Enhancing Genetics Understanding through Collaborative Learning using a Conversational Agent

Naomi Loffeld

S2474719

University of Twente

Department of Behavioural and Management Sciences (BMS)

Department of Educational Psychology (EP)

Bachelor Thesis

Dr. Hannie Gijlers

Dr. Lily Chen

Date: 23rd of January 2025

APA: 7th Edition

Abstract

Attempts to enhance genetics understanding must be made, as it holds a crucial position in secondary education. Yet, Dutch secondary schools often fail to foster a deep understanding of abstract concepts. Research has shown that innovative strategies focusing on active knowledge construction and collaborative learning effectively aid students in understanding abstract concepts. Within these strategies students must be steered towards Academically Productive Talk (APT) and collaboration using adequately timed prompts which a teacher alone cannot do. Hence, there is reason to believe that strategies such as Computer Supported Collaborative Learning (CSCL) using a Conversational Agent (CA) instead should enhance genetics understanding. The between-subjects factor was the presence or absence of CA CLAIR, and the within-subjects factor was the time point during which the level of genetics understanding was tested: pre-test, immediate post-test, and delayed post-test. The expectation was that the total scores on the post-tests would be higher compared to the scores on the pre-test, and that the scores of the intervention group would be higher than the control group. Additionally, the expectation was that the attitude towards the learning environment and peer collaboration would be higher for the intervention group than the control group. The learning environment overall produced positive learning results for both groups. Despite this, CLAIR's impact should be investigated by further analyzing the results and based on that finetune CLAIR's responses, as currently the intervention group did not score better on the knowledge tests nor attitude questionnaire than the control group. The research serves as a foundation for further research concerning CA implementation in Dutch secondary schools.

Keywords: Genetics understanding, Academically Productive Talk (APT), Active knowledge construction, Conversational Agents (CA), Collaborative Learning Agent for Interactive Reasoning (CLAIR)

Introduction

Secondary school education forms the foundation for understanding abstract concepts (Meelissen et al., 2012). However, as noted by Meelissen et al. (2012), over the last couple of years, the overall understanding portrayed by students with regards to abstract concepts has decreased significantly. This observation aligns with De Jong's (2019) findings that students often lack deep understanding of abstract scientific concepts. Research conducted by the Organisation for Economic Co-operation and Development (OECD, 2016), reports that "Dutch students are less willing to work through problems that are difficult, they do not remain interested in the tasks that they start, and, more than in other countries, they are likely to shy away from complex problems" (p.84). This poses interesting challenges for subjects such as math, physics, and biology. Within biology, genetics is among one of the concepts that stands out as particularly significant due to its central role in understanding an abundance of biological frameworks, such as evolution, biodiversity, and hereditary diseases. Moreover, early and adequate understanding of genetics creates the foundation for possible future medical advances (Meelissen et al., 2012; OECD, 2016). Adequate genetics understanding in secondary school is especially important given the trend identified by the OECD (2016), which highlights a shift in student interest from vocational to academic tracks. Therefore, genetics holds a crucial position in secondary education.

Given the essential role genetics plays in biological frameworks, it is crucial to innovate teaching methods to address challenges students face in secondary school to ensure that students gain a deep understanding of the topic and are genuinely prepared for further education (Meelissen et al., 2012; OECD, 2016). Although the decentral nature of the Dutch educational system provides ample opportunities to implement new teaching methods, current implemented methods often fail to foster a deep understanding of abstract concepts. To bridge this gap, innovative strategies focusing on active knowledge construction are proven to be among the most effective for teaching and developing a deep understanding of abstract subjects such as genetics (De Jong, 2019). Moreover, one promising teaching method worth reintroducing and further developing is Collaborative Learning (CL), in which students collaborate with one another to complete a task (De Jong, 2019; Karami et al., 2012). Provided that the students are willing to engage in the intervention, work cooperatively and engage in Academically Productive Talk (APT) (De Araujo et al., 2023; De Jong, 2019; Mao et al., 2021). Collaboration is

particularly effective in inquiry learning settings, as this mimics scientific methods, provides ample opportunities for discussing and sharing ideas, thus allowing for further active knowledge construction (Bell et al., 2009; De Araujo et al., 2023; Pedaste et al., 2015). Through this active knowledge construction and collaborative dialogues rooted in APT, students not only enhance their own learning, but also aid in their peers' learning experiences, thereby fostering mutual skill development (Karami et al., 2012).

Effective active knowledge construction using CL and inquiry learning is not just about pairing the students, they require guidance. Students need to receive adequate prompts at the right time and have a model for collaborative behaviour. Usually this is provided by a teacher, however, providing this might be an issue as they "do not work and learn in a collaborative culture" (OECD, 2016, p. 18). In addition, CL in the classroom can be quite time consuming for the teachers, as they need to mediate and guide all groups at the same time. Research has focused on using technology to create online environments that support both students and teachers in CL. These environments, known as Computer Supported Collaborative Learning (CSCL), are supported by Conversational Agents (CAs). Rather than being an information or answer supplier, CAs focus on enhancing collaboration between students by modelling collaborative behaviour and providing timely and adequate responses to stimulate APT and facilitate a deeper understanding of the topic— something a human teacher cannot achieve alone (De Araujo et al., 2023; Nguyen, 2022; OECD, 2016). Moreover, Cunningham-Nelson et al. (2019, as cited in Okonkwo and Ade-Ibijola, 2021, p1) found that "these systems can not only improve student engagement and support, but they can also greatly lessen lecturers' administrative workload, allowing them to focus on curriculum development and research''. Traditional CAs require students to pose questions themselves and actively seek help from the agent. More importantly, the traditional CAs do not prioritize productive collaborative dialogues, an essential component for the active knowledge construction in such learning environments (De Araujo et al., 2023; Sikström et al., 2022). Hence, the traditional CA should be improved upon to maximize effective collaboration by providing adequately timed prompts, model collaborative behaviour and foster productive collaborative dialogues to more effectively support students' learning processes.

An improved CA is Collaborative Learning Agent for Interactive Reasoning (CLAIR), a CA designed to be multilingual, applicable across many subjects and focuses heavily on steering students discussion towards productive collaborative dialogue, without intervening the students

in their learning process (De Araujo et al., 2023; Nguyen, 2022). CLAIR provides prompts tailored to the students, models collaborative behaviour and steers them back to productive collaborative dialogue, with the ultimate goal to foster active knowledge construction of the topic in the Computer Supported Collaborative Learning (CSCL) environment (De Araujo et al., 2023, 2024; Stahl et al., 2022). Where traditional CA's only relied on APT, CLAIR's utterances are designed to add another dimension as the triggering mechanism now also includes timing (De Araujo et al., 2023, 2024). As CLAIR has recently been developed, research with regards to the effectiveness of a CSCL with integrated CLAIR on students' understanding of genetics has not yet been conducted. This study aims to fill this gap in educational research by investigating the effectiveness of CLAIR in enhancing genetics understanding in Dutch secondary schools by answering the research question: How does collaborative learning with CLAIR influence Dutch secondary school students' understanding of genetic concepts?

This research is relevant for future curricula formation of secondary schools, as successful implementation of CLAIR could foster a deeper understanding of genetics as well as aiding teachers in successfully educating their students on genetics. The research will take place in a Dutch secondary school, and be experimental of nature, with a between group design in which high school students will work in an online environment either with integrated CA CLAIR, the intervention group or without, the control group. This online environment is created using Twente Go-Lab and is rooted in inquiry-based learning. The students will participate in two classes in which they will be asked to work collaboratively on gaining a deeper understanding of genetics. For all students the learning environment will be the same, the only difference will be the inclusion of CLAIR. To assess the student's understanding of genetics their knowledge will be assessed at several points of time using a knowledge test: prior to the sessions (pre-test), immediately after each session (immediate post-test), and at a delayed point in time (delayed post-test). Additionally, at the end of each session, students completed an immediate post-test questionnaire to evaluate their perception of the collaboration with their peers. For the intervention group, the questionnaire also included an assessment of their opinions regarding CLAIR's guidance. It was hypothesized that the total scores on the post-tests would be higher compared to the pre-test. Moreover, it was also hypothesized that the total scores of the intervention group would increase more than the total scores of the control group. Finally, it was hypothesized that the attitude towards the learning environment and the peer collaboration would

be higher for the intervention group than the control group due to the interventions from CLAIR. This study aims to contribute to research in the educational field, to test whether the conversational agent can enhance secondary school students' understanding of genetics and their engagement in the learning process. All in an attempt to shape the future of education.

Method

Design

For this study a mixed design was used. The intention was to measure the effect of CLAIR by comparing control (without CLAIR) and intervention (with CLAIR) groups' levels of genetics understanding throughout different points of time during the intervention: pre-test, immediate post-test, and delayed post-test. The dependent variable was the level of genetics understanding at the different time points. The independent variable was the usage of conversational agent CLAIR in the online learning environment. To test the participant's knowledge, the participants were required to complete several knowledge tests designed to measure their understanding of genetics. The pre and delayed post-test consisted of an eleven-item knowledge test addressing key concepts covered in the teaching sessions, the immediate post tests covered the items belonging to the classes. To assess the participant attitude towards the learning environment, and CLAIR, both the control and intervention groups completed an eleven-item Likert scale questionnaire assessing their attitudes towards the learning environment and the collaboration with their peers. The intervention group answered four extra questions designed to evaluate their perceptions of CLAIR.

Participants

Eighty participants (41 female, 39 male) aged 12 to 14 years (M=13.0) were recruited by purposive sampling from Isendoorn College, a Dutch secondary school in Warnsveld. All participants were second-year HAVO or VWO students without prior genetics instruction. Participants were randomly divided over control and intervention groups and then randomly assigned to a pair. The intervention group included more students, as this increased the chance of appearances from CLAIR. In total, the control group consisted of twenty-seven participants, divided over 13 groups, whereas the intervention group consisted of 57 participants which were divided over 26 intervention groups. Since the study involves minors, an informed consent letter (Appendix A) was provided to the participants prior to the start of the experiment. The sample

was obtained purposely to guarantee the smallest level of differences between the participants. There was one withdrawal from the study.

Materials

Learning Environment

The learning environments were created in the Twente Go Lab environment. For each of the classes two Go Lab environments rooted in inquiry learning were created, one control version and one intervention version with CLAIR. As the intervention phase consisted of two classes, there were a total of four online learning environments (Appendix B). Since the research revolved around collaborative learning with or without CLAIR the learning environment was designed in a way to allow participants to work collaboratively in the chat box on the right side of the learning environment. Due to time constraints, these chats were not analyzed during the scope of this research.

Both classes covered five key concepts related to genetics, the first class covered concepts regarding traits, gametes, dominant and recessive traits, gametes, dominance and recessivity, and scientific methods. The lesson mostly consisted of comprehensive reading, based on this text questions were posed. Students could work together on answering these by chatting with their partner in the chat box on the right-hand side. After all key concepts were covered, the students had to complete a questionnaire in Microsoft Forms to assess their gained knowledge on the concepts covered by the class.

The second class covered key concepts such as monohybrid crossings, purebred crossing, crossings, second generation crossings, and argumentation questions. In this lesson the participants started off with a summary of the first class, after which they continued with the new key concepts. During which the participants practiced with hypothesis formation and crossings by using collaborative tools from the Go-Lab environment. After all, ten key concepts were covered, the students had to complete a questionnaire in Microsoft Forms to assess their gained knowledge on the concepts covered by the class.

CLAIR

The learning environments for the intervention groups all included CLAIR. CLAIR's utterances were triggered by keywords relating to the concepts covered in the classes, namely: traits, gametes, dominant, recessive, phenotype, genotype, purebred, genetics, alleles,

homozygous and heterozygous. Moreover, CLAIR included specific APT talk moves, see Table 1.

Table 1

Talk Moves APT Goals

Goal	Dutch	English
Checking- In	"Ik merk dat het een beetje stil is. Kunnen jullie met elkaar delen waar je aan werkt en wat je volgende stap zal zijn?", "Het is nu al een tijdje stil. Kunnen jullie je huidige voortgang met elkaar delen? Is er iets dat je onzeker maakt over wat je hierna gaat doen?"	"I notice that it's a bit quiet. Could you share with each other what you're working on and what your next step will be?" "It's been quiet for a while now. Could you share your current progress with each other? Is there anything making you uncertain about what to do next?"
Reorienting	"Ik raak de draad een beetje kwijt. Zullen we het gesprek weer terugbrengen naar de les?" "Het is voor mij lastig deze discussie te volgen. Praten jullie nog wel steeds over de les? Misschien is het beter het nu weer over de les te hebben."	"I'm losing track a little. Shall we bring the conversation back to the lesson? "It's hard for me to follow this discussion—are you still talking about the lesson? Maybe it's better to focus on the lesson again now."
Peer Coordination	"<discussant>, denk je dat je <speaker> kunt helpen? Probeer met elkaar te delen wat een goede volgende stap zou zijn."	"<discussant>, do you think you could help <speaker>? Try to share with each other what a good next step might be."

	"<discussant>, wil je <speaker> helpen? Probeer met elkaar te delen wat een goede volgende stap zou zijn."	"<discussant>, would you like to help <speaker>? Try to share with each other what a good next step might be."
Peer Questioning	"<discussant>, heb je vragen voor <speaker> over iets dat je niet eenvoudig vindt te begrijpen?", "<discussant>, denk je dat er iets is waarbij <speaker> je kan helpen het te begrijpen? Als dat zo is, vraag het maar."	"<discussant>, do you have any questions for <speaker> about something you find difficult to understand?" "<discussant>, do you think there's something <speaker> could help you understand? If so, feel free to ask."
What- if	<speaker>, <what_if>"	<speaker>, <what_if>"
Press for Reasoning	"<discussant>, hoe zou je wat <speaker> zegt anders kunnen verwoorden, of kun je een voorbeeld geven om te illustreren wat <speaker> heeft gezegd?" "<discussant>, hoe komen jouw ideeën overeen met die van <speaker> en wat zou je nog toe willen voegen?"	"<discussant>, how could you rephrase what <speaker> said, or could you provide an example to illustrate what <speaker> has said?" "<discussant>, how do your ideas align with <speaker>'s, and what would you like to add?"

Assessment Materials

In order to assess the knowledge gained by the participants they had to complete several knowledge tests, which consisted solely of multiple-choice questions. These were all created in Microsoft Forms, and included a fifteen-minute time cap. The pre-test and the delayed post-test covered all eleven concepts: traits, gametes, dominant and recessive traits, gametes, dominance and recessivity, scientific methods, monohybrid crossings, purebred crossing, crossings, second generation crossings, argumentation questions, and inferring parentage. In addition to the eleven test items, participants had to answer demographic questions, regarding age, gender, class and student number during the pre-test. The maximum score on the pre and delayed post-test were eleven. The immediate post-tests each included five items, in which the test items matched the ones covered in the Go-Lab environment. The maximum score on each of the immediate post-tests was five. All knowledge test items were graded by assigning points to the question in Microsoft Forms, correct answers granted one point, incorrect answers zero points. To ensure that all knowledge test items tested the correct topic, questions were created to be similar.

Both the control and intervention groups completed an additional eleven-item Likert scale questionnaire assessing their attitudes towards the learning environment as well as the collaboration with their peers. Additionally, the intervention group answered four extra questions designed to evaluate their perceptions of CLAIR. The maximum scores of these questionnaires were 77 and 28, respectively. The Likert scale items were the same across all questionnaires and based on the Multicultural Personality Questionnaire (MPQ) from Van Der Zee and Brinkmann (2004) and Van Der Zee et al. (2012). These have proven to be reliable, as they have high levels of internal consistency as well as a high level of test-retest reliability. Moreover, for this study it is important to highlight that the MPQ is reliable in assessing social initiative, open mindedness as well as flexibility (Van der Zee & Brinkman, 2004). For all knowledge test and questionnaire items, see appendix C.

Hardware

The participants completed a pre-test on their school issued iPad. This same device was used for the experimental phase as well as the post-test. The specific type of device did vary per participant. The tests were shared with the participants through a QR code which led to a Microsoft Forms.

Procedure

The experiment was conducted over a four-week period, and comprised three primary components: a pre-test administered before interaction with the learning environment, an intervention phase during which participants engaged with the learning environment and completed knowledge tests as well as an attitude questionnaire, as well as a post-test. To maintain consistency across all classes, a standardized script was utilized throughout all stages of the experiment.

Pre-test

Participants completed the pre-test one week prior to the classes (details on the test in the material section) and was shared with the participants using a QR code. The test was completed within fifteen minutes and was under the supervision of the researcher as well as the teacher of the participants. Participants were ensured that in this phase of the classes served as a baseline and that the tested concepts are unfamiliar for them. Moreover, only questions with regards to technical difficulties were answered. The Microsoft Forms included a fifteen-minute timer to ensure all participants spent the same maximum amount of time on the test, regardless of completion of the questionnaire.

Intervention phase

The intervention consisted of two forty-five-minute sessions in the Twente Go-Lab environment over the course of two weeks. The concepts were distributed across the classes to fit the time frame allocated for each session, The eleventh item was not included in the teaching sessions as it was deemed too difficult to adequately incorporate into the classes. At the beginning of class, the participants were asked to log-in to the Twente Go-Lab environment using the created personalized cards with their name, student number and group number (Appendix D). Participants were instructed to work individually and converse via the chat box in the learning environment. Directly after the Go-Lab session, both the control and intervention groups completed a five-item knowledge test as well as an eleven-item Likert scale questionnaire assessing their attitudes towards the learning environment and the collaboration with their peers. Additionally, the intervention group answered four extra questions designed to evaluate their perceptions of CLAIR.

Post-test

Participants individually completed the post-test one week after the intervention phase. The test was only under the supervision of the teacher of the participants and was shared with the participants using a QR code. Similar to the pre-test, the knowledge test consisted of eleven items to assess knowledge gained during the intervention. The eleventh item was included in the delayed post-test to assess student's deeper understanding and application of knowledge. Additionally, the eleven item attitude questionnaire was filled in by all participants. The intervention group answered four additional questions designed to evaluate their perceptions of CLAIR.

Data analysis

In order to collect the data gathered during the intervention, several statistical methods were used. To use these statistical methods, the dependent and independent variables must be identified: the independent variable was the usage of conversational agent CLAIR in the online environment. The dependent variable was the level of genetics prior to, during, and after the intervention in the online environment. The data of the intervention was analyzed using the following data analysis techniques in R studio. For the full R-code used to analyze the data see appendix E.

First, the parametric assumptions were visualized and tested. This meant that the data was tested and visualized for scale, normality, homogeneity of variance. These were tested by using the Shapiro-Wilk Test and a Levene Test. For the Shapiro-Wilk Test, a test statistic (W-value) of one was looked for, as this indicates normality. Moreover, the data from the test items was visualized using a histogram to substantiate the test. After this an ANOVA was performed. An ANOVA was used to compare the total scores of the participants. Yet, all three of the ANOVA assumptions must also be met to ensure reliability of the test. If these were not met, then a non-parametric alternative, such as the Kruskal-Wallis Test was conducted, as this test is less sensitive to outliers. This test is less sensitive to outliers than the ANOVA and determined whether there was a statistically significant difference between the medians.

To conduct the ANOVA or a non-parametric alternative, the dataset was pivoted to a long format, after which the model was created. For the ANOVA model or to be considered significant, a p-value should be <0.05. This meant that in this study a significance level of 0.05 was used. If the outcome of the ANOVA or its non-parametric alternative was significant, it

indicated that the total scores over the different points of time were not the same, but it does not talk about how they are different. In order to see this, the data was visualized in boxplots, grouped per class per point of time. In this plot the differences between classes, groups and points of time could be seen. To determine the significance of the differences between the groups, a Friedman Test and a Wilcoxon Signed Rank Test were conducted. Finally, the Likert scale test items were split into two data sets: the eleven item test which all participants received, and the four item test for the intervention group only. On the eleven item Likert scale, a maximum score of seventy-seven could be obtained and indicates a highly positive attitude towards the Collaborative learning experience. On the four item Likert scale a maximum total score of twenty-eight could be obtained and indicates a highly positive attitude towards CLAIR. Both were analyzed by first visualizing the frequency of the scores in a histogram, testing for normality using the Shapiro-Wilk test on the total scores, and calculating the Cronbach's Alpha to check internal consistency, for which a value above 0.7 was considered acceptable. For the Shapiro-Wilk Test, a test statistic (W-value) of one was looked for, as this indicates normality. These were analyzed in order to assess how the participants experienced the intervention and CLAIR. By using the code presented in appendix E, this could be realized. While the participants' chats were not extensively analyzed due to time constraints, an overview of the amount of times CLAIR has entered the chats in the intervention group as well as an overview of reactions to CLAIR was gathered. This was done in order to grasp an understanding of the peer collaboration of the participants as well as their responses to CLAIR.

In this study the objective was to find a significant increase in total test scores of the intervention group after successful completion of the classes in the Twente Go-Lab environment. This increase would only occur when the participants have gained a deeper understanding of genetics, and thus answered more questions correctly. It was hypothesized that the total scores on the post-tests would be higher compared to the pre-test. Moreover, it was also hypothesized that the total scores of the intervention group would increase more than the total scores of the control group. Finally, it was hypothesized that the attitude towards the learning environment and the peer collaboration would be higher for the intervention group than the control group due to the interventions from CLAIR.

Results

Parametric Assumptions

First the parametric assumptions were assessed namely scale, normality, and homogeneity of variance. Outliers were identified, yet not removed as there was not a viable reason to do so. The scale of the data was identified, namely continuous. In Figure 1 the histogram used to visualize the distribution can be found. As seen, the distribution of the data is skewed to the right. In Figure 2 the data is separated by group and time during the study. In both tables can be seen that the data is positively skewed, which indicates that most participants had low scores over all tests. To substantiate this, the Shapiro-Wilk Test was conducted. With $p<0.05$, as well as a test statistic of 0.95 it can be stated that the data is not normally distributed. After which the homogeneity of variance was tested using the Levene Test, which resulted in $p<0.05$. Since the data is not distributed normally, nor is the variance equal across all levels, the dataset does not fulfill the parametric assumptions required for statistical tests.

Figure 1

Histogram of Total Scores over all Tests

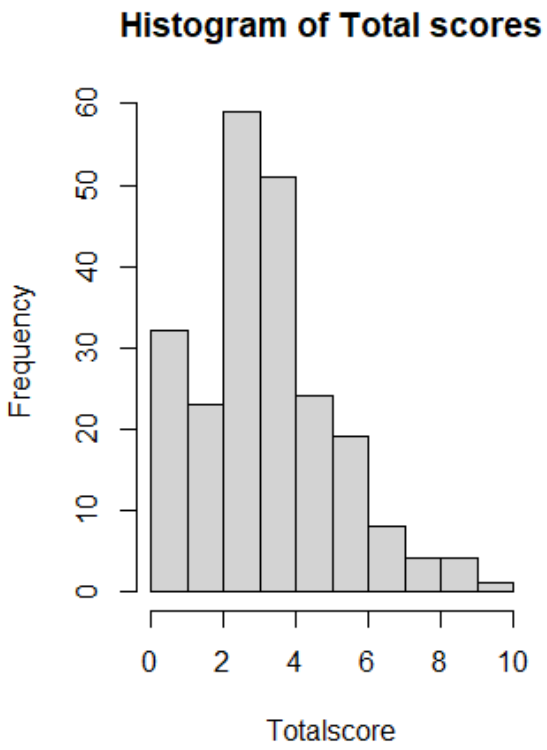
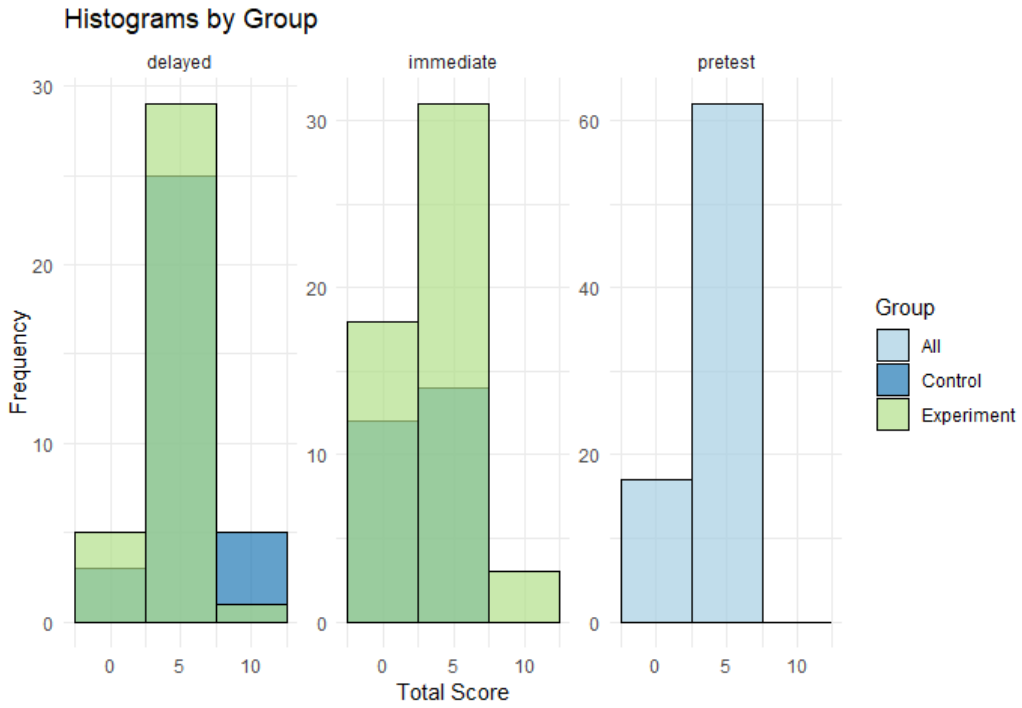


Figure 2

Histograms of Total Scores per Group and Time



CLAIR's Appearances and Responses to CLAIR

While the participant's chats were not analyzed in its entirety an overview of the amount of times CLAIR has entered the participants' peer collaboration in the chats can be seen in Table 3. During the sessions, CLAIR has appeared a total of 252 times for the intervention group, compared to the 0 times of the control group. In addition to this the participant's chats were looked into in order to get an idea of the participant's responses to CLAIR's appearances. The original response in Dutch, as well as the English translation used are listed in appendix F.

CLAIR has appeared a total of 148 times during the first session, while during the second session CLAIR appeared 104 times. Participants in the intervention group responded to CLAIR in varying ways. During this session, CLAIR motivated students to further explain their thoughts by saying: "[P1] How could you rephrase what [P2] is saying, or can you provide an example to illustrate what [P2] has said?" to which the participant responded: "I think so too, and 4 is the other one.". The participants were also steered back to working collaboratively by utterances of CLAIR when they have not sent chats in a while by saying: "I notice it's a bit quiet. Could you share with each other what you're working on and what your next step will be?." To which the

participants responded by asking each other where they were in the lesson and how to continue. At times, the participants did not respond to CLAIR in a kind manner. To illustrate, the participants were complaining that they did not understand much from the class. One participant (P1) said: “I do find it really difficult”, which the other (P2) confirmed by saying “Frrrr”. CLAIR then intervened and said:’’ I’m getting a bit lost. Shall we bring the conversation back to the lesson?”, to which P1 responded: “No, Claire”. After this response, the conversation side-tracked to other topics unrelated to the class. While in another instance the response to the same utterance of CLAIR was “Yes, sounds good!.”

In some cases, the participants asked CLAIR for help, and assumed it would provide the answer to them, this also led to varying responses. To illustrate, the participant (P1) asked CLAIR for help by saying: “Clair, will you help me?”, to which CLAIR responded “What kind of help do you need? Try asking your partner.” After which the participant did ask their partner (P2) for clarification. The participant (P2) responded by trying to help their partner (P1) find what they were looking for. In another instance, the participants were less willing to work together on the assignments, and at times a direct question concerning the lesson was posed to CLAIR: “Clair, do you know what to do when forming a hypothesis?”, To which CLAIR responded: “My name is Clair, I am a computer program, and my goal is to make discussions between people more productive.”. After which the participants responded with “Bro” (P1) and “Just answer or something” (P2).

Table 3

CLAIR Appearances per Group per Session

Group	Session 1	Session 2	Total
Intervention	148	104	252
Control	0	0	0

Total Scores of Intervention

To further visualize the data box plots were created. In these, it can be seen that regardless of the assigned group (intervention or control) the scores on the immediate post-test as well as the delayed post-test were higher than on the pre-test. This can be seen in Figure 3. Then, the data was further visualized in Figure 4 using boxplots to differentiate between intervention and control groups. In this plot, it can be seen that there was no significant difference between the control group and the intervention group. To substantiate this, a Kruskal-Wallis Test was conducted, of which the results can be seen in Table 4. The results of the Kruskal-Wallis test indicate that there is no statistically significant difference between the distributions of the total score across the groups. Thus, there is insufficient evidence to conclude that any group significantly differs from the others.

As seen in Figures 3 and 4, the scores of the delayed post-test are higher than the scores on the immediate post-test. To check the significance of this difference, a Friedman test was conducted, which indicates that there is a significant difference across the three time points, see Table 5. To further substantiate this, the Wilcoxon Signed-Rank test with a $p<0.05$ indicates that there is a significant difference between the immediate post-test and the delayed post-test scores.

Table 4

Kruskal-Wallis Test Results for Comparing Groups

	Chi-squared	DF	P-value
	3.14	2	0.21

Table 5

Friedman Test Results for Comparing Points of Time

	Chi-squared	DF	P-value
	14.8	2	6.1e-04

Table 6

Wilcoxon Signed Rank Test Results for Immediate and Delayed Post-Test

	V	P-value
	337	3.9e-06

Figure 3

Grouped Boxplot for Scores during Intervention

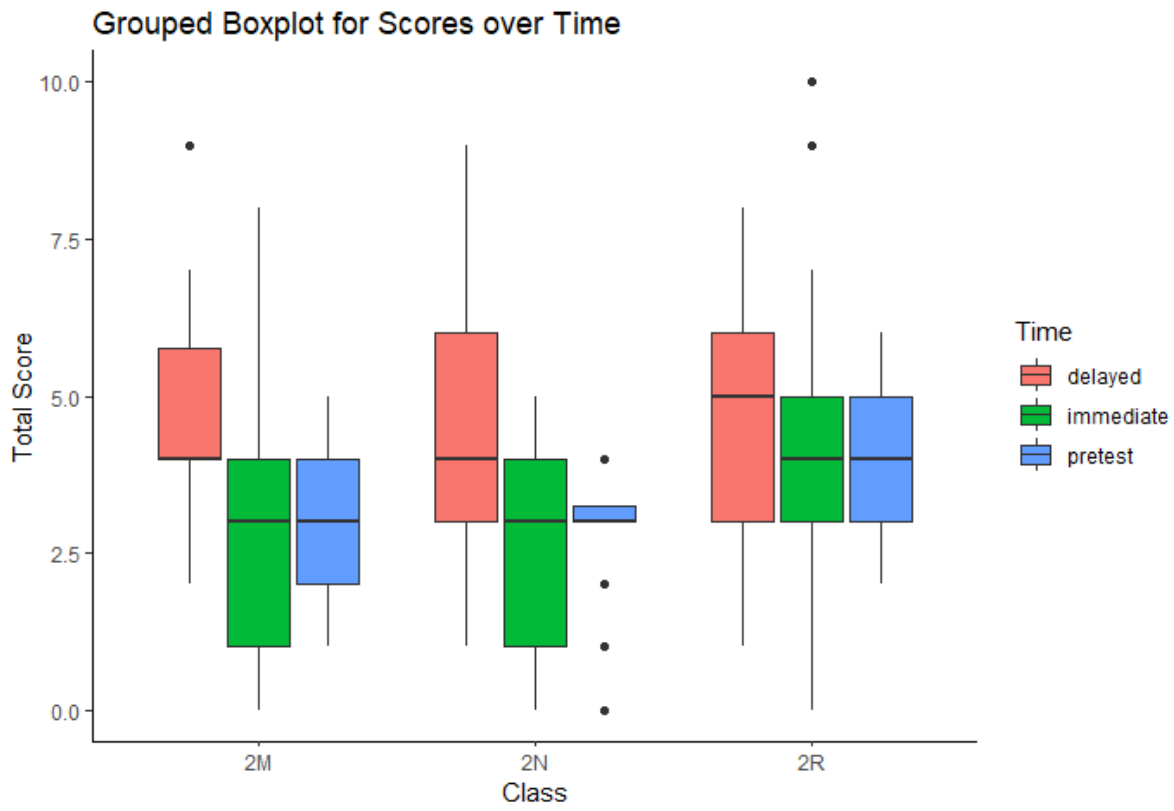
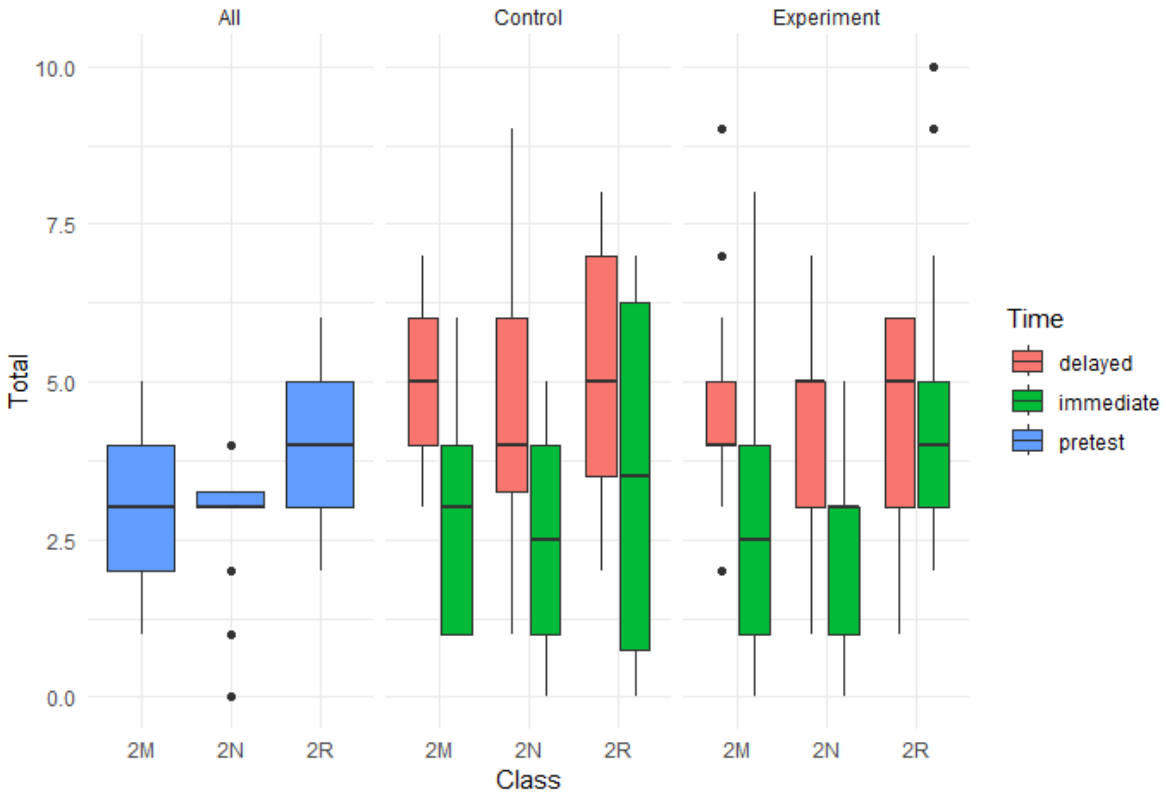


Figure 4

Grouped Boxplot for Scores in Control and Intervention Group



Note. All refers to the pre-test only as in this case the participants were not yet distributed over control or intervention group.

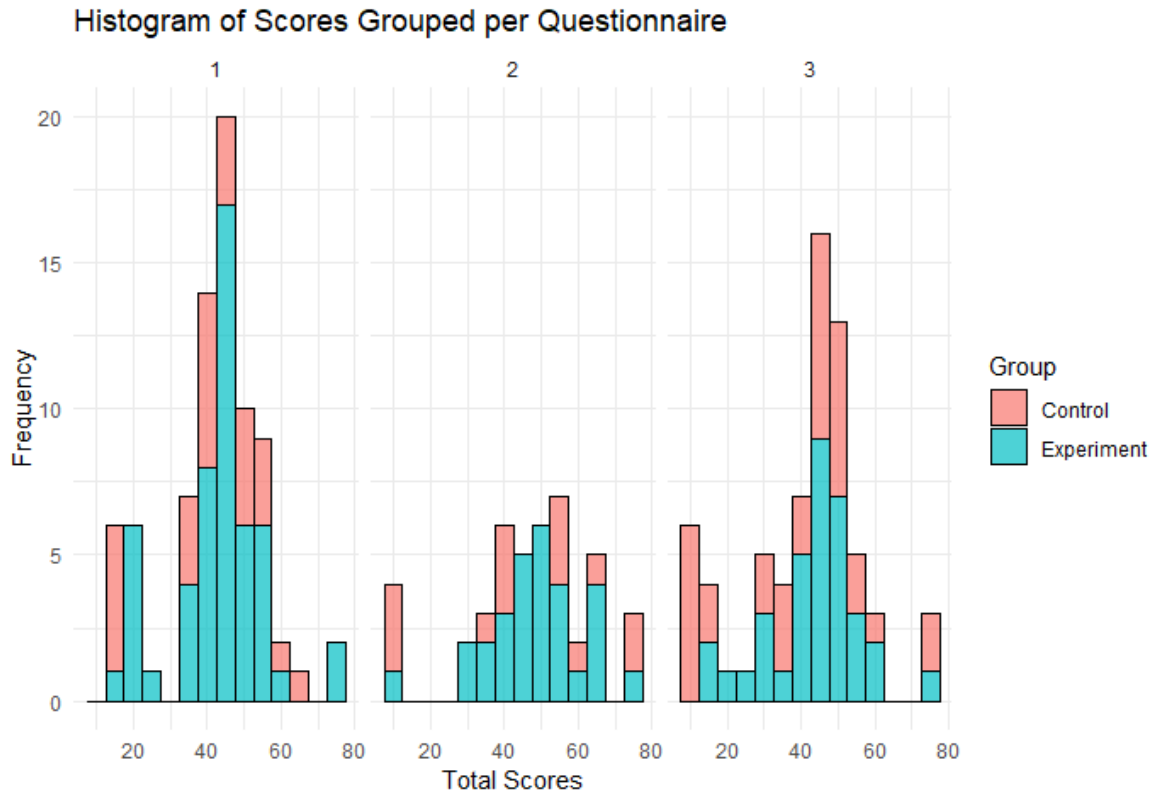
Attitude Towards Collaborative Learning Experience and CLAIR Interventions

To add another dimension to the research, the participants filled in an eleven item Likert scale questionnaire in which their experiences and attitude towards the intervention as a whole were tested (for all test items see Appendix C). To visualize the data a frequency histogram was created, see Figure 5. This figure shows the distribution of test scores, a maximum of 77 could be obtained and indicates a highly positive attitude towards the collaborative learning experience. It can be seen that in all cases the total scores for the control group are higher than for the intervention group. Moreover, the scores after the second class are higher than the scores after the first class. The scores after the delayed post-test are in turn higher than on both of the other test times. A Shapiro-Wilk test was conducted to further assess the normality of the data. Results

indicate that the scores deviate from a normal distribution, W=0.95, p<0.05. This suggests that the data were not normally distributed. To assess the reliability of the eleven-scale item questionnaire, a Cronbach's Alpha analysis was performed. The questionnaire demonstrates excellent internal consistency α= 0.97, across all 11 items. This indicates that the items reliably measure the participants attitudes towards their collaborative learning experience.

Figure 5

Frequency Histogram for Attitude Towards Collaborative Learning Experience



Note. 1 refers to the first time the items were tested after the first class, 2 and 3 refer to the later test moments after the second class and during the delayed post-test, respectively.

Table 7

Shapiro-Wilk Test Results for Attitude Towards Collaborative Learning Experience Items

W	P - value
0.94818	2.3e-06

The intervention group completed an additional four item Likert scale questionnaire to assess their attitude towards the interventions from CLAIR (for all test items see Appendix C). A Shapiro-Wilk test was conducted to assess the normality of the data. Results of this test indicate that there was a deviation from this, $W = 0.93$, $p < 0.05$. This implies that the data were not normally distributed. Moreover, to further visualize the data, a frequency histogram was created, see Figure 6. This figure shows the distribution of test scores, in this case a maximum of 28 could be obtained and indicates a highly positive attitude towards CLAIR. It can be seen that the total scores over the three points in time –after the first class, after the second class and during the delayed post-test – the total scores increased, which suggests a more positive attitude towards CLAIR as the intervention progressed. The reliability of the four-item questionnaire was assessed using Cronbach's Alpha. The questionnaire demonstrates excellent internal consistency $\alpha = 0.92$, across all four items. This indicates that the items reliably measure the participants attitudes towards CLAIR.

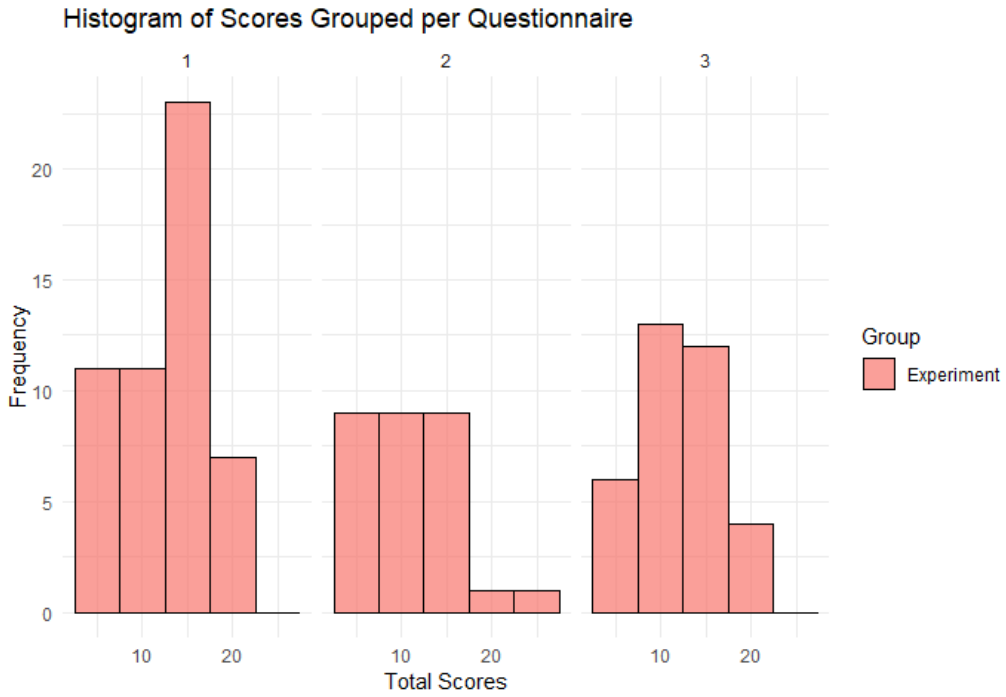
Table 8

Shapiro-Wilk Test Results for Attitude Towards CLAIR Test

W	P - value
0.93994	5.7e-05

Figure 6

Frequency Histogram of Attitude Towards CLAIR



Discussion

In this study the main objective was to find a significant increase in total test scores of the intervention group after completion of the classes in the Twente Go-Lab environment. This increase would only occur when the participants have gained a deeper understanding of genetics, and thus answered more questions correctly. Based on the results can be concluded that an overall increase in genetics understanding has occurred. For all participants, the scores on the knowledge tests in the delayed post-test are significantly higher than the total scores obtained in the pre-test as well as the immediate post-test. Yet, upon comparing the two conditions, the participants in the control group consistently obtained higher scores on the knowledge tests than the intervention group. The participants' attitude towards the collaborative learning experience was measured using test items adapted from the MPQ (Van Der Zee and Brinkmann, 2004; Van Der Zee et al., 2012) for this study specifically. These test items have proven to have excellent consistency, and thus measure the constructs adequately. The participants' attitudes towards the collaborative learning experience has shown to be moderately positive for both conditions, as the

most frequent scores have surpassed more than half of the available score of 77. In fact, as the intervention progressed, the overall attitudes became slightly more positive. Despite this, the attitudes of the control group were more positive than the intervention group.

To conclude, the hypothesis that *the total scores on the post-tests would be higher compared to the pre-test* is true, as the total scores on the delayed post-test are significantly higher than the pre-test scores. Next, the hypothesis that *the total scores of the intervention group would increase more than the total scores of the control group* is not true, as the intervention group scored lower than the control group. Finally, the hypothesis that *the attitude towards the learning environment and the peer collaboration would be higher for the intervention group than the control group due to the interventions from CLAIR* is also not true, as –throughout the three different measuring points– the control group scored higher on the attitude tests than the intervention group. Since previous research on enhancing genetics understanding using CLAIR has not been conducted, there is no research to compare these findings to. Despite this, the research has proven to be foundational for further research on the topic, as the main objective of obtaining a significant increase in genetics understanding as a result of the intervention was obtained. Yet, the scores of the intervention group on the knowledge test, as well as the attitude test are not significantly higher than the scores of the control group. Hence the limitations of this first attempt at researching the implementation of CLAIR into secondary education must be identified in order to successfully conduct further research.

Strengths, Limitations & Recommendations

Learning Environment and CLAIR

CLAIR and its interventions are a vital part of the research. The primary function of CLAIR is to steer participants' discussion towards productive collaborative dialogue, without intervening the students in their learning process (De Araujo et al., 2023; Nguyen, 2022). However, for CLAIR to intervene, participants did need to engage in collaborative talk, meaning that opportunities for collaboration needed to be created in the learning environments. The created learning environments did exactly that, as CLAIR has appeared a total of 252 times. Compared to one another, the second class should have provided a bit more opportunities for this as this was more inquiry based which allowed for even more collaboration tools to be incorporated in the environment (Bell et al., 2009; Sui et al., 2022). Despite this, CLAIR

appeared more often in the first session indicating that possibly the second class was more difficult. With the knowledge gained from (OECD, 2016), this could have resulted in less willingness to work through the problem resulting in less (collaborative) dialogue and thus less opportunities for CLAIR to intervene.

Based on responses from participants, it also seems that there have been some inadequately timed prompts; CLAIR was telling students to go back to the lesson while they were actively discussing the topics, which had the opposite effect. Resulting in participants responding to this, telling her to "shut up," rather than remaining focused on working collaboratively. These inaccurately timed and ill-received responses could have influenced the scores obtained by the intervention group both the knowledge test as well as the attitude questionnaire. A possible reason for the effectiveness of CLAIR's prompts are rooted in lexical alignment, which is the process of how two or more people in a conversation adapt their language to match one another (Srivastava et al., 2023). Thomas et al (2020, as cited in Srivastava et al., 2023) found that alignment of language in word choice and utterance length can also be used to ensure smooth conversation between conversational agents – and in this case CLAIR – and their users. Srivastava et al. (2023) then found that adequate lexical alignment between conversational agent and user has a positive effect on short-term information recall and understanding. So, it could be the case that CLAIR's prompts are misaligned with the language use of the participants, limiting them from reaching their full potential of genetics understanding. For future research it is recommended to look into whether this is the case for CLAIR, and if so, look into possibilities for fine tuning CLAIR based on lexical alignment with its users.

Another example of where lexical alignment plays an important role is in the login name to the learning environment. Participants were instructed to login to the Twente Go Lab environment using their school issued student numbers on their own iPad. This was done in order to ensure familiarity with the login name, aiming to limit invalid login attempts and ensuring anonymity. It was assumed that participants use these credentials more often, indicating familiarity with the login. Yet, some participants used other logins than this. As participants were assigned to a condition and a peer student using their student number, this resulted in a short inability to use the learning environment as intended for both them and their peers. These participants were instructed to login using their provided login credentials to complete the class as intended, however, it did limit several students from successfully participating in the

intervention, possibly limiting their potential genetics understanding. This suggests that in future learning environments a more natural login to the learning environment should be created to limit failed log-in attempts. Perhaps having them choose their own logins could limit failed login attempts and thus prevent it from hindering participants' time spent on the learning environment.

So, overall, the learning environments did provide ample opportunities for collaboration and thus for interventions from CLAIR. However, it could be the case that some topics or collaborative tools were simply too difficult for the participants. Future research would benefit from more explanation regarding the collaboration tools. Currently the tools were explained using videos, perhaps a classroom exercise for getting to know the learning environment would be more beneficial. Moreover, CLAIR's programmed responses should be finetuned in order to flow into the participants' collaboration more seamlessly. Which can be achieved by lexically aligning CLAIR's responses to its users. This would allow for smoother interventions of CLAIR to model collaborative behaviour and steering them back to productive collaborative dialogue.

Technical Difficulties

During the pre-test, the intervention phase in the Twente Go Lab environment, the immediate post-tests as well as the delayed post-test numerous participants have experienced technical difficulties. Some were inherent to their own devices, which is out of the researcher's control, while others were applicable for a multitude of participants. For example, participants received an error indicating that there was trouble connecting to the server. Moreover, during the second-class participants experienced Wi-Fi stability issues. This led to participants having difficulty working collaboratively and submitting answers in the Twente Go Lab environment as well as submitting the answers to the immediate post-test. Resulting in an incomplete dataset for the second class, and its corresponding immediate post-test. In general it seems that the participants' iPads struggled to adequately run the Twente Go Lab, impeding the collaborative learning experience of some participants. In general future research could opt to look into running the program on computers rather than on iPad, perhaps these can run the Twente Go Lab environment more efficiently.

Participants and Attitudes

All participants were required to complete the pre-test prior to starting with the first class. This meant that for all participants, data was available to assess genetics understanding prior to the intervention. As the phases of the research were spread over several weeks to accommodate

the curriculum of the school, there was a higher probability of a participant being absent. To ensure that all participants completed the pre-test prior to interacting with the learning environment, this meant that some participants had to complete the pre-test whilst the others were already working in the learning environment. Resulting in some participants and thus also their peers having less time to work collaboratively in the learning environment, leading to less possibilities for CLAIR to be prompted. Ultimately influencing their experience of the first part of the intervention. Furthermore, the increased possibilities for absenteeism resulted in some participants completing only one part of the intervention. As the second session built upon knowledge gained in the first, for some participants this meant that they were under prepared for the second class. Additionally, this led to an unequal amount of responses to the knowledge questions in each session. In general, absenteeism has affected the extent to which participants were able to adequately complete the intervention. For future research it is recommended to aim to limit the effect of absenteeism provided that the school curricula allow it.

Participants' attitude plays a big role in the success of the intervention, they need to be willing to engage in the intervention (Mao et al., 2021). So, the more positive their attitudes towards the intervention, the more likely the intervention is to be successful. While the overall attitude towards the learning experience was moderately positive, as measured by the attitude scale items, the students expressed negative behaviour and attitudes during the pre-test and the classes. To illustrate, participants behaved unruly during the pre-test. While it was stressed that we did not expect the participants to answer the questions correctly and that it would serve as a baseline to compare their progress to, the participants stated that the test was too difficult for them to complete. Some even showed an unwillingness to participate, resorting to guessing without reading the questions. Which aligns with the findings of OECD (2016), that Dutch students tend to shy away from difficult problems. In turn, this influenced the test scores used to assess the level of genetics understanding. To counteract this, future research can distinguish between participants who actively participate and those who do not. As currently the results still include participants who guessed answers. This can be done by implementing control questions throughout the questionnaires to check for attention. Which would allow for more true data.

Finally, the sample consisted of eighty participants, a sample of moderate size. Often this results in less reliable estimates of variance, and thus may not accurately reflect the population. Based on this, and previously mentioned recommendations, it is evident that more data should be

conducted. This can be achieved by increasing the sample size by for example repeating the research at the same secondary school over several years. The latter would also limit differences between groups and allow for the data to be compared. In addition to this, data could be gathered at different, yet similar secondary schools, this data can then eventually be compared to the data collected in this research. Together, this would allow for a better reflection of the population and allow for more accurate results.

Future Research and Practical Implications

While the current study is the first to research the possibility of enhancing genetics understanding through collaborative learning with CLAIR, research on the effectiveness of collaborative learning in online environments has been conducted. The findings of this study showcase that collaborative learning in an online learning environment is indeed effective for enhancing understanding of abstract concepts. Moreover, the current study does provide a firm foundation for further research on the implementation of CLAIR in these online collaborative learning environments. The results from the attitude questionnaires suggest that more extensive analysis of the chats could aid in uncovering the reason behind the moderately positive attitudes towards ClAIR. Based on this, more specific recommendations can arise to pinpoint where other improvements are required. Together with the previously made recommendations, possible subsequent research can gain a deeper understanding of participants' collaborative learning behaviour and experiences as well as their attitudes towards the learning environment and CLAIR. All in order to continue to further shape the future of secondary education.

References

Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2009). Collaborative Inquiry Learning: Models, tools, and challenges. *International Journal of Science Education*, *32*(3), 349–377. https://doi.org/10.1080/09500690802582241

De Araujo, A., Papadopoulos, P. M., McKenney, S., & De Jong, T. (2023). Supporting collaborative online science education with a transferable and configurable conversational agent. *Computer-supported Collaborative Learning/˜ the œComputer-Supported Collaborative Learning Conference*. https://doi.org/10.22318/cscl2023.469853

De Araujo, A., Papadopoulos, P. M., McKenney, S., & De Jong, T. (2024). A learning analytics-based collaborative conversational agent to foster productive dialogue in inquiry learning. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.13007

De Jong, T. (2019). Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning*, *35*(2), 153–167. https://doi.org/10.1111/jcal.12337

Karami, M., Pakmehr, H., & Aghili, A. (2012). Another view to importance of teaching methods in curriculum: collaborative learning and students' critical thinking disposition. *Procedia - Social and Behavioral Sciences*, *46*, 3266–3270. https://doi.org/10.1016/j.sbspro.2012.06.048

Mao, P., Cai, Z., He, J., Chen, X., & Fan, X. (2021). The Relationship between Attitude toward science and academic Achievement in Science: A Three-Level Meta-Analysis. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.784068

Meelissen, M. R. M., Netten, A., Drent, M., Punter, R. A., Droop, M., & Verhoeven, L. (2012). *PIRLS-en TIMSS-2011: Trends in leerprestaties in Lezen, Rekenen en Natuuronderwijs*. Radboud Universiteit / Universiteit Twente. https://doi.org/10.3990/1.9789036534758

Nguyen, H. (2022). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education, 192,* 104661. https://doi.org/10.1016/j.compedu.2022.104661

OECD. (2016). Netherlands 2016: Foundations for the Future, Reviews of National Policies for Education. In *Reviews of National Policies for Education/Reviews of national policies for education*. OECD Publishing. https://doi.org/10.1787/9789264257658-en

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education*, *188*, 104564. https://doi.org/10.1016/j.compedu.2022.104564

Srivastava, S., Theune, M., & Catala, A. (2023). The Role of Lexical Alignment in Human Understanding of Explanations by Conversational Agents. *IUI*, 423–435. https://doi.org/10.1145/3581641.3584086

Stahl, G., Koschmann, T., & Suthers, D. (2022). Computer-Supported collaborative learning. In *Cambridge University Press eBooks* (pp. 406–427). https://doi.org/10.1017/9781108888295.025

Sui, C., Chen, H., Cheng, P., & Chang, C. (2022). The Go-Lab Platform, an Inquiry-learning Space: Investigation into Students' Technology Acceptance, Knowledge Integration, and Learning

Outcomes. *Journal of Science Education and Technology*, *32*(1), 61–77. https://doi.org/10.1007/s10956-022-10008-x

Van Der Zee, K. I., & Brinkmann, U. (2004). Construct Validity Evidence for the Intercultural Readiness Check against the Multicultural Personality Questionnaire. *International Journal of Selection and Assessment*, *12*(3), 285–290. https://doi.org/10.1111/j.0965-075x.2004.283_1.x

Van Der Zee, K., Van Oudenhoven, J. P., Ponterotto, J. G., & Fietzer, A. W. (2012). Multicultural Personality Questionnaire: Development of a Short form. *Journal of Personality Assessment*, *95*(1), 118–124. https://doi.org/10.1080/00223891.2012.718302

Wang, L., & Calvano, L. (2022). Class size, student behaviors and educational outcomes. *Organization Management Journal*, *19*(4), 126–142. https://doi.org/10.1108/omj-01-2021-1139

Appendix A

Informed Consent Letter

Geachte ouders/verzorgers,

Het Isendoorn College doet mee aan een onderzoeksproject van de afdeling Instructietechnologie van de Universiteit Twente. De Universiteit Twente heeft een leeromgeving ontwikkeld voor het mogelijk verbeteren van biologie-onderwijs, waar onze leerlingen mee aan de slag zullen gaan. Deze leeromgeving is nog niet een onderdeel van het reguliere onderwijsaanbod, en zal daarom dus uitgeprobeerd worden.

Wij informeren u met deze brief over het onderzoek dat aan het gebruik van de leeromgeving gekoppeld is. Om te onderzoeken wat de leerlingen precies leren van deze leermiddelen, zullen de onderzoekers tijdens het onderzoek tweemaal een vragenlijst en een kennistoets afnemen. Ook zullen zij observeren op welke manier onze leerlingen de leermiddelen gebruiken. Wij willen als Isendoorn College graag bijdragen aan het evalueren en verbeteren van deze lesmaterialen.

Vrijwilligheid

We willen benadrukken dat deelname aan dit onderzoek volledig vrijwillig is. U als ouder/verzorger heeft te allen tijde het recht om uw toestemming in te trekken, zonder opgaaf van reden. Mocht u besluiten om de toestemming in te trekken, dan zullen de onderzoekers geen gegevens meer verzamelen voor de evaluatie van de leermiddelen.

Wij willen benadrukken dat er tijdens het onderzoek geen audio, video of fotomateriaal verzameld zal worden.

Ook als de leerling of ouder/verzorger zich terugtrekken zullen de leerlingen gebruik blijven maken van de leeromgeving, omdat deze onderdeel zijn van de reguliere lessen, er zullen dan echter geen gegevens verzameld worden. Tot twee weken na deelname (30 november 2024) kunnen de onderzoekers de onderzoeksgegevens en persoonsgegevens makkelijk verwijderen uit de database. Mocht u gegevens willen verwijderen dan kunt u dit doen door een mail te sturen naar [n.loffeld@student.utwente.nl en a.h.gijlers@utwente.nl].

Wat gebeurt er met de gegevens?

De onderzoeksgegevens die we in dit onderzoek verzamelen, zullen door wetenschappers gebruikt worden voor datasets, artikelen en presentaties. Alleen het onderzoeksteam heeft toegang tot de volledige gegevens. Deze gegevens worden veilig opgeslagen (in overeenstemming met de wetgeving inzake gegevensbescherming) op een versleutelde server van de Universiteit Twente. De bestanden worden geanonimiseerd en versleuteld. De anoniem gemaakte onderzoeksgegevens blijven tenminste 10 jaar beschikbaar voor onderzoek. Onderzoekers die de gegevens gebruiken, weten dus niet van wie deze gegevens afkomstig zijn.

Meer Informatie

Indien u meer informatie wenst over het onderzoek, kunt u contact opnemen met Hannie Gijlers via e-mail [a.h.gijlers@utwente.nl] of met Naomi Loffeld via [n.loffeld@student.utwente.nl]. Voor eventuele klachten over het onderzoek staat zij als verantwoordelijke onderzoeker tot uw beschikking.

Wij hopen dat u het waardevol vindt dat onze leerlingen deze unieke kans krijgen om deel te nemen aan dit onderzoeksproject. Samen kunnen we bijdragen aan de verbetering van het onderwijs en het verbeteren van abstracte kennis bij middelbare scholieren.

Instemming

Indien u instemt met deelname aan het onderzoek, hoeft u verder geen actie te ondernemen. Mocht u niet instemmen met deelname, dan verzoeken wij u dit kenbaar te maken voor 30 november 2024 door een mail te sturen naar [n.loffeld@student.utwente.nl].

Met vriendelijke groet,

[Name Teacher] namens het Isendoorn College

En

Naomi Loffeld, Hannie Gijlers en Lily Chen namens de Universiteit Twente

Appendix B

Go Lab Online Learning Environments

Table 1

Links to Twente Go Lab environments

Class	Condition	Link
1	Control	https://go-lab.bms.utwente.nl/ils-player/login/670e195373b17922dd49940a
1	Intervention	https://go-lab.bms.utwente.nl/ils-player/login/670e195373b17922dd49940a
2	Control	https://go-lab.bms.utwente.nl/ils-player/login/66f669e9e0194a73b80aa601
2	Intervention	https://go-lab.bms.utwente.nl/ils-player/login/66f669e9e0194a73b80aa601

Appendix C

All Test Items

Topic 1: Traits

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Wat is geen eigenschap?	De bloemkleur van azaleas	De hoogte van de erwten - plant	De oorvorm van de puppy	Het aantal chromosomen in mensen
Immediate post test 1	Welk van de volgende biologische kenmerken wordt beschouwd als een eigenschap?	De frequentie van het blaffen van een hond	De activiteits-tijd van een kikker in een vijver	De sociale gedrag - patronen van een dolfijn	De bloem bladkleur van een zonnebloem
Immediate post test 2	-	-	-	-	-
Delayed post test	Wat is geen eigenschap?	De kleur van een kersen-bloem	De vorm van erwten - zaden	De oorvorm van een kitten	De haarlengte van mensen

Topic 2: Gametes

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Als cellen in mensen wangen twee genetische factoren bevatten welke bepalen of je kuiltjes hebt. Hoeveel genetische actoren voor het bepalen van kuiltjes heeft Kim in haar eicellen, ervan uitgaande dat ze volledig gezond is?	0	2	23	1

Immediate post test 1	Een vrouwelijk konijn heeft een genotype van Rr, wat de vorm van haar oren bepaalt. Welke van de volgende uitspraken is correct?	De eicel van het konijn zal zowel genetische factoren van de moeder als van de vader bevatten.	De eicel van het konijn zal geen genetische factoren voor oorvorm bevatten.	Elke eicel van het konijn zal twee identieke genetische factoren bevatten.	De eicel van het konijn kan ofwel de R- of de r-genetische factor bevatten.
Immediate post test 2	-	-	-	-	
Delayed post test	Een mannetjes cavia heeft de genetische combinatie van Bb voor het bepalen van de vachtkleur. Welk van de volgende beweringen is juist?	De sperma van de cavia heeft geen genetische factoren voor het bepalen van de vachtkleur.	Deze cavia kan alleen spermacellen aanmaken met B als bepalende factor.	De cavia maakt alleen sperma aan met bepalende factor B en b.	De cavia kan sperma aanmaken met B of b als bepalende factor.

Topic 3: Dominant and Recessive

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Als een organisme een recessieve eigenschap heeft, hoe ziet de combinatie van genetische factoren er dan uit?	Beide factoren zijn dominant	Een factor is dominant, de ander is recessief.	Alle bovenstaande zijn mogelijk.	Beide factoren zijn recessief.

Immediate post test 1	Als een organisme een genotype van Bb heeft, waarbij B een dominant genetische factor is en b een recessief genetische factor, hoe zal de eigenschap van het organisme tot uiting komen?	Zowel de dominante als de recessieve eigenschap worden tot uiting gebracht	Alleen de recessieve eigenschap wordt tot uiting gebracht	Alle boven-staande opties zijn mogelijk	Alleen de dominante eigenschap wordt tot uiting gebracht
Immediate post test 2	-	-	-	-	-

Delayed post test	Welk van de volgende aannames komen overeen met Mendels principes van erfelijkheid als de allelen ongelijk zijn?	De aanwezige eigenschap wordt bepaald door het recessieve allel	De twee allelen mengen, dus geen van beide eigenschappen is aanwezig.	De persoon kan alle eigenschappen krijgen van beide allelen.	De aanwezige eigenschap wordt bepaald door het dominante allel.

Topic 4: Gametes, Dominance and Recessive

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Een lange erwt plant heeft een allelen combinatie van Tt. Welke van de volgende uitspraken is juist?	De erwt plant zal beide kenmerken tot uiting brengen welke gecontroleerd worden door de genetische factoren T en t.	De erwt plant is raszuiver.	De erwt plant zal alleen gameten produceren die de genetische factor T bevatten.	Na bestuiving kunnen de nakomelingen het kenmerk tot uiting brengen dat wordt gecontroleerd door de genetische factor t.

Immediate post test 1	Een muis met gele vacht heeft een genotype van Yy, waarbij Y een dominante genetische factor (geel) en y een recessieve genetische factor (wit). Welke van de volgende uitspraken is correct?	De muis zal alleen gameten produceren die Y bevatten	De muis zal kleur kenmerken vertonen die door zowel Y als y worden gecontroleerd	Alle nakomelingen van de muis zullen gele vacht hebben	De nakomelingen van de muis kunnen de eigenschap van witte vacht vertonen
Immediate post test 2	-	-	-	-	-
Delayed post test	Een zwarte cavia heeft allelen van Bb. Welke van de volgende uitspraken is juist?	De cavia zal alleen gameten aanmaken met genetische factor B.	De genetische factor b zal tot uiting komen in de cavia.	De genetische factor B zal zeker tot uiting komen bij de nakomelingen van de cavia.	De genetische factor b kan tot uiting komen bij de nakomeling - en van de cavia.

Topic 5: Scientific Methods

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Wetenschappers stellen de hypothese dat "het mogelijk gerelateerd is aan het broeikaseffect veroorzaakt door koolstofdioxide." In deze uitspraak, tot welke stap van het wetenschappelijke onderzoeksproces behoort het deel tussen	Observatie	Literatuur onderzoek	Gegevens inter-preteren	Vormen van een hypothese

anhalingstekens

?

Immediate post test 1	Wetenschappers hebben een afname van de bijenpopulatie in een bepaald gebied waargenomen en vermoeden dat dit mogelijk verband houdt met het gebruik van pesticiden. Om dit verder te onderzoeken, zijn ze van plan om omgevingen te creëren met verschillende pesticiden concentraties en veranderingen in de bijenpopulatie te	Hypotheses Vorming	Obser-vatie	Resultat Analyse	Experimen-teel ontwerp

observeren.

Welke fase van het wetenschappelijke onderzoeksproces vertegenwoordigt deze stap?

Immediate post test 2	-	-	-	-	-
Delayed post test	Virologen stellen de hypothese dat “het mogelijk te naken heeft met het ontwikkelen van de gewoonte bij kinderen om regelmatig hun handen te wassen.”. Deze uitspraak hoort in het wetenschappelijke onderzoeksproces tot:	Directe observatie	Ontwerp -en van een experiment	Analyser - en van gegevens	Vormen van een hypothese

Topic 6: Monohybrid crossings

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Een plant heeft twee eigenschappen: lange stelen en korte stelen. Welke van de volgende methoden kan de dominantie en recessiviteit van lange en korte stengels bepalen?	Observeren hoe vaak ange en korte stelen voorkomen in de natuur.	Observeren van het anpassingsvermogen van lange en korte stelen aan veranderingen in de omgeving.	De chromosomen van beide nemen en deze observeren onder een microscoop net speciale kleuring.	Het bestuiven van een raszuivere lange stelen plant met een raszuivere korte stelen plant.
Immediate post test 1	-	-	-	-	-

Immediate post test 2	Bij muizen is de kleur zwart (B) dominant over wit (b). Als een onderzoeker een zwarte muis heeft en wil bepalen of deze homozygoot dominant (BB) of heterozygoot (Bb) is, welke stap zou de onderzoeker moeten nemen	De zwarte muis kruisen met een andere zwarte muis	De zwarte muis alleen laten en wachten op spontane mutaties	Het genotype bepaalt door het uiterlijk van de muis te observeren.	De zwarte muis kruisen met een witte muis (bb)

Delayed post test	Het is bekend dat witte bloemen (a) in erwten recessief zijn ten opzichte van paarse bloemen (A). Als een onderzoeker wilt weten of paarse erwtenbloem genotype AA of Aa hebben, wat zou deze dan moeten doen?	De chromo-somen onder een micro - scoop bekijken	DNA extraheren met behulp van chemische middelen	De paars tint van de erwtenplant observeren.	De paarse erwten bloemen plant met een witte erwten bloemen plant kruisen en de eigenschappen van de nakomelingen observeren.

Topic 7: Purebred Crossing (First Generation)

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Wat is juist over raszuivere soorten?	Raszuivere soorten zijn organismen met dominante eigenschappen.	Na meerdere generaties van zelfbestuiving kunnen eigenschappen veranderen.	Dit verwijst naar heterozygoten.	Als raszuivere soorten gekruisd worden met een gelijke soorten met een recessieve eigenschap, dan zullen de nakomelingen de raszuivere eigenschappen uiten.
Immediate post test 1	-	-	-	-	-

Immediate post test 2	Wat gebeurde er met de erwt-nakomelingen in het experiment van Mendel toen de ronde erwt (RR) werd gekruist met de gerimpelde erwt (rr)?	Alle nakomel-ingen hadden gerimpelde zaden	De helft van de nakomelinge n had rond zaden en de andere helft had gerimpelde zaden.	De nakomeling en hadden zowel ronde als gerimpelde zaden in gelijke verhoudinge n.	Alle nakomeling-en hadden ronde zaden
Delayed post test	Wat gebeurde er met de erwt-nakomelingen in het erwten experiment van Mendel toen de gele erwt (YY) gekruisd werd met de groene erwt (yy)?	Alle erwten waren groen	De verhoud-ng gele erwt : Groene erwt was 1:1	De verhouding gele erwt : Groene erwt was 3:1	Alle erwten waren geel

Topic 8: Crossings

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Een erwtenplant met paarse bloemen heeft genotype Pp. Welk van de volgende beweringen is juist?	De erwtenplant maakt alleen voortplantingscellen aan met P als genetische factor.	Als de erwtenplant gekruisd is met een witte bloemen erwtenplant (pp), dan is de kans op nakomelingen met witte bloemen 1 op 4.	De erwten plant uit eigenschappen de gelinkt zijn aan genetische factoren P en p	Na zelfbestuiving kan de eigenschap met p als genetische factor waargenomen worden bij de nakomelingen.
Immediate post test 1	-	-	-	-	-

Immediate post test 2	Een muis met een zwarte vachtkleur heeft genotype Bb. Welke bewering is ONJUIST?	De muis vertoont de dominante eigenschap, een zwarte vacht.	De muis is heterozygoot wat de vachtkleur betreft.	De muis kan nakomeling en met witte vacht krijgen als deze wordt gekruist met een witte muis (bb)	De muis is homozygoot voor de dominante eigenschap

Delayed post test	Een erwtenplant met paarse bloemen heeft genotype Pp. Welke bewering is ONJUIST?	De erwten-plant zal voortplanting scellen aanmaken met P als genetische factor.	De bloem-kleur eigen-schap laat zien dat paars de dominante eigen-schap is.	De nakomel-ingen van zelf-bestuiv-ing kunnen eigen-schappen vertonen van de genetische factor P en p.	Als de erwtenplant gekruisd wordt met een erwtenplant met witte bloemen (pp), dan is de kans op nakomelingen met witte bloemen 1 op 4.

Topic 9: Second Generation

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Als lange stelen in ewten dominant (T) en korte stelen recessief (t) zijn en de eerste generatie nakomelingen van ouders TT x tt gekruist worden met een kortstelige erwtenplant (tt), wat is dan de kans op langstelige nakomelingen?	1	3 op 4	1 op 4	1 op 2

Immediate post test 1	-	-	-	-	-
Immediate post test 2	Een raszuivere paarse bloem is gekruist met een witte bloem. Alle bloemen in de eerste generatie waren paars. In de weede generatie waren er paarse en witte bloemen in een verhouding van 3:1. Welk van de volgende beweringen is ONJUIST?	Paarse bloemen zijn over witte bloemen.	De witte bloem is een recessieve eigen-schap.	In de tweede generatie zijn er meer paarse bloemen dan witte bloemen.	In de eerste generatie kwamen er zowel paarse als witte bloemen voor.

Delayed post test	Ethan heeft een genetsich experiment uitgevoerd met erwten. Hij heeft een raszuivere erwtenplant met paarse bloemen gekruisd met een erwtenplant met witte bloemen. De ereste generatie nakomelingen van deze kruising hadden paarse bloemen. Later heeft er zelfbestuiving plaatsgevonden bij de erwtenplant met paarse bloemen	Het hebben van paarse bloemen is de dominante eigenschap	Het hebben van witte bloemen wordt bepaald door de recessieve genetische factor.	De erwten-plant met paarse bloemen van de tweede generatie kunnen de recessieve genetische factor dragen.	De paarse bloemen van de eerste generatie dragen alleen een dominante genetische factor, en geen recessieve genetische factor.

van deze eerste generatie nakomelingen. Hieruit kwam de weede generatie nakomelingen, deze erwtenplanten hadden paarse en witte bloemen met respectivelijk een ratio van 3:1. Welk van de volgende beweringen is ONJUIST?

Topic 10: Argumentation

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer
Pre-test	Stel, het hebben van kuiltjes is een dominante eigenschap. Beide meneer en mevrouw Jansen hebben kuiltjes, wat is dan de kans dat hun kind kuiltjes heeft?	Het kind zal zeker kuiltjes hebben.	Het kind zal zeker geen kuiltjes hebben.	Het kind hoeft geen kuiltjes te hebben, maar de kans op het hebben van kuiltjes is lager.	Het kind hoeft geen kuiltjes te hebben, maar de kans op het niet hebben van kuiltjes is lager.
Immediate post test 1	-	-	-	-	-

Immediate post test 2	Bij mensen bepaalt het gen A een dubbele ooglidplooi en het gen a een enkele ooglidplooi. Emily heeft het genotype AA. Welke van de volgende beweringen is ONJUIST?	Emily heeft een dubbele ooglid-plooi	Emily heeft twee dominante genen voor ooglid-plooien	Emily's genotype is homozygoot dominant.	Emily kan het gen voor een enkele ooglidplooi doorgeven aan haar kinderen.

Delayed post test	Ervan uitgaand dat het hebben van een ooglidplooi in mensen bepaald wordt door een aantal genetische factoren. Hierin is de factor voor het hebben van een dubbele ooglidplooi A, en de factor voor het hebben van een enkele ooglidplooi a. Emily's genetische factoren voor ooglidplooien zijn AA. Welk van de volgende	Emily's kind zal zeker een dubbele ooglidplooi hebben.	Emily's kind kan de genetische factor combinatie hebben van AA als het gaat om ooglid-plooien.	Emily's kind kan de genetische factor combi-natie aa hebben als het gaat om ooglid-plooien.	Emily's kind kan een enkele ooglidplooi hebben.

beweringen is

ONJUIST?

Topic 11: Inferring Parentage

Questionnaire Type	Question	Answer	Answer	Answer	Correct Answer

›re-test	Het is bekend dat het kunnen oprollen van je tong een dominante eigenschap (A)s, en dat het niet ‹unnen oprollen van je tong een recessieve eigenschap (a)s. Sam kan haar tong niet oprollen, maar Sam’s vader en moeder kunnen lat wel. Wat zijn de genetische factor ɔombinaties van de ouders van Sam als het gaat ɔver het kunnen	AA x Aa	AA x aa	Aa x aa	Aa x Aa

oprollen van je

tong?

Immediate post test 1	-	-	-	-	-
Immediate post test 2	-	-	-	-	-
Delayed post test	Het is algemeen bekend dat het hebben van kuiltjes een dominante eigenschap (A) is, het niet hebben van kuiltjes is een recessieve eigenschap (a). Als Jan geen kuiltjes heeft en Emily wel, samen hebben zij een kind gekregen genaamd Michael die geen kuiltjes heeft. Welk van	Jan's genetische factoren voor het hebben van kuiltjes zijn aa	Michael's genetische factoren voor het hebben van kuiltjes is aa	De kans dat het volgende kind van Jan en Emily kuiltjes heeft is 1 op 2	Emily's genetische factoren voor het hebben van kuiltjes zijn AA

de volgende

beweringen is

ONJUIST?

Topic 12: Likert Scale Statements Collaboration

Statements		Helemaal nee oneens (1)	Sterk Oneens (2)	Oneens (3)	neutraal (4)	Eens (5)	Sterk Eens (6)	Helemaal mee eens (7)
Self explanation of Argument and Questions								
Active Questioning – self	Ik heb actief vragen gesteld aan mijn partner om een beter begrip van de leerstof te krijgen.	1	2	3	4	5	6	7
Asking for Explanation-self	Wanneer een uitleg van mijn partner niet duidelijk was, vroeg ik hen om een alternatieve uitleg te geven.	1	2	3	4	5	6	7
Reasoning check - self	Ik controleerde het redeneren van mijn groepsleden door hen te vragen hun ideeën uit te leggen.	1	2	3	4	5	6	7

Argument explanation and questioning of others								
Asking for Explanation - others	**Wanneer mijn uitleg niet duidelijk was, vroeg mijn partner mij om een alternatieve uitleg te geven.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Active Questioning - others	**Mijn partner stelde actief vragen om een beter begrip van de leerstof te krijgen.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Reasoning check - Others	**Mijn redenering werd gecontroleerd door mijn partner die mij vroeg mijn ideeën uit te leggen.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Statements	**Helemaal mee oneens**	**Sterk Oneens**	**Oneens**	**neutraal**	**Eens**	**Sterk Eens**	**Helemaal mee eens**

Cumulative Reasoning								
Self - argument construction	k bouwde voort op de bijdragen van mijn partner	1	2	3	4	5	6	7
Others- Argument Construction	Mijn partner bouwde voort op mijn bijdragen.	1	2	3	4	5	6	7
Self - proof	k zorgde ervoor dat ik de bijdragen die ik deed beargumenteerde	1	2	3	4	5	6	7
Others - proof	Mijn partner zorgde ervoor dat hij/zij bijdragen die hij/zij deed beargumenteerde	1	2	3	4	5	6	7
Group dynamics - Conclusions	Mijn partner en ik rokken conclusies die de discussie die we hadden weerspiegelden.	1	2	3	4	5	6	7

Topic 13: Likert Scale CLAIR Statements

Statements		Strongly Disagree	Disagree	somewhat disagree	ither agree or disagree	somewhat agree	Agree	Strongly Agree
CLAIR								
Using clair – self	Tijdens mijn chooljaren overweeg ik om Clair te gebruiken voor het eren van wetenschap.	1 1	2	3	4	5	6	7
Recommending clair	k ben bereid anderen aan te raden om Clair te gebruiken bij het eren van wetenschap.	1	2	3	4	5	6	7
illingness of future clair use	Over het algemeen is mijn bereidheid om Clair te gebruiken bij het leren van vetenschap vrij hoog.	1	2	3	4	5	6	7
Focus of clair	Het gebruik van Clair elpt me om me op de cursus te concentreren.	1	2	3	4	5	6	7

Appendix D

Example Personalised Cards for students

Naam: **Leerling nmr:** **Groep nummer:**	**Name:** **Student Number:** **Group number:**
QR Code to Learning environment	QR code to Learning Environment

Appendix E

R code

```
## loading packages##
library(readxl)
library(dplyr)
library(reshape2)
library (ggplot2)

## loading data set##
BSc_Naomi <- read_excel("BSc_Naomi.xlsx")
 View(BSc_Naomi)


## parametric assumptions ##

## normality

hist(BSc_Naomi$Total, main = "Histogram of Total scores", xlab = "Totalscore")

shapiro.test(BSc_Naomi$Total)


## homogeneity

leveneTest(Total ~ Group, data = BSc_Naomi)


##boxplots to further visualize data ##

## Total Scores over Time

ggplot(BSc_Naomi, aes(x = Class, y = Total, fill = Group)) +  geom_boxplot() + labs(title =
"Grouped Boxplot for Total Scores over Time", x = "Class", y = "Total Score") + theme_classic()

## Total Scores split by control  and intervention

ggplot(BSc_Naomi, aes(x = Class, y = Total, fill = Time)) + geom_boxplot() +
facet_wrap(~Group) + theme_minimal() +  labs(fill = "Time")
```

```
##Kruskal Wallis test

kruskal.test(Total ~ Group, data = BSc_Naomi)


##Friedman Test
##Transform to wide data set
Wide <- dcast(BSc_Naomi, Student ~ Time, value.var = "Total")
friedman.test(as.matrix(Wide[, -1]))

##Wilcoxon Signed Rank Test

## filter NA
Data  <- BSc_Naomi %>%group_by(Student) %>%filter(all(!is.na(Total)))

# Split data by condition focus on immediate and delayed
scores_IM <- Data$Total[Data$Time == "immediate"]
scores_DE <- Data$Total[Data$Time == "delayed"]

##Testing
Result <- wilcox.test(scores_IM, scores_DE, paired = TRUE)
print(Result)

## Likert scale items##

##loading packages##
library(readxl)
library (ggplot2)
library(psych)

## loading data sets##
##split into CLAIR questions and CL Questions

CLAIR_Likert <- read_excel("CLAIR_Likert.xlsx")
 View(CLAIR_Likert)

CL_Likert <- read_excel("CL_Likert.xlsx")
 View(CL_Likert)


## visualising frequency in histogram
##CL

 ggplot(CL_Likert, aes(x = Total, fill = Group)) +
```

```
 geom_histogram(binwidth = 5, color = "black", alpha = 0.7) +
 facet_wrap(~ Time) +
labs(title = "Histogram of Scores by Group per Questionnaire", x = "Total Scores", y =
"Frequency") +
theme_minimal()

##CLAIR
ggplot(CLAIR_Likert, aes(x = Total, fill = Group)) +
geom_histogram(binwidth = 5, color = "black", alpha = 0.7) +
facet_wrap(~ Time) +
labs(title = "Histogram of Scores Grouped per Questionnaire", x = "Total Scores", y =
"Frequency") +theme_minimal()

## checking normality
#CL
shapiro.test(CL_Likert$Total)

#CLAIR
shapiro.test(CLAIR_Likert$Total)


## Cronbach's Alpha

## CL

##removing columns
CLdata <- CL_Likert %>% select(-Time, -Group, - Class, -Student, -Total)
print(CLdata)


# Compute Cronbach's alpha
alpha_result_CL <- alpha(CLdata)

 # View results
 print(alpha_result_CL)


##CLAIR

CLAIRdata <- CLAIR_Likert %>% select(-Time, -Group, - Class, -Student, -Total)
 print(CLAIRdata)

# Compute Cronbach's alpha
 alpha_result_CLAIR <- alpha(CLAIRdata)
```

```
# View results
 print(alpha_result_CLAIR)
```

Appendix F

Chats From Twente Go-Lab Environments

Table 1

Responses to CLAIR Appearances during the Sessions

Chat	Original	Translated
1	**P2**: Ik denk dat 3 variëteiten is **CLAIR:** [P1] hoe zou je wat [P2] zegt anders kunnen verwoorden, of kun je een voorbeeld geven om te illustreren wat [P2] heeft gezegd? **P1**: Dat denk ik ook en 4 is die andere	**P2:** I think it's 3 varieties. **CLAIR:** [P1] How could you rephrase what [P2] is saying, or can you provide an example to illustrate what [P2] has said? **P1:** I think so too, and 4 is the other one.
2	**CLAIR:** Ik merk dat het een beetje stil is. Kunnen jullie met elkaar delen waar je aan werkt en wat je volgende stap zal zijn? **P1**: Waar ben jij? **P2**: Ik ben bezig met 1.1 **P1**: Ik ook	**CLAIR:** I notice it's a bit quiet. Could you share with each other what you're working on and what your next step will be? **P1:** Where are you? **P2:** I'm working on 1.1. **P1:** Me too.
3	**P1**: ik vind het wel echt lastig **P2**: Frrrr **P2:** ik doe ook niet zo veel oeps sorry	**P1:** I do find it really difficult. **P2:** Frrrr. **P2:** I'm not doing much either, oops, sorry.

	CLAIR: Ik raak de draad een beetje kwijt. Zullen we het gesprek weer terugbrengen naar de les? **P1:** Nee Claire	**CLAIR:** I'm getting a bit lost. Shall we bring the conversation back to the lesson? **P1:** No, Claire.
4	**CLAIR:** Ik raak de draad een beetje kwijt. Zullen we het gesprek weer terugbrengen naar de les? **P1:** Ja is goed!	**CLAIR:** I'm getting a bit lost. Shall we bring the conversation back to the lesson? **P1:** Yes, sounds good!
5	**P1:** clair wil je me helpen? **CLAIR:** Wat voor hulp heb je nodig? Probeer het eens bij je partner. **P1:** Oke. P2 weet jij vraag 1 van 1.3? **P2:** Ik ben daar nod niet sorryyy **P2:** Ik kijk ff	**P1:** Clair, will you help me? **CLAIR:** What kind of help do you need? Try asking your partner. **P1:** Okay. P2, do you know question 1 from 1.3? **P2:** I'm not there yet, sorryyy. **P2:** I'll check real quick.
6	**P1:** Clair weet jij wat je moet doen bij hypotese vormen? **CLAIR:** Mijn naam is Clair, ikbeneen computerprogramma en mijn doel is om discussies tussen mensen productiever te maken. **P1:** bro **P2:** Geef antwoord ofz	**P1:** Clair, do you know what to do when forming a hypothesis? **CLAIR:** My name is Clair, I am a computer program, and my goal is to make discussions between people more productive. **P1:** Bro **P2:** Just answer or something.