level of expectancy to higher levels of achievement (e.g., [16]). Finally, students who display higher levels of expectancy are also more likely to persist at difficult or uninteresting academic tasks [17].

Finally, the relatively low performance of the students could be caused by their own thinking skills (e.g., [18]). Kuhn [11] stated that although higher order thinking skills are an expected outcome of higher education, they are often not explicitly taught. Higher order thinking overlaps with levels that are above the comprehension level of Bloom's taxonomy [19]. It includes a broad range of competencies, including identifying questions, accessing reliable sources of information that address those questions, examining and evaluating patterns in data, and making and justifying arguments (e.g., [11,19,20]), all of which are relevant to the current writing assignment.

However, it is usually assumed that higher order thinking skills can only be addressed successfully when a certain level of mastery of the lower order thinking skills is achieved [21–23]. For example, Prakash and Ratnesh Litoriya [24] state that "A learner can upgrade competency by navigating various levels from lower order thinking skills to higher order thinking skills in particular" (p. 725). When that level is achieved, the higher order thinking skills can be used to further broaden and deepen one's knowledge base [23]. According to Newman [25], lower order thinking requires the routine or mechanical application of previously acquired information. Examples include listing memorized information and inserting numbers into previously learned formulas. Lower order thinking skills correspond with Bloom's [21] levels of remembering and understanding. As students will need these lower order thinking skills in order to apply the higher order thinking skills that are required for the assignment, it is important to determine their mastery of both levels. In that way, the level of support provided in the designed intervention can be determined.

Taken together, it was examined whether the clarity of the assignment, students' expectancy, and/or students' thinking skills influenced their performance.

4. Method Phase 1

4.1. Participants

In total, 13 students who were part of this course in Year 2 participated, which reflects a response rate of 72.22%. Seven participants were female (53.85%). An additional three students (two female) consented to the use of their performance data, but did not complete the questionnaire.

4.2. Procedure & Instruments

Quality assurance survey. The university's standard quality assurance survey was administered after the course ended in the summer of 2019. This survey includes questions about the clarity of the assignment. The resulting data represent the group level.

Online questionnaire. The remaining data was collected in November 2019 via an online survey. At the beginning of this questionnaire, students were informed about the goals of their study and had to provide active consent to the use of their data. This questionnaire included measures for student expectancy and students' lower and higher order thinking skills. All data were collected in English.

The student expectancy scale included nine statements such as 'I lack a real incentive to work' [15]. For each item, participants rated to what extent they agreed with the statement, ranging from *totally disagree* (1) to *totally agree* (5). Two items were removed from the original scale to substantially improve its Cronbach's alpha. After that, Cronbach's α was 0.71, which is acceptable. All items on this scale were negatively phrased, so that a low score represented a high level of expectancy.

Two self-constructed scales that measured students' perceived ability to use lower and higher order thinking skills were used which included eight and seven items, respectively. Example items include: 'I am able to paraphrase an article's most important findings in my own words' (lower order) and 'I find it difficult to evaluate the weight of certain arguments' (higher order). For each item, participants rated to what extent they agreed with the statement, ranging from *totally disagree* (1) to *totally agree* (5). Negatively phrased items were reverse scored so that a higher score indicates more mastery of the skill. The sample was too small to run a factor analysis. When the items were grouped based on theoretical reasons, the resulting Cronbach's alphas were sufficient and good, respectively, with $\alpha = 0.78$ for lower order thinking skills and $\alpha = 0.83$ for higher order thinking skills.

Finally, the questionnaire included three open-ended questions: what students found easiest and most difficult about writing their paper and any other remarks they wanted to share related to their writing of this paper. This last question was included to ensure that any other relevant causes could come to light. Students' answers were coded with a brief coding scheme that was based on Bloom's taxonomy; see Table 3. When in doubt, the answers were discussed with an expert on students' thinking skills.

Table 3. Coding Scheme for Qualitative Data.

Code	The Student's Answer Relates to His/Her Ability to ...	Examples of Students' Answers
Lower order	- demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating main ideas.	"I found it difficult to read academic texts"
In-between lower and higher order	- solve problems in new situations by applying acquired knowledge, facts, techniques and rules in a different way.	"I found it difficult to use an article's definition and translate that to a concrete example"
Higher order	- examine and break information into parts by identifying motives or causes; make inferences and find evidence to support generalizations. - present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.	"I found it difficult to evaluate the weight of certain arguments"

5. Results Phase 1

5.1. The Role of Student Expectancy

Overall, students displayed relatively high levels of expectancy, with a mean score of 1.87 ($SD = 0.45$). To test whether their level of expectancy influenced their performance on the final paper assignment, a linear regression analysis was conducted. The resulting model was not significant, $F(1, 10) = 1.21$, $p = 0.298$, meaning that students' expectancy was not a significant predictor of their performances.

5.2. The Role of the Clarity of the Assignment

Another possible cause for students' relatively poor performance on the paper assignment could be that the assignment was not clear enough. This indeed could have been a potential cause in year one, as students gave the clarity of the assignment and its assessment criteria a mean rating of 2.4 and 2.8, respectively, on a 5-point scale, which is below the mid-point. However, after year 1, substantial modifications were made to the assignment to simplify it in both content and phrasing. In year two, the clarity of the assignment and its assessment criteria were evaluated, on average, with a 4.0 and a 4.3 on a 5-point scale, respectively. Thus, the clarity of the assignment did not seem to play a role.

5.3. The Role of Lower and Higher Order Thinking Skills

Students' higher order thinking skills can only be taught when students have already mastered the lower order thinking skills. Therefore, it was hypothesized that if students

obtain passing scores on their written exam, which tests their lower order thinking skills, and score significantly higher on the exam than on their final paper, which tests their higher order thinking skills, it can be assumed that their higher order thinking skills are the cause of students' relatively poor performance on a written assignment.

To test this hypothesis, a paired samples *t*-test was conducted to determine whether students' written exam grades significantly differed from their final assignment grades. This test confirmed a significant difference in their performance, with $t(51) = 4.09$, $p < 0.001$. It appeared that on average, students scored higher on the written exam ($M = 6.85$, $SD = 1.02$) than on their paper ($M = 6.02$, $SD = 1.53$). Moreover, the descriptive statistics showed that students performed at a passing level for the written exam. Therefore, it could be concluded that the problem seemed to be caused by students' lack of higher order thinking skills.

To triangulate whether students' higher order thinking skills were the main problem, their mean scores on the lower and higher order thinking skills questionnaire scales were compared. It appeared that students performed above the mid-point on both of these scales with an average score of 4.05 ($SD = 0.49$) and 3.68 ($SD = 0.56$), respectively.

Moreover, the qualitative data illustrated that most students ($n = 7$) found it most easy to find relevant literature, which reflects lower order thinking skills. Two students also mentioned they found it easiest to write about their general idea behind the hype, which requires minimal analytical skills. For example, one student wrote: "The general description of the leadership/organizational change hype [was easiest to write]. This was more focused on the description of existing literature about the chosen subject (something that we have done multiple times in prior courses)." In addition, students reported that activities that required their use of higher order thinking skills were the most difficult. They especially found it difficult to find evidence that helps to argue for or against the hype and evaluate whether the overall line of reasoning in their paper was coherent. For example, one student stated that he/she found most difficult: "Staying critical [when writing the paper], so keeping in mind that people can be critical about your line of reasoning." This is in line with Table 2, which shows that students especially struggled with formulating a second argument for or against their chosen hype and drawing a coherent conclusion. Taken together, the results confirm that the main cause underlying students' relatively poor performance was their lack of ability to apply higher order thinking skills.

6. Redesign

Research shows that when students are too unfamiliar with higher order thinking skills, it is advisable to use a more teacher-centered approach to aid them (Marzano & Miedema, 2018). Various programs have been designed to address these higher order thinking skills, [23,26–28]. Although these programs could not be directly translated to the current context, important principles could be discerned that inspired part of the redesign. For example, Zohar and Nehet [28] used a combination of explicit instruction about argumentation skills and concrete cases on which students could practice these skills, which improved their overall domain knowledge as well as the quality of their argumentation. Moreover, it appears to be crucial that higher order thinking skills are modelled for the students. For example, Papathomas and Kuhn [29] found that students increased their argumentation skills significantly when an apprenticeship model was used by the teacher. Finally, it is also important to teach students how argumentation can be improved in their writing [30]. Based on these insights on explicit instruction, concrete exercises, modelling, and writing skills, a six-step guide to assist with the final writing assignment was developed for the students:

0. Select your hype.
1. Search for and read literature, and ask yourself who, what, why, how questions.
2. Organize info in a way that reveals the connections between the main topic and its various themes or categories.
3. Structure your reasoning: claim + argument.

4. Search for additional literature: which claims and arguments can you back up with literature?
5. Evaluate the claims, evidence, and conclusions.
6. Communicate your line of reasoning in the paper.

These steps were combined with an in-depth explanation. For example, during step 4, it was explained how to evaluate whether arguments support the claim, deepen the claim, lend weight (substantiate, in terms of methodological rigor, magnitude/size of findings, number of studies confirming the same perspective) and make the claim more plausible. This explanation was provided to them throughout the weeks of the course; see also Table 4. During the lectures, approximately half of the time was used for this. In doing this, each step was modelled by the teacher and exemplified using one concrete case that was used throughout the weeks. Moreover, students practiced these skills themselves as well via short, in-class assignments and polls of their answers to encourage all students to think actively about the assignment. For example, the following excerpt was given to the students:

Table 4. Overview of the Designed Activities.

Lecture	In-Class Activities
1	- Explanation of the relevance of higher order thinking skills - Explanation of the required steps, especially steps 1 and 2 - Modelling of steps 1 and 2 on a concrete case
2	- Brief recap of previous steps - Explanation and modelling of steps 3 and 4 - Students practice argument mapping in break-out rooms
3	- Brief recap of previous steps - Explanation and modelling of step 5 - Students practice step 5 - Answers are discussed and modelled in class
Work session	- Students practice step 5 twice - Answers are discussed and modelled in class - Explanation and modelling of step 6 - Dyads provide feedback on each other's Argument Map - Class discussion of dyads' peer feedback
4	- Provision of formative feedback - Students practice step 5 - Answers are discussed and modelled in class

"Is this a good claim and argument about Scrum? The core principle of Scrum methodology is the self-organization of a team and its involved members (Hoda et al., 2010). Within innovative product development, self-organization is often seen as premise for a fruitful outcome (Hoegl & Parboteeah, 2006), because it increases involvement, creative freedom and the overall autonomy of team members (Fenton-O'Creevy, 1998). Team members within Scrum teams do not make a sequence of isolated or unshared decisions, rather their decisions are inter-related, with each decision leading to further actions of the team (McAvoy & Butler, 2009). As a consequence, the leadership within Scrum teams has similarities with shared leadership, meaning that the leadership rotates to the persons with the particular skills and knowledge to face specific issues (Moe et al., 2009). This is especially beneficial in the decision making process, as members do not have to wait for a managers approval, but can rather quickly take actions themselves as they are close to the problem and product (Karhatsu, Ikonen, Kettunen, Fagerholm, & Abrahamsson, 2010). In this regard, Scrum leadership is a viable way to develop innovative products by providing team members with autonomy, take part in the decision making and generally act in quick succession when the product demands something else."

After everyone voted, an in-class discussion was held to discuss students' answers. Moreover, the teacher modelled her own answers.

In addition, a working session was designed, for which students had to prepare an Argument Map (See <https://www.reasoninglab.com/argument-mapping/>, accessed on 14 April 2020). During this session, break-out rooms were used where dyads provided each other with feedback on their map via a structured form. For example, they had to indicate to what extent "the supporting arguments for the first claim: Support the claim (are about same topic), deepen the claim (explain the 'why' behind your claim), lend weight (substantiate, in terms of methodological rigor, magnitude/size of findings, number of studies confirming the same perspective), and make the claim more plausible". After that, the peer feedback was discussed with the whole group. Finally, some time was also spent modelling and practicing steps 5 and 6.

To test whether the redesign of this course improved students' higher order thinking skills, the following research questions were posed:

1. How did the performance of the current cohort of students on the final writing assignment compare to that of the prior year's cohort of students?
2. To what extent were any differences in performance attributable to the alternative explanations: students' expectancy and the clarity of the assignment?
3. To what extent were any differences in performance attributable to the redesign, both in terms of its perceived quality and its implementation fidelity?

7. Method Phase 2

7.1. Participants

The redesign was tested for the 2019–2020 cohort of students which included 11 students, seven of whom were female (63.64%). Of those participants, all gave consent to the use of their performance data and 10 completed the questionnaire at both pre- and post-test.

7.2. Measures & Procedure: Effectiveness of the Redesign

Questionnaire. The online questionnaire was administered before the start of the course (April 2020) and at the end of the course (July 2020). At the pretest, students were informed about the goals of their study and had to provide active consent to the use of their data. The measures mentioned in the method section of phase 1 were used for the pre- and post-tests that helped to determine the effectiveness of the redesign. These measures addressed students' expectancy, students' lower and higher order thinking skills and what they found easiest/most difficult. At the post-test, the clarity of the assignment was measured as well. Although students' expectancy and the clarity of the assignment did not prove to play a significant role in predicting students' grades in phase 1, they were measured again because it was expected that expectancy might become significant due to the current pandemic, and to determine whether the assignment was now sufficiently clear.

Grades for students' papers. In addition, students' papers were graded using the same matrix as the previous year. These grades were compared to those of the previous cohorts of students. To ensure that the teacher's strictness in grading was comparable to the previous year, she first studied a relatively lower scoring paper (5.23) and a relatively high-scoring paper (8.83) from the previous year and its corresponding assessments. Based on that 'refresher,' the papers of the current cohort were graded in one sitting.

7.3. Measures: Implementation of the Redesign

Questionnaire. The aforementioned post-test questionnaire included various questions to determine the students' evaluation of the redesign, including how valuable the designed activities were for students, how much effort they put into it, and their overall perception of the course, the assignment, and their teacher.

Perception polls. During the course, students were asked to fill out a perception poll after each designed activity (nine times in total). This poll determined how valuable the

activity was to them. An example of such a poll is: ‘This break-out session with evaluation exercises is helpful to me’. Answers could range from *totally disagree* (1) to *totally agree* (5).

Observations. A structured observation form was created for evaluation of the fidelity of implementation of the activities per lecture and filled out by the teacher. Four types of implementation fidelity were used for this [31]:

- Adherence: did the teacher deliver the content of the assignments in the intended manner? (e.g., in terms of the instruction)
- Quality of delivery: aside from the content, how did the teacher deliver the assignments? (e.g., in terms of enthusiasm and preparedness)
- Exposure: were the activities implemented as expected in terms of quantity? (e.g., number of activities and duration of activities)
- Student’ responsiveness: did students actively and seriously perform the activities? (e.g., in terms of their participation and enthusiasm)

After the course ended, information concerning each type of implementation fidelity was analyzed systematically to determine the quality of implementation.

8. Results Phase 2

Evaluating Effectiveness: Students’ Performance

First, the descriptive statistics for students’ performance on the paper were calculated; see Table 5. It appears that on average, their scores were higher than during the previous years. In addition, the number of grades below the passing mark decreased. In the current cohort, only one dyad received an insufficient mark, which was also the main reason for the high standard deviation. This dyad was the only pair who did not use the opportunity to check their chosen hype up front, did not opt in for formative feedback on their paper, and were absent for some of the lectures.

Table 5. Number of Points Student Dyads ($n = 5$) Achieved per Element of the Assignment in the Current Year (all transformed to a 10-point scale).

	Overall Grade	Description of Vision 1	Description of Vision/Actions 2	Argument 1	Argument 2	Conclusion	Writing Style & Grammar	APA Style & Layout
<i>M</i>	7.25	6.67	6.53	6.95	6.55	6.73	8.00	8.60
<i>SD</i>	1.97	2.87	2.87	2.86	2.66	3.16	0.79	0.89
Min–Max	3.48–9.10	2.00–9.33	2.00–9.00	2.00–9.25	2.00–8.50	1.33–9.00	7.00–9.00	7.00–9.00

A detailed analysis of all dyads’ sub-grades confirmed that the current cohort of students outperformed the previous cohorts on all parts of the assignment; see Tables 2 and 5. On average, students in Year 2 scored a 5.23 on the core of the assignment (all elements except writing style and APA style/layout), whereas in the current year, students scored on average a 6.69 on those elements. This shows that on average, students improved more than 1 full grade point.

The qualitative data were consulted to triangulate these results. The pre- and post-test questionnaire included open-ended questions about what students found easiest and most difficult when writing the paper. For the pre-test which asked about students’ strengths and difficulties when writing papers in general, students almost exclusively provided answers about what was easiest that reflected their lower order thinking skills ($n = 10$). Examples included using APA style and following the general structure of an article. In terms of what they found most difficult, almost all students ($n = 10$), gave an answer that reflected higher order thinking skills such as: “Writing a solid in-depth argument and balancing your own view with that of other authors without it either being your own opinion or just paraphrasing authors.”

Similar to the pre-test, students’ answers on the post-test ($n = 10$) about what was easiest mainly reflected their lower order thinking skills. However, three answers that reflected higher order thinking skills were mentioned. Examples included writing the conclusion of the paper and writing down a line of reasoning. In reporting the things they

found most difficult, only six answers on the post-test reflected students' higher order thinking skills. Two of the remaining four answers reflected general struggles with the assignment. This comparison of the pre- and post-test seems to suggest that students improved their understanding of the higher order thinking skills, though the majority of the students still reported using them as the most challenging aspect of the assignment.

Taking all of this together, the current cohort of students improved their performance on the final writing assignment. To determine whether their improved performance could be attributed to the redesign, two alternative explanations were tested.

9. Exploring Two Alternative Explanations

Students' expectancy. Overall, students' expectancy levels were relatively high levels, with a mean score of 2.28, which was a bit lower than in the previous cohort (the items on this scale are negatively phrased, see method section), and a standard deviation of 0.41. To test whether their level of expectancy influenced their performance on the paper assignment, a linear regression analysis was conducted. The resulting model was not significant, $F(1, 8) = 0.78$, $p = 0.402$, meaning that students' expectancy was not a significant predictor of their achievement.

Clarity of the assignment. The clarity of the assignment and its assessment criteria were evaluated, on average, with a 3.2 and a 3.6 on a 5-point scale, respectively. Both scores were above the mid-point of the scale, indicating that the assignment was sufficiently clear.

Taken together, both alternative explanations are rejected, which makes it likely that students' improved performance was caused by the redesign. To further substantiate this claim, the implementation fidelity and quality of that implementation were studied as well.

10. Evaluating Implementation

Implementation fidelity. The implementation fidelity gives an indication of whether the redesign was implemented in an intended manner. Based on the classroom observations, it was determined that the content was mostly delivered in an intended manner and only relatively minor issues occurred (e.g., audio difficulties during the last 10 min of the class). Moreover, the quality of the delivery was as intended. In terms of exposure, all designed activities were used during the sessions. However, two out of five sessions were too tightly scheduled, due to which the lecture time had to be extended or the break had to be skipped. This was not desirable, as students' attention span might have been limited at that point in the session.

Finally, students' responsiveness was determined. During every assignment, all attending students participated in the perception polls. Moreover, some of them actively asked questions. Students' self-reported effort for various course elements mirrored the teacher's observations, with mean scores for their effort ranging from 4.00 to 4.90 ($SD = 0.32$ – 0.97), which signals a relatively high self-perceived invested effort. Moreover, almost all students were always present; one student missed one session, and one student missed two. These two students formed one dyad, and this is the only dyad that failed at writing the paper. Taken together, the data indicate that the redesign was implemented with high fidelity.

Quality of the redesign. In addition to establishing implementation fidelity, it was important to determine whether students perceived the redesign as helpful. To do so, their perception was polled each time an element of the redesigned course was tackled. The descriptive statistics for these perceptions indicated that all elements of the redesign were helpful to the students, with mean scores ranging from 3.56 to 4.36 ($SD = 0.45$ to 0.70). Students especially appreciated the steps and practice rounds that helped them evaluate claims, arguments and conclusions, as well as the writing tips. In contrast, they appreciated the steps on how to read literature and organize information and the practice of argument mapping relatively less. Similarly, their post-test ratings of the helpfulness of the redesign after writing the full paper confirmed the helpfulness of the redesign, with mean scores ranging from 3.60 to 4.10 ($SD = 0.57$ to 1.07).

11. Discussion

Improving Students' Higher Order Thinking Skills

The purpose of this study was to use an evidence-informed approach to improve students' performance on their final writing assignment by addressing their higher order thinking skills. Overall, the data indicated that the current cohort improved their performance and that this improvement was attributable to the redesign of the course. This shows the value of an evidence-informed approach to improving academic education. Given the small scale of the present study, however, one should be careful in providing scientific implications.

12. Limitations

The first limitation is that no existing instruments could serve as a pre-test/post-test for students' higher order thinking skills, especially not one that also drew a comparison with students' lower order thinking skills. Related instruments were found [32,33] but these did not match the scope of the current study well enough. Consequently, new measures were created and different types of data were collected to triangulate the data and substantiate the findings [34]. However, future research on a larger scale would help to determine the quality of these measures and, with that, substantiate the effectiveness of the redesign.

Second, the author of this paper was responsible for redesigning the course, and teaching it throughout all three years (including grading students' work). However, the author tried to remain as objective as possible; for example, during the grading process, she first determined what the papers were graded on the previous year, and then applied that way of working and standard to the current year. Moreover, an effort was made to triangulate the information from the grades by using open-ended and closed-ended questionnaire data.

Third, the educational intervention was on a relatively small scale. Examples exist in which the intervention was of a much bigger scale than executed here [26,27]. In addition, some researchers have included other relevant aspects in their approach. For example, Halpern [35] recommended paying attention to students' attitudes and dispositions (e.g., their willingness to engage in and persist at a complex task and their flexibility and open-mindedness) and the way in which students can transfer their skills outside of the context of a specific course. However, as the course objectives aimed beyond stimulating students' higher order thinking skills, it was not possible to increase the scale of the redesign.

13. Suggestions for Future Research

In addition to further substantiating the current insights by testing the redesign on a larger scale, additional areas for future research are likely to be fruitful. One of the most important ones is to dive deeper into students' learning process. The current study helped direct the way that the redesign was effective, but additional insight into why it was effective, in which instances it is effective, and what the underlying working mechanism is [36] is needed. One especially fruitful way to gain such insights would be to focus on developing teachers' pedagogical knowledge and pedagogical content knowledge. This will influence their teaching behavior in a way that is most likely to generate students' learning for understanding [37].

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the University of Twente ethics committee BMS on 06-02-2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bolander Laksov, K.; McGrath, C. Failure as a catalyst for learning: Towards deliberate reflection in academic development work. *Int. J. Acad. Dev.* **2020**, *25*, 1–4. [\[CrossRef\]](#)
2. Brown, C.; Schildkamp, K.; Hubers, M.D. Combining the best of two worlds: Integrating data-use with research informed practice for school improvement. *Educ. Res.* **2017**, *59*, 154–172. [\[CrossRef\]](#)
3. Coburn, C.E.; Turner, E.O. Research on data use: A framework and analysis. *Meas. Interdiscip. Res. Perspect.* **2011**, *9*, 173–206. [\[CrossRef\]](#)
4. Schildkamp, K.; Poortman, C.L. Factors influencing the functioning of data teams. *Teach. Coll. Rec.* **2015**, *117*, 1–42. [\[CrossRef\]](#)
5. Cain, T. Teachers' engagement with published research: Addressing the knowledge problem. *Curric. J.* **2015**, *26*, 488–509. [\[CrossRef\]](#)
6. Barends, E.; Rousseau, D.M.; Briner, R.B. *Evidence-Based Management. The Basic Principles*; Center for Evidence-Based Management: Amsterdam, The Netherlands, 2014.
7. Carlson, D.; Borman, G.D.; Robinson, M. A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educ. Eval. Policy Anal.* **2011**, *33*, 378–398. [\[CrossRef\]](#)
8. Van Geel, M.; Keuning, T.; Visscher, A.J.; Fox, J.P. Assessing the effects of a school wide data-based decision-making intervention on student achievement growth. *Am. Educ. Res. J.* **2016**, *53*, 360–394. [\[CrossRef\]](#)
9. Miri, B.; David, B.C.; Uri, Z. Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Reseach Sci. Educ.* **2007**, *37*, 353–369. [\[CrossRef\]](#)
10. Lewis, A.; Smith, D. Defining higher order thinking: Theory into practice. *J. Teach. High. Order Think.* **1993**, *32*, 131–137.
11. Kuhn, D. Do students need to be taught how to reason? *Educ. Res. Rev.* **2009**, *4*, 1–6. [\[CrossRef\]](#)
12. Shepard, L. Psychometricians' beliefs about learning. *Educ. Res.* **1991**, *20*, 2–9. [\[CrossRef\]](#)
13. Dowd, J.E.; Araujo, I.; Mazur, E. Making sense of confusion: Relating performance, confidence, and self-efficacy to expressions of confusion in an introductory physics class. *Phys. Rev. Spec. Top.—Phys. Educ. Res.* **2015**, *11*, 010107. [\[CrossRef\]](#)
14. Gorham, J.; Millette, D.M. A comparative analysis of teacher and student perceptions of sources of motivation and demotivation in college classes. *Commun. Educ.* **1997**, *46*, 245–261. [\[CrossRef\]](#)
15. Bruinsma, M. Motivation, cognitive processing and achievement in higher education. *Learn. Instr.* **2004**, *14*, 549–568. [\[CrossRef\]](#)
16. Weinstein, C.E.; Mayer, R.E. The teaching of learning strategies. In *Handbook of Research on Teaching*; Wittrock, M., Ed.; Macmillan: New York, NY, USA, 1986; pp. 315–327.
17. Pintrich, P.R.; De Groot, E.V. Motivational and self-regulated learning components of classroom academic performance. *J. Educ. Psychol.* **1990**, *82*, 33–40. [\[CrossRef\]](#)
18. Krathwohl, D.R. A revision of Bloom's taxonomy: An overview. *Theory Into Pract.* **2002**, *41*, 212–218. [\[CrossRef\]](#)
19. Zohar, A. Elements of teachers' pedagogical knowledge regarding instruction of higher order thinking. *J. Sci. Teach. Educ.* **2004**, *15*, 293–312. [\[CrossRef\]](#)
20. Kuhn, D.; Dean, D. Connecting scientific reasoning and causal inference. *J. Cogn. Dev.* **2004**, *5*, 261–288. [\[CrossRef\]](#)
21. Bloom, B.S. Taxonomy of educational objectives. In *Handbook 1: Cognitive Domain*; Longmans Green & Co.: London, UK, 1954.
22. Gagne, R.M. *The Conditions of Learning*, 2nd ed.; Holt, Rinehart & Winston: New York, NY, USA, 1974.
23. Marzano, R.; Miedema, W. *Leren in Vijf Dimensies: Moderne Didactiek Voor Het Voortgezet Onderwijs [Learning in Five Dimensions: Modern Didactics for Secondary Education]*; Koninklijke van Gorcum: Assen, The Netherlands, 2018.
24. Prakash, R.; Litoriya, R. Pedagogical transformation of Bloom Taxonomy's LOTs into HOTs: An investigation in context with IT education. *Wirel. Pers. Commun.* **2022**, *122*, 725–736. [\[CrossRef\]](#)
25. Newman, F.M. Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *J. Curric. Stud.* **1990**, *22*, 41–56. [\[CrossRef\]](#)
26. Hugerat, M.; Kortam, N. Improving higher order thinking skills among freshmen by teaching science through inquiry. *Eurasia J. Math. Sci. Technol. Educ.* **2014**, *10*, 447–454. [\[CrossRef\]](#)
27. Saido, G.A.M.; Siraj, S.; DeWitt, D.; Al-Amedy, O.S. Development of an instructional model for higher order thinking in science among secondary school students: A fuzzy Delphi approach. *Int. J. Sci. Educ.* **2018**, *40*, 847–866. [\[CrossRef\]](#)
28. Zohar, A.; Nemet, F. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *J. Res. Sci. Teach.* **2001**, *39*, 35–62. [\[CrossRef\]](#)
29. Papathomas, L.; Kuhn, D. Learning to argue via apprenticeship. *J. Exp. Child Psychol.* **2017**, *159*, 129–139. [\[CrossRef\]](#)
30. Voss, J.F.; Means, M.L. Learning to reason via instruction in argumentation. *Learn. Instr.* **1991**, *1*, 337–350. [\[CrossRef\]](#)
31. Dane, A.U.; Schneider, B.H. Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clin. Psychol. Rev.* **1998**, *18*, 23–45. [\[CrossRef\]](#)
32. Ghanizadeh, A. The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *High. Educ.* **2017**, *74*, 101–114. [\[CrossRef\]](#)
33. Halpern, D.F. *Halpern Critical Thinking Assessment: Test Manual*; Schuhfried GmbH: Mödling, Austria, 2012.
34. Poortman, C.L.; Schildkamp, K. Alternative quality standards in qualitative research? *Qual. Quant.* **2012**, *46*, 1727–1751. [\[CrossRef\]](#)
35. Halpern, D.F. Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *Am. Psychol.* **1998**, *53*, 449–455. [\[CrossRef\]](#)

-
36. Cohen, D.K.; Mehta, J.D. Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *Am. Educ. Res. J.* **2017**, *54*, 644–690. [[CrossRef](#)]
 37. Van Driel, J.H.; Berry, A. Pedagogical content knowledge. In *International Encyclopedia of Education*, 3rd ed.; Peterson, P., Baker, E., McGaw, B., Eds.; Elsevier: Amsterdam, The Netherlands, 2010; pp. 656–661.