Fakultät für Biologie, Chemie und Geowissenschaften Lehrstuhl Didaktik der Biologie

Große Hoffnungen, große Hürden: Der Einfluss von bilingualem Lernen im außerschulischen Lernort Labor auf Leistung, Verständnis, Selbstkonzept, und Kreativität

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth

vorgelegt von

Tamara Heidi Roth
geboren in Hof an der Saale

Bayreuth, September 2021

Die vorliegende Arbeit wurde in der Zeit von Juli 2019 bis September 2021 am Lehrstuhl für Didaktik der Biologie der Universität Bayreuth unter Betreuung von Herrn Prof. Dr. Franz X. Bogner angefertigt.

Dissertation eingereicht am: 15.09.2021
Dissertation verteidigt am: 21.01.2022

Amtierender Dekan: Prof. Dr. Matthias Breuning

Prüfungsausschuss:

Prof. Dr. Franz X. Bogner	(Gutachter)
Prof. Dr. Volker Ulm	(Gutachter)
Prof. Dr. Klaus Ersfeld	(Prüfer)
Prof. Dr. Stefan Schuster	(Vorsitz)

Danksagung

Gerne möchte ich ein paar Worte des Dankes an all die Personen richten, die durch ihre vielfältige Unterstützung über die letzten Jahre wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Mein Dank gilt besonders meinem Doktorvater Herrn Prof. Dr. Franz X. Bogner. Durch Sie habe ich nicht nur die Möglichkeit bekommen, im Elitestudiengang MINT-Lehramt Plus studieren zu dürfen, Sie haben mir auch mit der Promotion die Tür zu einer akademischen Laufbahn geöffnet. Ich danke Ihnen sehr für die wertvollen Lektionen beim Verfassen akademischer Veröffentlichungen und der Konzeptionierung sowie Präsentation wissenschaftlicher Projekte. Sie haben mir stets die nötige Freiheit gelassen, um selbstständig zu arbeiten, waren aber jederzeit verfügbar, wenn es um Ratschläge oder Feedback ging. Danke auch für die Möglichkeit, akademische Lehrerfahrung zu sammeln und die großzügige Unterstützung bei der materialintensiven Durchführung des bilingualen Genetikmoduls. In meiner Zeit an Ihrem Lehrstuhl durfte ich mich persönlich und akademisch weiterentwickeln.

Herzlichen Dank auch an meinen zweifachen Co-Autor, Herrn Dr. Franz-Josef Scharfenberg. Er stand mir vom ersten Tag an mit Rat und Tat zur Seite, sei es bei der Organisation des Moduls, Koordination der Lehrveranstaltungen oder statistischen Fragegestellungen. Auch möchte ich meinen besonderen Dank an Frau Petra Feuerstein und Frau Sabine Hübner richten. Egal mit welchem Anliegen man an beide herantrat, sie wussten immer eine Lösung. Ungeachtet der Natur des Problems, fand man stets eine offene Tür und ein offenes Ohr.

Ebenfalls möchte ich mich bei meinen Kolleginnen bedanken. Ihre vielseitige Unterstützung, Begleitung und Beratung hat die Promotion für mich zu einem wertvollen Erlebnis gemacht. Ich möchte unsere Unterhaltungen und besonders die Zeit mit meiner Bürokollegin Michaela Maurer nicht missen! Großer Dank gilt auch meinen wissenschaftlichen Hilfskräften Lisa Schmutzler und Stephanie Benedikt sowie den Masterandinnen Natalie Ultsch und Eva-Maria Feller für deren tatkräftige Unterstützung während und nach dem Schülerlabor.

Zudem möchte ich allen Schülern und Lehrkräften danken, die durch Ihre Teilnahme am bilingualen Schülerlabor die vorliegende Arbeit ermöglicht haben. Großer Dank gilt auch meiner Familie, meinen Freunden und besonders meinem Partner, der durch emotionale Unterstützung, viel Geduld, und die beständige Versorgung mit Kohlenhydraten und Koffein zur Fertigstellung der Arbeit beigetragen hat.

Inhaltsverzeichnis

Abbildungsverzeichnis

Hinweis auf diversitätsgerechten Sprachgebrauch:

Für personenbezogene Bezeichnungen wird in vorliegender Arbeit das generische Maskulinum verwendet, um eine bessere Lesbarkeit gegenüber diversitätssensiblen Bezeichnungen zu ermöglichen, die sich nicht auf die geschlechterübergreifenden Pluralformen beziehen. Entsprechend soll hier ausdrücklich darauf hingewiesen werden, dass das generische Maskulinum als geschlechtsunabhängig verstanden wird und alle Geschlechteridentitäten miteinbezieht. In Teilarbeiten, die explizit auf Geschlechterunterschiede eingehen, werden hingegen geschlechterspezifische Bezeichnungen verwendet.

1 SUMMARY

Bilingual education brings high hopes for the enhancement of English language skills, as already tightly knit curricula leave no room for additional language lessons. Yet, fluent everyday communication and discipline-specific professional English language skills are expected in most working environments. Professional language competencies are, however, highly contextualized and can only be developed in authentic, discipline-specific learning scenarios. This also applies to content-related scientific competencies. Both require innovative teaching concepts and modules to enable competency-oriented learning. One approach is through open and self-guided learning that helps students explore scientific content. In addition to fostering professional language skills and scientific competencies, these learning approaches may also promote creativity and a self-concept that recognizes learning capabilities. Both variables are becoming increasingly important in international comparisons of scientific education effectiveness.

Hence, it is necessary to design modules that provide a learning environment that strengthens students' self-concepts, promotes creativity, enhances learning success and supports natural language acquisition. Bilingual science laboratories are an example of such an approach in science education. Few reports on these laboratory modules have been published, as most literature focuses on the combination of science and CLIL within long-term CLIL modules. With this in mind, the present work has developed a bilingual genetics module, based on a module originally delivered in German. This study involved three samples of upper stratification grammar schools: (1) a sample of 296 ninth graders from legacy data (Paper A), (2) a sample of 538 students from different grades (Paper B) and (3) a sample of 316 ninth graders (Paper C-F). In a first paper (Paper A), knowledge acquisition and model learning were assessed, both qualitatively and quantitatively, after participation in the non-CLIL module. This assessment aimed to determine the success of a module design with only one or two model evaluation phases. The results revealed that, for both knowledge acquisition and model learning, the use of at least two model evaluation phases significantly improved learning success and overall modeling performance.

The questionnaire piloted in Paper B aimed to quantitively explore the interplay between personality and creativity variables. Significant correlations were observed between the subscales *Conscientiousness* and *Flow* as well as between the subscales *Extraversion* and *Neuroticism* of the personality scale. In addition, when comparing outlying age-groups, gender differences emerged in the stratification of the creativity subscales *Act* and *Flow*. Younger male

students were significantly more likely to report *Act* as their preferred form of creativity while older female students favored *Flow*.

Paper C used quantitative methods to compare the learning success of students of the non-CLIL and CLIL modules. Findings indicated that students of both modules increased their learning success after participation. Non-CLIL learners, however, received higher scores than CLIL learners, indicating that the mental load involved in learning content and language simultaneously may have led to a trade-off that is difficult to redress.

Regarding the influence of CLIL on model-related knowledge, similar results could be obtained using a qualitative and quantitative approach (Paper D). Non-CLIL learners also succeeded in the second model evaluation and produced significantly better handcrafted models. However, CLIL learners excelled in areas aimed at developing in-depth understanding. For example, the model understanding questionnaire revealed that, after participating in the module, CLIL learners abandoned the common misconception that models are exact replicas of observed phenomena. These learners also realized that models are only representations of current scientific research and may change with new scientific evidence. This was confirmed by the students' responses to the open questions on the model sketching sheet. CLIL learners were deemed to have performed better in the first model evaluation phase as their sketches were more accurate, albeit less detailed.

Paper E used quantitative and qualitative methods to investigate the impact of CLIL on English language acquisition. Both the Cloze test and the categorization of students' responses to the open questions in the manual indicated that, through participation, students were able to improve their English language skills. Gender differences commonly reported in literature could not be confirmed. Instead, a rapid decline of the language skills measured by the Cloze test was observed in the retention test. A possible explanation for this phenomenon could be the switch to online distance learning due to COVID-19 restrictions.

Finally, Paper F used a quantitative approach to investigate the influence that an open learning environment based on hands-on-minds-on activities and CLIL had on the development of creativity and school-related self-concept. This study revealed significant correlations between creativity and self-concept, indicating that both require similar learning environments to thrive. Moreover, the reciprocal relationship between academic achievement and self-concept, as it is reported in the wider literature, could be confirmed. Gender differences only existed for the subscale *Social* in the self-concept questionnaire. Female students perceived their performance to be worse when compared with the performances of other students, while male students did not.

2 ZUSAMMENFASSUNG

Bilingualer Unterricht ist mit großen Hoffnungen für die Verbesserung englischer Sprachkompetenz verbunden, ohne die Lehrpläne durch zusätzlichen Fremdsprachenunterricht weiter zu überladen. Denn fließendes Englisch ist in beinahe allen akademischen und nicht akademischen Berufen Grundvoraussetzung. Neben müheloser Kommunikation in englischer Alltagssprache wird insbesondere disziplinbezogene Fachsprachlichkeit erwartet. Diese Fachsprachlichkeit ist jedoch stark kontextbezogen und kann nur innerhalb authentischer, disziplinspezifischer Lernsituationen gefördert werden. Zusätzlich zur Sprachkompetenz wird auch inhaltlich verstärkt auf Kompetenzen gesetzt. Entsprechende Anpassungen des Lehrplans und damit einhergehende innovative Unterrichtskonzepte sollen kompetenzorientiertes Lernen ermöglichen. Diese Unterrichtskonzepte setzen besonders auf Hands-on-Minds-on Lernen durch offenen und selbstgesteuerten Erkenntnisgewinn in der Auseinandersetzung mit wissenschaftlichen Inhalten. Dies wiederum fördert neben Fachsprachlichkeit und wissenschaftlicher Kompetenz auch kreative Prozesse und das schulbezogene Selbstkonzept. Letztere sind inzwischen Indikatoren für die Effektivität naturwissenschaftlicher Bildung im internationalen Vergleich geworden.

Entsprechend müssen Module entworfen werden, die eine Selbstkonzept stärkende, Kreativität fördernde, Lernerfolg steigernde und natürlichen Spracherwerb verbessernde Lernumgebung unterstützen. Eine Möglichkeit für den naturwissenschaftlichen Unterricht stellen eintägige Experimentierlabore mit bilingualem Lernangebot dar. Bestehende Referenzen zu diesen Modulen sind jedoch rar, da sich der Großteil der Literatur über Kombinationen von Naturwissenschaften mit *Content and Language Integrated Learning* (CLIL) ausschließlich auf CLIL-Module in Langzeitform fokussiert. Vor diesem Hintergrund wurde in der vorliegenden Arbeit ein ursprünglich muttersprachlich deutsches Modul zu einem bilingual englischen Bio-/Gentechnik Modul weiterentwickelt. Für die Arbeit wurden insgesamt drei Stichproben verwendet: (1) In Teilarbeit A 296 Schüler der 9. Jahrgangsstufe, (2) in Teilarbeit B 538 Schülern verschiedener Jahrgangsstufen und (3) in Teilarbeiten C-F 316 Schüler der 9. Jahrgangsstufe bayerischer Gymnasien.

In einer ersten Arbeit (Teilarbeit A) wurde in einem qualitativen und quantitativen Verfahren der Wissenserwerb und das Modelllernen im muttersprachlich deutschen Vorgängermodul anhand bereits bestehender Daten erhoben. Die Ergebnisse dienten als Grundlage zur Auswahl des Moduldesigns mit nur einer oder zwei Modellevaluationsphasen. Sowohl für Wissen als auch Modelllernen zeigte sich, dass die Verwendung von mindestens zwei Modellevaluationsphasen den Lernerfolg signifikant verbessert.

In Teilarbeit B wurde ein Fragebogen pilotiert, der auf die Ergründung des Zusammenspiels zwischen Persönlichkeitsvariablen und Kreativität in einem quantitativen Verfahren zielte. Dabei konnten wesentliche Zusammenhänge zwischen den Subskalen *Gewissenhaftigkeit* und *Flow*-Erleben beobachtet werden, sowie eine Korrelation zwischen den Subskalen *Extraversion* und *Neurotizismus* innerhalb der Persönlichkeitsskala. Zudem wurden altersspezifische Geschlechterunterschiede in den Kreativitätsvariablen *Act* und *Flow*-Erleben beim Vergleich von Extremgruppen deutlich. Jüngere männliche Schüler gaben signifikant häufiger *Act* als bevorzugte Form der Kreativität an, und ältere weibliche Schüler *Flow*-Erleben.

Teilarbeit C war die erste Arbeit im Rahmen des bilingualen Moduls. Diese setzte sich insbesondere mit dem Vergleich des Lernerfolgs zwischen Schülergruppen des muttersprachlich deutschen und des bilingual englischen Moduls in einem quantitativen Verfahren auseinander. Die Schüler beider Module erzielten Lernerfolge nach ihrer Teilnahme. Die Schüler des muttersprachlich deutschen Moduls schnitten dabei jedoch besser ab. Die Doppelbelastung durch inhaltliches und sprachliches Lernen führte wahrscheinlich zu einem Zielkonflikt bei der bilingual englisch unterrichteten Schülergruppe.

In Teilarbeit D konnten ähnliche Ergebnisse bezüglich des Einflusses von CLIL auf Modellwissen in einem qualitativen und quantitativen Verfahren erzielt werden. Auch bei der zweiten Modellevaluation und der Qualität der selbstgebauten Modelle schnitten Schüler des muttersprachlich deutschen Moduls besser ab. Bei Fragen und Aufgaben, die Tiefenverständnis erforderten, bewiesen sich jedoch die Schüler des bilingual englischen Moduls. Diese konnten nach Teilnahme ihre falsche Vorstellung von Modellen als exakte Kopien des beobachteten Phänomens ablegen. Zudem erkannten die bilingual englischen Schüler, dass Modelle reine Repräsentationen des wissenschaftlichen Forschungsstands sind und sich mit neuen wissenschaftlichen Erkenntnissen verändern. Dies wurde ebenfalls durch die Schülerantworten auf die offenen Fragen des Modellskizzenblattes und ihre Leistung bei der ersten Modellevaluation bestätigt, da ihre Skizzen exakter, wenn auch weniger detailliert waren.

In Teilarbeit E wurde der Einfluss von CLIL auf englischen Spracherwerb in einem qualitativen und quantitativen Verfahren untersucht. Sowohl der Cloze-Test als auch die Kategorisierung der Schülerantworten auf die offenen Fragen im Kursheft zeigten, dass die Schüler durch Teilnahme am Modul ihre englischen Sprachkompetenzen verbesserten. In der Literatur allgemein postulierte Geschlechterunterschiede wurden nicht bestätigt. Dafür konnte ein rapides Absinken der zunächst im Cloze-Test gemessenen Sprachkompetenzen im Behaltenstest festgestellt werden. Eine mögliche Erklärung für dieses Phänomen ist der Wechsel auf Distanzunterricht im Zuge der COVID-19-Beschränkungen.

Abschließend wurde in Teilarbeit F der Einfluss von Hands-on-Minds-on-Aktivitäten und CLIL auf Kreativität und Selbstkonzept mit einem quantitativen Ansatz untersucht. Dabei konnte gezeigt werden, dass die Skalen von Kreativität und Selbstkonzept miteinander korrelieren und somit in einer ähnlichen Lernumgebung gefördert werden können. Außerdem wurde die in der Literatur geschilderte wechselseitige Beziehung zwischen Leistung und Selbstkonzept bestätigt. Geschlechterunterschiede zeigten sich nur für die Subskala *Social* des Selbstkonzepts. Hier nahmen sich weibliche Schüler im leistungsmäßigen Vergleich mit anderen Schülern leistungsschwächer wahr als sie tatsächlich waren und unterschieden sich damit von ihren männlichen Kollegen.

3 SYNOPSIS

3.1 Einleitung

Die Einführung von bilingualem Unterricht ist mit großen Hoffnungen verbunden, da die bereits dicht gepackten Lehrpläne kaum Spielraum für zusätzlichen Englischunterricht erlauben. Dabei ist gerade Englisch in vielen akademischen und beruflichen Kontexten auf einem Niveau erforderlich, das traditioneller Fremdsprachenunterricht nicht gewährleisten kann (Oktaviani & Fauzan 2017, Rodenhauser & Preisfeld 2014, Meskill & Oliveira 2019). Insbesondere fehlt häufig die Fachsprachlichkeit, die nur durch Kontextualisierung in authentischen und fachspezifischen Lernsituationen vermittelt werden kann (Gonzalez-Howard & McNeill 2016, Fazio & Gallagher 2019). Für mehr Authentizität und Fachlichkeit wurde *Content and Language Integrated Learning* (CLIL) in vielen verschiedenen Fächern, darunter auch den Naturwissenschaften, eingeführt (European Commission 2004). In den regulären Unterricht des Sachfachs integriert, soll im Rahmen von CLIL die erforderliche Fachsprachlichkeit über inhaltliches Wissen aufgebaut werden (Bonnet 2012, Henriksen et al. 2018). Konkrete Anleitungen oder eine gesonderte Ausbildung der Fachlehrer ist hierfür jedoch oft nicht vorgesehen (Fehling 2008). Entsprechend erfreut sich bilingualer Unterricht zwar großer Beliebtheit, kommt aber nur in ausgewählten Schulen und Zweigen zum Einsatz (Wildhage & Otten 2009).

Neben Fachsprachlichkeit in der Fremdsprache Englisch, müssen Schüler im Laufe ihrer Schulzeit zudem lernen, wissenschaftliche Texte zu lesen und deren Inhalte adressatengerecht zu kommunizieren sowie kritisch zu reflektieren und zu diskutieren (Glynn & Muth 1994). So wird neben CLIL auch Kreativität zunehmend in den naturwissenschaftlichen Unterricht aufgenommen, da reines Faktenwissen nicht ausreicht, um den Aufbau von Kompetenzen und Innovation voranzutreiben (Brynjolfsson & McAfee 2014). Im Gegenteil wird deutlich, dass das aktuell wenig flexible und stark standardisierte System nicht genügt, um die Gesellschaft für die Zukunft zu rüsten (Robinson 2011). In Schulen wird deshalb immer mehr auf Lernumgebungen gesetzt, die einen offenen und selbstgesteuerten Erkenntnisgewinn durch die Auseinandersetzung mit wissenschaftlichen Inhalten ermöglichen (Richards & Cotterall 2016, Ke et al. 2020). CLIL-Module bieten eine solche Lernumgebung und ermöglichen zusätzlich kreative Prozesse durch die Kombination von Hands-on-Minds-on-Aktivitäten (Fleith et al. 2002).

Diese Aktivitäten in CLIL-Lernumgebungen können, auf dem richtigen Schwierigkeitsgrad angewandt, auch das Selbstkonzept der Schüler stärken (Lian et al. 2018, Buse et al. 2018). Das Selbstkonzept gilt als eng verbunden mit Kreativität (z. B. Sisk 1972, Bournelli et al. 2009) und kann wesentlich zu individuellem akademischem Lernerfolg beitragen (z. B. Jansen et al.

2015, Hoferichter et al. 2018). Diese wechselseitig verstärkende Beziehung zwischen Lernerfolg und Selbstkonzept stellt einen weiteren essentiellen Aspekt schulischen Lernens dar (Marsh et al. 2015) und gilt als Schlüssel zur Entwicklung von Scientific Literacy (Wang et al. 2008).

Im Rahmen der vorliegenden Arbeit wurde daher ein ursprünglich muttersprachlich deutsches Modul (Langheinrich & Bogner 2015, Mierdel & Bogner 2020), das Selbstkonzept, Kreativität und Lernerfolg fördert, um CLIL erweitert und evaluiert. Das darin ermöglichte informelle Lernen schuf die Grundlage für Hands-on-Minds-on-Aktivitäten, die sowohl unbekannte wissenschaftliche Inhalte und Arbeitsweisen vermittelten als auch die kontextbezogene Verwendung wissenschaftlicher Fachtermini begünstigten (Rodenhauser & Preisfeld 2014, Buse et al. 2018). Zudem ließen sie die Entwicklung von Scientific Literacy zu (Glynn & Muth 1994, Gonzalez-Howard & McNeill 2016, Fazio & Gallagher 2019). Eventueller geistiger Überforderung der Schüler wurde durch Scaffolding vorgebeut (Meyerhöffer & Dreesmann 2019), sodass eine sichere Lernatmosphäre das Selbstkonzept bestärken und die Entfaltung von Kreativität ermöglichen konnte (Lian et al. 2018, Sisk 1972, Bournelli et al. 2009). Unter Scaffolding wird allgemein eine Art Lerngerüst verstanden, das den Schülern hilft, neue Inhalte auf verschiedenen Ebenen zu durchdringen und neue Fähigkeiten aufzubauen, die zur Lösung der wissenschaftlichen Fragestellung beitragen (Maybin et al. 1992).

Die einzelnen Teilarbeiten überprüfen kognitive und affektive Komponenten: Wissenserwerb und Modelllernen mit einer oder zwei Modellevaluationen, das Zusammenspiel zwischen Persönlichkeitsvariablen und Kreativität, den Zielkonflikt zwischen Wissenserwerb und CLIL, den Einfluss von CLIL auf Modellverständnis, Modellwissen und Modellbauen im Vergleich zur muttersprachlich deutschen Schülergruppe, genderneutrales Sprachenlernen und den Aufbau sprachlichen Wissens durch CLIL sowie CLIL für Kreativität und Selbstkonzept fördernde Lernumgebungen. Die Synopsis der Teilarbeiten gliedert sich in einen theoretischen und einen empirischen Teil. Zunächst erfolgt eine Auseinandersetzung mit dem aktuellen Forschungsstand zu inhaltlichem und sprachlichem Lernen, Scientific Literacy, Modellverständnis und Modelllernen, schulbezogenem Selbstkonzept, Kreativitäts- und Individualitätsförderung durch die Lernumgebung und der Implementierung bilingualer Schülerlabore (Kapitel 3.2). Anschließend werden, ausgehend von den wissenschaftlichen Fragestellungen (Kapitel 3.3) und der verwendeten Methodik (3.4), die Ergebnisse der Teilarbeiten dargestellt und vor dem Hintergrund didaktisch-psychologischer Literatur diskutiert (Kapitel 3.5).

3.2 Theoretischer Hintergrund

3.2.1 Lernprozesse

Lernen wird oft als eine auf Erfahrungen beruhende Änderung des Verhaltens beschrieben (De Houwer et al. 2011). Die Annahme, dass Erfahrungen Grundlage verschiedener Lernprozesse sind, reflektiert konstruktivistische Ansätze, in denen der Wissenserwerb als rein situativ wahrgenommen wird (Korthagen & Lagerwerf 1995). Derartige Betrachtungen erfordern die Verwendung entsprechender Lehr-Lernmethoden, um durch den Lehrplan festgelegte Lerninhalte im naturwissenschaftlichen Unterricht erfahrungsbasiert zu vermitteln. Lernen ist jedoch mehr als nur das Ergebnis von Erfahrungen und nicht jede Lehr- und Lernmethode führt bei allen Schülern zu vergleichbaren Erfahrungen (Hodson 2014). Ganz im Gegenteil sind Erfahrungen sehr individuell und können durch Faktoren wie Vorwissen und Alltagswissen oder unterschiedliche Lernprofile beeinflusst werden. Folglich führt der Einsatz einer Lehr- oder Lernmethode in für gewöhnlich sehr heterogenen Schülergruppen selten zu denselben Lernergebnissen. Dies gilt insbesondere für viele Methoden, die allgemeinhin als effektiv gelten, wie beispielweise kooperatives Lernen und Projektarbeit, jedoch ebenfalls nicht alle Lerntypen berücksichtigen (Dunlosky et al. 2013). Viele Lehrkräfte sind sich dessen inzwischen bewusst und setzen auf differenziertere Methoden. Dennoch schleicht sich gelegentlich noch immer die Metapher vom Lernenden als *tabula rasa* in den modernen Unterricht ein (Stahl et al. 2014).

Um derartigen Rückfällen in antiquierte Unterrichtsmuster vorzubeugen, sollten gerade bei der Auswahl der richtigen Lernmethode Schüler allgemein stärker in die Pflicht genommen werden. Im Rahmen solcher Unterrichtskonzepte treten Lehrkräfte als Lernbegleiter oder Mentoren in den Hintergrund und bieten Lernenden den Freiraum ihre Lernmethoden und, bis zu einem gewissen Grad, auch ihre Lerninhalte selbst zu bestimmen (Donaldson & Allen-Handy 2020). Als effektiv gelten dabei Lernkonzepte, die die Selbstreflexion anregen (Dunlosky et al. 2013) sowie generell offene Lernumgebungen, in denen Lernmethoden nicht vorgegeben, sondern selbst erprobt werden (Stahl et al. 2014). Dieser Ansatz fördert insbesondere Deep-Learning, da Schüler selbst Prozesse des Erkenntnisgewinns anstoßen und das Ergebnis materialgestützt oder in der Gruppe reflektieren. Gerade das Reflektieren über den eigenen Gedankengang sowie die gegebenenfalls ergänzende Diskussion mit anderen, geben Schülern Kontrolle über ihren Lernfortschritt und fördern die vertiefte Auseinandersetzung mit Inhalten (Lee et al. 2015).

Moderne Auffassungen von Lernen sehen also eine gewisse Handlungsmacht und Selbstbestimmung der Schüler vor. Dies bedeutet nicht, dass die Lernenden vollkommen frei über die Lerninhalte entscheiden können. Die jeweiligen Lehrpläne (ISB 2021) der Bundesländer geben klare inhaltliche und kompetenzorientiere Lernziele vor[1]. So ist es den Schülern lediglich überlassen, die Methoden und über die Mindestanforderung hinaus wählbare Lerninhalte frei zu definieren (Dunlosky et al. 2013, Donaldson & Allen-Handy 2020). Dies ermöglicht eine aktive Mitgestaltung des Unterrichtsgeschehens und der Lernumgebung in gegenseitiger Absprache mit Klassenkameraden und Lehrkräften (Stahl et al. 2014). Auf diese Weise sollen die Schüler Kontrolle über ihren Lernprozess gewinnen (Lee et al. 2015), was langfristig intrinsische Motivation fördert und zu einem positiven Lernergebnis führen kann (Renninger et al. 2018).

3.2.2 Content and Language Integrated Learning

Heute ist es unerlässlich, souverän in einer international verständlichen Sprache in Wort und Schrift kommunizieren und Inhalte aus in dieser Sprache geschriebenen Texten extrahieren zu können. Gerade im Bereich der Wissenschaft ist eine gemeinsame Sprache Grundvoraussetzung, um Ideen auszutauschen, diese kritisch zu diskutieren, das Verständnis zu vertiefen und eventuelle Probleme zu lösen (Krajcik & Sutherland 2010). Da Englisch die *lingua franca* der Wissenschaft ist und auch in anderen Bereichen eine zentrale Rolle einnimmt, sind schriftliche und mündliche Sprachkompetenzen im Englischen elementar, um auf dem internationalen wissenschaftlichen Parkett bestehen zu können (Yore & Treagust 2006). Für Schulen bedeutet dies, dass sie sich um eine intensivere Förderung englischer Sprachkompetenzen unter Berücksichtigung der Scientific Literacy des jeweiligen naturwissenschaftlichen Sachfachs bemühen sollten (Meyerhöffer & Dreesmann 2019). Die Europäische Kommission unterstützt derartige Bestrebungen und ermutigt Schulen, das Erlernen der englischen Sprache außerhalb des regulären Englischunterrichts durch sogenanntes *Content and Language Integrated Learning* (CLIL) in den Lehrplan der Sachfächer zu integrieren (European Commission 2012). Effektiv eingesetzt, verbindet CLIL Sprachenlernen mit inhaltlichem Lernen, sodass beides gleichzeitig erlernt werden kann, ohne dass das eine das andere voraussetzt (European Commission 2004, Stoddard et al. 2002).

CLIL beschreibt damit eine Lernsituation, in der die Sprache Lerninhalt und zugleich Werkzeug zum Erlernen anderer sachfachlicher Inhalte ist. Diese Verflechtung von Sprache und Inhalt (Marsh 2002) widerspricht traditionellen Unterrichtskonzepten, wonach fundierte

[1] Hinweis zur Einführung eines neuen Lehrplans an bayrischen Gymnasien: Aufgrund der Rückkehr zum neunjährigen Gymnasium, wird seit Beginn des Schuljahres 2017/18 sukzessive der neue LehrplanPLUS eingeführt (ISB, 2017).

Sprachkenntnisse eine wesentliche Voraussetzung von effektivem inhaltlichem Lernen sind (Cummins 1981, Stoddart et al. 2002). Entsprechend wird in vielen CLIL-Modulen die Sprache noch immer als reiner Vermittler von Wissen betrachtet, obgleich es zunehmend Forderungen gibt, Sprache und Inhalt gleichberechtigt in die Unterrichtsplanung aufzunehmen (Rodenhauser & Preisfeld 2014). Erschwerend für CLIL-Ansätze kommt hinzu, dass nur wenige Lehrkräfte überhaupt die komplexe Verbindung zwischen Sprache und Inhalt wahrnehmen (Prinsloo & Harvey 2020, Meskill & Oliveira 2019). Mit dieser eindimensionalen Betrachtungsweise verkennen sie, dass es zwar Inhalte sind, die Sprechakte induzieren, aber nur die Sprache diesen Inhalten Bedeutung verleihen vermag (Stoddart et al. 2002). Durch die wechselseitige Beziehung zwischen Sprache und Inhalt ist es entsprechend unabdingbar, dass authentische Sprachanlässe geschaffen werden müssen, die als Grundlage natürlicher Sprachentwicklung dienen (Stoddart et al. 2002, Krashen 1985, Meskill & Oliveira 2019).

Authentische Sprechanlässe in Naturwissenschaften bedingen jedoch auch das Erlernen fachspezifischen Vokabulars und den Aufbau von Scientific Literacy (Gonzalez-Howard & McNeill 2016). Bei der Erkundung wissenschaftlicher Phänomene in authentischen Lernumgebungen findet ein Teil der Aufarbeitung wissenschaftlicher Inhalte durch Kommunikation, Diskussion und gemeinsame Reflexion statt (Nation & Newton 2008, Bradbury 2014, Fazio & Gallagher 2019). Diese vertiefte Auseinandersetzung mit dem naturwissenschaftlichen Phänomen auf zahlreichen Ebenen gilt als Grundlage für die Entwicklung von Scientific Literacy (Walker & Sampson 2013, Gonzalez-Howard & McNeill 2016). Beispiele aus der Praxis (Lam et al. 2012, Erkol et al. 2010) bestätigen diese Vermutung. Sie zeigen, dass authentische Lernumgebungen einen positiven Effekt auf den Aufbau von Scientific Literacy und inhaltliches Lernen haben, wenn sie den natürlichen wissenschaftlichen Diskurs nachahmen (Quarderer & McDermott 2020).

Gleichzeitig ist CLIL geistig sehr anspruchsvoll, da inhaltliches Lernen und Sprachenlernen simultan erfolgen. Um trotz dieser doppelten Belastung Kompetenzen im Sachfach und der Sprache aufzubauen, bedarf es effektiver Scaffolding-Materialien, die die kognitive Belastung verringern (Coyle 2007, Grandinetti et al. 2013). Gottlieb (2016) beschreibt vier verschiedene Formen des Scaffolding: linguistisch, grafisch, sensorisch und interaktiv. In naturwissenschaftlichen Fächern wie der Biologie kommt häufig grafisches oder visuelles Scaffolding in der Form von Diagrammen, Graphen oder Tabellen zum Einsatz (Buxton et al. 2019). Für naturwissenschaftlichen CLIL-Unterricht ist jedoch auch das linguistische Scaffolding unabdingbar, um durch die Definition von Schlüsselbegriffen oder die Bereitstellung sprachlicher

Brücken den Redefluss aufrecht zu erhalten und Scientific Literacy zu fördern (Finn 2012, Unsal et al. 2018, France 2019).

In Deutschland wird CLIL entweder in bilingualen Zügen oder vereinzelten CLIL-Modulen realisiert. Während es sich bei bilingualen Zügen um regelmäßigen CLIL-Unterricht in einem Sachfach handelt, sind CLIL-Module in ihrer Dauer oft sehr unterschiedlich (Diehr 2012, Krechel 2013). Unabhängig von Zeit und Häufigkeit wird aber davon ausgegangen, dass sich jede sinnvolle Beschäftigung mit Sprache und Inhalt positiv auf die Scientific Literacy auswirkt (Bonnet 2004, Rodenhauser & Preisfeld 2014, Buse & Preisfeld 2018).

3.2.3 Sprachbewusstsein und Scientific Literacy im naturwissenschaftlichen Unterricht

Sprachbewusstsein beschreibt das natürliche Verständnis der Wirkweise von Sprache und ihrer Funktion in Bezug auf menschliche Interaktion unabhängig von der bewussten Auseinandersetzung mit linguistischen Theorien (Little 1997). Es ist dabei sowohl für das Lernen als auch Lehren einer Sprache unverzichtbar (Lo 2017, Lindahl 2020). Entsprechend bauen Schüler mit grundlegendem Sprachbewusstsein sehr viel schneller Sprachkompetenzen auf und können selbstbestimmt ihren Sprachlernprozess steuern. Lehrende hingegen profitieren von der instrumentellen Funktion des Sprachbewusstseins mit seinen drei Dimensionen (Little 1997, Lindahl 2020): (1) Bewusstsein dafür, dass Sprache Inhalten Bedeutung verleiht, (2) Bewusstsein dafür, dass das Erlernen einer Sprache Komplexitäten mit sich bringt, die durch entsprechendes Scaffolding adressiert werden können und (3) Bewusstsein für Theorien und pädagogische Herangehensweisen, die dem Sprachenlernen zugrunde liegen (Lo 2017).

Neben dem Sprachbewusstsein ist besonders die Förderung von Scientific Literacy ein wesentlicher Aspekt von CLIL. Scientific Literacy kann dabei auf zwei Arten bedeutsam sein. Einerseits besitzt Scientific Literacy wesentliche Relevanz für den akademischen Lernerfolg. Konkret ist Scientific Literacy unabdingbar für das vertiefte Verständnis von wissenschaftlichen Inhalten sowie daraus abgeleiteten wissenschaftlichen Arbeitsweisen. Andererseits besitzt Scientific Literacy eine wichtige Bedeutung bei der Entwicklung von mündigen Bürgern, die verantwortungsvoll und informiert handeln können (Ke et al. 2020). Krajcik & Sutherland (2010) ergänzen diese zwei Bedeutungsaspekte von Scientific Literacy mit fünf grundlegenden Prinzipien, die die Förderung von Scientific Literacy im naturwissenschaftlichen Unterricht ermöglichen: (1) Anknüpfung an das Vorwissen der Schüler und ihre Erfahrungswelt, (2) Schaffen von Alltagsrelevanz, (3) Darstellungsvarianz, (4) offene Lernumgebungen, in der Schüler ihre Ideen austesten können und (5) Förderung von wissenschaftlichen Diskursaktivitäten und kritischer Reflexion.

Entsprechend ist es nicht ausreichend, wie oft fälschlicherweise angenommen, dass Scientific Literacy ausschließlich durch Lesen von und Schreiben über wissenschaftliche Inhalte aufgebaut wird. Scientific Literacy erfordert Lernumgebungen, die eine Betrachtung der komplexen Beziehung zwischen Sprache, Inhalt und gelebter Praxis erlauben. Dies ermöglicht das Reifen und Kommunizieren von Erkenntnissen durch die Anwendung kritischer Reflexion (Moje et al. 2001, Fazio & Gallagher 2019). Demnach ist die Kombination von disziplinspezifischen wissenschaftlichen Inhalten mit Sprachenlernen, wie sie in CLIL erfolgt, zur Förderung von Scientific Literacy unerlässlich (Bradbury 2014, Fazio & Gallagher 2018). Zum einen müssen sich Schüler intensiv mit Fachsprache und wissenschaftlichen Inhalten auseinandersetzen und zum anderen werden die gewonnen Erkenntnisse bei der Diskussion und Reflexion oft in verschiedene Darstellungsformen überführt (Glynn & Muth 1994, Fazio & Gallagher 2019, Quarderer & McDermott 2020).

3.2.4 Modellverständnis und Modellernen im naturwissenschaftlichen Unterricht

Da Modelllernen die Übertragung von Wissen in verschiedene Darstellungsformen erfordert, hat es sich inzwischen zu einer beliebten Methode zum Aufbau wissenschaftlicher Kompetenz entwickelt (Akerson 2009, Ke et al. 2020). Modelle werden meist zur Visualisierung und Erklärung schwer beobachtbarer oder komplexer wissenschaftlicher Phänomene herangezogen (Krajcik & Merritt 2012, Windschitl & Thompson 2006). Generell wird dabei in materielle und ideelle Modelle unterschieden. Materielle Modelle sind Anschauungsmodelle, die Strukturen entweder möglichst real wiedergeben, sichtbare und unsichtbare Dinge verbildlichen, oder dabei helfen, Funktionen zu analysieren. Bei ideellen Modellen handelt es sich hingegen um Denkmodelle, wodurch sehr komplexe Phänomene verständlich und greifbar gemacht werden sollen (Hager & Hörz 1977, Parthey 1983). Um ein Modell überhaupt erstellen zu können, muss zunächst das Phänomen grundlegend verstanden werden. Anschließend können mögliche Zusammenhänge zwischen den identifizierten Strukturen entweder der Beobachtung entnommen oder durch das Aufstellen und Testen von Hypothesen erklärt werden (Akerson 2009). Sollte das vorgeschlagene Modell der nachfolgend kritischen Diskussion durch Kollegen nicht standhalten oder neue wissenschaftliche Erkenntnisse das Modell in der aktuellen Form unbrauchbar machen, kann das Modell entsprechend angepasst und verbessert werden (Passmore et al. 2009, Carpenter et al. 2018).

Gerade der Aspekt, dass Modelle stets nur den aktuellen Forschungsstands repräsentieren und jederzeit mit dem Aufkommen neuer wissenschaftlicher Erkenntnisse verändert werden können, ist ein wesentlicher Bestandteil des SUMS-Fragebogens (Students' Understanding of

Models in Science). Dieser wurde entwickelt, um das Modellverständnis von Schülern abzuprüfen, insbesondere die latente Variable der „Changing Nature of Models" (CNM; Treagust et al. 2002). In Anbetracht der weit verbreiteten Fehlvorstellung, dass Modelle exakte Nachbildungen eines realen Phänomens darstellen („Models as Exact Replicas“; EM) und nicht die Manifestation wissenschaftlicher Erkenntnisse sind, ist es wichtig, dass Schüler den Erkenntnisprozess selbst durchlaufen und dadurch ihre Fehlvorstellung revidieren (Grosslight et al. 1991). Der aktuelle Einsatz von Modellen im naturwissenschaftlichen Unterricht berücksichtigt zumeist diesen Erkenntnisprozess nicht und begnügt sich mit einer oft wenig kontextbezogenen, oberflächlichen Verwendung von Modellen. Entsprechend verstehen Schüler den wissenschaftlichen Mehrwert der Verwendung von Modellen zur Erklärung komplexer Phänomene nicht und erkennen auch weder deren Abhängigkeit vom Forschungsstand noch grundlegende Unterschiede zwischen Modell und realem Phänomen (Krajcik & Merritt 2012, Ke et al. 2020).

Da Modelle inzwischen jedoch essenzieller Bestandteil naturwissenschaftlichen Lernens sind, ist es umso wichtiger, dass sie im Unterricht fachgemäß eingesetzt werden. Ihre Verbindung von Theorie mit Beobachtung und Prozesswissen macht Modelle für Schüler zu wertvollen Ressourcen des selbstgesteuerten Erkenntnisgewinns (Jackson et al. 2008, Passmore et al. 2009). Naturwissenschaftliche Phänomene sind oft komplex und besonders auf molekularer Ebene nur schwer zu beobachten. Modelle können unter Berücksichtigung der entsprechenden Theorien eine ebenso authentische, wenn nicht sogar gehaltvollere Erfahrung bieten als beispielsweise das Lesen eines wissenschaftlichen Textes. Modelle sind häufig, je nachdem wie erfolgreich ihre didaktische Reduktion war, leichter nachzuvollziehen als das natürliche Phänomen (Windschitl und Thompson 2006, Carpenter et al. 2018). So ist es auch nicht verwunderlich, dass es einen positiven Zusammenhang zwischen der Verwendung von Modellen und Lernerfolg gibt. Gerade die Erstellung eigener Modelle erfordert Tiefenverständnis und die Überführung von Wissen in eine andere Darstellungsform. Diese Form der Metakognition ist ein wesentlicher Bestandteil von Deep Learning Strategien (Mendonça & Justi 2013). Auch die kritische Reflexion und Auswertung von Ergebnissen sowie das Testen möglicher Erklärungen für beobachtete Phänomene zählen dazu (Chin & Brown 2000, Lee et al. 2015). Diese Strategien können auch den Aufbau von Scientific Literacy anregen, da sie den wissenschaftlichen Diskurs zwischen den Schülern fördern. Schüler müssen ihre Überlegungen kommunizieren, durch Argumentation ihre Mitschüler überzeugen und gegebenenfalls valide Überlegungen anderer Schüler in ihre Argumentation integrieren (Passmore & Svoboda 2012, Ke et al. 2020). Der Einsatz von Modellen im naturwissenschaftlichen Unterricht bietet den Schülern damit die

Möglichkeit, Erfahrungen im Kontext authentischer wissenschaftlicher Praxis zu sammeln (Mendonça & Justi 2013, Ke & Schwarz 2020).

3.2.5 Schulbezogenes Selbstkonzept

Neben dem Einsatz individueller Lernmethoden besitzt auch das schulbezogene Selbstkonzept einen erheblichen Einfluss auf den Lernerfolg von Schülern (Marsh & Martin 2011, Jansen et al. 2015). Je nachdem, wie Schüler ihre fachbezogenen Fähigkeiten wahrnehmen, kann diese Einschätzung positive oder negative Einflüsse auf die schulische Leistungsmotivation sowie die Demonstration und Weiterentwicklung der individuellen Fertigkeiten haben. Entsprechend ist es wichtig, eine Lernumgebung zu schaffen, die es Schülern ermöglicht, ein gesundes und lernförderliches Selbstkonzept aufzubauen (Patall et al. 2014, Bakadorova & Raufelder 2020). Denn weder ein übersteigertes noch ein zu geringes Selbstkonzept tragen zu nachhaltigem Lernerfolg bei. Lernerfolg ist dabei aber individuell definiert und unterscheidet sich von Schüler zu Schüler (Jansen et al. 2015). Wichtig ist hier eine realistische Zielsetzung zu fördern, die ein ebenso realistisches Selbstkonzept begünstigt. Denn allgemein gelten Selbstkonzept und Leistung als voneinander abhängig und sich gegenseitig verstärkend (Arens et al. 2016). Das bedeutet, dass ein positives Selbstkonzept, positives Verhalten bewirkt und damit auch den Lernerfolg positiv beeinfluss. Der erlebte Lernerfolg wiederum bestätigt das positive Verhalten und verstärkt das positive Selbstkonzept (Hoferichter et al. 2018).

TIMSS (Trends in International Mathematics and Science Study) Studien im Zeitraum von 2015 bis 2019 weisen zudem darauf hin, dass das Selbstkonzept über fachbezogene Fertigkeiten (Mullis et al. 2020, Bakadorova & Raufelder 2020) auch Karriereentscheidungen nachhaltig beeinflussen kann (Mejía-Rodríguez et al. 2020). So sehen Schüler trotz überdurchschnittlich guter Leistungen in MINT-Fächern von einer MINT-Karriere ab, wenn sie ihre wahrgenommene Leistung als deutlich geringer einschätzen als ihre tatsächliche Leistung (Goldman & Penner 2016). Dies könnte auch ein möglicher Grund für noch immer bestehende Geschlechterunterschiede in der Wahl von MINT-Karrieren sein (Mejía-Rodríguez et al. 2020). Entsprechend sollte ein besonderes Augenmerk auf die Entwicklung des individuellen Selbstkonzepts in der Schule gelegt werden, um möglichen Einflüssen vorzubeugen, die Geschlechterunterschiede fördern könnten (Hardy 2014, Mawang et al. 2018). Denn bislang besitzt das Selbstkonzept für Lehrkräfte oft keine unmittelbare Relevanz bei der Planung und Ausgestaltung von Unterricht (Patall et al. 2014).

Gleiches gilt auch für das Zusammenspiel von Kreativität und Selbstkonzept, obwohl dies in der Literatur, insbesondere bei begabten Schülern, immer wieder beschrieben wird

(Fleith et al. 2002). Allgemein wird Kreativität nur peripher mit naturwissenschaftlichen Prozessen des Erkenntnisgewinns in Verbindung gebracht (Hadzigeorgiou et al. 2012). Gerade Lehrkräfte verfolgen noch immer auf logischem Denken und Wissenserweiterung beruhende Ansätze, um das Selbstkonzept zu fördern, anstatt kreative Lernumgebungen zu schaffen, in denen die Schüler auch durch unkonventionellere Herangehensweisen Lernerfolg erzielen (Fleith et al. 2002).

3.2.6 Kreativitäts- und Individualitätsförderung durch die Lernumgebung

Die Integration von Kreativität in den naturwissenschaftlichen Unterricht erfordert bestimmte Rahmenbedingungen, innerhalb derer kreative Prozesse ermöglicht werden können, die unmittelbare Relevanz für den wissenschaftlichen Erkenntnisgewinn besitzen (Kind & Kind 2007). Obwohl Kreativität wesentlich ist für anhaltenden Lernerfolg (Gadja 2017), können diese Rahmenbedingungen im regulären Klassenunterricht oft nicht ohne Weiteres geschaffen werden (Runco et al. 2017). Entsprechend wurden Richtlinien für Lehrkräfte entworfen, um mit überschaubarem Aufwand kreative Lernumgebungen zu gestalten (Richards & Cotterall 2016). Dies umfasst insbesondere die Schaffung von Freiräumen zur Förderung der geistigen Flexibilität, also der Fähigkeit, neue Ideen aus der Betrachtung von bestehenden Konzepten auf verschiedenen geistigen Ebenen abzuleiten (Filippetti & Krumm 2020). So können abhängig vom Lerntyp individuelle und kollaborative Ansätze mit offenen Diskursaktivitäten die natürliche Lernumgebung für flexible und kreative Lernprozesse bereiten (Miller 2014). Wichtig ist dabei, dass die Schüler nicht gezwungen sind, einen vorgegebenen Ansatz zu verfolgen, sondern selbst wählen können, wie sie an die Lösung einer wissenschaftlichen Fragestellung herangehen (Lian et al. 2018). Obwohl beispielsweise Gruppenarbeit insgesamt positiv auf die Kreativität wirkt, da sie zum gemeinsamen Reflektieren und Diskutieren einlädt, vermuten manche Forscher, dass erzwungene Gruppenarbeit kreatives Denken einschränken kann (Csikszentmihalyi 2000; Schmidt 2011).

Kreativität lässt sich zudem in vier große Kategorien unterteilen: (1) Mini-C, das frühe und explorative Ansätze zur Entdeckung individueller Kreativität beschreibt, (2) Little-C, das sich aus der Wiederholung von Mini-C-Prozessen und deren Bestätigung entwickelt, (3) Pro-C, das das Individuum befähigt komplexere Probleme und Ideen zu lösen, (4) und Big-C, das nach vielen Jahren kreativer Prozesse erreicht wird. Soziale Interaktion im Klassenzimmer kann maximal Mini-C und Little-C fördern, da die Diskussion und Bewertung der individuellen Kreativität zur Lösung wissenschaftlicher Fragestellungen durch Mitschüler und Lehrer erfolgt.

Zudem werden Schüler dazu angehalten, durch ihre kreativen Herangehensweisen mögliche Interpretationen wissenschaftlicher Inhalte zu reflektieren (Kaufman & Beghetto 2009).

Neben der kritischen Reflexion ist es besonders wichtig, Situationen zu schaffen, die intrinsische Motivation fördern und nicht versuchen, den kreativen Prozess durch extrinsische Motivation zu erzwingen. Letzteres ist jedoch repräsentativ für die gängige Praxis an Schulen, die auf Prüfungen, der Bewertung von Teilarbeiten oder der Belohnung von Lernerfolg basiert. Negative Auswirkungen auf die Entstehung kreativer Prozesse werden oft übergesehen (Baer 2010, Amabile 1983). Kreativität und Lernerfolg sind seit jeher über intrinsische Motivation miteinander verbunden (Gajda 2016). Deshalb ist es für eine kreativitätsfördernde Lernumgebung unerlässlich, extrinsische Motivatoren zu reduzieren. Nur so kann den Schülern der Freiraum zur kreativen Auseinandersetzung mit naturwissenschaftlichen Inhalten ermöglicht werden (Amabile & Pratt 2016, Beghetto 2015).

Zudem sind kommunikative Praktiken ein wichtiger Bestandteil von Kreativität. Da Sprache ein Teil von Kultur ist und Kultur die Definition von Kreativität wesentlich beeinflusst, ist Sprache aus kreativen Prozessen nicht wegzudenken. Dieser Umstand macht Kreativität auch in bilingualen Unterrichtssituationen unmittelbar relevant (Csikszentmihalyi 1988). Die Erforschung von Kreativität in diesen Kontexten hat bereits ergeben, dass bilingual unterrichtete Schüler vermehrt unkonventionelle Denkansätze und Handlungsweisen zeigen. Als Begründung dafür werden eine gesteigerte Toleranz für Ambivalenzen und erhöhte geistige Flexibilität aufgeführt (Fleith et al. 2002).

Wie bereits erwähnt, ist Kreativität ebenso wie Lernen stark vom betreffenden Individuum abhängig. So haben zahlreiche Studien bereits verallgemeinerbare Korrelationen zwischen Persönlichkeitsvariablen wie *Extraversion* und *Offenheit* mit Kreativität nachgewiesen (Furnham et al. 2013, Antinori et al. 2017, Hoseinifar et al. 2011). *Extraversion* steht im Zusammenhang mit kreativen Leistungen und *Offenheit* ist eng mit der Einschätzung der eigenen Kreativität verbunden (Batey et al. 2010). *Offenheit* hat beispielsweise einen unmittelbaren Einfluss darauf, wie neugierig und kreativ Individuen ihre Welt erkunden und wie sie diese aus neurokognitiver Sicht wahrnehmen. Trotz ihres nachgewiesenen Einflusses auf kreative Prozesse, ist *Offenheit* noch immer die am schlechtesten verstandene Persönlichkeitsvariable (Antinori et al. 2017). Doch auch negativ assoziierte Persönlichkeitsvariablen, wie *Neurotizismus* (Batey et al. 2010) mit seinen zugehörigen Subskalen, beeinflussen Kreativität (Watrin 2019). *Neurotizismus* bezieht sich auf eine Veranlagung zu bestimmten Gefühlszuständen, wie Angst, Depression und Wut. Auch kann *Neurotizismus* mit Reaktionen auf Verlust oder Frustration in Zusammenhang gebracht werden. Frauen zeigen dabei oft eine stärkere Ausprägung dieser

Persönlichkeitsvariable als Männer (Tackett & Lahey 2017). Lange Zeit wurde Persönlichkeit als stabiles Konstrukt betrachtet. Neuere Studien, wie z. B. Kitamura et al. (2015), zeigten jedoch, dass sich bestimmte Persönlichkeitsfaktoren als Reaktion auf verschiedene Einflüsse verändern können. Infolgedessen wurden pädagogische Konzepte entworfen, die gezielt Verhaltensweisen fördern, die in direktem Zusammenhang mit Kreativität stehen (Hoseinifar et al. 2011).

3.2.7 Implementierung bilingualer Schülerlabore und Vorgängermodule

Ein pädagogisches Konzept, das viele der in den Vorkapiteln beleuchteten Anforderungen an ganzheitliches Lernen umfasst, ist das Schülerlabor. Wie bereits in der Masterarbeit der Autorin dargelegt, ist es gerade die Authentizität des außerschulischen Lernorts Labor, die ein wesentliches Unterscheidungsmerkmal zum Experimentieren in Klassenzimmer darstellt. Universitäten öffnen ihre Forschungseinrichtungen für Schüler, sodass diese ein möglichst reales Erleben des Forscheralltags erhalten können (Scharfenberg, 2005). Dabei werden Lerninhalte und zugehörige Experimente gemäß des nach KMK (2005) definierten Anforderungsniveaus der Teilnehmer angepasst, sodass die Gefahr der kognitiven Überlastung reduziert werden kann. Die Lernenden erhalten aber dennoch die Möglichkeit, sich mit authentischen Problemen und Fragestellungen des Laboralltags auseinanderzusetzen (Scharfenberg 2005).

Schülerlabore müssen gleichwohl nicht an den Lehrplan gebunden sein oder in ihrem Experimentdesign bestimmte Bezüge zu diesem aufweisen. Je besser sie jedoch angepasst sind, desto wahrscheinlicher ist ihre Akzeptanz und Integration in den Schulalltag. Da sie aber zumeist den aktuellen Stand der Wissenschaft und Technik vermitteln, Interesse wecken oder die Akzeptanz neuer Technologien fördern wollen, besitzen sie Gesellschaftsrelevanz (KMK 2005). Losgelöst von konkreten unterrichtlichen Vorgaben des jeweiligen Fachgebiets, werden am außerschulischen Lernort Labor ganz andere Möglichkeiten zur Erweiterung des wissenschaftlichen und persönlichen Horizontes geboten (Haupt et al. 2013).

Lernintensiv ist in besonderem Maße auch die Abhaltung des Experimentierlabors in einem bilingualen Format mit der Arbeitssprache Englisch. Dieses Design ist immer noch sehr selten, obgleich sich bilingualer Unterricht an Schulen großer Beliebtheit erfreut (Rodenhauser & Preisfeld 2014, Buse & Preisfeld 2018). So hat zwar sprachsensibler Unterricht seinen Weg auch in Schülerlabore gefunden, nicht aber basierend auf CLIL mit Englisch als Wissenschaftssprache (Schülerlabor-Atlas). Gerade in Deutschland gibt es dazu bislang nur wenig Forschung. Das Schülerlabor BeLL-Bio der Bergischen Universität Wuppertal und die daraus entstandenen

Forschungsarbeiten (Rodenhauser & Preisfeld 2014, Buse et al. 2018) gelten daher als Vorreiter.

Die Schülerlabore der Didaktik der Biologie an der Universität Bayreuth haben bereits eine lange Tradition. So konnte aus den über die Jahre hinweg gewonnenen Erfahrungen und Erkenntnissen ein bilinguales Modul geschaffen werden. Franz-Josef Scharfenberg (2005) entwickelte in seiner Dissertation das erste exemplarische Modul für ein molekularbiologisches Demonstrationslabor, das den selbstständigen Erkenntnisgewinn durch Schüler in den Vordergrund stellte. In der Form eines eintägigen Labormoduls konnten Schüler der gymnasialen Oberstufe (12. Jahrgangsstufe) authentische Experimente aus der Bakteriengenetik durchführen. In einer darauffolgenden Dissertation von Gaitano Franke (2010) wurde dieses Modul weiterentwickelt und an die Bedürfnisse von Schülern der 10. Jahrgangsstufe bayerischer Realschulen angepasst. Im Rahmen dieser Arbeit wurden insbesondere Schülervorstellungen und die Lernemotionen vor und während der Teilnahme am Demonstrationslabor untersucht. Ebenfalls für die 10. Jahrgangsstufe bayerischer Realschulen entwickelte Marlen Goldschmidt (2015) in ihrer Dissertation ein Demonstrationslabor für grüne Gentechnik. Neben der geistigen Anstrengung beleuchtete die daraus hervorgegangene Arbeit Hoffnungen und Befürchtungen gegenüber grüner Gentechnik. Der erste Entwurf eines Moduls, das in Struktur und Ablauf unmittelbarer Vorläufer der vorliegenden Arbeit war, wurde in der Dissertation von Jessica Langheinrich (2015) für die 11. Jahrgangsstufe bayerischer Gymnasien als ein eLearning-gestütztes Modul entwickelt. Besonderes Augenmerk wurde dabei auf die Schülervorstellungen bezüglich des Aufbaus der DNA gelegt sowie die Erforschung des Computerselbstkonzepts und die Einflüsse selbstständigen Experimentierens. Ohne eine eLearning-Phase, jedoch mit einer Erweiterung des Moduls um eine Modellbauphase und zunächst eine und später zwei Modellevaluationsphasen, wurde das Modul in der Dissertation von Julia Mierdel (2019c) weiterentwickelt. Neben Lernerfolg und kognitiver Belastung wurden Schüler insbesondere hinsichtlich ihres Modellverständnisses und ihrer Kreativität untersucht. Das daraus entstandene Schülerlabor „Einfach GENial“ mit zwei Modellevaluationsphasen diente zum einen als Grundlage der vorliegenden Arbeit und gleichzeitig als muttersprachliche Vergleichsgruppe. In die Konzeptionierung des in dieser Arbeit aufgeführten Moduls flossen aber insgesamt die Erfahrungen und Erkenntnisse aus allen aufgelisteten Bio-/Gentechnik Laboren ein.

3.3 Ziele und Fragestellungen der Teilarbeiten

Die vorliegende Arbeit beschäftigt sich mit dem Einfluss von bilingualem Lernen auf Leistung, Verständnis, Selbstkonzept und Kreativität (Abb. 1). Dabei dienen die ersten beiden Arbeiten der Festlegung des Modulaufbaus und der Pilotierung möglicher Fragebögen (Teilarbeit A und Teilarbeit B). Aus den daraus gewonnenen Erkenntnissen konnte ein auf *Content and Language Integrated Learning* (CLIL) aufbauendes englischsprachiges Modul für den Lernort Labor entwickelt werden. Dies untersuchte insbesondere inhaltliches Lernen von molekularbiologischen Methoden und Eigenschaften der DNA (Teilarbeit C), sprachliches Lernen (Teilarbeit E), Modellverständnis und Modellbau (Teilarbeit D) sowie Kreativität und Selbstkonzept (Teilarbeit F). Der Aufbau des Moduls orientierte sich dabei am alten Lehrplan der 9. Jahrgangsstufe bayerischer Gymnasien.

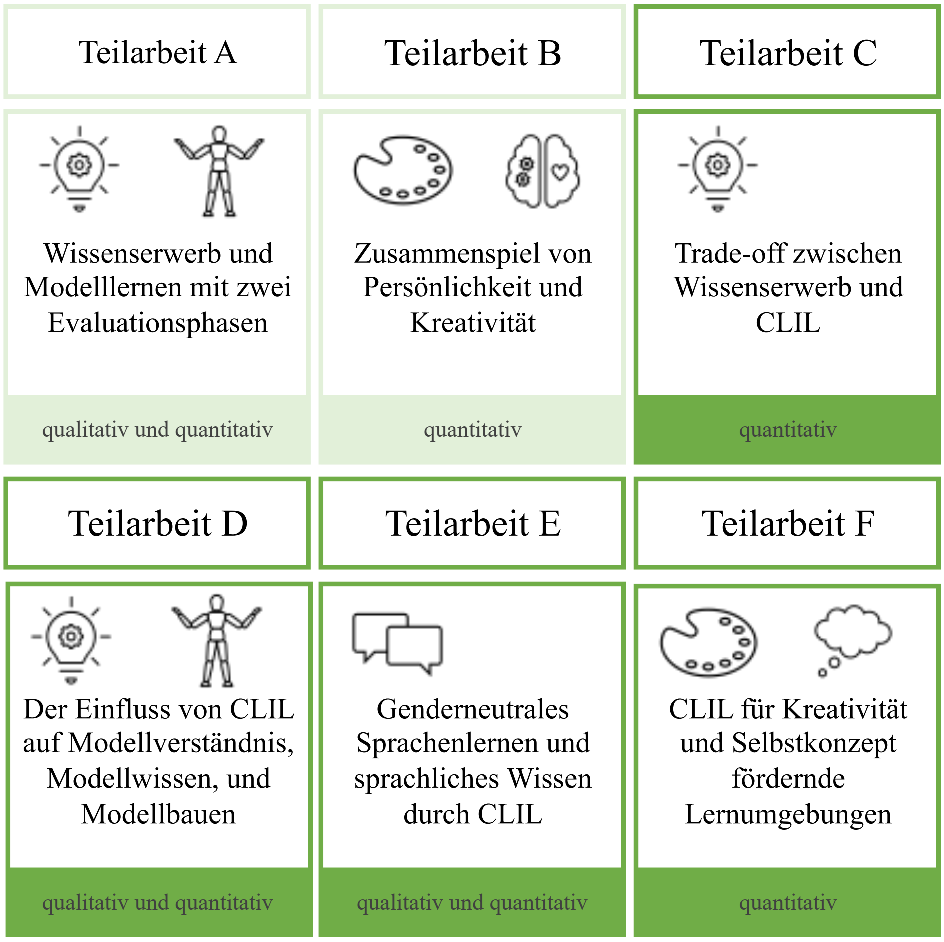


ABBILDUNG 1 Überblick über die sechs Teilarbeiten der Gesamtstudie.

Teilarbeit A: Wissenserwerb und Modelllernen mit zwei Evaluationsphasen

Modelle sind, wie bereits beschrieben, unerlässlich zur Erklärung komplexer oder schwer beobachtbarer naturwissenschaftlicher Phänomene (Kind 2015). Obwohl durch die Einführung des neuen Kompetenzkatalogs der Kultusministerkonferenz die Bedeutung von Modelllernen für wissenschaftliche Denkprozesse zunehmend ins Bewusstsein der Fachlehrkräfte dringt, werden Modelle im Unterrichtsalltag oft falsch eingesetzt (Werner et al. 2017). So sollen

Schüler Modelle meist ohne vorherige Einführung über ihre Funktion verwenden, was natürlich zu entsprechenden Fehlvorstellungen führt (Krajcik & Merritt 2012, Ke et al. 2020). Das Ziel von Teilarbeit A war es, durch eine zweite Modellevaluationsphase die kritische Reflexion und Diskussion anzuregen, die beide Teil jedes wissenschaftlichen Modellkonstruktionsprozesses sind (Acevedo 2008). Konkret hatte Teilarbeit A drei Teilziele:

1. Untersuchung des Einflusses der zweiten Modellevaluationsphase auf den Lernerfolg der Schüler durch einen Vergleich zwischen modellbezogenem Wissen der Schülergruppen mit nur einer und zwei Modellevaluationsphasen.
2. Bestimmung der Qualität der Selbstevaluation der Schüler sowie Erfassung der richtig identifizierten Bestandteile der DNA in den Schülermodellen.
3. Aufdeckung möglicher Korrelationen zwischen der Qualität der zweiten Selbstevaluation und dem Lernerfolg der Schüler, die an beiden Modellevaluationsphasen teilnahmen.

Entsprechend versuchte Teilarbeit A die folgenden Forschungsfragen zu beantworten:

1. Wie beeinflusst der Einsatz von einer und von zwei Modellevaluationsphasen den Lernerfolg der Schüler während ihrer Teilnahme im Lernlabor, besonders in Bezug auf Modellwissen?
2. Wie beeinflusst die zweite Modellevaluationsphase den generellen Lernerfolg der Schülergruppe, die an beiden Evaluationsphasen teilnahm?

Teilarbeit B: Zusammenspiel von Persönlichkeit und Kreativität

Die Förderung von Kreativität im naturwissenschaftlichen Unterricht kann dabei helfen, neue Lösungswege für schwierige wissenschaftlichen Problemstellungen zu finden (Nation & Newton 2008). Obwohl es noch immer keine einheitliche Definition von Kreativität gibt und das Konstrukt schwer zu fassen ist, gilt in Hinblick auf die Forschung, dass Kreativität die kognitive Leistungsfähigkeit in formellen und informellen Lernumgebungen steigert (Gajda 2016). Entsprechend birgt die erfolgreiche Integration von Kreativität in den regulären naturwissenschaftlichen Unterricht zahlreiche Hoffnungen und Herausforderungen. Dies umfasst u.a. die Förderung von Kreativität durch das Identifizieren und Schaffen entsprechender Lernsituationen (Amabile & Pratt 2016, Beghetto 2015). Auch sollten darin Persönlichkeitsvariablen berücksichtigt werden, die mit kreativem Verhalten korrelieren (Kaufman et al. 2008, Barron & Harrington 1981). Folglich wurden zwei Messinstrumente herangezogen, die beide Konstrukte – Kreativität und Persönlichkeit – messen sollten. Da diese in ihrer ursprünglichen Form zu lang waren, entschied man sich für den etablierten *BFI-10* (Rammstedt & John 2007) und eine eigene Kurzfassung des *CPAC* (Miller 2014). So war es Ziel der Teilarbeit B ein Messinstrument zu identifizieren und zu evaluieren das zuverlässig Kreativität und Persönlichkeitsvariablen erfasst, ohne den Rahmen des regulären Klassenunterrichts zu sprengen. Die folgenden zwei Teilziele sollten in Teilarbeit B beantwortet werden:

1. Überprüfung des Einsatzes der Messinstrumente für Kreativität und Persönlichkeit und Identifikation möglicher Korrelationen zwischen diesen.
2. Untersuchung möglicher Geschlechterunterschiede in der Verteilung von Kreativitäts- und Persönlichkeitsvariablen.

Entsprechend versuchte Teilarbeit B die folgenden Forschungsfragen zu beantworten:

1. Wie erfassen die gekürzten Fragebögen für Kreativität und Persönlichkeit die jeweiligen Konstrukte im Klassenzimmer?
2. Wie korrelieren die im *BFI-10* beschriebenen fünf Persönlichkeitsvariablen mit den latenten Variablen *Act* und *Flow* des Kreativitätstests?
3. Wie zeigen sich Geschlechterunterschiede für Kreativität und Persönlichkeitsvariablen in nach dem Alter der Schüler bestimmten Extremgruppen?

Teilarbeit C: Trade-off zwischen Wissenserwerb und CLIL

Bilinguales Lernen ist eine Möglichkeit, den Herausforderungen zu begegnen, die durch die disziplinübergreifende Verbreitung der englischen Sprache entstehen. Besonders das *Content and Language Integrated Learning* (CLIL) vermag die Sprachkompetenzen zu vermitteln, die notwendig sind, um in diesem Umfeld zu bestehen (Lemke 1990, Klieme et al. 2010). Zahlreiche Schulen bieten entsprechend bereits bilinguale Programme an. Zumeist handelt es sich dabei um bilingualen Unterricht in Langzeitform, da hier, im Vergleich zu bilingualem Unterricht in Kurzzeitform, der positive Einfluss auf das fachliche und sprachliche Lernen bereits nachgewiesen ist (Meyerhöffer & Dreesmann 2019). Viele Schulen verfügen jedoch nicht über die Mittel, bilingualen Unterricht in Langzeitform anzubieten (Ngai 2002). Entsprechend ist es wichtig zu untersuchen, inwiefern auch bilingualer Unterricht in Kurzzeitform für den naturwissenschaftlichen Unterricht geeignet ist. Die Teilarbeit C wertete daher den Lernerfolg von bilingualem Unterricht in Kurzzeitform anhand des bilingual adaptierten Moduls mit zwei Modellevaluationsphasen aus. Dabei standen folgende drei Teilziele im Vordergrund:

1. Bewertung des allgemeinen Lernerfolges der Schüler während ihrer Teilnahme am Modul.
2. Ermittlung möglicher Unterschiede zwischen dem Lernerfolg der Schülergruppen, die an der muttersprachlich deutschen und der CLIL Intervention teilgenommen haben.

3. Untersuchung möglicher Korrelationen zwischen dem Lernerfolg der CLIL-Schülergruppe und ihren Biologie- sowie Englischnoten.

Dies führte zu insgesamt zwei Forschungsfragen:

1. Wie beeinflusst der Besuch eines eintägigen CLIL-Forschungsmoduls den Lernerfolg der Schüler?
2. Wie beeinflusst der Besuch eines eintägigen CLIL-Forschungsmoduls den Lernerfolg der Schüler im Vergleich zur Schülergruppe, die das strukturell identische, muttersprachliche deutsche Labor besucht hat?

Teilarbeit D: Der Einfluss von CLIL auf Modellverständnis, Modellwissen, Modellbauen

Modelle werden besonders im naturwissenschaftlichen Unterricht gerne eingesetzt (Passmore & Svoboda 2012). Sie adressieren und fördern wissenschaftliche Kompetenzen auf verschiedenen Ebenen (Schwarz et al. 2009). Schüler werden durch die Verwendung von Modellen dazu angeregt, aktiv über wissenschaftliche Erkenntnisse zu diskutieren und reflektieren, weshalb Modelle ein wichtiges Werkzeug zur Schulung von wissenschaftlichen Kompetenzen sind (Grandinetti et al. 2013, Lee et al. 2015). Besonders in Verbindung mit CLIL wird der wissenschaftliche Austausch stärker angeregt, was sich wiederum positiv auf das Tiefenverständnis auswirkt (Passmore & Svoboda 2012, Mendonça & Justi 2013, Ke et al. 2020). Entsprechend wurden in Teilarbeit D vier Teilziele verfolgt:

1. Bewertung des inhaltlichen Wissens der Schüler über die DNA als Modell während der Teilnahme am muttersprachlichen und CLIL-Modul.
2. Identifikation möglicher Unterschiede im Modellverständnis zwischen den muttersprachlichen und den CLIL-Schülergruppen.
3. Identifikation möglicher Unterschiede in der Qualität der selbstgebauten Modelle zwischen den muttersprachlichen und den CLIL Schülergruppen.
4. Identifikation möglicher Unterschiede in der ersten und zweiten Modellevaluationsphase zwischen den muttersprachlichen und den CLIL Schülergruppen.

Daraus ergaben sich folgende Forschungsfragen:

1. Wie beeinflusst die Teilnahme an einem eintägigen CLIL-Labormodul das inhaltliche Wissen der Schüler über die DNA als Modell?
2. Wie beeinflusst die Teilnahme an einem eintägigen CLIL-Labormodul das Modellverständnis im Vergleich zur muttersprachlichen Schülergruppe?
3. Wie beeinflusst die Teilnahme an einem eintägigen CLIL-Labormodul die Qualität der selbstgebauten Modelle im Vergleich zur muttersprachlichen Schülergruppe?

4. Wie beeinflusst die Teilnahme an einem eintägigen CLIL-Labormodul die erste und zweite Modellevaluationsphase im Vergleich zur muttersprachlichen Schülergruppe?

Teilarbeit E: Genderneutrales Sprachenlernen und sprachliches Wissen durch CLIL

Neben der Vermittlung sachfachlicher Inhalte über praktische Laborexperimente sollte das CLIL-Modul auch sprachliche Kompetenzen fördern. Teilarbeit E fokussierte sich entsprechend auf die Auswirkungen der Integration von wissenschaftlichen Texten, verbalen oder schriftlichen Anweisungen, Schreibarbeiten und Diskursaktivitäten in englischer Sprache auf die sprachliche Kompetenz und Scientific Literacy. Zudem griffen kurz nach der Teilnahme am Labormodul die COVID-19 bedingten Lockdown-Beschränkungen (Hoenig & Wenz 2020), was u.a. auch die Ergebnisse des Behaltenstests beeinflusst haben könnte. Denn zum Ausbau und Erhalt sprachlicher Fähigkeiten gehören insbesondere wiederholte Diskursaktivitäten und repetitives Üben des neuen Wortschatzes (Thürmann 2008). Entsprechend zielte die Teilarbeit E vornehmlich auf die Bewertung des sprachlichen Lernerfolgs und der Scientific Literacy während und kurz nach der Teilnahme am Labor. Zudem sollten mögliche Auswirkungen des Distanzunterrichts nach Einführung der COVID-19 bedingten Lockdown-Beschränkungen auf die sprachliche Kompetenz untersucht werden. Konkret wurden zwei Teilziele verfolgt:

1. Untersuchung des Einflusses von CLIL auf den Spracherwerb und Scientific Literacy.
2. Identifikation möglicher Auswirkungen der Lockdown-Beschränkungen auf die Sprachkompetenz.

Daraus ergaben sich folgende Forschungsfragen:

1. Wie beeinflusst die Teilnahme an einem eintägigen CLIL-Labormodul die sprachlichen Kompetenzen und Scientific Literacy?
2. Wie beeinflussen die COVID-19 bedingten Lockdown-Beschränkungen den Erhalt bzw. Ausbau sprachlicher Kompetenzen?

Teilarbeit F: CLIL für Kreativität und Schulselbstkonzept-fördernde Lernumgebungen

Praktische und geistig fordernde Lösungsstrategien zu naturwissenschaftlichen Fragestellungen zu finden, erfordert bereits ein hohes Maß an Kreativität. Diese Lösungsstrategien zusätzlich in einer Fremdsprache auszuarbeiten und zu diskutieren, bedarf entsprechend gänzlich neuer Wege der Sinnbildung (Fleith et al. 2002). Diese sind jedoch notwendig, um erfolgreich Experimente durchzuführen, Modelle zu konstruieren und sachfachliches sowie sprachliches Wissen aufzubauen. Dadurch wird insbesondere Kreativität gefordert und gefördert, was erheblich zum Lernerfolg in formellen und informellen Lernumgebungen beitragen kann (Gajda 2016).

Zudem wird angenommen, dass Kreativität mit dem schulischen Selbstkonzept der einzelnen Schüler in Zusammenhang steht (z.B. Sisk 1972, Mawang et al. 2018). Die Entwicklung beider Konstrukte hängt dabei von zahlreichen Faktoren ab. Grundsätzlich scheint jedoch das Schaffen einer anregenden Lernumgebung geeignet zu sein, Ängste abzubauen und das Gefühl von Sicherheit zu vermitteln aber gleichzeitig Raum für eigenständiges Erkunden zu lassen (Justo 2008, Marsh et al. 2015). Entsprechende Skalen zur Messung von Kreativität (*CPAC*, zwei Faktoren: *Act* und *Flow*) und Selbstkonzept (*SESSKO*, vier Faktoren: Kriterial, Individuell, Sozial und Absolut) sollen Faktoren identifizieren, die eine solche Lernumgebung ermöglichen. Daraus entwickelten sich die insgesamt fünf Teilziele der Teilarbeit F:

1. Nachweisen der Vier-Faktor-Struktur des schulbezogenen Selbstkonzepts für die Fächer Biologie und Englisch.
2. Nachweis der Zwei-Faktor-Struktur des Kreativitätsfragebogens in Abhängigkeit des schulbezogenen Selbstkonzepts für die Fächer Biologie und Englisch.
3. Erfassen möglicher Korrelationen zwischen dem schulbezogenen Selbstkonzept für die Fächer Biologie und Englisch und den Subskalen des Kreativitätsfragebogens.
4. Erfassen möglicher Korrelationen zwischen dem schulbezogenen Selbstkonzept für die Fächer Biologie und Englisch und den Vor-, Nach- und Behaltenstests zur Messung des sachfachlichen und sprachlichen Lernerfolgs.
5. Identifikation möglicher Geschlechterunterschiede im schulbezogenen Selbstkonzept für die Fächer Biologie und Englisch.

Entsprechend ergaben sich folgende Forschungsfragen:

1. Wie beeinflussen sich das schulbezogene Selbstkonzept und Kreativität während der Teilnahme an einem eintägigen CLIL-Labormodul?
2. Wie beeinflusst das schulbezogene Selbstkonzept den sprachlichen Lernerfolg während der Teilnahme an einem eintägigen CLIL-Labormodul?
3. Wie beeinflusst die Geschlechterzugehörigkeit das schulbezogene Selbstkonzept während der Teilnahme an einem eintägigen CLIL-Labormodul?

3.4 Methoden

3.4.1 Stichprobe und Studiendesign

Insgesamt wurden für die vorliegende Arbeit Daten zu drei unterschiedlichen Zeitpunkten mit unterschiedlichen Schülergruppen erhoben. Für Teilarbeit A wurde auf Daten zurückgegriffen, die während der praktischen Moduldurchführung mit Schülern der 9. Jahrgangsstufe am Ende des Wintersemesters 2016/2017 und 2017/2018 von Julia Mierdel erhoben wurden. Daten aus der Intervention von 2017/2018 dienten zudem für die Teilarbeiten C und D als Vergleichsgruppe. Bei Teilarbeit B handelte es sich um eine vom praktischen Modul unabhängige Datenerhebung zur Pilotierung eines Fragebogens. Hierfür wurden die Schülerdaten unterschiedlicher Jahrgangsstufen herangezogen. Die letzte große Datenerhebung, die relevant war für die Ausarbeitung der Teilarbeiten C, D, E und F, fand am Ende des Wintersemester 2019/2020 statt und umfasste die praktische Durchführung des CLIL-Moduls mit Schülern der 9. Jahrgangsstufe unter der Leitung der Autorin.

Für die Beteiligung am muttersprachlich deutschen Genlabor (Teilarbeit A) im Winterhalbjahr 2017/2018 konnten insgesamt 296 Neuntklässler (Mädchen 52.0%, Jungen 48.0%; $M_{Klassengröße} = 22.8$, $SD = 6.2$; $M_{Alter} = 14.6$, $SD = 0.8$) von verschiedenen oberfränkischen Gymnasien gewonnen werden. Sechs Klassen nahmen dabei an der einfachen (N = 151) und sieben Klassen (N = 145) an der zweifachen Modellevaluation teil. Grundlage der Datenerhebung war das Ausfüllen von Papierfragebögen zu drei Testzeitpunkten. Der erste Fragebogen wurde in der Regel ein bis zwei Wochen vor dem Laborbesuch ausgefüllt (Vortest oder T0), der zweite direkt im Anschluss an das Modul (Nachtest oder T1) und der dritte etwa acht Wochen nach Teilnahme (Behaltenstest oder T2). Für das Ausfüllen der Fragebögen wurde kein zeitliches Limit gesetzt, doch betrug die Bearbeitungszeit im Durchschnitt zwischen 20 und 30 Minuten. In Einklang mit der Datenschutzgrundverordnung erfolgte die Anonymisierung der Teilnehmenden über einen Code bestehend aus Geschlecht (M/W/D), Geburtsmonat und -jahr, den beiden ersten Buchstaben des Vornamens der Mutter sowie der Hausnummer. Unter Berücksichtigung der Deklaration von Helsinki wurde zudem das schriftliche Einverständnis der Erziehungsberechtigten an der Teilnahme des Moduls sowie der Erhebung von Daten durch Fragebögen eingeholt. Für die Auswertung der gesammelten Daten wurden stets nur vollständige Datensätze herangezogen, weshalb die Stichprobenzahl in den einzelnen Teilarbeiten variieren kann, je nachdem ob einzelne Teile des Fragebogens nur an einem oder allen drei Testzeitpunkten abgefragt wurden. Die Legitimation des Unterrichtsmoduls sowie der Fragebögen erfolgte durch das Bayerische Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst mit dem Schreiben vom 16.02.2017 unter Aktenzeichen X.7-BO5106/149/10.

An der Pilotierung des gekürzten Fragebogens zur Erfassung von Kreativität und Persönlichkeit für das bilinguale Genlabor (Teilarbeit B) im Winterhalbjahr 2019/2020 nahmen insgesamt 538 Schüler (Mädchen 65.4 %, Jungen 34.6 %; $M_{Alter} =16{,}96$; $SD = 2{,}99$) aus drei weiterführenden Schulen teil. Die Datenerhebung erfolgte zu nur einem Testzeitpunkt mittels Papierfragebögen. Das Ausfüllen der Fragebögen erfolgte ohne Zeitlimit, nahm aber in etwa zehn Minuten in Anspruch und wurde von den Lehrern der jeweiligen MINT-Fächer beaufsichtigt. Auch hier wurden die Grundsätze der Datenschutzgrundverordnung und der Deklaration von Helsinki berücksichtigt. Die Legitimation des Unterrichtsmoduls sowie der Fragebögen erfolgte durch das Bayerische Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst mit dem Schreiben vom 04.10.2019 unter Aktenzeichen IV.7-BO5106/149/18.

An der eigentlichen Datenerhebung im Winterhalbjahr 2019/2020, die relevant für die Ausarbeitung der Teilarbeiten C bis F war, nahmen insgesamt 316 Neuntklässler (Mädchen 50.8%, Jungen 49.2%; $M_{Klassengröße}$ = 21.9, *SD* = 4.6; M_{Alter} = 14.7, *SD* = 0.7) aus 14 Klassen von zehn verschiedenen oberfränkischen Gymnasien teil. Die Daten von 149 Neuntklässlern aus Intervention vom Winterhalbjahr 2017/2018 dienten als Vergleichsgruppe. Die Datenerhebung erfolgte wieder zu drei Testzeitpunkten analog zum Testdesign vom Winterhalbjahr 2017/2018. Das Ausfüllen der Fragebögen nahm in etwa 30 bis 35 Minuten in Anspruch, wobei sowohl für die Teilnahme als auch Datenerhebung die Grundsätze der Datenschutzgrundverordnung und der Deklaration von Helsinki berücksichtigt wurden. Darüber hinaus erfolgte die Auswertung der gesammelten Daten nur mit vollständigen Datensätzen, weshalb die Stichprobenzahl in den einzelnen Teilarbeiten (C bis F) variierte, je nachdem ob einzelne Teile des Fragebogens nur an einem oder allen drei Testzeitpunkten abgefragt wurden. Auch konnten im Modul verwendete Arbeitshefte und Scaffolding-Aufgaben sowie die Modellskizzen, selbst gebastelte DNA-Modelle und Kopien der Modellevaluationsblätter zur Auswertung herangezogen werden. Die Legitimation des Unterrichtsmoduls sowie der Fragebögen erfolgte durch das Bayerische Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst mit dem Schreiben vom 04.10.2019 ebenfalls unter Aktenzeichen IV.7-BO5106/149/18.

3.4.2 Ausführliche Beschreibung des Unterrichtsmoduls

Das Unterrichtsmodul „Einfach GENBILIal! Ein bilinguales Modul zur DNA als Träger der Erbinformation“ wurde für zehn Schulstunden (etwa 450 min) konzipiert. Da für die Durchführung der einzelnen Teilexperimente Equipment benötigt wird, das selbst in gut ausgestatteten Schullaboren in der erforderlichen Anzahl nicht vorhanden ist (Sarmouk et al. 2019), wie beispielsweise Zentrifugen, Mikropipetten, Gelelektrophoresekammern und UV-Geräte, erfolgte das Modul ausschließlich am außerschulischen Lernort-Labor der Universität Bayreuth. Vor der Teilnahme wurden die verantwortlichen Fachlehrkräfte aufgefordert, das Themengebiet Genetik bis nach dem Laborbesuch aufzuschieben und auch eine Nachbereitung der Inhalte des Moduls erst nach Bearbeitung des Behaltenstests vorzunehmen.

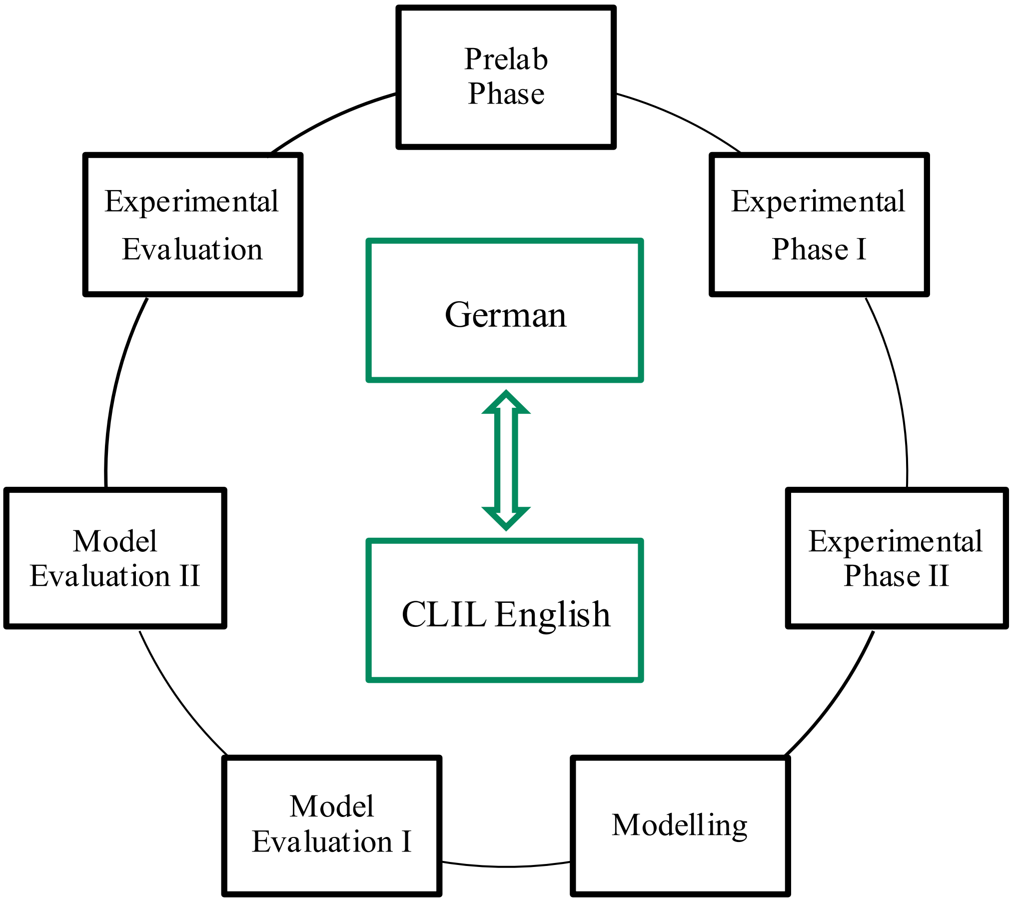


ABBILDUNG 2 Einzelne Phasen des bilingualen Unterrichtsmoduls „Einfach GENBILIal!“ mit sprachlicher Vermittlung im Zentrum und zweifacher Modellevaluation.

Insgesamt gliederte sich das Unterrichtsmodul „Einfach GENBILIal!“ in sieben aufeinanderfolgende Phasen (Abb. 2). Damit war das CLIL-Modul strukturell identisch zu der im Winterhalbjahr 2017/2018 durchgeführten und rein monolingual deutschen Implementierung mit der doppelten Modellevaluation (Teilarbeit A). Auch die Kursanleitung zu den einzelnen experimentellen Schritten und die entsprechenden fachlichen Erklärungen blieben weitestgehend gleich. Die Inhalte wurden für das CLIL-Modul lediglich ins Englische überführt, wobei für einzelne wichtige Vorgangsverben die deutsche Übersetzung in Klammern erhalten blieb. Für weiteres sprachliches Scaffolding, mit dem das Verständnis von fachlichen Inhalten erleichtert werden sollte (Gottlieb 2016, Fernández-Fontecha et al. 2020), wurden eigens Übungshefte mit Vokabelrätseln für das CLIL-Modul angefertigt, die analog zu den individuellen Modulphasen entsprechend relevantes Vokabular enthielten. Dabei sollten beispielsweise Kreuzwort- oder Wortsuchrätsel gelöst werden, indem zu vorgegebenen Definitionen die dazugehörigen Fachwörter aus der Kursanleitung herausgesucht werden mussten. Zusätzlich sollten die Lernenden anhand von Definitionen die deutsche Übersetzung des Wortes ableiten oder eine Definition zu einem bereits übersetzen Wort finden. Wörterbücher, die das Lösen der Rätsel erleichterten und gegebenenfalls bei der Formulierung eigener Sätze halfen, wurden sowohl papierbasiert als auch digital zur Verfügung gestellt.

Zusammenarbeit und wechselseitiger Austausch innerhalb und zwischen den Schülergruppen zur Lösung der Aufgaben war im Rahmen der Modulteilnahme ausdrücklich erwünscht. Die Unterrichtssprache, besonders in Bezug auf Instruktion und Betreuung durch die Autorin, war im CLIL-Modul ausschließlich Englisch. Dadurch sollten die Schüler ermutigt werden ebenfalls in der Fremdsprache zu kommunizieren (Fehling 2008). Zur Aufrechterhaltung des Redeflusses oder bei Verständnisproblemen durften jedoch auch einzelne Wörter oder Satzbestandteile gemäß der Praktik des Code-Switchings in Deutsch geäußert oder schriftlich festgehalten werden (Königs 2013). Auf Code-Switching wurde besonders bei der Einführung der Gelelektrophorese zurückgegriffen. Zwar war das Lernplakat auf Englisch, doch wurden die durch Diskussion und Demonstration erarbeiteten Stichpunkte oder Schlagwörter in Deutsch festgehalten. Nach jeder Einführung von theoretischen Inhalten zu einer der sieben Phasen konnten die Lernenden zudem gemäß eines Ampelsystems eine Karte in Grün, Gelb oder Rot hochhalten, um mitzuteilen, wie gut sie das Gesagte verstanden hatten. In seltenen Fällen musste dann durch die Autorin eine Wiederholung der Inhalte in Englisch oder auf Deutsch erfolgen.

Inhaltlich setzte sich das Labor mit den Eigenschaften, der molekularbiologischen Analyse und dem Aufbau von DNA auseinander. Zur Kontextualisierung und Schaffung eines Bezugs zur Lebenswelt der Schüler schlüpften diese zunächst in die Rolle von Kriminalisten, um mögliche Herangehensweisen an die Aufklärung des Isarmords (Mierdel & Bogner 2020) zu identifizieren und später in die Rolle der Wissenschaftler Watson und Crick, um das Rätsel um die Struktur der DNA zu lösen (Usher 2013). Im Sinne des schülerzentrierten, forschend-entdeckenden Lernens wurde zur Umsetzung der Lerninhalte eine Mischung aus praktischem Experimentieren mit geistig fordernden Aufgaben zur kritischen Reflexion und Erklärungsfindung gewählt (Glynn & Muth 1994, France 2019). Da neben dem inhaltlichen Lernen auch „Scientific Literacy“ und englische Sprachkompetenz gefördert werden sollten, arbeiteten die Schüler weitestgehend selbstständig in Zweiergruppen unter Zuhilfenahme der Kursanleitung. Die Kursanleitung enthielt neben allgemeinen Erklärungen und expliziten Arbeitsaufträgen auch Verständnisfragen, die jeweils vor der Durchführung des Arbeitsauftrags diskutiert und beantwortet wurden. Das inhaltliche und sprachliche Scaffolding dient dabei zur inneren Differenzierung und ermöglicht eine Berücksichtigung der individuellen Stärken und Schwächen der Schüler sowie ein Angleichen des Arbeitstempos (Coyle 2007, Grandinetti 2013). Der Inhalt des Moduls entsprach dem bayerischen Lehrplan (ISB 2021) und folgte den nationalen

Kompetenzanforderungen (KMK 2005). Die Zuordnung der Schülerklassen zu den jeweiligen Interventionsvarianten erfolgte zufällig und Schülerklassen aus der muttersprachlichen Schülergruppe nahmen nicht erneut an der CLIL-Intervention teil. Dadurch wurden die für quasi-experimentelle Untersuchungen erforderlichen intakten Klassenverbände gewahrt (Cook & Campbell 1979) und die muttersprachlichen und CLIL-Schülergruppen konnten als unabhängige Variable behandelt werden.

Beide Gruppen starteten vor Beginn der Experimentier- und Modellphasen mit einer ausführlichen **Einführungsphase** (Prelab-Phase, Abb. 2). Viele Schüler waren vor der Teilnahme am Modul noch nie in einem Labor und hatten in der Schule kaum Experimentiererfahrung sammeln können. Folglich waren die meisten Arbeitsweisen und Arbeitsgeräte im Labor unbekannt. Entsprechend zielt die Einführungsphase in solchen Kontexten besonders auf die Vermittlung von theoretischem und praktischem Wissen im Umgang mit Laborequipment und naturwissenschaftlichen Arbeitsweisen (Sarmouk et al. 2019). So lernten die Schüler zunächst losgelöst von den eigentlichen Experimenten, durch intensive Übung mit den für die Experimente notwendigen Geräten umzugehen, wie beispielsweise Mikropipetten und Zentrifugen. Die Autorin nahm dabei die Rolle der Lernbegleiterin ein und trat, abgesehen von der Vermittlung theoretischer Konzepte über eine interaktive Smartboard-Präsentation sowie einer kurzen Demonstration der nachfolgenden experimentellen Schritte, komplett in den Hintergrund. So sollte sichergestellt werden, dass sich die Schüler weitestgehend selbstständig an die Experimente herantasten konnten (Glynn & Muth 1994, France 2019) ohne sich von dieser Situation überfordert zu fühlen (Kalyuga 2009).

Außerhalb der Einführungsphase wurden jedoch zwischen den jeweiligen Experimentier- und Modellphasen immer wieder relevante theoretische Inhalte in Bezug auf die Bedeutung von DNA, den zugrundeliegenden wissenschaftlichen Vorgängen zur DNA-Isolierung und DNA-Analyse mittels Gelelektrophorese vermittelt. Die Inhalte bauten dabei auf dem Vorwissen der Schüler über den Aufbau und die Funktion von Zellbestandteilen aus der achten Jahrgangsstufe auf (ISB 2021). Durch die Aktivierung dieses Vorwissens sollten die Schüler befähigt werden, eigene Hypothesen über den Ausgang der Experimente aufzustellen (Mierdel & Bogner 2019a). Faktenwissen in einer Einführungsphase geballt, aber danach themenbezogen über den gesamten Kurstag verteilt zu vermitteln, stellte sicher, dass die Schüler von der Vielzahl neuer Informationen nicht überwältigt wurden. Auch konnten sie sich dadurch besser auf die bevorstehenden Experimente vorbereiten und konzentrieren (Rodenhauser & Preisfeld 2014, Mierdel & Bogner 2019a).

Für die beiden **Experimentierphasen** (Abb. 2) wurde ein evidenzbasierter, zweistufiger Untersuchungsansatz gewählt (Mierdel & Bogner 2019a). Entsprechend beantworteten die Lernenden zunächst offene Verständnisfragen zum theoretischen Hintergrund in ihren Kursanleitungen und überlegten darauf basierend ihr Vorgehen für den weiteren Versuchsablauf. Die Arbeit in Zweier- oder Dreiergruppen ermöglichte dabei den notwendigen Austausch vor der Durchführung der Experimente, wodurch sich die Schüler bereits theoretisch mit den nachfolgenden Experimenten auseinandersetzten. So sollte verhindert werden, dass die Schüler einfach nur Anweisungen befolgen, anstatt selbst nachzudenken (Hofstein & Lunetta 2004). Um dies zu ermöglichen, bot das Kurshandbuch ausführliches inhaltliches und sprachliches Scaffolding, sodass Schüler selbstständig experimentieren und ihre Beobachtungen protokollieren konnten. Auch wurde das Modellieren der DNA fragenbasiert gelenkt, ohne die Schüler in ihren geistigen Denk- und Verarbeitungsprozessen zu bevormunden (Schwichow et al. 2016). Die Kursanleitung wurde zuvor im Rahmen von vier konsekutiven Laborauswertungen konzipiert und weiterentwickelt (Goldschmidt & Bogner 2016, Langheinrich & Bogner 2015, Mierdel & Bogner 2020, Teilarbeit A).

Die beiden **Modellphasen** erfolgten direkt im Anschluss an die Experimentierphasen (Abb. 2). Da Modelle zur Visualisierung und Erklärung beobachteter Phänomene unverzichtbar sind (Krajcik & Merritt 2012), griff das Modul für die Modellphasen auf drei Teilschritte zurück: (1) Mentale Modellphase basierend auf der Analyse eines Textes (Usher 2013), (2) Modellbauphase unter Zuhilfenahme diverser Bastelmaterialien und (3) Zwei Modellevaluationsphasen. Orientiert wurde sich bei der Konzeptionierung der Modellphasen am Model of Modelling von Gilbert und Justi (2002), das vier Phasen der Modellerstellung unterscheidet: (1) Sammeln von Informationen und Erkenntnissen über das zu modellierende Phänomen, (2) Erstellung eines mentalen Modells, (3) Auswahl der passenden Darstellungsweise und (4) Evaluation des erstellten Modells und Analyse möglicher Grenzen der Darstellungsform. Das mentale Modell gilt als Grundlage, um Beobachtungen des zu untersuchenden Phänomens theoretisch zu fundieren. Für das vorliegende Unterrichtsmodul wurde ein Brief von Crick an seinen Sohn verwendet, der alle relevanten Informationen über die Bestandteile der DNA sowie deren Anordnung enthält (Usher 2013). Die Schüler konnten entsprechend ihr über den Kurstag aufgebautes mentales Modell einer ersten theoretischen Fundierung und Validierung unterziehen (Mierdel & Bogner 2019a, Franco & Colinvaux 2000). Nach Festlegung der Darstellungsform und Anfertigung eines darauf basierenden Modells erfolgte die erste Modellevaluation (Abb. 2) in der Form einer reziproken Selbstevaluation. Dabei fertigten die Schüler eine beschriftete Skizze ihres DNA-Modells an, was sie zu einem erneuten Vergleich ihres Modells mit dem im

Text beschriebenen Modell ermutigte (Prabha 2016). Zusätzlich regten offene Fragen im Kursheft zu den einzelnen Komponenten der DNA und ihrer Repräsentation im Modell zum reflektierten Schreiben an (Kovanović et al. 2018). Schüler konnten dabei bestimmte Entscheidungen bei der Entwicklung und Anfertigung ihres Modells überdenken und gegebenenfalls verändern (Mierdel & Bogner 2019a). In einer zweiten Modellevaluation verglichen die Schüler ihr handgefertigtes Modell mit einem kommerziell erhältlichen Schulmodell. Ein Evaluationsblatt mit einem Foto des Schulmodells und entsprechender Beschriftung der einzelnen Komponenten erleichterten die Auswertung. Die Schüler mussten lediglich alle Bestandteile, die sie in ihrem handgefertigten Modell berücksichtigt hatten auf dem Evaluationsblatt abhaken. Anschließend konnte das Evaluationsblatt gewendet werden, wodurch eine Fotografie des im Brief beschriebenen Originalmodells von Watson und Crick sichtbar wurde. Die Lernenden sollten mit offenen Fragen auf der Rückseite des Skizzenblattes gezielt Unterschiede und Gemeinsamkeiten zwischen ihrem Modell, dem Schulmodell und dem Modell von Watson und Crick herausarbeiten sowie über die Funktion und Natur von Modellen nachdenken (Mierdel & Bogner 2020).

In der letzten Phase, der **Interpretationsphase**, wurden Erkenntnisse aus den Modellphasen kurz zusammengefasst sowie das Originalmodell von Watson und Crick erklärt. Zudem wurden weitere DNA-Modelle gezeigt und die Lernenden diskutierten über die Vielzahl an verschiedenen Modellen, die allein im Zuge des Moduls entstanden sind. Ferner wurde das Ergebnis der Gelelektrophorese unter UV-Licht sichtbar gemacht. Die Schüler konnten so erkennen, ob sie die experimentellen Schritte sauber durchgeführt hatten und ihre Hypothesen überprüfen, die sie in Bezug auf die Darstellung ihres genetischen Fingerabdrucks aufgestellt hatten.

3.4.3 Erhebungsinstrumente und Datenauswertung

Die statistische Auswertung der Ergebnisse erfolgte vorwiegend mit dem Programm *IBM SPSS Statistics Version 26* (IBM Corp. 2020a). Für Faktorenanalysen, wie sie in Teilarbeit B, Teilarbeit D und Teilarbeit F Verwendung fanden, wurde zunächst in Teilarbeit B das Statistik-Programm R Studio Team (2020), in Teilarbeit D *IBM SPSS Statistics Version 26* (IBM Corp. 2020a) und in Teilarbeit F *IBM SPSS Amos Version 26* (IBM Corp. 2020b) herangezogen. Die Daten waren nach Auswertung mit dem Shapiro-Wilk-Test über alle Teilarbeiten hinweg nicht normalverteilt ($p < .001$). Trotz der Stichprobengröße wurde die Annahme des zentralen Grenzwertsatzes abgelehnt, da zum einen die Stichprobengröße je nach Fragebogen und Gruppenvergleich stark schwankte und zum anderen die Messwerte aufgrund der Likert-Skalierung keine eindeutig interpretierbaren relativen Abstände aufwiesen (Bortz & Schuster 2010).

Entsprechend wurden die nichtparametrischen Verfahren Friedman-Test, Wilcoxon-Test und Mann-Whiteny-U-Test für die Inner- und Zwischengruppenvergleiche angewandt (Field 2012). Aufgrund von Mehrfachtestung wurde zudem die Signifikanzgrenze über eine Bonferroni-Korrektur auf $p < .017$ angehoben (ebd.). Zusätzlich wurden für signifikante Ergebnisse die entsprechende Effektstärke r (Lipsey & Wilson 2001) mit kleiner (> 0.1), mittlerer (> 0.3) und großer (> 0.5) Effektstärke berechnet. Für Korrelationsanalysen kamen Spearmans Rangkorrelationen mit den jeweiligen Spearmans-Rho-Werten zum Einsatz (Field 2012).

Die zur Auswertung relevanten Daten der Teilarbeiten wurden über Fragebogen erhoben, wobei die Beantwortung vieler Forschungsfragen das Austeilen der Fragebogen zu allen drei Testzeitpunkten erforderte. Entsprechend wurde die Reihenfolge der abgeprüften Items im Nach- und Behaltenstest verändert, sodass systematischen Antworten durch Fragereihenfolgeeffekten vorgebeugt werden konnte (Schnell et al. 2018). Modelle, zugehörige Modellskizzen sowie die jeweiligen Evaluationsblätter wurden nach jedem Kurstag abfotografiert und archiviert. Zudem konnte mit insgesamt 179 ausgefüllten Kursanleitungen eine weitere Datengrundlage für Teilarbeit E geschaffen werden.

In **Teilarbeit A** wurde relevantes Wissen über die molekularbiologischen Eigenschaften der DNA und ihrer Repräsentation als Modell mittels eines Wissensfragebogen abgefragt sowie eine Analyse der gebastelten Modelle und eine Modellevaluation vorgenommen, um Unterschiede zwischen den Schülergruppen mit nur einer oder zwei Modellevaluationsphasen zu erfassen. Teilarbeit A befasste sich mit dem Vergleich ausschließlich monolingual deutscher Module. Zur Analyse des Wissenszuwachses und des langfristigen Lernerfolgs wurden vollständige Datensätze herangezogen, weshalb ausschließlich Daten von Schülern berücksichtigt werden konnten, die zu allen drei Testzeitpunkten (T0, T1, T2) die Fragebögen ausgefüllt hatten. Insgesamt verblieben nach Ausschluss aller unvollständigen Datensätze 296 Schülerdaten (Mädchen 52.0%, Jungen 48.0%; $M_{Klassengröße} = 22.8$, $SD = 6.2$; $M_{Alter} = 14.6$, $SD = 0.8$).

Der Wissensfragebogen umfasste 30 selbsterstellte Multiple-Choice-Fragen zu den molekularbiologischen Eigenschaften der DNA (12 Fragen) und ihrer Repräsentation als Modell (18 Fragen), wobei es zu jeder Frage je vier Antwortmöglichkeiten gab, von denen aber nur eine richtig war. Die Fragen bezogen sich dabei auf verschiedene Inhalte der einzelnen Modulphasen. Inhaltsvalidität war gewährleistet, da die Items mit dem Lehrplan übereinstimmten. Im Hinblick auf die Konstruktvalidität bestätigten Inter-Item-Korrelationen unter .20 (T0 = .08, T1 = .19, T2 = .18), dass sich jedes Item auf einen anderen Teilbereich bezog. Die Heterogenität der einzelnen Fragen bei Überprüfung komplexer und abstrakter Konstrukte, wie beispielsweise der kognitiven Leistung, unterstützt zudem die Konstruktvalidität (Rost 2004). Cronbachs Alpha-Werte von .71 (T0), .64 (T1), und .70 (T2) ließen auf eine akzeptable innere Konsistenz schließen, da der Wert über dem generellen Limit von .70 lag. Nach Lienert & Raatz (1998) erlauben Werte zwischen .50 und .70 die Differenzierung von Gruppen. Die Itemschwierigkeiten (Prozentsatz der richtigen Antworten, Bortz & Döring 1995) lagen zwischen 7% (hohe Schwierigkeit) und 88% (niedrige Schwierigkeit). Während der Intervention invertierten sich die Itemschwierigkeiten von T0 zu T1 und nahmen generell ab.

Für die statistische Analyse wurden die Antworten zunächst binär in „1 = richtig " und „0 = falsch" kodiert. Danach erfolgte die Bildung individueller Summenwerte des molekularbiologischen und modellspezifischen Wissens zur DNA für die einzelnen Testzeitpunkte beider Schülergruppen. Der maximal erreichbare Summenwert lag entsprechend bei 30 Punkten. Darauf beruhend wurden anschließend der Wissenszuwachs (T1 minus T0) und die Behaltensleistung (T2 minus T0) berechnet sowie der tatsächliche Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T1-T0) x (T1/30)) und der langfristige Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T2-T0) x (T2/30)). Der Wissenszuwachs wurde also in Abhängigkeit vom tatsächlichen Wissen der Schüler gemessen, sodass ein Vergleich der Lernleistung möglich war, ungeachtet der erreichten Punktezahl (Scharfenberg et al. 2007). Zur Berechnung von Wissensunterschieden zwischen den drei Testzeitpunkten in beiden Schülergruppen wurden die nichtparametrischen Friedman- und Wilcoxon-Tests herangezogen. Zwischengruppenvergleiche erfolgten mittels Mann-Whitney-U-Test. Bei signifikanten *p*-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012). Korrelationsanalysen zwischen Biologienoten und den Ergebnissen des Nachtests (T1) wurden mit Spearman-Rho durchgeführt (ebd.).

Zur Analyse der ersten Modellevaluationsphase wurden die Modellskizzen der Schüler nach Langheinrich & Bogner (2015) bewertet. Hierfür wurden aus 150 Zeichnungen der Schülergruppe, die auch an der zweiten Modellevaluationsphase teilnahm, zufällig 26 für eine zweite Bewertung ausgewählt (17.3 %). Die Cohens-Kappa-Koeffizienten (Cohen 1968) dieser Analyse betrugen .88 und .82 für die Intra-Rater- und Inter-Rater-Reliabilität, was für eine sehr objektive Interpretation der Zeichnungen sprach (Wolf 1997). Ein Vergleich zwischen den Modellevaluationsgruppen ergab keinen signifikanten Unterschied (MWU: $Z = -.745$, $p = .456$). Eine Inhaltsanalyse (Bos & Tarnai 1999) half dabei, die Schülerantworten auf die offenen Fragen über iterative Kategorisierung auszuwerten. Für die erste Frage „Welche Merkmale des ursprünglichen DNA-Modells sind in eurem Modell vereinfacht?" konnten insgesamt vier verschiedene Kategorien identifiziert werden: Ebene der DNA, Stoffebene, Teilchenebene und

Strukturebene. Von 384 Antworten wurden 58 (15.1%) für eine zweite Bewertung ausgewählt. Die dabei errechneten Cohens-Kappa-Koeffizienten (Cohen 1968) von .98 und .78 für die Intra-Rater- und Inter-Rater-Reliabilität ergaben eine substanzielle bis herausragende Objektivität (Wolf 1997). Zur Auswertung der zweiten Frage „Erläutert, warum man unterschiedliche Modelle eines biologischen Originals (in unserem Fall die Struktur der DNA) erstellen kann?“ wurde das adaptierte Kategoriensystem von Mierdel & Bogner (2019) herangezogen, wobei fünf verschiedene Kategorien identifiziert werden konnten: Individualität der DNA, unterschiedliche Interpretation, unterschiedliches Modelldesign, unterschiedlicher Forschungsschwerpunkt und unterschiedlicher Forschungsstand. Aus insgesamt 150 Antworten wurden zufällig 27 (18.0%) ausgewählt und die Cohens-Kappa-Koeffizienten (Cohens 1968) berechnet. Mit Werten zwischen .75 und .70 für die Intra-Rater- und Inter-Rater-Reliabilität konnte insgesamt eine akzeptable Objektivität erzielt werden (Wolf 1997). Um Voreingenommenheit in der Bewertung vorzubeugen, wurden die zwei Modellevaluationsgruppen bezüglich der Kategorienhäufigkeiten der Antworten auf beide offene Fragen verglichen. Es konnten keine signifikanten Abhängigkeiten festgestellt werden (adjustiertes Pearsons $C \leq .192$; $p \geq .065$; Pearson 1900).

Die zweite Modellevaluationsphase wurde über ein dreistufiges Verfahren ausgewertet. Zunächst erfolgte die Dokumentation der Eigenbewertung der Schülermodelle. Dabei ergab jedes gesetzte Kreuz auf den Evaluationsblättern der Schüler einen Punkt. Die maximale Punktzahl lag bei 14 Punkten. Im Anschluss wurde eine Auswertung dieser Eigenbewertung vorgenommen. Für jedes gesetzte Kreuz, das tatsächlich mit einem im Modell berücksichtigten Bestandteil der DNA übereinstimmte, gab es einen Punkt. Wenn die Schüler alle Kreuze richtig gesetzt hatten, erreichten sie die auf ihren Evaluationsblättern angegebene Punktezahl. Zusätzlich wurden die Modelle losgelöst von den Evaluationsblättern bewertet. Für korrekt dargestellte Merkmale der DNA gab es jeweils einen Punkt, unabhängig davon, ob Schüler diese bei ihrer Selbstevaluation identifiziert hatten. Auch hier lag die maximale Punktezahl bei 14 Punkten. Ein anschließender Vergleich zwischen der Eigenbewertung der Schüler, der Auswertung der Eigenbewertung und der Modellbewertung ermöglichte eine Einschätzung der Fähigkeiten von Schülern, selbstreflexiv ihre Ergebnisse beurteilen zu können. Niedrigere Punktezahlen bei der Eigenbewertung als der Modellbewertung deuteten häufig auf Schwierigkeiten der Schüler hin, die Qualität ihrer Modelle richtig einzuschätzen. Vergleichsweise niedrigere Punktezahlen bei der Modellbewertung als bei der Eigenbewertung verwiesen wiederum darauf, dass Schüler Merkmale dokumentiert hatten, die im selbstgebauten Modell nicht vorhanden waren. Höhere

Punktezahlen bedeuteten hingegen, dass Schüler Merkmale in ihrem Modell aufgeführt hatten, die sie aber nicht als solche erkannt und bei der Eigenbewertung dokumentiert hatten.

Bei **Teilarbeit B** handelte es sich um eine Pilotstudie. Da in der Literatur häufig Zusammenhänge zwischen Kreativität und Persönlichkeit postuliert werden (Kaufman et al. 2008, Barron & Harrington 1981), wurde ein Fragebogen zusammengestellt, der eine auf 15 Fragen gekürzte Version des *Cognitive Processes Associated with Creativity* (*CPAC*) Fragebogens von Miller (2014) umfasste sowie den aus zehn Items bestehenden *Big Five Inventory 10 (BFI-10)* Fragebogen (Rammstedt & John 2007), der aus dem Persönlichkeitsmessinstrument Big Five von Caprara (1993) abgeleitet wurde. So sollte der wechselseitige Einfluss von Kreativität und individuellen Persönlichkeitsfaktoren der Schüler auf kreatives Empfinden und Handeln untersucht werden (Kaufman et al. 2008). Der *BFI-10* umfasste dabei je zwei Items für jede im Big Five (Caprara 1993) aufgeführte Persönlichkeitskategorie, wobei jeweils ein Item revers kodiert war: Extraversion (Items 6, 36), Verträglichkeit (Items 2, 22), Gewissenhaftigkeit (Items 3, 23), Neurotizismus (Items 9, 39) und Offenheit (Items 20, 41). Jedes Item konnte über eine fünfstufige Likert-Skala von "stimme überhaupt nicht zu" (1) bis "stimme voll zu" (5) bewertet werden (Rammstedt & John 2007). Dieselbe Likert-Skalierung wurde für die Einschätzung der *CPAC*-Items verwendet. Diese umfasste die unterschiedlichen Kreativitätsdomänen Ideenmanipulation (Items 7, 15), Bildhaftigkeit/Sensorik (Items 1, 10), *Flow* (Items 2, 3, 9, 13, 14), Metaphorisches/Analoges Denken (Item 5) und Ideengenerierung (Items 4, 6, 8, 11, 12). Inter-Item-Korrelationen unter .20 sowie die Heterogenität der einzelnen Fragen in Bezug auf das latente Konstrukt Kreativität bestätigten die Konstruktvalidität des *CPAC* (Rost 2004). Die Items aus *CPAC* und *BFI-10* addiert, bestand der Fragebogen nun aus insgesamt 25 Items und wurde zu einem Testzeitpunkt abgefragt. Entsprechend konnten für die statistische Auswertung beinahe alle 538 ausgefüllten Fragebögen berücksichtigt werden (Mädchen 65,4 %, Jungen 34,6 %; M_{Alter} =16,96; SD = 2,99). Zusätzlich wurden die demografischen Variablen Alter und Geschlecht erfragt.

Zur Verifizierung der latenten Variablen *Flow* und *Act* (übergreifende Variable für beobachtbare kreative Handlungen) des gekürzten *CPAC* Fragebogens, erfolgte eine konfirmatorische Faktorenanalyse mittels R Studio Team (2020). So konnte gezeigt werden, dass die vorhergesagte Zwei-Faktoren-Struktur aus früheren Befragungen (Conradty & Bogner 2018, Conradty & Bogner 2019) replizierbar war (Thompson 2004). Entsprechend berechnete Kriterien, die eine Einschätzung über die Qualität des Strukturgleichungsmodells erlauben, bestätigten, dass die Zwei-Faktoren-Modellstruktur zur Erfassung des latenten Konstrukts Kreativität geeignet waren. So zeigten der Comparative Fit Index (CFI = .837), der Tucker-Lewis Index (TLI = .808) sowie der Root Mean Square Error of Approximation (RMSEA < .05, p = .492), dass das Modell akzeptabel war. Das Ladungsmuster der jeweiligen Faktoren bestätigte zusätzlich die Modellstruktur.

Die Analyse des *BFI-10* erfolgte über eine Hauptkomponentenanalyse mit Oblimin zur Dimensionsreduktion der Daten unter Beibehaltung ihrer natürlichen Varianz (Bro & Smilde 2014). Dabei entsprachen die Faktorladungen und die Mustermatrix den fünf latenten Variablen des Persönlichkeitstests. Dieses dem Kaiser-Guttman-Kriterium (Kaiser 1970) entsprechende Fünf-Faktor-Modell erreichte einen Kaiser-Meyer-Olkin-Wert (KMO = .54, χ^2 = 484,1), der gerade noch akzeptabel war.

Mögliche Geschlechterunterschiede, besonders in Bezug auf Extremgruppenvergleiche nach Alter der Schüler, wurden mit dem Mann-Whitney-U-Test (Field 2012) untersucht. Vor Durchführung einer anschließenden Korrelationsanalyse erfolgte zudem die zufällige Auswahl einer Stichprobe von 340 Teilnehmern, um möglichen Verzerrungen der Ergebnisse durch Geschlechterverteilung und Altersgruppen vorzubeugen. Bei signifikanten Ergebnissen wurden die Effektgröße r (Lipsey & Wilson 2001) mit kleinen (> 0,1), mittleren (> 0,3) und großen (> 0,5) Effektstärken berechnet. Die Erstellung entsprechender Boxplots wurde mit dem Programm *IBM SPSS Statistics Version 26* (IBM Corp. 2020a) vorgenommen.

In **Teilarbeit C** wurde ähnlich wie in Teilarbeit A relevantes Wissen mittels eines Wissensfragebogen über die molekularbiologischen Eigenschaften der DNA abgefragt, um Unterschiede zwischen den Schülergruppen mit muttersprachlich deutschem (Gruppe mit zwei Modellevaluationen) oder CLIL englischem Modul zu erfassen. Zur Analyse des Wissenszuwachses und des langfristigen Lernerfolgs wurden nur vollständige Datensätze herangezogen, weshalb ausschließlich Schülerdaten berücksichtigt werden konnten, die zu allen drei Testzeitpunkten (T0, T1, T2) die Fragebögen ausgefüllt hatten. Insgesamt verblieben nach Ausschluss aller unvollständigen Datensätze 252 Schülerdaten (Mädchen 52.4%, Jungen 47.6%; M_{Alter} = 14.6, SD = 0.7).

Der Wissensfragebogen umfasste 30 selbsterstellte Multiple-Choice-Fragen zu diversen molekularbiologischen sowie struktur- und modellbezogenen Eigenschaften der DNA, wobei es zu jeder Frage je vier Antwortmöglichkeiten gab, von denen aber nur eine richtig war. Die Fragen bezogen sich dabei auf verschiedene Inhalte der einzelnen Modulphasen. Inhaltsvalidität war gewährleistet, da die Items mit dem Lehrplan übereinstimmten. Im Hinblick auf die Konstruktvalidität bestätigten Inter-Item-Korrelationen unter .20 (T0 = .08, T1 = .19, T2 = .18), dass sich jedes Item auf einen anderen Teilbereich bezog. Die Heterogenität der einzelnen Fragen bei Überprüfung komplexer und abstrakter Konstrukte, wie beispielsweise der kognitiven Leistung, unterstützt zudem die Konstruktvalidität (Rost 2004). Cronbachs Alpha-Werte von .75 (T0), .76 (T1) und .78 (T2) lassen auf eine sehr gute innere Konsistenz schließen (Lienert & Raatz 1998). Die Itemschwierigkeiten (Prozentsatz der richtigen Antworten, Bortz & Döring 1995) lagen zwischen 5% (hohe Schwierigkeit) und 90% (niedrige Schwierigkeit). Während der Modulteilnahme wurde anhand der Verschiebung der Itemschwierigkeiten von T0 zu T1 ein erster Lernerfolg ersichtlich.

Für die statistische Analyse wurden die Antworten zunächst binär in „1 = richtig " und „0 = falsch" kodiert. Danach erfolgte die Bildung von Summenwerten des Wissens zur DNA für die einzelnen Testzeitpunkte beider Schülergruppen. Der maximal erreichbare Summenwert lag entsprechend bei 30 Punkten. Darauf beruhend wurden anschließend der Wissenszuwachs (T1 minus T0) und die Behaltensleistung (T2 minus T0) berechnet sowie der tatsächliche Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T1-T0) x (T1/30)) und der langfristige Lernerfolg ((T2-T0) x (T2/30)) (Scharfenberg et al. 2007). Der Wissenszuwachs wurde also in Abhängigkeit vom tatsächlichen Wissen der Schüler gemessen, sodass ein Vergleich der Lernleistung ungeachtet der erreichten Punktezahl möglich war. Zur Berechnung von Wissensunterschieden zwischen den drei Testzeitpunkten in beiden Schülergruppen wurden die nichtparametrischen Friedman- und Wilcoxon-Tests herangezogen. Zwischengruppenvergleiche erfolgten mittels Mann-Whitney-U-Test. Bei signifikanten p-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012). Im Falle signifikanter Ergebnisse wurden Effektgrößen r (Lipsey & Wilson 2001) mit kleinen (> 0,1), mittleren (> 0,3) und großen (> 0,5) Effektstärke berechnet.

In **Teilarbeit D** wurde ähnlich wie in Teilarbeit A Wissen über die Repräsentation der DNA als Modell (18 Fragen) abgefragt, eine Analyse der Modelle und Modellevaluation vorgenommen sowie das Modellverständnis nach Treagust et al. (2002) abgefragt, um Unterschiede zwischen den Schülergruppen mit muttersprachlich deutschem oder CLIL englischem Modul zu erfassen. Zur Analyse des Wissenszuwachses und des langfristigen Lernerfolgs sowie des Modellverständnisses wurden nur vollständige Datensätze herangezogen, weshalb ausschließlich Schülerdaten berücksichtigt werden konnten, die zu allen drei Testzeitpunkten (T0, T1, T2) die Fragebögen ausgefüllt hatten. Insgesamt wurden 465 Schülerdatensätze (Mädchen 50.8%, Jungen 49.2%; $M_{Klassengröße} = 21.9$, $SD = 4.6$; $M_{Alter} = 14.7$, $SD = 0.7$) verwendet.

Der Wissensfragebogen umfasste insgesamt 30 selbsterstellte Multiple-Choice-Fragen, von denen aber nur 18 Fragen relevant waren, um die Repräsentation der DNA als Modell abzufragen. Zu jeder Frage gab es je vier Antwortmöglichkeiten, von denen aber nur eine richtig war. Die Fragen bezogen sich dabei besonders auf die Modellphase des Moduls. Inhaltsvalidität war gewährleistet, da die Items mit dem Lehrplan übereinstimmten. Im Hinblick auf die Konstruktvalidität bestätigten Inter-Item-Korrelationen unter .20, dass sich jedes Item auf einen anderen Teilbereich bezog. Die Heterogenität der einzelnen Fragen bei Überprüfung komplexer und abstrakter Konstrukte, wie beispielsweise der kognitiven Leistung, unterstützt zudem die Konstruktvalidität (Rost 2004). Cronbachs Alpha-Werte von .70 (T0), .71 (T1) und .69 (T2) ließen auf eine akzeptable innere Konsistenz schließen (Lienert & Raatz 1998). Für die statistische Analyse wurden die Antworten zunächst binär in „1 = richtig " und „0 = falsch" kodiert. Danach erfolgte die Bildung individueller Summenwerte des modellspezifischen Wissens zur DNA für die einzelnen Testzeitpunkte beider Schülergruppen. Der maximal erreichbare Summenwert lag entsprechend bei 18 Punkten. Darauf beruhend wurden anschließend der Wissenszuwachs (T1 minus T0) und die Behaltensleistung (T2 minus T0) berechnet sowie der tatsächliche Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T1-T0) x (T1/18)) und der langfristige Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T2-T0) x (T2/18)). Der Wissenszuwachs wurde also in Abhängigkeit vom tatsächlichen Wissen der Schüler gemessen, sodass ein Vergleich der Lernleistung möglich war, ungeachtet der erreichten Punktezahl (Scharfenberg et al. 2007). Zur Berechnung von Wissensunterschieden zwischen den drei Testzeitpunkten in beiden Schülergruppen wurden die nichtparametrischen Friedman- und Wilcoxon-Tests herangezogen. Zwischengruppenvergleiche erfolgten mittels Mann-Whitney-U-Test. Bei signifikanten p-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012).

Das Modellverständnis der Schüler wurde über eine verkürzte Version des *Students' Understanding of Models Fragebogen (SUMS)* (Treagust et al. 2002) abgeprüft, wobei der Fokus vornehmlich auf den Subskalen „Models as Exact Replicas (ER)" und „Changing Nature of Models (CNM)", mit je vier und drei Items, lag. Zur Beantwortung des Fragebogens konnten die Schüler auf einer fünfstufigen Likert-Skala zwischen „lehne ab (1)" bis „stimme zu (2)" auswählen. Der Fokus auf die Subskalen ER und CNM erfolgte in Einklang mit gesetzten Schwerpunkten in den Modellphasen des Moduls (Treagust et al. 2002, Mierdel & Bogner 2019b). Bei ER handelt es sich um einen konzeptionellen Fehler, der bei Schülern aller Jahrgangsstufen fest verankert ist (Ke et al. 2020, Schwarz et al. 2009). So waren besonders niedrige Werte bei ER und besonders hohe Werte bei CNM, das auf Modellverständnis hinweist, auf der

Likert-Skala wünschenswert (Treagust et al. 2002). Cronbachs Alpha-Werte von über .70 (T0 = .81, T1 = .76, T2 = .70) ließen auf eine gute interne Konsistenz schließen (Lienert & Raatz 1998). Die statistische Auswertung des *SUMS* erfolgte in zwei Schritten: die Bestätigung der beiden Faktoren ER und CNM sowie die Veränderungen der beiden Faktoren über die drei Testzeitpunkte und zwischen den Schülergruppen. Die Hauptkomponentenanalyse mit Varimax über *IBM SPSS Statistics Version 26* (IBM Corp. 2020a) zur Dimensionsreduktion unter Beibehaltung der natürlichen Varianz (Bro & Smilde 2014) produzierte die erwartete Zwei-Faktor-Lösung mit einem akzeptablen bis guten Kaiser-Meyer-Olkin-Wert (KMO = .80, $\chi 2$ = 599,2). Die anschließende Berechnung von Unterschieden im Modellverständnis zwischen den drei Testzeitpunkten in beiden Schülergruppen erfolgte mittels der nichtparametrischen Friedman und Wilcoxon Tests. Zwischengruppenvergleiche wurden über den Mann-Whitney-U-Test vorgenommen. Bei signifikanten p-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012).

Zur Analyse der ersten Modellevaluationsphase wurden die Modellskizzen der Schüler nach einem von Langheinrich & Bogner (2015) geänderten Schema bewertet. Hierfür wurden aus 231 Zeichnungen der Schülergruppe, die am bilingual englischen Modul teilnahm, zufällig 38 für eine zweite Bewertung ausgewählt (16.4 %). Mithilfe der Inhaltsanalyse (Bos & Tarnai 1999) wurden Schülerantworten auf die offenen Fragen über iterative Kategorisierung ausgewertet. Für die erste Frage „Welche Merkmale des ursprünglichen DNA-Modells sind in eurem Modell vereinfacht?“ konnten insgesamt vier verschiedene Kategorien identifiziert werden: Ebene der DNA, Stoffebene, Teilchenebene und Strukturebene. Von 474 Antworten wurden 76 (16.0%) für eine zweite Bewertung ausgewählt. Die dabei errechneten Cohens-Kappa-Koeffizienten (Cohen 1968) von .97 und .84 für die Intra-Rater- und Inter-Rater-Reliabilität ergab eine sehr gute bis herausragende Objektivität (Wolf 1997). Zur Auswertung der zweiten Frage „Erläutert, warum man unterschiedliche Modelle eines biologischen Originals (in unserem Fall die Struktur der DNA) erstellen kann?“ wurde das adaptierte Kategoriensystem von Mierdel & Bogner (2019) herangezogen, wobei fünf verschiedene Kategorien identifiziert werden konnten: Individualität der DNA, unterschiedliche Interpretation, unterschiedliches Modelldesign, unterschiedlicher Forschungsschwerpunkt und unterschiedlicher Forschungsstand. Aus insgesamt 416 Antworten wurden zufällig 61 (14.7%) ausgewählt und die Cohens-Kappa-Koeffizienten (Cohens 1968) berechnet. Mit Werten zwischen .89 und .72 für die Intra-Rater- und Inter-Rater-Reliabilität konnte insgesamt eine sehr gute Objektivität erzielt werden (Wolf 1997).

Die zweite Modellevaluationsphase wurde über ein dreistufiges Verfahren ausgewertet. Zunächst erfolgte die Dokumentation der Eigenbewertung der Schülermodelle. Dabei ergab jedes Kästchen, das die Schüler auf ihren Evaluationsblättern angekreuzt hatten, einen Punkt. Die maximale Punktzahl lag bei 14 Punkten. Im Anschluss wurde eine Auswertung dieser Eigenbewertung vorgenommen. Für jedes gesetzte Kreuz, das tatsächlich mit einem im Modell berücksichtigten Bestandteil der DNA übereinstimmte, gab es einen Punkt. Wenn die Schüler alle Kreuze richtig gesetzt hatten, erreichten sie die auf ihren Evaluationsblättern angegebene Punktezahl. Zusätzlich wurden die Modelle losgelöst von den Evaluationsblättern bewertet. Für korrekt dargestellte Bestandteile der DNA gab es jeweils einen Punkt, unabhängig davon, ob Schüler diese selbst identifiziert hatten. Auch hier lag die maximale Punktezahl bei 14 Punkten. Ein anschließender Vergleich zwischen der Eigenbewertung der Schüler, der Auswertung der Eigenbewertung und der Modellbewertung ermöglichte eine Einschätzung der Fähigkeiten von Schülern, selbstreflexiv ihre Ergebnisse beurteilen zu können. Niedrigere Punktezahlen bei der Eigenbewertung als der Modellbewertung deuteten häufig auf Schwierigkeiten der Schüler hin, die Qualität ihrer Modelle richtig einzuschätzen. Niedrigere Punktezahlen bei der Modellbewertung im Vergleich zur Eigenbewertung verwiesen wiederum darauf, dass Schüler Bestandteile der DNA dokumentiert hatten, die im selbstgebauten Modell nicht vorhanden waren. Höhere Punktezahlen bedeuteten hingegen, dass Schüler Bestandteile der DNA in ihrem Modell aufführten, die sie aber nicht als solche erkannt und bei der Eigenbewertung dokumentiert hatten. Für nachfolgende Kontingenzanalysen wurde der angepasste Pearsons Kontingenzkoeffizient (C; Pearson 1900) berechnet. Da Pearsons C ein Mitglied der r-Effektstärkenfamilie ist (z. B. Ellis 2010), wird er auch als Effektstärkenwert behandelt.

In **Teilarbeit E** wurde sprachliches Wissen während der Teilnahme am bilingual englischen Modul über einen Cloze-Test erfasst. Zur Analyse des Wissenszuwachses und des langfristigen Lernerfolgs wurden ausschließlich vollständige Datensätze herangezogen, weshalb nur Schülerdaten berücksichtigt werden konnten, die zu allen drei Testzeitpunkten (T0, T1, T2) die Fragebögen ausgefüllt hatten. Insgesamt verblieben nach Ausschluss aller unvollständigen Datensätze 179 Schülerdaten (Mädchen 48.5%, Jungen 51.5%; $M_{Alter} = 14.5$, $SD = 0.6$).

Der selbst erstellte Cloze-Test bestand aus Sätzen, bei denen Wörter systematisch entfernt wurden und Schüler diesen Lückentext mit den richtigen Wörtern aus einer Auswahl von vier möglichen Optionen füllen mussten. Dadurch sollen die natürlichen linguistischen Redundanzen reduziert werden und die Schüler anhand des Aufbaus die Bedeutung des Textes schlussfolgern. Je mehr sprachliche Kompetenzen die Schüler haben, desto eher gelingt es ihnen, die Lücken mit den fehlenden Wörtern zu füllen (Ajideh & Mozaffarzadeh, 2012). Der eingesetzte Cloze-Test umfasste 14 Multiple-Choice-Items, deren inhaltliche Validität durch Übereinstimmung mit dem bayerischen Lehrplan gewährleistet war. Ebenso konnte die Konstruktvalidität bestätigt werden, da die Items zur Beschreibung des sehr komplexen Konstruktes „Spracherwerb" heterogen sein müssen (Rost 2004). Cronbachs Alpha-Werte von .61 (T0), .63 (T1), und .66 (T2) ließen auf eine akzeptable innere Konsistenz schließen, da Werte zwischen .50 und .70 die Differenzierung von Gruppen erlauben Werte (Lienert & Raatz 1998). Die Itemschwierigkeiten (Prozentsatz der richtigen Antworten, Bortz & Döring 1995) lagen zwischen 20 % (höhere Schwierigkeit) und 73 % (niedrigere Schwierigkeit). Während der CLIL-Intervention verbesserten sich die Itemschwierigkeiten geringfügig von T0 zu T1. Für die statistische Analyse wurden die Antworten zunächst binär in „1 = richtig " und „0 = falsch" kodiert. Danach erfolgte die Bildung individueller Summenwerte des sprachlichen Wissens für die einzelnen Testzeitpunkte. Der maximal erreichbare Summenwert lag entsprechend bei 14 Punkten. Darauf beruhend wurden anschließend der Wissenszuwachs (T1 minus T0) und die Behaltensleistung (T2 minus T0) berechnet sowie der tatsächliche Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T1-T0) x (T1/14)) und der langfristige Lernerfolg in Relation zur maximal erreichbaren Punktezahl ((T2-T0) x (T2/14)) (Scharfenberg et al. 2007). Zur Berechnung von Wissensunterschieden zwischen den drei Testzeitpunkten in beiden Schülergruppen wurden die nicht parametrischen Friedman- und Wilcoxon-Tests herangezogen. Zwischengruppenvergleiche in Bezug auf Geschlechterunterschiede erfolgten mittels Mann-Whitney-U-Test. Bei signifikanten *p*-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012). Korrelationen zwischen den Englischnoten und Nachtestergebnissen (T1) wurden mit Spearman-Rho bestimmt (ebd.).

Weitere Analysen sprachlicher Fähigkeiten erfolgten über die Auswertung der Kurshefte. Hierfür wurden Mithilfe der Inhaltsanalyse (Bos & Tarnai 1999) Schülerantworten auf die offenen Fragen über iterative Kategorisierung aufbereitet. Anschließend wurden die Antworten in einemdreistufigen Verfahren weiter ausgewertet. Die Verwendung der englischen Sprache ergab je zwei Punkte, Code-Switching auf Deutsch einen Punkt und nicht Beantworten der Frage null Punkte. Grammatik- oder Rechtschreibfehler wurden nicht berücksichtigt, solange die Aussage eindeutig interpretierbar war. Anschließend erfolgte eine unabhängige Bewertung des Inhalts der Schülerantworten. Für jede richtige Teilantwort gab es je einen Punkt, was zu variierenden Maximalpunktezahlen pro Frage zwischen zwei und sechs Punkten führte. Danach wurden zufällig 52 von 179 Schülerantworten für eine zweite Bewertung ausgewählt und die Cohens-Kappa-Koeffizienten (Cohens 1968) berechnet. Mit Werten zwischen .87 und

.69 für die Intra-Rater- und Inter-Rater-Reliabilität konnte insgesamt eine akzeptable bis sehr gute Objektivität erzielt werden (Wolf 1997). Danach wurden die Punkte für Sprache und Inhalt addiert, um den individuellen Gesamtpunktestand der Schüler zu ermitteln. So konnte das inhaltliche Verständnis und Sprachverständnis der Schüler während des bilingual englischen Moduls analysiert werden. Niedrigere Punktezahlen deuteten häufig auf Schwierigkeiten der Schüler hin, die Inhalte des Moduls richtig zu verstehen und zu kommunizieren. Hohe Punktezahlen hingegen verwiesen darauf, dass Schüler sowohl die Inhalte verstanden als auch fachgerecht kommuniziert hatten. Zwischengruppenvergleiche in Bezug auf Geschlechterunterschiede erfolgten mittels Mann-Whitney-U-Test. Bei signifikanten p-Werten hob die Bonferroni-Korrektur die Signifikanzgrenze an (Field 2012). Im Falle signifikanter Ergebnisse wurden Effektgrößen r (Lipsey & Wilson 2001) mit kleinen (> 0,1), mittleren (> 0,3) und großen (> 0,5) Effektstärken berechnet.

In **Teilarbeit F** wurden die in der Literatur häufig postulierten Zusammenhänge zwischen Kreativität und Selbstkonzept (Fleith et al. 2002, Bournelli et al. 2009) sowie Selbstkonzept und Leistung (Marsh & Martin 2011, Jansen et al. 2015) untersucht. Entsprechend wurde ein Fragebogen zusammengestellt, der eine auf 15 Fragen gekürzte Version des *Cognitive Processes Associated with Creativity (CPAC)* Fragebogens von Miller (2014) umfasste sowie die aus 22 Items bestehenden *Skalen zur Erfassung des schulischen Selbstkonzepts* (*SESSKO*; Schöne et al. 2002). So sollte der wechselseitige Einfluss von Kreativität und Selbstkonzept der Schüler auf kreatives Handeln und Leistungsfähigkeit untersucht werden (Bournelli et al. 2009, Jansen et al. 2015). Die *SESSKO* umfassten dabei vier latente Variablen – *Kriterial, Individuell, Sozial* und *Absolut* – die jeweils zur Analyse des Selbstkonzepts in Biologie und Englisch Verwendung fanden. Zur Beantwortung des Fragebogens konnten die Schüler auf einer fünfstufigen Likert-Skala zwischen verschiedenen, auf die latenten Variablen bezogenen Extremen auswählen. Inhaltsvalidität wurde über das Erfüllen von faktoranalytisch extrahierten Referenznormen gewährleistet (Schöne et al. 2002). Cronbachs Alpha-Werte von .80 (*SESSKO* Biologie) und .88 (*SESSKO* Englisch) erlauben nach Lienert & Raatz (1998) eine sehr gute Differenzierung von Gruppen.

Die *Skala Cognitive Processes Associated with Creativity (CPAC)* von Miller (2014) hat sich als besonders geeignet erwiesen, um die sechs Dimensionen der Kreativität - Ideenmanipulation (Items 7, 15), Bildhaftigkeit/Sensorik (Items 1, 10), *Flow* (Items 2, 3, 9, 13, 14), Metaphorisches/Analoges Denken (Item 5) und Ideengenese (Items 4, 6, 8, 11, 12) - mit insgesamt 28 Items zu erfassen (Tsai 2018). Da die Verwendung einer Skala mit 28 Items jedoch mehr Zeit in Anspruch nehmen würde, als für die Durchführung solcher Tests im naturwissenschaftlichen Unterricht zur Verfügung steht, wurde von Conradty & Bogner (2018) ein verkürzter Fragebogen mit nur 15 Items vorgeschlagen. Dieser umfasst noch immer alle relevanten Dimensionen der erweiterten Variante und wurde bereits im naturwissenschaftlichen Unterricht validiert (Mierdel & Bogner 2019b). Items konnten über eine fünfstufige Likert-Skala von "stimme überhaupt nicht zu" (1) bis "stimme voll zu" (5) bewertet werden (Miller 2014). Inter-Item-Korrelationen unter .20 sowie die Heterogenität der einzelnen Fragen in Bezug auf das latente Konstrukt Kreativität bestätigten die Konstruktvalidität des *CPAC* (Rost 2004). Der nun aus insgesamt 37 Items bestehende Fragebogen wurde zu allen drei Testzeitpunkten (T0, T1, T2) erfasst. Dies ließ jedoch nur Fragebögen von Schülern zu, die zu allen drei Testzeitpunkten die Fragebögen ausgefüllt hatten (252, Mädchen 48.5%, Jungen 51.5%; $SD_{Geschlecht} = 6.2$; $M_{Alter} = 14.5$, $SD_{Alter} = 0.6$).

Zur Verifizierung der latenten Variablen *Flow* und *Act* (übergreifender Begriff für beobachtbare Variablen kreativen Handelns) des gekürzten *CPAC* Fragebogens in Kombination mit den jeweiligen Selbstkonzept Subskalen erfolgte eine konfirmatorische Faktorenanalyse mittels *IBM SPSS Amos Version 26* (IBM Corp. 2020b). Dabei konnte gezeigt werden, dass die vorhergesagte Zwei-Faktoren-Struktur aus früheren Befragungen (Conradty & Bogner 2018, Conradty & Bogner 2019) replizierbar war (Thompson 2004). Entsprechend berechnete Kriterien, die eine Einschätzung über die Qualität des Strukturgleichungsmodells erlauben, bestätigten, dass die Zwei-Faktoren-Modellstruktur zur Erfassung des latenten Konstrukts Kreativität (Kaiser 1970, Conradty & Bogner 2018) und Vier-Faktoren-Modellstruktur zur Erfassung des jeweiligen Selbstkonzepts (Schöne et al. 2002) in Biologie und Englisch geeignet waren. So zeigten der Comparative Fit Index ($CFI_{Biologie} = .943$; $CFI_{Englisch} = .943$), der Tucker-Lewis Index ($TLI_{Biologie} = .936$; $TLI_{Englisch} = .808$) sowie der Root Mean Square Error of Approximation ($RMSEA_{Biologie} < .05$, p = .110; $RMSEA_{Biologie} < .05$, p = .006) und *CMIN/DF* < 5.00 (Schermelleh-Engel et al. 2003), dass das Modell akzeptabel war. Da jedoch der χ^2-Wert sowohl für Biologie ($\chi^2 = 1425.59$, $df = 587$, $p < .001$) und Englisch ($\chi^2 = 1513.42$, $df = 587$, $p < .001$) signifikant war, musste eine zusätzliche Power Analyse vorgenommen werden, um die Modellstruktur zu verifizieren (MacCallum et al. 2006). Werte über $\lambda = .84$ für das Selbstkonzept in Biologie und $\lambda = .96$ für das Selbstkonzept in Englisch zeigten jedoch, dass das Modell akzeptabel war (Cohen 1992). Zwischengruppenunterschiede bezüglich Jungen und Mädchen wurden über den Mann-Whitney-U-Tests (MWU) bestimmt. Zusätzlich erfolgte eine Bonferroni-Korrektur $p < .017$ für signifikante Ergebnisse sowie eine Bestimmung der Effektgröße r (Lipsey & Wilson 2001) mit kleinen (> 0,1), mittleren (> 0,3) und großen (> 0,5) Effektstärken. Für

Korrelationsanalysen wurden Spearmans Rangkorrelationen sowie entsprechende Spearmans Rho-Werte herangezogen (Field 2012).

3.5 Ergebnisse und Diskussion

Grundsätzlich fokussierte sich die Gesamtstudie auf die Auswahl eines geeigneten Moduls und passender Messinstrumente, um ein auf Content and Lanugage Integrated Learning beruhendes Modul erfolgreich zu implementieren und Effekte des kombinierten Sprachen- und Inhaltlernens zu messen. In der interaktiven und authentischen Lernumgebung des Schülerlabors sollten die Schüler über Hands-on-Minds-on Experimente und wissenschaftliches Modellieren inhaltliches Wissen zur DNA-Struktur und parallel Scientific Literacy in Englisch und Deutsch aufbauen.

Entsprechend befasste sich **Teilarbeit A** mit der Auswahl eines Moduls, das entweder eine oder zwei Modellevaluationsphasen umfasste. Ein Fragebogen zur Erfassung von Kreativität und Persönlichkeit wurde in **Teilarbeit B** pilotiert. **Teilarbeit C** war die erste Studie im Rahmen des CLIL-Moduls und verglich den Lernerfolg des monolingual deutschen Moduls mit dem CLIL-Modul. Spezifisches Modellwissen, Modellverständnis, sowie die Ergebnisse des Modellierens und der Modellevaluationsphasen wurde in **Teilarbeit D** erfasst und mit der Schülergruppe des monolingual deutschen Moduls verglichen. **Teilarbeit E** beleuchtete den Aufbau von Scientific Literacy über die Auswertung des Cloze-Tests sowie der offenen Fragen im Kursheft. In **Teilarbeit F** wurde der wechselseitige Einfluss von Lernerfolg, Selbstkonzept und Kreativität genauer betrachtet.

3.5.1 Teilarbeit A

Im Rahmen der Teilarbeit A wurden mögliche Unterschiede zwischen den Schülergruppen untersucht, die an dem muttersprachlich deutschen Modul mit nur einer und mit zwei Modellevaluationsphasen teilnahmen. Die Ergebnisse der Teilarbeit deuteten darauf hin, dass die zweite Modellevaluationsphase den kurzfristigen und langfristigen Lernerfolg in Bezug auf molekularbiologisches und modellbezogenes Wissen zur DNA erhöhte.

Dies zeigt, dass wiederholtes Reflektieren über die Qualität des Modells den Prozess des Erkenntnisgewinns unterstützt. Nach Ainsworth (2008) kann die Erstellung eines Modelles folgende Auswirkungen auf das Lernen haben: (1) ermöglicht den Schülern, inhaltliches Wissen zu abstrahieren und in verschiedene Darstellungsformen zu übertragen, (2) fördert die Anknüpfung an Vorwissen und (3) hilft Schülern, verschiedene Aspekte des beobachteten

Phänomens angemessen darzustellen. Insgesamt kann das Erstellen von eigenen Modellen so dem Aufbau von Tiefenverständnis dienen (Oh und Oh 2011, Ke et al. 2020).

Die Ergebnisse dieser Teilarbeit stimmen mit den Ergebnissen einschlägiger Referenzstudien überein. So haben bereits Mierdel & Bogner (2019a) festgestellt, dass die Evaluation von Modellen in Kombination mit praktischem Experimentieren und der Konzeptionierung eigener Modelle den Lernerfolg bei molekulargenetischen Inhalten fördert. Das Aufbereiten und Argumentieren komplexer naturwissenschaftlicher Themen als Modelle im Biologieunterricht scheint also effektiver zu sein als klassische Lehrmethoden (Peel et al. 2019). Gleichzeitig weisen Mierdel & Bogner (2019a) darauf hin, dass eine weitere Modellevaluation zu noch besseren Lernerfolgen und Modellverständnis führen könnte, weshalb das Modul in Teilarbeit A eine zweite Modellevaluationsphase berücksichtigte. Es konnte eindeutig festgestellt werden, dass der Lernerfolg kurzfristig und langfristig höher war als bei der Teilnahme an dem Modul mit nur einer Modellevaluationsphase. Bei beiden Modulen entwickelten die Schüler zunächst ein mentales Modell, woraufhin sie entweder gleich mit dem Bau eines dreidimensionalen Modells begannen oder zunächst eine grobe Skizze ihres Vorhabens malten. In einer ersten Modellevaluationsphase fertigten die Schüler eine detaillierte und beschriftete Skizze ihres Modells an und werteten in Iteration mit dem Text ihr Modell aus. Zudem konnten sie innerhalb der Gruppe oder mit anderen Teilnehmern den Austausch suchen und mit diesen ihr Modell kritisch diskutieren. In der zweiten Modellevaluationsphase erfolgte der Vergleich mit einem Demonstrationsmodell, das als Grundlage diente, um Übereinstimmungen und Unterschiede mit dem eigenen Modell zu diskutieren. Diese kritische Reflexion des eigenen Entwurfs ist in der Wissenschaft unerlässlich und ein wesentlicher Aspekt erfolgreichen Modelllernens. Im Rahmen der Teilarbeit A konnte so eine möglichst authentische wissenschaftliche Lernsituation geschaffen werden, innerhalb derer die Schüler die Konzeptionierung ihres Modells erklärten, argumentierten und evaluierten (z. B. Passmore et al. 2009, Schwarz et al. 2009).

Insgesamt war der kurzfristige und langfristige Lernerfolg des Moduls mit zwei Modellevaluationsphasen deutlich höher. Dennoch fanden sich quantifizierbare Unterschiede in den dargestellten Elementen der DNA. Nur wenige Schüler identifizierten und beschrifteten alle wesentlichen Bestandteile der DNA, viele modellierten nur einige wenige Elemente und manche Schüler beschränkten sich auf die Grundstruktur der DNA. Diese Leistungsunterschiede in der Konzeptionierung von Modellen stimmen mit einschlägigen Referenzstudien überein. So beschreiben beispielsweise Howell et al. (2019) und Kim et al. (2015), dass Schüler Schwierigkeiten beim Verständnis der Struktur-Funktions-Beziehungen der DNA hatten. Grundsätzlich wäre auch denkbar, dass die Schüler zum Teil nicht in der Lage waren, ihr zuvor erworbenes Faktenwissen in eine andere Darstellungsform zu überführen. Dennoch konnten sehr gute Ergebnisse bei den selbstgebauten Modellen erzielt werden. Derartige Leistungen häuften sich besonders bei Schülern, die in Biologie als leistungsschwächer galten, was Erkenntnisse aus anderen Studien bestätigt (Bamberger & Davis 2011).

Die insgesamt in Teilarbeit A vermutete Verbindung von Modelllernen mit Lernerfolg konnte durch Quigley et al. (2017) belegt werden, die ein sogenanntes EcoSurvey-Tool verwendeten, um verschiedene Arten von Modellen im naturwissenschaftlichen Unterricht zu testen. In ihrer Studie zeigte sich zudem, dass vorhergehende Erfahrungen mit Modellieren und Modelllernen den Lernerfolg beeinflussten. Dies konnte in Teilarbeit A bestätigt werden, wenn auch die Effektstärke zwischen Modellqualität und der Modellerfahrung klein blieb. Die Modellqualität war allgemein besonders gut, wenn Schüler entweder ein allgemein besseres Modellverständnis besaßen, neue Informationen besser verarbeiteten konnten oder den Text sorgfältiger lasen (Krajcik & Merritt 2012, Ke et al. 2020). Die Analyse der Schülerantworten auf die offenen Fragen des Selbstevaluationsblattes der ersten Modellevaluation unterstützt diese Vermutung. Beispielantworten zeigten, dass Modellverständnis zu besseren DNA-Modellen führte. So erkannten die Schüler, dass auf Teilchenebene die Zucker- und Phosphatmoleküle vereinfacht dargestellt wurden und hielten dies in ihren Modellen fest. Entsprechend scheint die Modellqualität eng mit dem Prozess des Erkenntnisgewinnung der Schüler verbunden zu sein (Quigley et al. 2017), weshalb möglichst authentische wissenschaftliche Anlässe zum Modellieren geschaffen werden sollen und Modelle nicht zum Selbstzweck werden (Schwarz et al. 2009).

Die Schülergruppe mit zwei Modellevaluationsphasen schien ein besseres Verständnis und eine gesteigerte Behaltensleistung zu erzielen, was sich mit Ergebnissen einiger Referenzstudien deckt. Schwarz et al. (2009) und Bryce et al. (2016) weisen beispielsweise auf einen Zusammenhang zwischen der wissenschaftlichen Verwendung von Modellen im Unterricht und dem Lernerfolg der Schüler hin. Wissen über die Konzeptionierung von Modellen und daraus abgeleitete Erkenntnisse über die allgemeinen Eigenschaften von Modellen könnten also den in Teilarbeit A gemessenen Lernerfolg beeinflusst haben (Gilbert & Justi 2016). Substanzielle Korrelationen zwischen der Modellbewertung und dem kurzfristigen und langfristigen Lernerfolg bestätigen diese Vermutung. Daraus lässt sich zudem schließen, dass Modellverständnis eng mit Modellwissen verbunden ist (Peel et al. 2019). Insgesamt halfen die Konstruktion und Evaluation der Modelle den Schülern dabei, die gesammelten Informationen über die DNA zu organisieren, zu verstehen und zu abstrahieren. Diese Phasen förderten signifikant das

Tiefenverständnis der Lerninhalte (Bryce et al. 2016, Grünkorn et al. 2014) und begünstigten den Erkenntnisgewinn der Schüler über die DNA als Modell.

3.5.2 Teilarbeit B

Im Rahmen der Teilarbeit B wurden mögliche Messinstrumente identifiziert und evaluiert, die zuverlässig die beiden latenten Konstrukte Kreativität und Persönlichkeit erfassen, ohne den Rahmen des regulären Klassenunterrichts zu sprengen. Etablierte Fragebögen zu beiden Konstrukten sind sowohl inhaltlich als auch zeitlich zu anspruchsvoll (Fraser 1982). Die Ergebnisse der Teilarbeit deuten darauf hin, dass eine Kombination aus dem gekürzten *CPAC* Fragebogen und der *BFI-10* Skala in der Lage ist, diese Konstrukte zu messen und entsprechende Anregungen für Kreativität und Persönlichkeit integrierende Lernumgebungen abzuleiten.

Vorherige Studien haben bereits den gekürzten *CPAC* Fragebogen mit verschiedenen Altersgruppen und in unterschiedlichen Lernumgebungen getestet (Conradty & Bogner 2018, Mierdel & Bogner 2019b). Trotz dieser Vorstudien zeigt die in Teilarbeit B durchgeführte konfirmatorische Faktorenanalyse der Kreativitätsskala, dass einige der beobachtbaren Variablen des *CPAC* Fragebogens besser zur Beschreibung des latenten Konstrukts Kreativität geeignet sind als andere. Insbesondere die Variablen FL3, MA5, IN6, FL9 und IN12 zeigen sehr niedrige Faktorladungen, obwohl sie einen signifikanten Beitrag zur Beschreibung des latenten Konstrukts leisten. Die Faktorladungen von FL3 und IN6 sind dabei am niedrigsten. Bei FL9 erschweren Querladungen zusätzlich eine eindeutige Zuordnung zu den latenten Variablen *Flow* oder *Act*. Die Variable IN6 zeigt zwar keine Querladungen, ist aber doppeldeutig: Schüler gelangen entweder zur Lösung des Problems, indem sie unbewusst weiter daran arbeiten, ohne sich explizit damit auseinanderzusetzen oder zunächst pausieren und sich später mit frischem Kopf erneut damit befassen. Die Fragestellung ermöglicht beide Interpretationen, spricht jedoch gänzlich andere Dimensionen der Kreativität an, was die niedrige Faktorladung des latenten Konstrukts *Act* erklären könnte. Entsprechend sollten gerade beim gekürzten Fragebogen solche doppeldeutigen Variablen ausgeschlossen werden.

Die Kurzfassung des Kreativitätsfragebogens scheint jedoch grundsätzlich die latenten Konstrukte *Act* und *Flow* im naturwissenschaftlichen Unterricht zu erfassen. Sie kann folglich dazu beitragen, den Effekt kreativer Handlungen (*Act*) und die emotionalen Faktoren, die kreatives Handeln bedingen (*Flow*), im Klassenzimmer zu messen. Besonders *Flow*-Erleben gilt als Schlüssel in der Entwicklung von Kreativität. *Flow* beschreibt das positive Gefühl, das mit dem vollständigen Eintauchen in die Lernsituation einhergeht (Csikszentmihalyi 2000). Um dieses *Flow*-Erleben möglich zu machen, lassen sich aus den beobachtbaren Variablen bestimmte Anforderungen an die Lernumgebung ableiten. So ist es unerlässlich, dass beispielsweise eine Differenzierung der im Fachunterricht gestellten Lernaufgaben nach Schwierigkeit erfolgt. Schüler müssen sich gefordert aber nicht überfordert fühlen, um *Flow* erleben zu können (Lian et al. 2018). Da die wahrgenommene Schwierigkeit von Aufgaben jedoch individuell variiert, müssen Aufgaben mit unterschiedlichen Schwierigkeitsstufen zur Verfügung gestellt werden (e.g., Grandinetti 2013, Gottlieb 2016). Für die vollkommene Immersion der Lernenden in Aufgaben ist zudem das Gefühl von Sicherheit eine weitere Grundvoraussetzung (Csikszentmihalyi 2015, Conradty & Bogner 2018). Das erforderliche Maß an Sicherheit kann insbesondere durch offene Lernumgebungen erzielt werden, die selbstgesteuertes Lernen ermöglichen, nicht an feste Unterrichtszeiten gebunden sind und der Lehrkraft die Rolle des Lernbegleiters oder Mentors ermöglichen (Justo 2008). Gerade in dieser Rolle können Lehrkräfte Sicherheit auch ohne konkrete Anleitung geben, was die Schüler automatisch dazu ermutigt, individuelle Lern- und Problemlösungsansätze zu testen (Csikszentmihalyi 2015).

Diese Art der offenen Lernumgebung berücksichtigt auch die Bedürfnisse unterschiedlicher Persönlichkeitstypen (Safwat et al. 2020). Dies kann wiederum Einfluss auf die individuelle Kreativität haben (Kaufman et al. 2008). Im Rahmen der Teilarbeit B wurden diese Persönlichkeitstypen mit der *BFI-10* Skala erfasst, die bereits in vielen Unterrichtssituationen und Lernumgebungen, einschließlich dem naturwissenschaftlichen Unterricht, getestet wurde (Conradty & Bogner 2016, Carciofo et al. 2016). Es wird jedoch stets empfohlen, die Skala vor ihrer Verwendung zu prüfen, da viele der abgefragten beobachtbaren Variablen kulturabhängig sind und ein verzerrtes Bild über die Stratifikation von Persönlichkeitstypen in einer Klassengemeinschaft geben könnten (Carciofo et al. 2016, Rammstedt & John 2007). In Teilarbeit B wurde jedoch mit einer Hauptkomponentenanalyse die für die Kurzskala *BFI-10* charakteristische Faktorenverteilung trotz verschiedener Altersgruppen der Schüler erzielt. Zudem konnte neben substanzieller Faktorladungen für alle fünf latenten Variablen (Extraversion, Neurotizismus, Gewissenhaftigkeit, Freundlichkeit, Offenheit) eine oft berichtete positive Korrelation zwischen den latenten Variablen Neurotizismus und Extraversion beobachtet werden (z.B. Gomez et al. 1999, Verduyn et al. 2012). Die zugrundeliegenden Zusammenhänge zwischen den beiden Variablen werden durch das Arousal-Modell und das Yerkes-Dodson-Gesetz erklärt (Aguilar-Alonso 1996).

Zudem konnten in Teilarbeit B weitere Korrelationen zwischen den latenten Variablen der *CPAC* und *BFI-10* Skala erfasst werden. So zeigt *Flow* eine starke Beziehung zu dem Persönlichkeitsfaktor Gewissenhaftigkeit. Dieses Ergebnis deckt sich mit Erkenntnissen von Chui

& Lee (2012), die den Einfluss von *Flow*-Erleben auf Arbeitsleistung untersuchen. Entsprechend wurde in Teilarbeit B die gängige *Flow*-Theorie bestätigt. Auch kann das mit zielorientierter und fleißiger Arbeit verbundene *Flow*-Erleben einen positiven Einfluss auf den Lernerfolg (Demerouti 2006) sowie auf das Engagement und die Motivation haben (Cacioppe 2017). Wesentlich für das *Flow*-Erleben ist jedoch das offene und flexibles Umfeld, um Möglichkeiten zur individuellen Entfaltung zu bieten. Daher wurde *Flow* ursprünglich vorwiegend im Kontext von Aktivitäten wie Tanzen, Lesen, Online-Kommunikation und Kunst untersucht (Lian et al. 2018). Diese Aktivitäten erfordern neben Engagement auch ein hohes Maß an Gewissenhaftigkeit (Chui & Lee 2012).

Im naturwissenschaftlichen Unterricht kann *Flow* am besten im Rahmen von praktischen Experimenten erlebt werden, die die Schüler optimal einbinden und ihnen gleichzeitig Kontrolle über ihre Handlungen geben. Dieses Zugeständnis an Freiheit bedarf aber auch eines gewissenhaften Handelns (Hoseinifar et al. 2011). Klassische Unterrichtsmethoden, die wenig schülerzentriert sind und Schüler eher in die passive Rolle des Zuhörers und des Mitschreibenden drängen, sind hingegen nachteilig für das *Flow*-Erleben (Miller & Dumford 2016, Csikszentmihalyi 2015). Auch die zeitliche Begrenzung des Unterrichts auf maximal zwei Doppelstunden hindert Schüler oft daran, vollständig in wissenschaftliche Fragestellungen einzutauchen (Boyer & Lamoreaux 1997). Dabei ist gerade der naturwissenschaftliche Unterricht prädestiniert für *Flow*-Erleben. Ähnlich wie Tanz und Kunst erfordert er Fähigkeiten, die Kreativität fördern. Dies umfasst beispielsweise das Lösen komplexer Fragestellungen, das Treffen schneller Entscheidungen oder das Ableiten neuer Ideen aus Vorwissen (Yager 2005).

Zudem wurden in Teilarbeit B Geschlechterunterschiede in nach Alter gebildeten Extremgruppen untersucht. Dabei wiesen insbesondere *Act* und *Flow* signifikante Unterschiede auf. In der unteren Altersgruppe erzielten Jungen höhere Werte für *Act* als Mädchen der gleichen Altersgruppe, weshalb angenommen wurde, dass Jungen besonders durch praktische Aktivitäten ihre Kreativität ausleben. Folglich sind insbesondere Gruppenarbeiten zur Förderung dieser kreativen Aktivitäten geeignet, da sie den Austausch und die Diskussion kreativer Ideen fördern sowie Raum für das Ausprobieren von Lösungsansätzen und die Reflexion eigener Strategien bieten (Sawyer 2012). Zudem birgt entdeckendes Lernen mit vielen praktischen Experimenten oder Lernstationen Potenziale, um die notwendige Freiheit für kreatives Denken und Handeln zu schaffen (Schneiderhan-Opel & Bogner 2020). In höheren Altersgruppen glich sich die ursprüngliche Diskrepanz von *Act* zwischen den Geschlechtern wieder aus. Stattdessen zeigten sich jedoch Unterschiede im *Flow*-Erleben. So schienen Mädchen der oberen Altersgruppe ein signifikant besseres *Flow*-Erleben zu haben als männliche Altersgenossen. Laut eigenen Angaben lösten sie komplexe Fragestellungen mehrheitlich dadurch, dass sie ohne Einfluss von außen komplett in die Aufgabe eintauchten. Ein wesentlicher Grund für diese Beobachtung könnte in Sozialisationsprozessen und kulturellen Einflüssen während der Adoleszenz liegen (Iiamura & Taku 2018). So wird davon ausgegangen, dass gesellschaftliche Einflüsse bereits in der Kindheit unterschiedliche Herangehensweisen in dem Erleben und Bewältigen von alltäglichen Ereignissen zwischen Männern und Frauen verstärken (Goodwin & Gotlib 2004).

3.5.3 Teilarbeit C

Die Teilarbeit C fokussierte sich auf die Auswertung des Lernerfolgs der bilingualen Schülergruppe im Vergleich zur muttersprachlichen Vergleichsgruppe mit zwei Modellevaluationsphasen während der Teilnahme am Schülerlabor. Unabhängig von der Sprache des Moduls erzielten beide Schülergruppen signifikante Lernerfolge. In beiden Fällen profitierten Schüler von der Kombination aus Hands-on-Aufgaben mit Minds-on-Aktivitäten. Dieser informelle Hands-on-Minds-on-Ansatz hat sich bereits in früheren Studien bei der Vermittlung wissenschaftlicher Konzepte aus der Molekularbiologie bewiesen (z.B. Langheinrich & Bogner 2015, Mierdel & Bogner 2019a).

In der Konzeptionierung des Moduls wurde gezielt darauf geachtet, dass die Instruktionen im Kursbuch kognitive und affektive Dimensionen berücksichtigten und zu viele Vorgaben machten (Hofstein & Lunetta 2004). Neben dem sicheren Umgang mit Laborgeräten beinhaltete das wissenschaftliche Arbeiten auch die Identifikation möglicher fachbezogener Probleme, die Aufstellung von Hypothesen, die Analyse und Interpretation von Lerninhalten sowie die Verwendung von Modellen, um für komplexe und schwer beobachtbare Phänomene Erklärungen zu finden (Carmel et al. 2020). Die Kombination dieser wissenschaftlichen Arbeitsweisen kann jedoch, wenn nicht regelmäßig angewandt, schnell individuelle kognitiven Kapazitäten übersteigen und den Lernerfolg beeinträchtigen (z.B. Goldschmidt 2016, Langheinrich 2016, Mierdel 2019). Auch andere Studien berichten von negativen Auswirkungen einer zu hohen kognitiven Belastung auf den Lernerfolg. So beschreibt die Cognitive-Load-Theory den schmalen, für Erinnerungs- und Lernprozesse notwendigen Grat zwischen geistiger Anregung und Überforderung durch zu hohe intrinsische und extrinsische Belastung (Sweller 2015).

Die Cognitive-Load-Theory könnte eine mögliche Erklärung für die Unterschiede im Lernerfolg zwischen muttersprachlichen und CLIL-Schülergruppen sein. Demnach könnten eine zu hohe Informationslast und resultierende Überlastung des Arbeitsgedächtnisses insbesondere bei den CLIL-Schülergruppen zu Lernschwierigkeiten geführt haben (Johnstone &

Wham 1982). Ergebnisse von Reynolds (1991) bestätigen beispielsweise diese Vermutung. Das Verarbeiten der Lerninhalte in einer Fremdsprache scheint dabei die ohnehin bereits gesteigerte mentale Anstrengung durch Hands-on-Lernen weiter zu erhöhen, was eine geistige Überlastung nach sich zieht (Rodenhauser & Preisfeld 2014, Roussel et al. 2017). Frühere Studien haben sogar gezeigt, dass Hands-on-Lernen allein bereits eine kognitive Überlastung hervorrufen kann (z. B. Goldschmidt 2016, Langheinrich 2016, Mierdel 2019).

Trotz einer vergleichsweise höheren kognitiven Auslastung, die mit dem Lernen wissenschaftlicher Inhalte in einer Fremdsprache verbunden ist, hilft der wissenschaftliche Diskurs den Schülern beim Wissenserwerb und Verständnis (z.B. Evnitskaya & Morton 2011, Kress et al. 2001). Denn ungeachtet einer möglichen kognitiven Überlastung zeigen die Ergebnisse aus Teilarbeit C, dass die Kombination von Hands-on-Aufgaben mit Minds-on-Aktivitäten beim CLIL-Lernen (Glynn & Muth 1994) grundlegende wissenschaftliche Konzepte vermitteln konnte. Da beispielsweise alle Fragebögen, abgesehen vom Cloze-Test, in der Muttersprache gestellt wurden, musste automatisch auch ein Transfer des Wissens von der Fremdsprache in die Muttersprache erfolgen (z.B. Dalton-Puffer 2007, Grandinetti 2013).

Wie in der funktionalen Linguistik beschrieben wird, hilft Sprache dabei Realität zu schaffen (Maxwell-Reed 2017). So ist das Spachenlernen und das inhaltliche Lernen durch Sprache untrennbar miteinander verwoben und sich gegenseitig bedingend (z.B. Maxwell-Reed 2017, Halliday 2003). Die Hypothese wird besonders durch allgemeine Theorien über die Entstehung von Scientific Literacy gestützt. Diese besagen, dass wissenschaftliche Inhalte nur dann richtig verstanden und effektiv kommuniziert werden, wenn die zugrunde liegende wissenschaftliche Fachsprachlichkeit vorhanden ist. Umgekehrt braucht es jedoch zunächst einmal wissenschaftliche Inhalte, um überhaupt Fachsprachlichkeit aufzubauen. Diese gegenseitige Abhängigkeit setzt also voraus, dass eine basale Fachsprachlichkeit notwendig ist, um wissenschaftliche Inhalte zu vermitteln, wobei jedoch der vermittelte Inhalt zugleich Grundlage einer verbesserten Fachsprachlichkeit ist (Sutton 1996, Yore & Treagust 2006). Entsprechend kann ein Unvermögen, wissenschaftliche Inhalte in der Muttersprache und Fremdsprache zu kommunizieren, das inhaltlichen Lernen erheblich beeinträchtigen (Rodenhauser & Preisfeld 2014).

Gerade im Kontext von CLIL erfordert die gegenseitige Abhängigkeit von Sprache – Muttersprache und Fremdsprache - und Inhalt eine zeitgleiche Verarbeitung des Gelernten auf verschiedenen mentalen Ebenen. Dies könnte eine Erklärung für die in Teilarbeit C festgestellten Unterschiede zwischen muttersprachlich deutsch unterrichteten und CLIL Schülergruppen sein (Craik & Lockhart 1972). Piesche et al. (2016) nehmen ebenfalls an, dass gleichzeitiges

Lernen von Sprache und Inhalt die mentalen Kapazitäten und den resultierenden Lernerfolg beeinträchtigt.

Obwohl die CLIL-Schülergruppe deutlich schlechter abschnitt als die muttersprachliche Vergleichsgruppe, konnte bei beiden Schülergruppen ein temporärer sowie langfristiger Lernerfolg erzielt werden. Die vertiefte kognitive Verarbeitung von Inhalten auf mehreren Ebenen durch den aktiven Einsatz einer weiteren Sprache, führte also auch bei CLIL-Lernenden zu einer Verbesserung ihres inhaltlichen Wissens (Meyerhöffer & Dreesmann 2019). Gerade aufgrund erschwerter Bedingungen durch die Verwendung von CLIL mussten wichtige Inhalte wiederholt gelesen werden, um sie zu verstehen, wobei Wissen verstetigt wurde (z. B. Rodenhauser & Preisfeld 2014, Marian & Fausey 2006). Piesche et al. (2016) sowie Fernández-Sanjurjo et al. (2017) bedienen sich ähnlicher Erklärungen, um den Lernerfolg sowohl für CLIL als auch muttersprachliche Schülergruppen zu begründen. Admiraal et al. (2006) oder Haagen-Schützenhöfer et al. (2011) erzielten jedoch davon abweichende Ergebnisse und zogen andere Erklärungsansätze heran, weshalb Erkenntnisse über den tatsächlichen Lernerfolg von CLIL unklar bleiben.

Scaffolding beispielsweise kann CLIL unterstützen, da es sowohl die sprachliche als auch die inhaltliche kognitive Belastung deutlich verringert (z. B. Gottlieb 2016, Fernández-Fontecha et al. 2020). Das in Teilarbeit C verwendete sprachliche, grafische und interaktive Scaffolding scheint jedoch die kognitive Belastung nicht so weit reduziert zu haben, dass der Lernerfolg der CLIL Schülergruppe dem der muttersprachlich deutschen Schülergruppe ebenbürtig war. Besonders beim sprachlichen Scaffolding ist die Qualität der Materialien ausschlaggebend, um einer Fossilierung sprachlicher Fehlvorstellungen vorzubeugen, die am Ende das Verständnis wissenschaftlicher Inhalte beeinträchtigt (Buxton et al. 2019). Deshalb muss das Scaffolding einer ständigen systematischen Analyse, Reflexion und Anpassung unterliegen. Gemäß McGuiness (1999) konzentrierte sich das in Teilarbeit C umrissene Scaffolding vorwiegend darauf, sprachlich-kognitive Anforderungen niedrig zu halten. Dies erfolgte durch die Bereitstellung von Vokabelquizzen in einem separaten Übungsheft sowie Übersetzungen einzelner wissenschaftlicher Termini im Kursbuch.

Da die Schüler jedoch weder mit den wissenschaftlichen Grundlagen der Genetik noch mit Laborexperimenten vertraut waren, wurde durch die Verknüpfung von wissenschaftlichem Wortschatz mit wissenschaftlichen Arbeitsweisen die Scientific Literacy nicht schnell genug aufgebaut, um inhaltliches Lernen auf demselben Level wie bei der monolingual deutschen Vergleichsgruppe zu fördern. Gerade bei CLIL-Modulen in Kurzeitform ist die inhaltliche Dichte hoch und stellt entsprechende Anforderungen an die naturwissenschaftliche

Sprachkompetenz. Folglich bedarf es Hilfen jenseits des klassischen inhaltlichen und sprachlichen Scaffoldings, um bei CLIL Schülergruppen einen vergleichbaren Lernerfolg zu erzielen (z.B. Coyle 2007, Grandinetti et al. 2013). Fernández-Fontecha et al. (2020) schlagen multimodales Scaffolding mit z.B. Bildern vor, um den Schülern Inhalte in anderen semiotischen Formen als Text zu präsentieren. Neben verschiedenen Arten von Scaffolding ist es auch denkbar, dass Schüler elementares Vokabular vor der Teilnahme an CLIL-Modulen in Kurzzeitform erlernen. Dies könnte die mentale Belastung reduzieren, da die grundlegende wissenschaftliche Terminologie bereits vorhanden ist und nur noch in einen inhaltlichen Kontext gebracht werden muss (McGuiness 1999).

3.5.4 Teilarbeit D

In Teilarbeit D konnte gezeigt werden, dass das Modul einen grundlegend positiven Einfluss auf den modellbezogenen Lernerfolg, das Modellverständnis, die Erstellung von Modellen sowie die beiden Modellevaluationen hatte. Die Hands-on-Minds-on-Aktivitäten führten im Rahmen des Moduls zu einem besseren Verständnis der sachfachlichen Inhalte und Charakteristika von Modellen (Schwarz et al. 2009) sowie zum Aufbau von Scientific Literacy (Ke et al. 2020). Die Experimente zur Isolierung und Visualisierung von DNA aus Mundschleimhautzellen dienten als Grundlage für wissenschaftliche Diskussionen über die Struktur und die nachfolgende Modellkonstruktion der DNA. Kritische Reflexion und Evaluation der Ergebnisse vermittelte zusätzlich authentische wissenschaftliche Alltagspraktiken (Lemke 1990, Berland & Reiser 2009).

Bereits aus Teilarbeit A ist der generell positive Einfluss des Moduls auf modellbezogenen Lernerfolg bekannt. Teilarbeit D ergänzte nun Teilarbeit A um die Erkenntnis, dass der modellbezogene Lernerfolg zwar grundsätzlich unabhängig von der Kommunikationssprache war, insgesamt jedoch die muttersprachlich deutsche Schülergruppe besser abschnitt. Dies kann dem im Vergleich höheren mentalen Aufwand beim Erlernen naturwissenschaftlicher Inhalte in einer Fremdsprache geschuldet sein (Rodenhauser & Preisfeld 2014, Sweller 2015). Trotz des geringeren Lernerfolgs konnte der intensive Diskurs zur Konstruktion und Evaluation eines passenden DNA-Modells auch im CLIL-Modul einen Zuwachs inhaltlichen Wissens und sogar Tiefenverständnis bewirken (Krell et al. 2015, Ke et al. 2020, Ke & Schwarz 2020). Besonders die Kombination von Hands-on-Minds-on-Aktivitäten (Glynn & Muth 1994) mit kommunikativen Praktiken zur Informationsverarbeitung steigerte den Lernerfolg (Evnitskaya & Morton 2011, Kress et al. 2001).

Da die Kommunikation in einem auf Sprachenlernen fokussierten Modul intensiver war, zeigten sich die Vorteile von CLIL besonders beim Modellverständnis. Hier schnitten Schülergruppen des CLIL-Moduls besser ab als Schülergruppen des muttersprachlich deutschen Moduls. Dies unterstützt die Annahme, dass die verstärkten Diskursaktivitäten des CLIL-Moduls in Verbindung mit Modellkonstruktion und -evaluation Tiefenverständnis förderten (Ke et al. 2020, Ke & Schwarz 2020). Schüler, die am CLIL-Modul teilnahmen, erkannten sehr schnell, dass Modelle die Ergebnisse wissenschaftlicher Diskussion und wissenschaftlichem Erkenntnisgewinns sind und sich mit dem aktuellen Forschungsstand verändern können (CNM). Die Lernenden nahmen also Abstand von der verbreiteten Fehlvorstellung, dass Modelle exakte Abbilder der Wirklichkeit sind (EM), die sich nicht evolutiv entwickeln und unabhängig von wissenschaftlichen Erkenntnissen sind (Grosslight et al. 1991, Ke & Schwarz 2020). Tatsächlich sind sich nur wenige Schüler bewusst, dass die Validität eines Modells in beständiger Diskussion mit Fachexperten bestätigt oder verbessert wird (Treagust et al. 2002, Mendonça & Justi 2013, Ke et al. 2020). Nach Besuch des muttersprachlich deutschen Moduls hielten die Schüler noch immer an der Fehlvorstellung fest und bestätigen damit den grundlegenden Mangel an Modellverständnis.

Ein wesentlicher Unterschied zwischen dem CLIL und dem muttersprachlich deutschen Modul bestand in der stärkeren Betonung wissenschaftlicher Diskursaktivitäten. Da die wissenschaftliche Argumentation und Reflexion ein tieferes Verständnis der Inhalte verlangt (Passmore & Svoboda 2012), könnte dies das Modellverständnis bezüglich CNM verbessert haben (Schwarz et al. 2009, Ke et al. 2020). Die kritische Reflexion des Modells war jedoch bereits Teil der zweiten Modellevaluation im muttersprachlich deutschen Modul. Entsprechend bedarf es einer anderen Erklärung, die in den von Gilbert & Justi (2002) deklarierten Entwicklungszyklen zur Validierung eines Modells zu finden ist. Um ein sinnvolles Modell zu konstruieren, das empirischer Prüfung standhält, müssen mehrere dieser Modellphasen durchlaufen werden. Dies erfordert u.a. die Iteration zwischen beobachtetem Phänomen, wissenschaftlicher Theorie und dem daraus abgeleiteten Modell (Schwarz et al. 2009, Passmore & Svoboda 2012, Lee et al. 2015). Da die Teilnehmer des CLIL-Moduls die Fachtermini in englischer Sprache noch nicht beherrschten und auch das wissenschaftliche Arbeiten erst erlernen mussten, war eine Iteration zwischen Text, Theorie und mentalem Modell unerlässlich, bevor sie überhaupt genug Informationen zusammengetragen hatten, um ein Modell zu konstruieren (Oh & Oh 2011, Gilbert & Justi 2002). Während dieser zahlreichen mentalen Iterationen war es möglich, dass die Bedeutung von CNM durch den stetigen Informationsaustausch in und zwischen den Teilgruppen bewusstwurde (Treagust et al. 2002).

Dies ist nur eine mögliche Erklärung, gleichzeitig wird sie durch einschlägige Theorien über die Entwicklung von Scientific Literacy und Deep Learning unterstützt. Demnach soll die Verwendung von Modellen Denkprozesse über die grundlegenden Mechanismen und wiederkehrenden Muster eines beobachteten Phänomens anregen (Ke & Schwarz 2020). Indem nun die Natur des Phänomens selbst hinterfragt wird, bewegt sich die Verwendung von Modellen weg von rein inhaltlichem Lernen und rückt authentische wissenschaftliche Praktiken in den Vordergrund. Schüler werden dabei ermutigt, selbst wissenschaftliche Hypothesen aufzustellen und entsprechende Argumente zu ihrer Verteidigung zu finden. Dies fördert nicht nur das vollkommene Durchdringen der Lerninhalte, sondern bewirkt zusätzlich die Entwicklung von Scientific Literacy (Quarderer & McDermott 2020, France 2019) und Deep Learning (Biggs 1987, Chin & Brown 2000, Lee et al. 2015). Deep Learning bezieht sich dabei nicht auf künstliche Intelligenz, sondern auf das Deep-Level-Processing von Informationen (Case & Gunstone 2010, S. 461), das durch CLIL in Verbindung mit naturwissenschaftlichen Arbeitsweisen angeregt wird. Denn gerade die offenen Fragen der Kursbücher ermutigten Schüler ihre Beobachtungen zu reflektieren und zu kontextualisieren. Zudem erwies sich die verbale Repräsentation des physischen Modells durch intensiven Diskurs über seine Validität als mindestens ebenso wichtig (Campbell & Fazio 2020, S. 2302).

Bei der Beantwortung der offenen Fragen auf dem Modellskizzenblatt bewiesen die CLIL-Schülergruppen trotz der allgemein schlechteren Modellleistung ein besseres Modellverständnis. Diese nannten insbesondere die Strukturebene der DNA und den aktuellen Forschungsstand als wichtigste Grundlagen der Modellkonstruktion. Darin wird erneut die Anpassbarkeit von Modellen (CNM) in Abhängigkeit wissenschaftlicher Erkenntnisprozesse betont (Grosslight et al. 1991, Ke & Schwarz 2020). Ergebnisse der ersten Modellevaluation unterstützen diesen Eindruck, da die Skizzen der CLIL-Schülergruppen viel genauer und besser beschriftet waren als die der monolingual deutschen Vergleichsgruppe. Die Skizzen der ersten Modellevaluation erforderten zudem ein tieferes Verständnis der Struktureigenschaften von DNA als der Vergleich zwischen den selbstgebastelten Modellen und einem Demonstrationsmodell aus der zweiten Modellevaluation (Teilarbeit A). Die CLIL-Lernenden verstanden also nicht nur die Bedeutung von CNM, sondern waren sich auch der Funktion und Anordnung der Komponenten bewusst, die sie in ihr Modell übernommen hatten.

In Bezug auf die zweite Modellevaluation waren jedoch die muttersprachlich deutschen Schülergruppen sowohl in den Modellskizzen als auch den dreidimensionalen Modellen den CLIL-Schülergruppen überlegen. Dies kann ähnlich wie der Lernerfolg auf die höhere kognitive Belastung durch Sprachenlernen und inhaltliches Lernen zurückgeführt werden (Sweller

2015). Da die Konstruktion von Modellen das inhaltliche Verständnis des Textes erforderte (Gilbert & Justi 2002, Passmore & Svoboda 2012, Zangori et al. 2017), könnten die durch mangelnde Sprachkenntnisse erforderlichen Iterationen des Inhalts (Johnstone & Wham 1982) zu weniger Zeit für die Konstruktion der Modelle geführt haben. Jedoch hatten sowohl die monolingual deutsche also auch CLIL Schülergruppe Schwierigkeiten, die im Modell dargestellten Elemente der DNA in der Modellskizze und den Selbstevaluationsblättern korrekt wiederzugeben. So wurden vermeintliche Elemente identifiziert, die nicht im Modell enthalten waren, und andere Elemente nicht identifiziert, die aber im Modell berücksichtigt wurden. Dieses Ergebnis deckt sich mit Erkenntnissen früherer Studien (Howell et al. 2019, Kim et al. 2015), in denen insbesondere Verständnisprobleme der Struktur-Funktions-Beziehungen beschrieben wurden. Ähnliche Schwierigkeiten wurden bereits in Teilarbeit A diskutiert.

3.5.5 Teilarbeit E

In Teilarbeit E konnte die positive Auswirkung des eintägigen CLIL-Moduls auf das sprachliche Wissen der Schüler gezeigt werden. Auch wurde deutlich, dass diese Auswirkung unabhängig vom Geschlecht war. Gleiches gilt für das Tiefenverständnis der wissenschaftlichen Inhalte und der damit verbundenen Fachterminologie (z.B. Glynn & Muth 1994), die in der Kombination mit Hands-on-Minds-on-Aktivitäten die Entwicklung von Scientific Literacy begünstigten (Ke et al. 2020). Praktisches Experimentieren und inhaltliches sowie sprachliches Scaffolding ermöglichte Erkenntnisgewinn und die Diskussion dieser Erkenntnisse bei der Erstellung entsprechender Modelle (Lemke 1990, Berland & Reiser 2009).

So steigerte sich das Sprachverständnis im Cloze-Test von anfänglich vier auf über sieben richtig beantwortete Fragen nach dem Besuch des Moduls. Derartige Effekte auf den Spracherwerb sind selten bei CLIL-Modulen in Kurzzeitform (z. B. Rodenhauser & Preisfeld 2014, Buse et al. 2018). Zumeist wird bei CLIL-Modulen in Langzeitform davon berichtet (z.B. Lorenzo et al. 2010, Dalton-Puffer 2011). Dies lässt entsprechend darauf schließen, dass bereits weniger repetitive Interventionen zu positiven Ergebnissen führen. Dieser anfänglich positive Effekt war jedoch im Behaltenstest nach acht Wochen nicht mehr ersichtlich und die Leistung viel auf ein Niveau ab, das deutlich unter dem des Vortests lag. Dies ist nach derart intensiven Sprachmodulen in Anbetracht des anfänglich festgestellten Erfolgs eher ungewöhnlich (Thürmann 2008), da allgemein davon ausgegangen wird, dass sich die Fremdsprache gerade am Anfang stark kontextbezogen und langsamer entwickelt (z.B. Rosenthal 1993). Zudem werden viele der im Modul aufgeführten Fachtermini in der Schulgenetik als Anglizismen wieder aufgegriffen (Meyerhöffer & Dreesmann 2019). Somit kann eine kognitive Überforderung als

Erklärung für den beobachteten Leistungsabfall ausgeschlossen werden (Sweller 2015, Ryoo 2015). Viel wahrscheinlicher ist, dass der Spracherwerb durch die landesweiten Lockdown-Maßnahmen im Zuge der COVID-19 Pandemie in direktem Anschluss an das Modul beeinträchtigt wurde. Da alle Schulen in Deutschland für mehrere Wochen auf Distanzunterricht umstellten, wurde betont auf digitalen Frontalunterricht und Selbstlernaktivitäten gesetzt. Dies bot Schülern jedoch wenig Anlässe zur Sprachpraxis (König et al. 2020, Hoenig & Wenz 2020).

Gerade dies ist jedoch unerlässlich für effektives Sprachenlernen, da das ständige Üben der Fremdsprache Fehler reduziert, das Wissen verstetigt und die Sprachkompetenz verbesset (Thürmann 2008, Pérez-Vidal & Roquet 2015). Deutlich reduzierte sprachpraktische Übungen in der Fremdsprache während der Lockdown-Maßnahmen könnten so eine mögliche Erklärung für die niedrigen Werte im fachspezifischen Wortschatz und der Grammatik sein. Eine Studie von Wößmann et al. (2020) berichtete zudem von einer Verringerung der Lernzeit während des Distanzunterrichts von durchschnittlich 7,4 auf 3,6 Stunden pro Tag. Obwohl natürlich versucht wurde, die Lernzeit durch sinnvolle Aufgabenstellungen zu erhöhen, ohne die Schüler mit der neuen Situation zu überlasten (Huber & Helm 2020), birgt das Online-Sprachenlernen Nachteile (Gao & Zhang 2020). So kann der Rückgang des in Teilarbeit E geschilderten Sprachverständnisses darauf hindeuten, dass Fernunterricht den Präsenzunterricht beim Sprachenlernen nicht effektiv ersetzen kann.

Die offenen Fragen des Kursbuchs wurden nur von 48% der Teilnehmer beantwortet, von denen 2% in ihrer Muttersprache antworteten. Dieses Ergebnis stimmt mit Cheshire & Gardner-Chloros (1998) überein, die postulieren, dass das Verständnis von Inhalten unabhängig von der Sprache ist, in der sie kommuniziert werden. Antworten in einer anderen Sprache zu geben als der aktuellen Kommunikationssprache wird als Code-Switching bezeichnet. Dies kann zugunsten des Gesprächsflusses einzelne Wörter oder ganze Sätze in der Muttersprache umfassen (Königs 2013). Gerade beim CLIL-Lernen werden die Schüler zum Teil sogar dazu ermutigt, auf Deutsch zu antworten, wenn ihnen der erforderliche Wortschatz im Englischen fehlt (KMK 2013). Das Lernen wissenschaftlicher Inhalte und den Spracherwerb zu parallelisieren, kann jedoch kognitiv sehr anspruchsvoll sein und ohne ausreichendes Scaffolding zu geistiger Überlastung führen (Ryoo 2015, Sweller 2015). Gerade CLIL und die damit verbundenen Anforderungen bedürfen einer größeren mentalen Kapazität als inhaltliches Lernen in der Muttersprache, um die gleiche Menge an Informationen zu extrahieren und zu verarbeiten (Bogner 1995). So kann Scaffolding helfen, die kognitive Belastung zu reduzieren (Coyle 2007, Grandinetti 2013, Gottlieb 2016) und inhaltlichen sowie sprachlichen Lernerfolg zu ermöglichen. Da Schüler sowohl den wissenschaftlichen Inhalt als auch die Fachsprache verstehen müssen, um Scientific Literacy zu erlangen, müssen beide gleichwertig Berücksichtigung finden (Unsal et al. 2018, France 2019).

Scientific Literacy erfordert vor allem Minds-on-Hands-on-Aktivitäten und die entsprechende Fachsprachlichkeit, um wissenschaftliche Inhalte erklären und abstrahieren zu können (Glynn & Muth 1994, France 2019). Das vorliegende Modul berücksichtigte dieses Zusammenspiel auf drei Ebenen: (1) Verstehen und Durchführen von praktischen Experimenten unter Verwendung von Informationen aus der Einführungsphase und dem Kursbuch, (2) Übertragung von Informationen, die während der Experimente gesammelt wurden in verschiedene Darstellungsformen und (3) effektive Kommunikation eigener Hypothesen und Erkenntnisse. Zusätzlich wurden disziplinspezifische Fachtermini und die generelle Fachsprachlichkeit berücksichtigt (Norris & Phillips 2003, Prain 2004), die sich oftmals deutlich von der Alltagssprache unterscheiden (Wang & Chen 2014). Dies war besonders für die Minds-on-Aktivitäten notwendig, bei denen die Schüler ihre Beobachtungen und Reflexionen aus den praktischen Experimenten zusammenfassten. Hier beantworteten aber nur etwa 50% aller Schüler des CLIL-Moduls die Fragen zuverlässig, was darauf hindeutete, dass das Verstehen und Reflektieren der Inhalte so viel Zeit in Anspruch nahmen, dass sie nur auf eine Auswahl von Fragen antworten konnten (Glynn & Muth 1994, Meyerhöffer & Dreesmann 2019). Denn insbesondere das Erklären des "Wie" und "Warum" erforderte Zeit zum Nachdenken. Die Schüler mussten hierfür zunächst wissenschaftliche Hypothesen und Argumente entwickeln, was eine vertiefte Auseinandersetzung mit dem Thema anregte (Bogner 1995), aber kognitiv sehr anspruchsvoll war. Zudem wurde dadurch auch die Entwicklung von Fachsprachlichkeit gefördert sowie das Verständnis wissenschaftlicher Inhalte und schlussendlich die Scientific Literacy (Quarderer & McDermott 2020, France 2019).

In der aktuellen PISA-Studie gibt es keine Geschlechterunterschiede bezüglich Scientific Literacy (Reiss et al. 2018). Im Allgemeinen deckt sich dieses Fehlen von Geschlechterunterschieden mit den Ergebnissen der Teilarbeit E. Auch wurden in keinem der Vorläufermodulen unseres Moduls signifikante Geschlechterunterschiede festgestellt (z.B. Scharfeberg & Bogner 2013, Franke & Bogner 2013, Goldschmidt & Bogner 2016, Langheinrich & Bogner 2015, Mierdel & Bogner 2019c). Gerade in Bezug auf den Lernerfolg gab es keine Unterschiede zwischen männlichen und weiblichen Schülern im Rahmen der verschiedenen Module: weder in den 12. Jahrgangsstufen (Scharfenberg & Bogner 2013), noch in den 10. oder 11. Jahrgangsstufen (Franke & Bogner 2013, Goldschmidt & Bogner 2016, Langheinrich & Bogner 2015) oder den 9. Jahrgangsstufen (Mierdel & Bogner 2019c, Teilarbeit A). Während solche Ergebnisse ermutigend sind, gibt es jedoch in zahlreichen anderen Studien noch immer

Geschlechterunterschiede beim sprachlichen oder naturwissenschaftlichen Lernen (z. B. Andreou et al. 2005, van der Slik et al. 2015). Entsprechend wird davon ausgegangen, dass bestimmte soziale Faktoren, auch im Klassenzimmer, das Sprachenlernen bei weiblichen Schülern besser unterstützt als bei männlichen. Zum Teil können diese Faktoren bei männlichen Schülern sogar zu einem signifikanten Rückgang der intrinsischen Motivation und des sprachlichen Selbstkonzepts führen (Mady & Seiling 2017). Entsprechend bedarf es einer anregenden und motivierenden Lernumgebung, die beide Geschlechter gleichermaßen anspricht und Lernen in kleinen Gruppen ermöglicht (Bećirović 2017, Mady & Seiling 2017, Mierdel & Bogner 2019a, Mierdel & Bogner 2019b). Dies gilt nicht nur für den Spracherwerb, sondern auch für naturwissenschaftliches Lernen, bei dem weibliche Schüler typischerweise ein deutlich niedrigeres Selbstkonzept aufweisen und folglich – trotz guter Leistungen – von einer MINT-Karriere absehen (Delaney & Devereux 2019). Da das in der vorliegenden Arbeit präsentierte CLIL-Modul die erforderliche offene Lernumgebung bereitstellte, konnten offenbar beide Geschlechter gleichermaßen gefördert werden, was zu insgesamt vergleichbaren Lernerfolgen bei sprachlichem und inhaltlichem Lernen führte.

3.5.6 Teilarbeit F

In Teilarbeit F wurde festgestellt, dass der Besuch eines eintägigen CLIL-Moduls insbesondere das Selbstkonzept von leistungsschwachen Schülern in Biologie und Englisch unabhängig vom Geschlecht positiv beeinflusst. Dieses Ergebnis stimmt mit einschlägiger Referenzliteratur überein (z.B. Patall et al. 2014). Ebenso wurde deutlich, dass das Selbstkonzept in beiden Fächern stark mit Kreativität korrelierte, wie u.a. durch Autoren wie Sisk (1972) oder Mawang et al. (2018) bestätigt wird. Zudem gab es keine signifikanten Geschlechterunterschiede bezüglich des Selbstkonzepts mit Ausnahme der Subskala *Social.* Weibliche Schüler betrachteten ihre wahrgenommene Leistung als schlechter im Vergleich zu ihrer tatsächlichen Leistung, insbesondere wenn sie sich mit anderen vergleichen sollten (z.B. Mullis et al. 2020, Mejía-Rodríguez et al. 2020). Generell sollte jedoch vermieden werden, dass Schüler zur Bestimmung ihres leistungsbezogenen Selbstwertes in den Vergleich mit anderen Schülern gezwungen werden. Denn Selbstwert und Selbstkonzept müssen unabhängig von externen Faktoren in einem inneren Erkenntnisprozess reifen (Hoferichter et al. 2018, Bournelli 2009). Abhängig davon mit wem sich Schüler vergleichen, kann dies positive oder negative Auswirkungen auf ihr Selbstkonzept haben. Wenn sich beispielsweise leistungsstarke Schüler mit anderen leistungsstarken Schülern vergleichen, kann das ihre eigene Leistung bestätigen oder zu noch besseren Leistungen anspornen. Wenn sich jedoch leistungsschwache Schüler - die viel häufiger vorkommen- mit möglicherweise besseren Schülern vergleichen, kann dies nachteilig für das Selbstkonzept sein (Hoferichter et al. 2018). Zudem wird durch solche Vergleiche eine toxische Kommunikation begünstigt, da der Vergleich mit anderen stets eine Form der (Ab-)Wertung darstellt. Nach Rosenberg (2012) birgt dies also sehr große Konfliktpotenziale.

Kreativität und CLIL sind grundsätzlich miteinander verbunden. Gerade die mit dem *Flow*-Erleben verbundene positive Emotion (Csikszentmihalyi 2000) bedarf offener Lernumgebungen, um eine Differenzierung der Aufgabenschwierigkeit zu erlauben (Lian et al. 2018). Das in dieser Arbeit behandelte CLIL-Modul stellt eine solche Lernumgebung dar und ermöglicht beispielsweise das Basteln eines dreidimensionalen DNA-Modells basierend auf einem Brief von Crick an seinen Sohn (Usher 2013, Mierdel & Bogner 2019b, Teilarbeit A). Dabei wurden explizit kreative Lösungsstrategien ermutigt und von einer konkreten Beurteilung des Ergebnisses abgesehen, um das Selbstkonzept zu stärken (Bournelli et al. 2009). Den Schülern wurde also die Verantwortung für das Finden einer Lösung übertragen, was sich in positiven Korrelationen zwischen *Flow*-Erleben und den latenten Variablen des Selbstkonzepts zeigte. Das völlige Eintauchen in den Lösungsfindungsprozess stärkte die fachspezifische und kompetenzspezifische Selbstwirksamkeit, wodurch auch das Selbstkonzept positiv beeinflusst wurde (Justo 2008). Die Lösung naturwissenschaftlicher und sprachspezifischer Probleme erfolgte zumeist in kleinen Teams mit zwei oder drei Schülern. So wurde eine inklusive Arbeitsatomsphäre geschaffen, die sowohl die Kreativität (Mierdel & Bogner 2019b) als auch das schulische Selbstkonzept in Biologie und in Englisch unterstützte. Das umfangreiche Scaffolding, das für eine erfolgreiche Durchführung des CLIL-Moduls notwendig war, ermöglichte zudem eine flexible Anpassung der Aufgabenschwierigkeit in Abhängigkeit von individuellen Präferenzen (z. B. Grandinetti 2013, Gottlieb 2016). Dies führte zu einem grundlegenden Gefühl der Sicherheit (Csikszentmihalyi 2000).

Das Gefühl von Sicherheit gilt als Voraussetzung für das *Flow*-Erleben und den Aufbau eines stabilen Selbstkonzeptes (Conradty & Bogner 2018, Justo 2008). Sicherheit unterstützt dabei die vollkommene Immersion in Aufgaben (Csikszentmihalyi 2000) und ermöglicht die Lösungsfindung ohne Angst vor Bewertung. Daher wird Lehrkräften allgemein empfohlen, die Rolle des Lernbegleiters und Gestalters der Lernumgebung einzunehmen, sodass Schüler ihre individuellen Stärken und Schwächen ausleben können (Justo 2008). Lehrkräfte in der Funktion des Lernbegleiters erhöhen zusätzlich das Sicherheitsempfinden der Schüler und fördert individuelle Lern- und Problemlösungsstrategien (Csikszentmihalyi 2000). Dies kann die Motivation (Marsh et al. 2015, Conradty et al. 2016), Kreativität (Conradty & Bogner 2018) und sogar den Lernerfolg (Conradty & Bogner 2018) positiv beeinflussen. Insgesamt soll Lernprozess dabei von den Schülern angestoßen und gesteuert werden (Justo 2008, Marsh et al. 2015). Dieser Ansatz wurde auch während des CLIL Moduls verfolgt, ist aber keinesfalls repräsentativ für den traditionellen Klassenunterricht, der sich noch immer auf Vorbereitung, Inkubation, Evaluation und Elaboration stützt (Csikszentmihalyi 1988, Justo 2008, Bournelli et al. 2009).

Neben Kreativität und Selbstkonzept konnte in Teilarbeit F auch ein Zusammenhang zwischen fachspezifischem Selbstkonzept und den Wissensvariablen der Biologie- und Englisch-Tests festgestellt werden. Dieses Ergebnis stimmt mit einschlägiger Referenzliteratur überein (Jansen et al. 2015, Arens et al. 2016, Wang et al. 2008, Mejía-Rodríguez et al. 2020). So wirkt sich ein positives fachspezifisches Selbstkonzept typischerweise verstärkend auf die individuellen Fähigkeiten aus und resultiert in entsprechenden Lernerfolgen (Patall et al. 2014). Dies bekräftigt wiederum das Selbstkonzept und führt zu mehr Selbstvertrauen sowie fachlicher Sicherheit (Justo 2008). Doch auch die Rolle von Gleichaltrigen und Lehrern in der Entwicklung eines gesunden Selbstkonzepts darf nicht unterschätzt werden. Abhängig von den Erwartungen und dem Vertrauen anderer in die eigenen Fähigkeiten kann das Selbstkonzept entweder gestärkt oder geschwächt werden, was sich in den schulischen Leistungen niederschlägt. Szumski & Karwowski (2019) benennen dieses Phänomen nach dem viktorianischen Bühnenstück von George Bernard Shaw "Pygmalion-Effekt". Daher ist es wichtig, ein bestätigendes und ermutigendes Lernumfeld zu schaffen (Blöte 1995, Burnett 2003), wie beispielsweise das praktische Experimentieren (Rodenhauser & Preisfeld 2014, Buse et al. 2018, Mierdel & Bogner 2019a, Teilarbeit A).

Studien, die das Selbstkonzept beim CLIL-Lernen untersuchen, sind rar und konzentrieren sich insbesondere auf Grund- oder Hochschulen. Im Gegensatz zu den Ergebnissen der Teilarbeit F konnten andere Studien feststellen, dass das Selbstkonzept bei CLIL-Lernenden niedriger ist als bei muttersprachlich unterrichteten Vergleichsgruppen (Roiha & Mäntylä 2019). Mögliche Gründe für die vergleichsweise niedrigeren Selbstkonzeptwerte könnten in den hohen Anforderungen an Sprachkompetenz liegen, durch die sich Schüler weniger kompetent in der Unterrichtssprache fühlen (Doiz et al. 2014). Im vorliegenden CLIL-Modul wurde dies jedoch mit entsprechenden Scaffolding-Maßnahmen (Coyle 2007, Grandinetti 2013) und Code-Switching (Königs 2013) vermieden. Schüler sollten inhaltliches Lernen in einer Fremdsprache erleben können, ohne an ihren Fähigkeiten zu zweifeln. Ganz im Gegenteil wirkten sich die intrinsische Motivation und das positive Feedback (Roiha & Mäntylä 2019) nach erfolgreicher Teilnahme am Modul besonders bestärkend auf das Selbstkonzept eigentlich leistungsschwacher Schüler aus. Bei leistungsstarken Schülern hatte die Teilnahme keinen signifikanten Einfluss. Dennoch konnte das CLIL-Modul Selbstkonzept und akademischen

Lernerfolg aufgrund der Wechselwirkung von Leistung und Selbstkonzept insgesamt positiv beeinflussen (Jansen et al. 2015, Arens et al. 2016).

Die bereits erwähnten Geschlechterunterschiede innerhalb der latenten Variable *Social* der Selbstkonzeptskala für das Fach Biologie sind – obwohl es nur wenige empirische Studien gibt (z.B. Wang et al. 2008) – gerade in den naturwissenschaftlichen Fächern bekannt. Weibliche Schüler nehmen sich trotz vergleichbarer Leistungen mit ihren männlichen Kollegen oft als weniger leistungsfähig wahr (Mullis et al. 2020, Mejía-Rodríguez et al. 2020). Obwohl für den Spracherwerb eine umgedrehte Verteilung der Geschlechterunterschiede erwartet wurde (z.B. Andreou et al. 2005, van der Slik et al. 2015), erzielten weibliche und männliche Schüler ähnliche Ergebnisse bezüglich ihres Selbstkonzepts. Diese Ausgeglichenheit der Geschlechter weist darauf hin, dass das CLIL-Modul gerade in Bezug auf das sprachliche Selbstkonzept weibliche und männliche Schüler gleichermaßen förderte.

3.6 Schlussfolgerungen und Ausblick

Die Ergebnisse der vorliegenden Arbeit sind auf den nur spärlich untersuchten Teilbereich der bilingualen Schülerlabore begrenzt (Rodenhauser & Preisfeld 2014, Buse et al. 2018) und können entsprechend schwer auf den Erfolg von CLIL im Schulalltag übertragen werden. Dennoch ist es möglich, aus den Ergebnissen Anhaltspunkte für die Unterrichtspraxis und fachdidaktische Forschung zur Förderung von Lernerfolg, Verständnis, Selbstkonzept und Kreativität durch bilinguale Module abzuleiten. Eine nähere Betrachtung derartiger Implikationen wird im Folgenden vorgenommen.

Ausgehend von einem eher konstruktivistischen Verständnis von Lernen sind besonders Erfahrungen und die daraus resultierende Konstruktion von Wissen ausschlaggebend für Lernerfolg (De Houwer et al. 2011, Hodson 1998, Lachman 1997). Entsprechend eignen sich Unterrichtskonzepte zur Lernförderung, die selbstgesteuerte Erkenntnisprozesse der Schüler durch das Aufstellen und Testen von Hypothesen zulassen (Schwarz et al. 2009, Abdulwahed & Nagy 2013, Lee et al. 2015). Das in der vorliegenden Arbeit erläuterte bilinguale Modul stellt ein solches Unterrichtskonzept dar. In Übereinstimmung mit einschlägiger Referenzliteratur (Rodenhauser & Preisfeld 2014, Buse et al. 2018) hat die Teilnahme an solch einem Modul, das zudem Hands-on-Minds-on-Aktivitäten zur Erkenntnisgewinnung einsetzt, positive Auswirkungen auf den Lernerfolg (Ke et al. 2020). Dies konnte in Teilarbeit A und Teilarbeit C bestätigt werden. Wo in Teilarbeit A jedoch Hands-on-Minds-on-Aktivitäten losgelöst vom Spracherwerb betrachtet wurden, konnte ein insgesamt besserer sowohl kurzfristiger als auch langfristiger Lernerfolg erzielt werden im Vergleich zur CLIL-Gruppe aus Teilarbeit C. In der Literatur wird von einer Überanstrengung des Arbeitsgedächtnisses ausgegangen, das infolge der hohen kognitiven Belastung (Sweller 2015) durch gleichzeitiges inhaltliches und sprachliches Lernen weniger inhaltliche Informationen ins Langzeitgedächtnis überführen kann (Johnstone & Wham 1982, Coyle 2007, Buse et al. 2018). Für die konkrete Umsetzung von CLIL in naturwissenschaftlichen Sachfächern bedeutet dies, dass es einen gewissen Zielkonflikt zwischen Inhalt und Sprache gibt, der in der Unterrichtsplanung berücksichtigt werden muss (Dalton-Puffer 2007, Grandinetti 2013). Da dennoch ein messbarer Lernerfolg im CLIL-Modul erzielt werden konnte, sowohl für inhaltliches Lernen (Teilarbeit C) als auch für Sprachenlernen (Teilarbeit E), sollte CLIL weiterhin im Lehrplan verbleiben. Gerade auch die Kommunikation über vorausgegangene sachfachliche Inhalte (Evnitskaya & Morton 2011, Kress et al. 2001) und der mögliche Einfluss auf die Scientific Literacy (Teilarbeit E) machen CLIL zu einer Brücke zwischen Sprache und Sachfach. Demzufolge sollte das Bewusstsein über einen existierenden Zielkonflikt zu einer Anpassung der kognitiven Belastung führen. Das heißt, der sachfachliche Inhalt könnte beispielsweise so reduziert werden, dass die Parallelisierung von inhaltlichem Lernen und Sprachenlernen funktioniert (Teilarbeit C, Teilarbeit E), ohne zu einer sich gegenseitig verstärkenden Benachteiligung durch geistige Überlastung zu führen. Auch könnten die Differenzierung der Aufgabenschwierigkeiten und das inhaltliche sowie sprachliche Scaffolding verbessert werden (Maybin et al. 1992, Gottlieb 2016, Fernández-Fontecha 2020). Bei inhaltlich besonders komplexen Themen könnte ein Teil der relevanten Fachtermini auch bereits vorab eingeführt werden, um eine gewisse Latenz zwischen beiden Lerndimensionen zu erzielen.

Aufgrund der Ergebnisse von Teilarbeit D ist zudem zu überlegen, CLIL-Module in Kurzzeitform insbesondere bei Themen einzusetzen, die nicht auf Faktenwissen, sondern auf Tiefenverständnis zielen. Denn wie in Teilarbeit D gezeigt wird, scheint der Besuch des CLIL-Moduls der Fehlvorstellung von Modellen als exakte Kopie der Realität (Grosslight et al. 1991, Krajcik & Merritt 2012) langfristig entgegengewirkt zu haben. Stattdessen verstanden die Schüler, dass der Stand der Forschung und der zugehörige wissenschaftliche Diskurs ausschlaggebend für die Erstellung und Veränderung von Modellen sind (Treagust et al. 2002, Mendonça & Justi 2013, Ke et al. 2020). Dies wurde auch bei der Kategorisierung der offenen Fragen mehrfach wörtlich genannt und unterstützt damit die durch den Modellverständnisfragebogen identifizierten Tendenzen. In der einschlägigen Referenzliteratur wird das Verständnis komplexer Phänomene durch Deep-Learning-Prozesse erklärt (Lee et al. 2015, Case & Gunstone 2010). Diese Prozesse werden während der Übertragung von inhaltlichem Wissen in eine andere Darstellungsform, dem wiederkehrenden diskursiven Austausch (Quarderer & McDermott 2020, France 2019) und der kritischen Reflexion eigener Erkenntnisse angestoßen (Ke &

Schwarz 2020). Da gerade beim Sprachenlernen sachfachliche Inhalte aufgrund der Sprachbarriere oft nicht sofort erfasst werden, müssen Schüler Strategien anwenden, um an die Informationen zu gelangen, die für Deep-Learning wichtig sind. Dazu zählt das Aufstellen von Hypothesen, das Herantasten an Erklärungsansätze und die kritische Evaluation der erzielten Ergebnisse in Reflexion mit den Ergebnissen anderer Schüler oder Lösungshilfen (Lee et al. 2015). So kann CLIL beispielsweise bei kontroversen naturwissenschaftlichen Themen zum Einsatz kommen, um eine rege Diskussion anstoßen. Der Inhalt sollte jedoch vorab bereits vollständig durchdrungen sein. Auch für das Verständnis von komplexen naturwissenschaftlichen Phänomenen, wie der DNA, die zusätzlich den Einsatz von Modellen benötigen, könnte CLIL gewinnbringender sein als der klassische muttersprachliche Unterricht.

Neben dem objektiv messbaren Lernerfolg und dem Verständnis von Inhalt und Sprache (Teilarbeit E) zeichnet sich akademischer Erfolg besonders durch subjektive Faktoren aus, die oft in der Wahrnehmung der eigenen Leistung liegen (Hoferichter et al. 2018). Dies wird allgemein als Selbstkonzept bezeichnet und kann die akademische Leistung positiv oder negativ beeinflussen. Dabei gelten Selbstkonzept und Leistung als sich gegenseitig verstärkend und voneinander abhängig (Arens et al. 2016, Jansen et al. 2015). Entsprechend ist es wichtig, in der Schule Lernumgebungen zu schaffen, die das Selbstkonzept positiv beeinflussen, sodass sich die Leistung ebenfalls verbessert (Blöte 1995, Burnett 2003). Da im CLIL-Modul besonders das Selbstkonzept leistungsschwächerer Schüler angesprochen wurde (Teilarbeit F), scheint diese Lernumgebung besonders förderlich zu sein. Die einschlägige Literatur äußert sich jedoch generell eher kritisch gegenüber CLIL-Modulen (Roiha & Mäntylä 2019, Doiz et al. 2014), da die sprachlichen Fähigkeiten, um souverän sachfachliche Inhalte zu kommunizieren, noch vollkommen nicht ausgereift sind. Während der Teilnahme am CLIL-Modul konnten aber in Teilarbeit E keine negativen Effekte für leistungsstarke Schüler festgestellt werden. Konkrete Implikationen für Selbstkonzept-förderliche Lernumgebungen umfassen demnach in einem ersten Schritt viele selbstgesteuerte Erkenntnisprozesse ohne sofortige Bewertung durch die Lehrkraft (Csikszentmihalyi 2000, Marsh et al. 2015). Darüber hinaus bietet die Kombination aus Sprachlernen und naturwissenschaftlichen Inhalten eine Bandbreite an möglichen Situationen für Kompetenzerleben. Wer sich eher als sprachlich begabt identifiziert, könnte über dieses positive Selbstkonzept auch die Wahrnehmung der eigenen Leistung in naturwissenschaftlichen Fächern verbessern und umgekehrt (Jansen et al. 2015, Arens et al. 2016, Wang et al. 2008, Mejía-Rodríguez et al. 2020). Daneben bietet das CLIL-Modul Entfaltungsmöglichkeiten für eher praktisch orientierte Schüler, die durch Experimente und die zugehörige Modellkonstruktion Bestätigung erfahren und sich motiviert fühlen (Rodenhauser & Preisfeld

2014, Buse et al. 2018). Aber auch Theoretiker, die sich eher in mentalen Modellen und dem Generieren von Hypothesen verlieren, können in diesem Modul ihre Stärken ausspielen (Justo 2008). Insgesamt scheint es also wichtig zu sein, dass der dargebotene Unterricht variable Anknüpfungsmöglichkeiten für oft sehr heterogene Schülergruppen bietet. Eine Möglichkeit, dies im Klassenunterricht zu bewerkstelligen, sind offene Lernumgebungen, in denen Schüler ihre Lernmethoden selbst bestimmen können (Blöte 1995, Burnett 2003, Justo 2008, Marsh et al. 2015).

Eine ähnliche Lernumgebung wird auch für kreativitätsfördernden Unterricht vorgeschlagen. Wie in Teilarbeit F beleuchtet, korrelieren Selbstkonzept und Kreativität auf mehreren Ebenen (Fleith et al. 2002). Aber auch verschiedene Persönlichkeitsfaktoren, wie Teilarbeit B zeigt, werden mit Kreativität in Verbindung gebracht (Kaufman et al. 2008, Barron & Harrington 1981). Beide Studien haben gemein, dass Kreativität nur unter bestimmten Voraussetzungen erlebt werden kann. Besonders die positive Emotion *Flow*, die mit dem kompletten Eintauchen in eine Aufgabe verbunden ist und oft so weit führt, dass die betreffende Person das Gefühl für Raum und Zeit verliert (Cseh 2016, Piniel & Alber 2020), ist von zahlreichen intrinsischen und extrinsischen Faktoren abhängig (Richards & Cotterall 2016, Baer 2010, Amabile 1983). Neben den bereits für das Selbstkonzept genannten Ansprüchen an eine förderliche Lernumgebung kommt für *Flow*-Erleben noch das Bedürfnis nach Sicherheit hinzu. Dies kann besonders dann erzielt werden, wenn Lehrkräfte die Rolle von Lernbegleitern einnehmen, die Bewertung von Lernergebnissen nicht immer sofort erfolgt und auch Stundenbegrenzungen der einzelnen Fächer aufgehoben werden (Justo 2008, Marsh et al. 2015). Je mehr Kontrolle über den eigenen Lernprozess entsprechend in die Hände der Schüler gegeben wird, desto höher ist auch die intrinsische Motivation (Marsh et al. 2015, Conradty & Bogner 2016).

So zeigt das CLIL-Modul insgesamt eine Reihe an positiven Implikationen für den naturwissenschaftlichen Unterricht und fördert Lernerfolg, Verständnis, Selbstkonzept und Kreativität. Da viele der vorgeschlagenen Änderungen jedoch eine grundlegende Umgestaltung der Lernumgebung und des Verständnisses von Unterricht erfordern, ist es eher unwahrscheinlich, dass diese, abgesehen von vereinzelten Interventionen, in der Breite angenommen werden. Häufig ist das aber noch nicht einmal notwendig. Denn wie zuvor erläutert, kann auch der gelegentliche, wenngleich gezielte Einsatz von CLIL-Modulen im Kurzformat, bei wichtigen und auf Verständnis abzielenden Themen bereits genügen.

4 LITERATURVERZEICHNIS DER SYNOPSIS

Abdulwahed, M., & Nagy, Z. K. (2013). Applying Kolb's experiential learning cycle for laboratory education. *Journal of English Education*, 98, 283–294.

Acevedo, J. A. (2008). El estado actual de la naturaleza de la ciencia en la didáctica de lasciencias [The state of the art on nature of science in Science Education]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias,* 5(2), 134–169.

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation,* 12, 75–93.

Aguilar-Alonso, A. (1996). Personality and creativity. *Personality and Individual Differences,* 21(6), 959-969.

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). London: Springer.

Ajideh, P., & Mozaffarzadeh, S. (2012). C-test vs. multiple-choice cloze test as tests of reading comprehension in Iranian EFL context: Learners' perspective. *English Language Teaching*, 5(11), 143-150.

Akerson, V.L., Townsend, J.S., Donnelly, L.A., Hanson, D.L., Tira, P., & White, O. (2009). Scientific modeling for inquiry teachers network (SMIT'N): the influence on elementary teachers' views of nature of science, inquiry, and modeling. *Journal of Science Teachers Education,* 20, 21-40.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology,* 45, 357-376.

Amabile, T.M., & Pratt, M.G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior,* 36, 157-183.

Andreou, G., Vlachos, F., & Andreou, E. (2005). Affecting factors in second language learning. *Journal of Psycholinguistic Research,* 34, 429-438.

Antinori, A., Carter, O. L., & Smillie, L. D. (2017). Seeing it both ways: Openness to experience and binocular rivalry suppression. *Journal of Research in Personality,* 68, 15–22.

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., and vom Hofe, R.

(2016). Math self-concept, grades, and achievement test scores: long-term reciprocal effects across five waves and three achievement tracks. *Perspectives on Psychological Science,* 133–163.

Baer, J. (2010). Gender differences in the effects of anticipated evaluation on creativity. *Creativity Research Journal,* 10(1), 25-31.

Bakadorova, O., & Raufelder, D. (2019): The relationship of school self-concept, goal orientations and achievement during adolescence. *Self and Identity,* DOI: 10.1080/15298868.2019.1581082.

Bamberger, Y. M., & Davis, E. A. (2011). Middle-school science students' scientific modelling performances across content areas and within a learning progression. *International Journal of Science Education,* 35(2), 213–238.

Barron, F. & Harrington, D. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology,* 32, 439–76.

Bećirović, S. (2017). The relationship between gender, motivation and achievement in learning English as a foreign language. *European Journal of Contemporary Education,* 6(2), 210-220.

Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences,* 20(5), 532–535.

Beghetto, R. A. (2015). *Teaching creative thinking in K12 schools. The Routledge International Handbook of Research on Teaching Thinking.* New York: Routledge.

Berland, L.K., & Reiser, B.J. (2009). Making sense of argumentation and explanation. *Science Education,* 93(1), 26–55.

Biggs, J.B. (1987). *Student approaches to learning and studying.* Melbourne: Australian Council for Educational Research.

Blöte, A.W. (1995). Students' self-concept in relation to perceived differential teacher treatment. *Learning and Instruction,* 5, 221-236.

Bogner, F.X. (1995). Eine bilinguale Unterrichtseinheit zum Thema Glas (mit konsequenter Umsetzung des Erkundungsprinzips vor Ort). *Chemkon,* 2(1), 9-16.

Bonnet, A. (2004). *Chemie im bilingualen Unterricht: Kompetenzerwerb durch Interaktion.* [Chemistry in bilingual education: competencies through interaction]. Opladen: Leske & Budrich.

Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal,* 1(4), 65–78.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation* [Research methods and evaluation]. Berlin: Springer.

Bortz, J., & Schuster, C. (2010). Faktorenanalyse [Factor Analysis]. In: Bortz, J., & Schuster, C. (Eds.). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social scientists] (pp. 385-426). Berlin, Heidelberg, New York: Springer Verlag.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research,* 31, 659–671.

Bournelli, P., Makri, A., & Mylonas, K. (2009). Motor creativity and self-concept. *Creativity Research Journal,* 21(1), 104-110.

Boyer, S.E., & Lamoreaux, C.D. (1997). Flow theory as a construct for analyzing learning environments in a 7th-grade science classroom. *Annual Meeting of the American Educational Research Association, Chicago, IL.* (ERIC Document Reproduction Service No.ED409215).

Bradbury, L. (2014). Linking science and language arts: a review of the literature which compares integrated versus non-integrated approaches. *Journal of Science Teacher Education,* 25(4), 465–488.

Bro, R., & Smilde, A.K. (2014). Principal component analysis. *Analytical Methods,* 6, 2812-2831.

Bryce, C. M., et al. (2016). Exploring models in the biology classroom. *American Biology Teacher,* 8, 35–42.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* New York, NY: WW Norton & Company.

Buse, M., Damerau, K. & Preisfeld, A. (2018). A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts. *International Journal of Environmental & Science Education,* 13(8), 647-660.

Burnett, P.C. (2003). The impact of teacher feedback on student self-talk and self-concept in reading and mathematics. *The Journal of Classroom Interaction,* 38(1), 11-16.

Buxton et al. (2019). Understanding science and language connections: new approaches to assessment with bilingual learners. *Research in Science Education,* 49, 977-988.

Cacioppe, R. (2017). Integral mindflow: A process of mindfulness-in-flow to enhance individual and organization learning. *The Learning Organization,* 24(6), 408–417.

Campbell, T., & Fazio, X. (2020). Epistemic frames as an analytical framework for understanding the representation of scientific activity in a modeling-based learning unit. *Research in Science Education,* 50, 2283-2304.

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The "big five questionnaire": A new questionnaire to assess the five-factor model. *Personality and Individual Differences,* 15(3), 281–288.

Carciofo, R., Yang, J., Song, N., Du, F., & Zhang, K. (2016). Psychometric evaluation of Chinese-language 44-item and 10-item big five personality inventories, including correlations with chronotype, mindfulness and mind wandering. *PLoS One,* 11(2), e0149963. https://doi.org/10.1371/journal.pone.0149963.

Carmel, J.H., et al. (2020). Helping students to „do science": characterizing scientific practices in general chemistry laboratory curricula. *Journal of Chemical Education,* 96(3), 423-434.

Carpenter, S.L., et al. (2018). Models are a "metaphor in your brain": How potential and preservice teachers understand the science and engineering practice of modeling. *Science Education,* 119, 275–286.

Case, J., & Gunstone, R. (2002). Metacognitive development as a shift in approach to learning: an in-depth study. *Studies in Higher Education,* 27(4), 459–470.

Cheshire, J., & Gardner-Chloros, P (1998). Code-switching and the sociolinguistic gender pattern. *International Journal of the Sociology of Language,* 129, 5–34.

Chin, C., & Brown, D.E. (2000). Learning in science: a comparison of deep and surface approaches. *Journal of Research in Science Teaching,* 37(2), 109–138.

Chui, L., & Lee, C. (2012) Exploring the impact of flow experience on job performance. *Journal of Global Business Management,* 8 (2), 150-158.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70, 213–220.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science,* 1(3), 98–101

Conradty, C., & Bogner, F.X. (2016). Hypertest or textbook: Effects on motivation and gain in knowledge. *Education Sciences,* 6(3), 29, 1-16.

Conradty, C., & Bogner, F.X. (2018). From STEM to STEAM: How to monitor creativity. *Creativity Research Journal,* 30(3), 233-240.

Conradty, C., & Bogner, F.X. (2019). From STEM to STEAM: Cracking the code? How creativity & motivation interacts with inquiry-based learning. *Creativity Research Journal,* 31(3), 284-295.

Cook, T. D., & Campell, D. (1979). *Quasi-experimentation. Design & analysis issues for field settings.* Chicago: Rand McNally College Publishing Company.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism,* 10(5), 543-562.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior,* 11(6), 671-684.

Cseh, G. M. (2016). Flow in creativity: a review of potential theoretical conflict. In *Flow Experience* (pp. 79–94). Cham: Springer.

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In R. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 325–339). New York, NY: Cambridge University Press.

Csikszentmihalyi, M. (2000). Happiness, flow, and economic equality. *American Psychologist,* 55(10), 1163–1164.

Csikszentmihalyi, M. (2015). *Flow und Kreativität: Wie Sie Ihre Grenzen überwinden und das Unmögliche schaffen* [Flow and creativity: How to overcome your limits and achieve the impossible]. Stuttgart: Klett-Cotta.

Cummins, J. (1981). *Bilingualism and minority-language children.* Toronto: OISE Press.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated learning (CLIL) Classrooms.* Amsterdam: John Benjamins.

De Houwer, J., Barnes Holmes, D., & Moors, A. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science,* 6, 202–209.

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review,* 72, 219–238.

Demerouti, E. (2006). Job Characteristics, Flow, and Performance: The Moderating Role of Conscientiousness. *Journal of Occupational Health Psychology,* 11, 266-280.

Diehr, B. (2012). What´s in a name? Terminologische, typologische und programmatische Überlegungen zum Verhältnis der Sprachen im Bilingualen Unterricht. [What's in a name? Reflections on terminology, typology and programme regarding the relation of languages in bilingual education]. In B. Diehr & L. Schmelter (Eds.), *Bilingualen*

Unterricht weiterdenken. Programme, Positionen, Perspektiven (pp.17-36). Frankfurt a.M.: Peter Lang.

Doiz, A., Lasagabaster, D., & Sierra, J.M. (2014). CLIL and motivation: the effect of individual and contextual variables. *The Language Learning Journal,* 42(2), 209-224.

Donaldson, J.P. & Allen-Handy, A. (2020). The nature and power of conceptualizations of learning. *Educational Psychology Review,* 32, 545-570.

Dunlosky, J., et al. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest,* 14(1), 4-58.

Ellis, P. D. (2010). *The essential guide to effect sizes statistical power, meta-analysis, and the interpretation of research results.* Cambridge: Cambridge University Press.

European Commission (2004): *Promoting language learning and linguistic diversity. An action plan 2004-06.* Luxembourg, European Communities.

European Commission (2012). *FAQs on multilingualism and language learning.* Brussels. Retrieved from http://europa.eu/rapid/press-release_MEMO-12-703_en.pdf

Erkol, M., Kışoğlu, M., & Büyükkasap, E. (2010). The effect of implementation of science writing heuristic on students' achievement and attitudes toward laboratory in introductory physics laboratory. *Procedia-Social and Behavioral Sciences,* 2(2), 2310–2314.

Evnitskaya, N. & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language and Education,* 25(2), 109-127.

Fazio, X., & Gallagher, T.L. (2018). Bridging professional teacher knowledge for science and literary integration via design-based research. *Teacher Development,* 22(2), 267-280.

Fehling, Sylvia (2008): *Language Awareness und bilingualer Unterricht – Eine komparative Studie.* Frankfurt am Main: Peter Lang.

Fernández-Fontecha, A., O'Halloran, K.L., Wignell, P., & Tan. S. (2020). Scaffolding CLIL in the science classroom via visual thinking: a systematic functional multimodal approach. *Linguistics and Education,* 55, 1-10.

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J.M. (2017). Analysing students' content-learning in science in CLIL vs. Non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism,* 22(6), 661-674.

Field, A. (2012). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.

Filippetti, V.A., & Krumm, G. (2020). A hierarchical model of cognitive flexibility in children: Extending the relationship between flexibility, creativity and academic achievement. *Child Neuropsychology,* DOI: 10.1080/09297049.2019.1711034.

Finn, Alke (2012): *Bio bilingual. Entwicklung und Durchführung eines Unterrichtskonzeptes zu bilingualem Unterricht.* Marburg: Tectumverlag.

Fleith, D. S., Renzulli, J. S., & Westberg, K. L. (2002). Effects of a creativity training program on divergent thinking abilities and self-concept in monolingual and bilingual classrooms. *Creativity Research Journal,* 14, 373-386.

France, A. (2019). Teachers using dialogue to support science learning in the primary classroom. *Research in Science Education.* https://doi.org/10.1007/s11165-019-09863-3.

Franke, G. (2010). *Untersuchungen zum Wissenserwerb, zur kognitiven Belastung und zu emotionalen Faktoren im experimentellen Unterricht über Grundlagen der Gentechnik im Lernort Labor unter besonderer Berücksichtigung von Schülervorstellungen.* Dissertation. Bayreuth.

Franke, G., & Bogner, F.X. (2013). How does integrating alternative conceptions into lessons influence pupils' situational emotions and learning achievement. *Journal of Bilingual Education,* 47(1), 1-11.

Franco C. & Colinvaux D. (2000). Grasping Mental Models. In: Gilbert J.K., Boulter C.J. (Eds.) *Developing models in science education* (pp. 93-118). Springer.

Fraser, B. J. (1982). The development of short forms of several classroom environment scales. *Journal of Educational Measurement*, 19, 143–154.

Furnham, A., Hughes, D. J., & Marshall, E. (2013). Creativity, OCD, narcissism and the Big Five. *Thinking Skills and Creativity,* 10, 91-98.

Gadja, A. (2017). The relationship and moderators of school achievement and creativity at different educational stages. *Thinking Skills and Creativity,* 24, 250-267.

Gilbert, J. K., & Justi, R. S. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education,* 24, 369–387.

Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: achieving scientific literacy. *Journal of Research in Science Teaching,* 31(9), 1057-1073.

Gao, L.X., & Zhang, L.J. (2020). Teacher learning in difficult times: examining foreign language teachers' cognitions about online teaching to tide over COVID-19. *Frontiers in Psychology,* DOI: 10.3389/fpsyg.2020.549653.

Goldman, A. D., & Penner, A. M. (2016). Exploring international gender differences in mathematics self- concept. *International Journal of Adolescence and Youth,* 21(4), 403–418.

Goldschmidt, M. (2015). *Unterricht zum Thema Grüne Gentechnik im außerschulischen Lernort Labor: Schülervorstellungen, unterrichtliche Umsetzung von Schülerdiskussionen und geschlechtsspezifische Unterschiede.* Dissertation. Bayreuth.

Goldschmidt, M., & Bogner, F.X. (2016). Learning about genetic engineering in an outreach laboratory: Influence of motivation and gender on students' cognitive achievement. *International Journal of Science Education,* 6(2), 166-187.

Gomez, R., Holmberg, K., Bounds, J., Fullarton, C., & Gomez, A. (1999). Neuroticism and extroversion as predictors of coping styles during early adolescence. *Personality and Individual Differences,* 27, 3–17.

Gonzalez-Howard, M. & McNeill, K.L. (2016). Learning in a community of practice: factors impacting English-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching,* 53(4), 527-553.

Goodwin, R. D., & Gotlib, I. H. (2004). Gender differences in depression: the role of personality factors. *Psychiatry Research,* 126(2), 135–142.

Gottlieb, M. (2016). *Assessing English language learners: Bridges for language proficiency to academic achievement.* Thousand Oaks: Corwin.

Grandinetti, M., Langellotti, M., & Teresa Ting, Y.L. (2013). How CLIL can provide a pragmatic means to renovate science education – even in a sub-optimally bilingual context. *International Journal of Bilingual Education and Bilingualism,* 16(3), 354-374.

Grosslight, L., Unger, C., & Jay, E. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *Journal of Research in Science Teaching,* 28(9), 799–822.

Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education,* 36(10), 1651–1684.

Haagen-Schützenhöfer, C., Mathelitsch, L., & Hopf, M. (2011). Fremdsprachiger Physikunterricht: Fremdsprachlicher Mehrwert auf Kosten fachlicher Leistungen? Auswirkungen fremdsprachenintegrierten Physikunterrichts auf fachliche Leistungen [Foreign language physics lessons: Added value for language skills at the cost of content achievement? Effects of content-language-integrated physics classes on content achievement]. *Zeitschrift für Didaktik der Naturwissenschaften,* 17, 223–260.

Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. *Creative Education,* 3(5), 603-611.

Hager, N., & Hörz, H. (1977). Modelle und Modellmethode in der wissenschaftlichen Erkenntnise. *Deutsche Zeitschrift für Philosophie,* 25(2), 164-179.

Halliday, M. A. K. (2003). Introduction: On the 'Architecture' of Human Language. In J.J. Webster (eds.) *Collected Works of M.A.K. Halliday* (pp. 1–35). London: Continuum.

Hardy, G. (2014). Academic self-concept: Modeling and measuring for science. *Research in Science Education,* 44, 549-579.

Haupt, Olaf J., et al. (2013): Schülerlabor – Begriffsschärfung und Kategorisierung. In: *Der Mathematische und Naturwissenschaftliche Unterricht,* 66(6), 324-330.

Henriksen, D., Mishra, P., & Fisser, P. (2018). Infusing Creativity and Technology in 21st Century Education: A Systemic View for Change. *Educational Technology & Society,* 19(3), 27–37.

Hodson, D. (1998). *Teaching and Learning Science - a personalized approach.* Buckingham: Open University Press.

Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education,* 36(15), 2534-2553.

Hoening, K., & Wenz, S.E. (2020). Education, health behavior, and working conditions during the pandemic: evidence from a German sample. *European Societies,* DOI: 10.1080/14616696.2020.1824004.

Hoferichter, F., Lätsch, A., Lazarides, R., & Raufelder, D. (2018). The big-fish-little-pond effect on the four facets of academic self-concept. *Frontiers in Psychology,* 9(1247), 1-11.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education,* 88, 28–54.

Hoseinifar, J., Siedkalan, M. M., Zirak, S. R., Nowrozi, M., Shaker, A., Meamar, E., et al. (2011). An investigation of the relation between creativity and five factors of personality in students. *Procedia-Social and Behavioural Sciences,* 30, 2037–2041.

Howell, M. E., Booth, C. S., Sikich, S. M., Helikart, T., Roston, R. L., Couch, B. A., & van Dijk, K. (2019). Student understanding of DNA structure–function relationships improves from using 3D learning modules with dynamic 3D printed models. *Biochemistry and Molecular Biology Education,* 47(3), 303–317.

Huber, S.G., & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises - reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability,* 32, 237-270.

Iimura, S., & Taku, K. (2018). Gender differences in relationship between resilience and Big Five personality traits in Japanese adolescents. *Psychological Reports,* 121(5), 920-931.

IBM Corp. (2020a). *IBM SPSS Statistics for Windows (Version 26).* IBM Corp. Armonk, New York.

IBM Corp. (2020b). *IBM SPSS Amos for Windows (Version 26).* IBM Corp. Armonk, New York.

ISB (2021). *Staatsinstitut für Schulqualität und Bildungsforschung: LehrplanPLUS.* URL: https://www.lehrplanplus.bayern.de/bildungs-und-erziehungsauftrag/textabsatz/44372 (acessed: 01.03.2021).

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Science Educator,* 17, 10–17.

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology,* 41, 13-24.

Johnstone, A.H. & Wham, A.J.B. (1979). A model for undergraduate practical work. *Education in Chemistry,* 16, 16–17.

Justo, C.F. (2008). Creative relaxation, motor creativity, self-concept in a sample of children from early childhood education. *Electronic Journal of Research in Educational Psychology,* 14(6), 29-50.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika,* 35(4), 401–415.

Kalyuga S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction,* 19, 402-410.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. John Wiley & Sons.

Kaufman, J.C., & Beghetto, R.A. (2009). Beyond big and little: the four C model of creativity. *Review of General Psychology,* 13(1), 1-12.

Ke, L., & Schwarz, C. (2020). Using epistemic considerations in teaching: Fostering students' meaningful engagement in scientific modeling. In A. Upmeier zu Belzen, D. Krüger, & J. Van Driel (Eds.), *Towards a Competence-based View on Models and Modeling in Science Education* (pp. 181–199). Springer International Publishing.

Ke, L., Zangori, L., Sadler, T.D., & Friedrichsen, P.J. (2020). Integrating scientific modeling and socio-scientific reasoning to promote scientific literacy. In Warden A. Powell (Ed.), *Socioscientific issues-based instruction for scientific literacy development* (pp. 31-56). Hershey: IGI Global.

Kim, et al. (2015). Cycles of exploration, reflection, and consolidation in model-based learning of genetics. *Journal of Science Education and Technology,* 24(6), 789– 802.

Kind, V. (2015). Preservice science teachers' science teaching orientations and beliefs about science. *Journal of Science Education,* 100, 122–152.

Kind, P.M., & Kind, V. (2007). Creativity in science education: perspectives and challenges for developing school science. *Studies in Science Education,* 43(1), 1-37.

Kitamura, S., et al. (2015). A structural model of age, grey matter volumes, education, and personality traits. *Psychogeriatrics,* 16(1), 46–53.

Klieme, E., et al. (Eds.) (2010). *PISA 2009: Bilanz nach einem Jahrzehnt* [PISA 2009: Review after a decade]. Münster: Waxmann.

KMK (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

KMK (2013). *Konzepte für den bilingualen Unterricht - Erfahrrungsbericht und Vorschläge zur Weiterentwicklung* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany – concepts for bilingual teaching – progress report and suggestions for further development]. Munich: Luchterhand.

Königs, F.G. (2013): Einsprachigkeit, Zweisprachigkeit und Code-Switching. In: Wolfgang Hallet und Frank G. Königs (Eds.): *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning.* Seelze: Friedrich Verlag GmbH, S. 174–180.

König, J., Jäger-Biela, D.J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education,* 43(4), 608-622.

Korthagen, F. & Lagerwerf, B. (1995). Levels in learning. *Journal of Research in Science Teaching,* 32(1), 1011-1038.

Kovanović, V., et al. (2018). Understanding the relationship between technology-use and cognitive presence in MOOCs. *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge,* 582-583. New York: Association for Computing Machinery.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom?. *Science and Children,* 49, 10–13.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science,* 328(5977), 456–459.

Krashen, S.D. (1985). Inquiries and insights: *Second language teaching immersion & bilingual education literacy.* Englewood Cliffs, NJ: Alemany Press.

Krechel, H.-L. (2013). Organisationsformen und Modelle in weiterführenden Schulen. [Organisation and models of bilingual education]. In W. Hallet & F. G. Königs (Eds.), *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (pp.74–80). Seelze: Friedrich Verlag.

Kress, G., C., Ogborn, J.J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom.* London: Continuum.

Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *Journal of Psychology,* 131(5), 477–480.

Lam, S. F.,Wong, B. P. H., Yang, H., & Liu, Y. (2012). Understanding student engagement with a contextual model. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 403-420). Boston, MA: Springer US.

Langheinrich, J. (2015). *eLearning-gestützter Unterricht am außerschulischen Lernort Labor: Schülervorstellungen und kognitives Lernen.* Dissertation. Bayreuth

Langheinrich, J., & Bogner, F.X. (2015). Students conceptions about the DNA structure within a hierarchical organizational level: Improvement by experiment- and computer-based outreach learning. *Biochemistry and Molecular Biology Education,* 43(6), 393-402.

Langheinrich, J., & Bogner, F.X. (2016). Computer-related self-concept: The impact on cognitive achievement. *Studies in Educational Evaluation,* 50, 46-52.

Lee, S., Kang, E., & Kim, H-B. (2015). Exploring the impact of students‘ learning approach on collaborative group modeling of blood circulation. *Journal of Science Education Technology,* 24, 234-255.

Lemke, J. L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Lian, B., Kristiawan, M., & Fitriya, R. (2018). Giving creativity room to students through the friendly school's program. *International Journal of Scientific and Technology Research,* 7(7), 1-7.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lindahl, K. (2020). Connecting ideology and awareness: critical multilingual awareness in CLIL contexts. *English Teaching & Learning,* 44, 211-228.

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Little, D. (1997). Language awareness and the autonomous language learner. *Language Awareness,* 6(2-3), 93-104.

Lo, Y.Y. (2017). Development of the beliefs and language awareness of content subject teachers in CLIL: does professional development help?. *International Journal of Bilingual Education and Bilingualism,* DOI: 10.1080/13670050.2017.1318821.

Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning (CLIL) in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics,* 31, 418–442.

Mady, C., & Seiling, A. (2017). The coupling of second language learning motivation and achievement according to gender. *Theory and Practice in Language Studies,* 7(12), 1149-1159.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods,* 1(2), 130–149.

Marian, V. & Fausey, C.M. (2006). Language-dependent memory in bilingual learning. *Applied Cognitive Psychology,* 20, 1025-1047.

Marsh, H.W., Hau, K.T., & Kong, K.W. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English vs. Chinese) for Hong Kong students. *American Educational Research Journal,* 39, 727– 763.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: relations and causal ordering. *British Journal of Educational Psychology,* 81, 59–77.

Marsh, H.W., et al. (2015). The internal/external frame of reference model of self-concept and achievement relations: age-cohort and cross-cultural differences. *American Educational Research Journal,* 52(1), 168-202.

Mawang, L.L., Kigen, E.M., & Mutweleli, S.M. (2018). The relationship between musical self-concept and musical creativity among secondary school music students. *International Journal of Music Education,* 1-13.

Maxwell-Reed, C. (2017). Classroom discourse in bilingual secondary science: language as medium or language as dialectic? *International Journal of Bilingual Education and Bilingualsim,* 23(4), 499-512.

Maybin, J., Mercer, N. & Stierer, B. (1992). Scaffolding: learning in the classroom. In Kate Norman (Ed.), *Thinking Voices. The Work of the National Oracy Project* (pp.186–95). London: Hodder & Stoughton.

McGuiness, C. (1999) From Thinking Skills to Thinking Classrooms: A Review and Evaluation of Approaches for Developing Pupils' Thinking. *Research Report*, 115, DfEE: HMSO.

Mejía-Rodríguez, A.M., Luyten, H., & Meelissen M.R.M. (2020). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. *International Journal of Science and Mathematics Education,* DOI: org/10.1007/s10763-020-10100-x.

Mendonça, P.C.C., & Justi R. (2013). The relationships between modelling and argumentation from the perspective of the model of modelling diagram. *International Journal of Science Education,* 35(14), 2407–2434.

Meskill, C., & Oliveira, A.W. (2019). Meeting the challenges of English learners by pairing science and language educators. *Research in Science Education,* 49, 1025-1040.

Meyerhöffer, N. & Dreesmann, D.C. (2019). English-bilingual Biology for standard classes development, implementation and evaluation of an English-bilingual teaching unit in standard German high school classes. *International Journal of Science Education,* 41(10), 1366-1386.

Mierdel, J., & Bogner, F.X. (2019a). Comparing the use of two different model approaches on students' understanding of DNA models. *Education Sciences,* 9(2), 115-133.

Mierdel, J., & Bogner, F. X. (2019). Is creativity, hands-on modeling and cognitive learning gender-dependent? *Thinking Skills and Creativity,* 31, 91–102.

Mierdel, J. (2019c). *Kreatives Modellieren im Lernort Labor. Eine vergleichende Studie zu Wissenserwerb, Modellverständnis und dem Einfluss individueller Kreativität.* Dissertation. Bayreuth.

Mierdel, J., & Bogner, F.X. (2020). Simply InGEN(E)ious! How creative DNA modeling can enrich classic hands-on experimentation. *Journal of Microbiology and Biology Education,* 21(2), 21-45.

Miller, A. L. (2014). A self-report measure of cognitive processes associated with creativity. *Creativity Research Journal,* 26(2), 203–218.

Miller, A. L., & Dumford, A. D. (2016). Creative cognitive processes in higher education. *The Journal of Creative Behavior,* 50(4), 282–293.

Moje, E.B., Collazo, T., Carrillo, R., & Marx, R.W. (2001). "Maestro, what is 'quality'?": language, literacy, and discourse in project-based science. *Journal of Research in Science Teaching,* 38(4), 469-498.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science.* URL: https:// timssandpirls.bc.edu/timss2019/international-results/ (accessed 01.03.2021).

Nation, I. S. P., & Newton, J. (2008). *Teaching ESL/EFL listening and speaking.* London: Routledge.

Norris, S.P., & Phillips, L.M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224-240.

Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education,* 33(8), 1109–1130.

Oktaviani, A., & Fauzan, A. (2017). Teachers perceptions about the importance of English for young learners. *Linguistic, English Education and Art (LEEA) Journal,* 1(1), 1-15.

Parthey, H. (1983). Experiment – Modell – Theorie. *Deutsche Zeitschrift für Philosophie,* 31(5), 632-633).

Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. *School Science and Mathematics,* 109, 394–402.

Passmore C., & Svoboda J. (2012). Exploring opportunities for argumentation in modelling classrooms. *International Journal of Science Education,* 34(10), 1535–1554.

Patall, E.A., Awad, G.H., & Cestone, C.M. (2014). Academic potential beliefs and feelings: Conceptual development and relations with academic outcomes. *Self and Identity,* 13(1), 58-80.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine,* 50, 157–175.

Peel, A., Zangori, L., Friedrichsen, P., Hayes, E., & Sadler, T. (2019). Students‘ model-based explanations about natural selection and antibiotic resistance through socio-scientific issues-based learning. *International Journal of Science Education,* 41(4), 510–532.

Pérez Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: an analysis measuring relative gains. *System.* DOI: 10.1016/j.system.2015.05.004

Piesche et al. (2016). CLIL for all? A randomised controlled field experiment with sixth grade students on the effects of content and language integrated science learning. *Learning and Instruction,* 44, 108–116.

Piniel K., & Albert Á. (2020) Motivation and Flow. In: Lamb M., Csizér K., Henry A., Ryan S. (Eds.) *The Palgrave Handbook of Motivation for Language Learning* (p. 579-597). Cham: Palgrave Macmillan.

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. *European Scientific Journal,* 12, 235–250.

Prain, V. (2004). *The role of language in science learning and literacy. In Writing and learning in the science classroom* (pp. 33–45). Dordrecht: Springer.

Prinsloo, C.H., & Harvey, J.C. (2020). The differing effect of language factors on science and mathematics achievement using TIMSS 2015 data: South Africa. *Research in Science Education,* 50, 2207-2226.

Quarderer, N.A., & McDermott, M.A. (2020). Examining science teacher reflections on argument-based inquiry through a critical discourse lens. *Research in Science Education,* 50, 2483-2504.

Quigley, D., Ostwald, J., & Sumner, T. (2017). Scientific modelling: Using learning analytics to examine student practices and classroom variation. *Proceedings of the Seventh International Learning Analytics and Knowledge Conference,* 329–338. New York: Association for Computing Machinery.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality,* 41, 203−212.

Reiss, K., Weiß, M., & Klieme, E. (2018). PISA 2018 - Grundbildung im internationalen Vergleich [PISA 2018 – International comparison of basic education]. In Olaf Köller (Ed.) *OECD Programme for International Student Assessment (PISA).*

Renninger, K. A., Ren, Y., & Kern, H. (2018). Motivation, engagement, and interest: “In the end, it came down to you and how you think of the problem”. In C. E. Hmelo-Silver, S. R. Goldman, P. Reimann, & F. Fischer (Eds.), *International handbook of the learning sciences* (pp. 116–126). Milton: Routledge Ltd..

Reynolds, A.G. (1991). The cognitive consequences of bilingualism. In A.G. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning* (pp. 145-182). Hillsdale NJ: Lawrence Erlbaum Associates.

Richards, J., & Cotterall, S. (2016). Exploring creativity in language teaching. In R. H. Jones & J. C. Richards (Eds.), *Creativity in language teaching* (pp. 97–113). New York, NY: Routledge.

Robinson, K. (2011). *Out of our minds: Learning to be creative.* Oxford, UK: Capstone.

Rodenhauser, A. & Preisfeld, A. (2014). Bilingual (German-English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. *International Journal of Environmental and Science Education,* 10(1), 99-110.

Rohia, A., & Mäntylä, K. (2019). 'It has given me this kind of courage …': the significance of CLIL in forming a positive target language self-concept. *International Journal of Bilingual Education and Bilingualism,* DOI: 10.1080/13670050.2019.1636761.

Rosenberg, M.B. (2012). *Gewaltfreie Kommunikation. Eine Sprache des Lebens* [Peaceful communication. The language of life]. Paderborn: Junfermann.

Rosenthal, J.W. (1993). Theory and practice: science for undergraduates of limited English proficiency. *Journal of Science Education and Technology,* 2(2), 435-443.

Rost, J. (2004). *Lehrbuch Testtheorie–Testkonstruktion* [Textbook test theory–test construction]. Bern: Hans Huber.

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: a cognitive load theory approach. *Learning and Instruction,* 52, 69-79.

RStudio Team (2020). *RStudio: Integrated Development for R. RStudio,* PBC, Boston, MA URL: http://www.rstudio.com/ (accessed: 01.03.2021).

Runco, M.A., Ancar, S., & Cayirdag, N. (2017). A closer look at the creativity gap and why students are less creative at school than outside of school. *Thinking Skills and Creativity,* 24, 242-249.

Ryoo, K. (2015). Teaching science through the language of students in technology-enhanced instruction. *Journal of Science Education Technology,* 24, 29-42.

Safwat S., El Bolock A., Alaa M., Faltaous S., Schneegass S., Abdennadher S. (2020). The effect of student-lecturer cultural differences on engagement in learning environments - a pilot study. In: De La Prieta F.et al. (Eds.) *Highlights in practical applications of agents, multi-agent systems, and trustworthiness.* Cham: Springer.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. *Innovations in Education and Teaching International,* 1–12.

Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation.* New York, NY: Oxford University Press.

Scharfenberg, Franz-Josef (2005): *Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse. (am Beispiel des Demonstrationslabors Bio- / Gentechnik der Universität Bayreuth mit Schülern aus dem Biologie-Leistungskurs des Gymnasiums).* Dissertation. Bayreuth.

Scharfenberg, F. J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus—Results of a teaching unit with authentic experiments. Biochemistry and Molecular Biology Education, 35(1), 28–39.

Scharfenberg, F-J., & Bogner, F.X. (2013). Teaching gene technology in an outreach lab: Students' assigned cognitive load clusters and the clusters' relationships to learner characteristics, laboratory variables, and cognitive achievement. *Research in Science Education,* 43, 141-161.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness of-Fit Measures, Methods of Psychological Research Online, 8(2), 23-74.

Schmidt, A.L. (2011). Creativity in science: Tensions between perceptions and practice. *Creative Education,* 2, 435-445.

Schneiderhan-Opel, J., & Bogner, F. X. (2020). FutureForest: Promoting Biodiversity Literacy by Implementing Citizen Science in the Classroom. *American Biology Teacher,* 82(4), 234–240.

Schnell, R., Hill, P. B., & Esser, E. (2018). *Methoden der empirischen Sozialforschung* (11., überarbeitete Auflage). De Gruyter Studium. De Gruyter Oldenbourg. URL: http://www.degruy- ter.com/search?f_0=isbnissn&q_0=9783110577327&searchTitles=true (accessed: 01.03.2021).

Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO* [Scales to capture school self-concept (SESSKO)]. Göttingen: Hogrefe.

Schülerlabor-Atlas. URL: https://www.schuelerlabor-atlas.de/schuelerlabore/in/Bayern (accessed: 01.03.2021).

Sweller, J. (2015). In academe, what is learned and how is it learned? *Current Directions in Psychological Science,* 24, 190-194

Schwichow, M., Zimmerman, C., Croker, S., Härtig, H. (2016). What students learn from hands-on activities. *Journal of Research in Science Education,* 53(7), 980-1002.

Sisk, D. (1972). Relationship between self-concept and creativity: Theory into practice. *Gifted Child Quarterly,* 16, 229-234.

Stahl, G., Koschmann, T., & Suthers, D. (2014). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 479–500). New York, New York: Cambridge University Press.

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integration inquiry science and language development for English language learners. *Journal of Research in Science Teaching,* 39(8), 664-687.

Szumski, G., & Karwowski, M. (2019). Exploring the Pygmalion effect: The role of teacher expectations, academic self-concept, and class context in students' math achievement. *Contemporary Educational Psychology,* 59, DOI: 10.1016/j.cedpsych.2019.101787.

Sutton, C. (1996). Beliefs About Science and Beliefs About Language. *International Journal of Science Education,* 18 (1), 1–18.

Tackett, J. L., & Lahey, B. B. (2017). Neuroticism. In T. A. Widiger (Ed.), *Oxford library of psychology. The Oxford handbook of the Five Factor Model* (p. 39–56). Oxford University Press.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington. DC: American Psychological Association.

Thürmann, E. (2008): Eine eigenständige Methodik für den bilingualen Sachfachunterricht? [Methodology for bilingual learning] In: *Gerhard Bach und Susanne Niemeier (Hg.): Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* Frankfurt am Main: Peter Lang, S. 71–89.

Treagust, D.F., Chittleborough, G.D., & Mamiala, T.L (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education,* 24, 357–368.

Tsai, K.-C. (2018). Investigating the empirical links between creative and critical thinking. *Psychology, Society, & Education*, 11(3), 267–280.

Unsal, Z., Jakobson, B., Molander, B.-O., & Wickman, P.-O. (2018). Language use in a multilingual class: A study of the relation between bilingual students' languages and their meaning-making in science. *Research in Science Education,* 48(5), 1027–1048.

Usher, S. (2013). *Letters of Note. Correspondence Deserving of a Wider Audience.* Edinburgh, Scotland: Canongate.

van der Slik, F.W.P., van Hout, R.W.N.M., & Schepens, J.J. (2015). The gender gap in second language acquisition: gender differences in the acquisition of dutch among immigrants from 88 countries with 49 mother tongues. *PLOS ONE,* 10(11), DOI: org/10.1371/journal.pone.0142056.

Verduyn, P., & Brans, K. (2012). The relationship between extraver- sion, neuroticism and aspects of trait affect. *Personality and Individual Differences,* 52, 664–669.

Walker, J.P. & Sampson, V. (2013). Learning to argue and arguing to learn – driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching,* 50(5), 561-596.

Wang, J-R., & Chen, S-F. (2014). Exploring mediating effect of metacognitive awareness on comprehension of science tests through structural equation modeling analysis. *Journal of Research in Science Teaching,* 50(2), 175-191.

Wang, J., Oliver, J.S., & Staver, J.R. (2008). Self-concept and science achievement: Investigating a reciprocal relation model across the gender classification in a cross-cultural context. *Journal of Research in Science Teaching,* 45(6), 711-725.

Watrin, L., Geiger, M., Spengler, M., & Wilhelm, O. (2019). Forced-choice versus Likert responses on an occupational Big Five questionnaire. *Journal of Individual Differences,* 40(3), 134–148.

Werner, S., et al. (2017). Investigating how German biology teachers use three-dimensional physical models in classroom instruction: A video study. *Research in Science Education,* 49, 1–27.

Wildhage, Manfred & Otten, Edgar (2009). *Praxis des bilingualen Unterrichts.* Berlin: Cornelsen Verlag.

Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understandings of model-based inquiry. *American Educational Research Journal,* 43(4), 783–835.

Wößmann, L., Freundl, V.Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2020). Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen? [Education in the corona crisis: how have school children spent the period of school closures, and what educational measures do Germans support]. *Ifo Schnelldienst,* 73 (9), S. 25–39.

Wolf, R. (1997). Rating scales. In J. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook* (pp. 958–965). Oxford, UK: Elsevier.

Yager, R. E. (2005). A vision for what science education should be like for the twenty-five years of a new millennium. *School Science and Mathematics,* 100, 327–341.

Yore, L.D., & Treagust, D.F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *International Journal of Science Education,* 28(2–3), 291–314.

Zangori, L., Peel, A., Kinslow, A., Friedrichsen, P., & Sadler, T. D. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching.* DOI: 10.1002/tea.21404.

5 TEILARBEITEN

5.1 Publikationsliste

Teilarbeit A: Roth, T.; Scharfenberg, F-J.; Mierdel J.; Bogner, F.X. (2020). Self-evaluative Scientific Modeling in an Outreach Gene Technology Laboratory. Journal of Science Education and Technology, DOI: 10.1007/s10956-020-09848-2. (Scopus: 95% Percentile, 53/1254; Impact Factor: 1.64)

Teilarbeit B: Roth, T., Conradty, C., & Bogner, F.X. (2021). Testing Creativity and Personality to Explore Creative Potentials in the Science Classroom, Research in Science Education, DOI: 10.1007/s11165-021-10005-x. (Scopus: 86%; Impact Factor: 2.25)

Teilarbeit C: Roth, T., & Bogner, F.X. (2021). The Trade-Off Between STEM Knowledge Acquisition and Language Learning in Short-Scale Bilingual Implementations. Revision Round 1 (Completed) International Journal of Science Education (Scopus: 83%; Impact Factor: 2.24)

Teilarbeit D: Roth, T., Scharfenberg, F-J., & Bogner, F.X. (2021). Content and Language Integrated Scientific Modelling: A Novel Approach to In-depth Model Learning. Submitted to Interactive Learning Environments (Scopus: 93%; Impact Factor: 3.9)

Teilarbeit E: Roth, T., & Bogner, F.X. (2021). Influence of COVID-19 Measures on CLIL Language Learning in an Outreach Gene Laboratory. Revision Round 1 (Again Under Review) The Language Learning Journal (Scopus 88%; Impact Factor: 1.72)

Teilarbeit F: Roth, T., Conradty, C., & Bogner, F.X. (2021). The Relevance of School Self-Concept and Creativity for CLIL Outreach Learning. Revision Round 1 (Again Under Review) Studies in Educational Evaluation (Scopus 89%; Impact Factor: 1.95)

5.2 Darstellung des Eigenanteils

Das Unterrichtsmodul, auf dem Teilarbeiten C bis F beruhen, wurde in seiner Grundform von Frau Dr. Langheinrich und Frau Dr. Mierdel entworfen. Aufbauend auf ihrem erprobten Design, entwickelte die Autorin der vorliegenden Arbeit eigenständig ein bilinguales Modul. Dies umfasste selbst erstellte Materialien für adressatengerechtes sprachliches Scaffolding sowie die Umwandlung des ehemals muttersprachlich deutschen Moduls in ein bilinguales Bio-/Gentechnik Modul mit vorwiegend englischer Unterrichts- und Schriftsprache. Zudem erfolgte die Durchführung des Moduls sowie die Erhebung aller empirischen Daten ausschließlich durch die Autorin. Lediglich in Teilarbeit A war die Autorin noch vor Beginn ihrer Promotion gemeinschaftlich an der Erhebung der Daten als Hilfswissenschaftliche Mitarbeiterin von Frau Dr. Mierdel beteiligt.

Die zur Erhebung der empirischen Daten notwendigen Skalen aus Teilarbeiten B-F wurden beinahe ausschließlich der Literatur entnommen und entsprechend den Fragestellungen adaptiert. Einzig der Cloze-Test zur Erhebung von Sprachwissen wurde nach in der Literatur vorgegebenen Kriterien eigens von der Autorin und zugeschnitten auf das bilinguale Bio-/Gentechnik Modul erstellt. Zur Bewertung der Modellqualität aus Teilarbeiten A und D wurde ein bestehendes und in vorhergehenden Arbeiten des Lehrstuhls erprobtes Kategoriensystem verwendet. In Ermangelung bereits bestehender Kategoriensysteme, leitete die Autorin zur qualitativen Auswertung der Sprachfertigkeiten in Teilarbeit E induktiv ein Kategoriensystem ab und brachte dies zur Anwendung.

Bei der Erstellung der aufgeführten Publikationen erfolgte die Erhebung aller empirischen Daten (abgesehen von Teilarbeit A) und die Entwicklung entsprechender Fragestellungen ausschließlich durch die Autorin. Selbiges galt für die statistischen Analysen und die nachfolgende Interpretation der Ergebnisse. Alle Teilarbeiten wurden zudem von der Autorin als Erstautorin eigenständig konzipiert und in einem ersten Entwurf verfasst. Dieser Entwurf wurde dann mit den jeweiligen Co-Autoren, Herr Dr. Scharfenberg und Frau Dr. Conradty, aufgrund ihrer jeweiligen Expertise in Modell- und Kreativitätsforschung iteriert und unter Einbezug von Herrn Prof. Dr. Bogner auf ein publikationsfähiges Niveau gebracht. Die Einarbeitung des Feedbacks der Peer-Reviewer in den internationalen Fachjournalen erfolgte unter vorheriger Absprache mit Herrn Professor Dr. Bogner und den Co-Autoren allein durch die Autorin.

5.3 Teilarbeit A

Journal of Science Education and Technology, DOI: 10.1007/s10956-020-09848-2, 2020

Self-evaluative Scientific Modeling in an Outreach Gene Technology Laboratory

Tamara Roth[1*], Franz-Josef Scharfenberg[1], Julia Mierdel[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

The integration of scientific modelling into science teaching is key to the development of students' understanding of complex scientific phenomena, such as genetics. With this in mind, we conducted an introductory hands-on module during an outreach gene technology laboratory on the structure of DNA. Our module examined the influence of two model evaluation variants on cognitive achievement: evaluation 1, based on students' hand-drawn sketches of DNA models and two open questions; and evaluation 2, based on students' own evaluations of their models in comparison to a commercially available DNA model. We subsequently subdivided our sample (N = 296) into: modellers-1 (n = 151) and modellers-2 (n = 145). Analyses of cognitive achievement revealed that modellers-2 achieved higher scores than modellers-1. In both cases, low achievers, in particular, benefitted from participation. Assessment of modellers-2' self-evaluation sheets revealed differences between self-evaluation and independent reassessment, as non-existent model features were tagged as correct whereas existent features were not identified. Correlation analyses between the models' assessment scores and cognitive achievement revealed small-to-medium correlations. Consequently, our evaluation-2 phase impacted students' performance in overall and model-related cognitive achievement, attesting to the value of our module as a means to integrate real scientific practices into science teaching. Although it may increase the workload for science teachers, we find that the potential scientific modelling holds as an inquiry-based learning strategy is worth the effort.

Keywords: Cognitive knowledge. Scientific models and modelling. Gene technology outreach learning. Science education.

Introduction

Hands-on experiments in the classroom are key for conveying knowledge of complex natural phenomena that might otherwise remain inaccessible to students (Kind 2015). Yet, for many years, a clear focus on instructional and theoretical teaching has dominated the educational landscape. Today, new national educational standards (KMK 2005) replace these long-established methods and raise hope of more practical approaches to science teaching. However, one must bear in mind that some schools may lack the financial means to build and stock laboratories (e.g., Raviv et al. 2019) which may require expensive and specialist equipment.

For schools without the necessary means, outreach laboratories (e.g., at universities) offer opportunities for students to experience science in authentic settings. Many such laboratories have been developed over the last two decades and offer opportunities to foster students' interest and motivation (Glowinski & Bayrhuber 2011). Hereby, solution-oriented approaches to scientific problems and an adequate learning environment are important (e.g., Nasir et al. 2006, Brody et al. 2007). Our outreach program, *Simply inGEN(E)ious! DNA as a carrier of genetic information,* meets these requirements: Designed as a one-day-long module, it involves hands-on experiments, prediction of potential results, and a modelling phase (Mierdel and Bogner 2019a).

In our previous study, we compared two modelling variants: model viewers and modellers. Model viewers worked "with a commercially available school model of DNA structure" while modellers "were required to generate a DNA model using assorted handcrafting materials" (Mierdel and Bogner 2019a, p. 1). Most of the existing literature suggests that hands-on modelling contributes to cognitive achievement (e.g., Jackson et al. 2008, Passmore et al. 2009), yet we found that model viewers had a higher mid-term increase in knowledge. Thus, "we dispensed a more detailed model evaluation for modellers" (Mierdel and Bogner 2019a, p. 12). To fill this gap in our research, we selected two variants related to students' model evaluation: modellers-1, who only participated in one evaluation phase, and modellers-2, who additionally completed a second evaluation phase. We based the latter on our model viewing approach.

Our study aims to observe the potential influences of both model evaluation variants on cognitive achievement. We first outline the relevant theoretical background, including knowledge of learning and model learning which influenced the way we developed, conducted, and evaluated our intervention. We then provide a brief explanation of the phases of our intervention before we present our research questions.

Theoretical Background

Learning

Despite numerous suggestions, a universally accepted definition of ‘learning’ has not yet been developed (e.g., Domjan 2015, Haselgrove 2016). Learning is often summarized as “a change in behavior [depending on] experience” (De Houwer et al. 2013, p. 631), a view which is supported by constructivists (Hodson 1998). Thereby, students autonomously verify or falsify hypotheses and modify preconceptions accordingly or develop new ideas (Schwarz et al. 2009). These competencies also reflect Kolb’s (1984) categorization of learners’ abilities: concrete experience ability, reflective observation ability, abstract conceptualization ability, and active experimentation ability. Hence, the construction of knowledge from chunks of information requires experiential processes to transform such information into a coherent mental model.

Real experiments in class or outreach settings provide the necessary basis for students to build up knowledge and allow them to manage scientific equipment. Such first-hand experience is vital to understanding the nature of the physical world and science, where experimental outcomes are not always as expected (Loveys and Riggs 2019). Solomon (1980, p.13) even described hands-on experimenting as key to increasing scientific knowledge: “Science teaching must take place in the laboratory; science simply belongs there as naturally as cooking belongs in the kitchen and gardening in the garden.” Different levels of scientific knowledge can, thus, be achieved by combining hands-on experimentation with the development of scientific thinking skills. This approach significantly enhances skills set by the national standards (KMK 2005), such as the understanding of scientific processes. Together with an open and flexible learning environment, critical thinking skills and self-evaluation may be fostered. This is especially important when it comes to reassessing results. An effective method for encouraging self-evaluation is reflective writing (Kovanović et al. 2018) about particular experimental phases. This approach can encourage students to rethink and reassess certain steps in the experimental process (Scharfenberg and Bogner 2013; for details, see below). In this “self-directive process”, students “transform their mental abilities into academic skills” (Zimmermann 2002, p. 65).

Model Learning

Models are tools commonly used to visualize and explain phenomena (Krajcik and Merritt 2012). Windschitl and Thompson (2006, p. 796), for instance, described models as “hypothesized relationships among objects, processes, and events”. Thereby, models can reveal underlying mechanisms, show causal links, raise questions, and test multiple hypotheses. Whenever a prediction proves incorrect or new evidence emerges, a model can be adapted and refined (e.g., Passmore et al. 2009, Carpenter et al. 2018). Models are regarded as the cornerstones of every scientific discipline, combining theory with modelled processes and functions to explain how things really are. Thus, modelling in science classes is invaluable for student learning (e.g., Jackson et al. 2008, Passmore et al. 2009). Although natural phenomena are often difficult to observe, models can provide an authentic alternative experience, are often easier to understand, and do not require exhaustive preparatory work (e.g., Windschitl and Thompson 2006, Carpenter et al. 2018). Students, *inter alia*, gained better cognitive achievement scores when applying three-dimensional DNA models rather than two-dimensional DNA models (e.g., Rotbain et al. 2006, Saka et al. 2006).

Modelling may lead to more productive student activity in the classroom, emphasizing vital scientific practices, including the need to "engage in inquiry other than controlled experiments, [to] use existing models in their inquiries, [to] engage in inquiry that leads to revised models, [to] use models to construct explanations, [to] use models to "unify" their understanding, and [to] engage in argumentation" (Passmore et al. 2009, p. 397). We decided to include these activities in our three-dimensional modelling approach (for details, see below).

Intervention Phases

We divided our one-day module into four different intervention phases (Table 1)

TABLE 1 Quasi-experimental study design.

Intervention phases	Evaluation variants	
	Modellers-1 (n = 151)	Modellers-2 (n = 145)
Pre-lab	+	+
DNA-related theoretical and experimental phases		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases		
Mentally modelling: text analysis	+	+
DNA modelling with craft materials	+	+
Model evaluation-1: Drawing a paper & pencil version of the crafted model and answering questions	+	+
Model evaluation-2: Comparing the crafted model with a scientific demonstration model	-	+
Interpretation	+	+

Pre-lab phase

As many students are not familiar with laboratory equipment or scientific work, we provide an appropriate pre-lab phase. This phase addresses three scientific aspects, in particular: affective dimensions, introduction to laboratory techniques, and introduction to theoretical concepts (e.g., Sarmouk et al. 2019). Students, thus, learn how to handle the equipment necessary for experimentation, such as micropipettes and centrifuges. Teachers act as external guides and introduce theoretical concepts via presentations and demonstrations, which are key to understanding the subsequent experiments. This phase should also prevent students from feeling overwhelmed in response to the experimental situation (e.g., Kalyuga 2009).

DNA-related theoretical and experimental phases

Theoretical phases. Having provided a real-life background to the subsequent experiments (DNA relevance, Table 1), the underlying scientific basics of DNA isolation and gel electrophoresis were introduced. Students were then encouraged to connect new information with prior knowledge and hypothesise about the experiments' potential outcomes (Mierdel & Bogner 2019a). Our decision to provide information in a series of theoretical phases ensures that students are not overwhelmed by too much new information at once and allows them to focus on the experimental phase ahead. This is particularly important when explaining the difficult concept of gel electrophoresis which not only involves chemical knowledge about the DNA's contents but also abstract thinking in order to imagine the processes involved.

Experimental Phases. Ours is an evidence-based, two-step approach (Scharfenberg and Bogner 2013), in which students answer questions in their individual workbooks and think about subsequent experimental procedures. They work in pairs to discuss every step before carrying out experiments that effectively combine hands-on and minds-on activities and require students to do more than simply follow instructions (Scharfenberg and Bogner 2013).

Model-related Phases

Both model-related phases directly followed the experimental, DNA-related phases. As we regard models as vital to visualizing and explaining phenomena in science and science education (Krajcik & Merritt 2012), we subdivided our model-related phases into a mental modelling phase involving text analysis, a modelling phase involving craft materials, a model evaluation-1 phase, and a model evaluation-2 phase. Only modellers-2 participated in all four phases.

As in our previous study, we based our model-related phases on the four main stages of the *Model of Modelling* (Justi and Gilbert 2002, p. 370 ff.): (1) "experience[s] of the phenomenon being modelled", (2) "forming a mental model", (3) "decision … about the mode of representation in which it is to be expressed", and (4) "testing … scope and limitations of the model". Mental modelling is, thereby, key to providing a theoretical basis for experimental findings. According Franco & Colinvaux (2000) building mental models involves reasoning about previously obtained knowledge to make predictions and derive new ideas from it. A text about the discovery of the DNA's structure (Usher 2013), thus, provides fundamental knowledge about the necessary components of the DNA, which will be applied to the students' simplified mental model (e.g., Mierdel and Bogner 2019b, Franco & Colinvaux 2000).

After determining the DNA's representation and building the hand-crafted model, model evaluation-1 phase was organized as a reciprocal self-evaluation. As combining sketching and handcrafting models proved to be important (e.g., Prabha 2016, Orhan and Sahin 2018), modellers-1 and modellers-2 evaluated their hand-crafted DNA model based on their earlier paper-and-pencil version. Another effective method for encouraging self-evaluation is reflective writing (Kovanović et al. 2018). Open-ended questions about model-related components encourage students to rethink and reassess certain steps and decisions in their development of the mental model into its physical counterpart (Mierdel and Bogner 2019a). Modellers-2 also assessed their hand-crafted DNA models using a comparison-based self-evaluation with a commercially available DNA demonstration model. In other words, we based this second evaluation phase on our previous model viewing approach.

Interpretation Phase

Here, students compared their hypothesis about the outcome of their experiments with the gel electrophoresis' images. They also discussed their individual models in class and compared these to the molecular DNA model by Watson & Crick which, in most cases, differed from the students' models (Mierdel and Bogner 2019b).

Objectives of the Study

The present study tries to answer the following research questions:

1. How does the application of one or two model evaluation phases influence overall cognitive achievement and, in particular, the model-related knowledge developed during the hands-on laboratory?

2. How does the additional evaluation-2 phase of modellers-2' three-dimensional models influence their overall learning?

Thus, we had three specific objectives:

- to assess students' overall cognitive achievement and the model-related knowledge of modellers-2, which we would compare with that of modellers-1;
- to determine the quality of the evaluation-2 phase and the DNA component that modellers-2 correctly identified in their handcrafted models;
- to examine the potential correlations between modellers-2' performance in the model evaluation-2 phase and modellers-2' cognitive achievement.

Materials and Methods

After a brief introduction to our educational intervention, its design, and the independent variable, we explain the modelling phases and the model evaluation phases. We then describe the students' sample, discuss the dependent variables, and outline the statistical methods applied.

Educational Intervention, Design, and Independent Variable

The one-day, hands-on module offered inquiry-based learning activities focused on the structure of DNA, aimed at ninth graders. Students worked in pairs to complete their tasks with guidance provided in a workbook (for a detailed module description, see Mierdel and Bogner 2019a). The content of the module is in line with the state's syllabus and follows national competency requirements (KMK 2005).

We conducted two versions of the intervention which differed concerning students' evaluations of their models. Since often only quasi-experimental designs are feasible for students in intact class groups (Cook and Campell 1979), student classes were randomly assigned to each of the evaluation variants. The students were, thus, divided into modellers-1 and modellers-2, respectively, as the independent variable (Table 1).

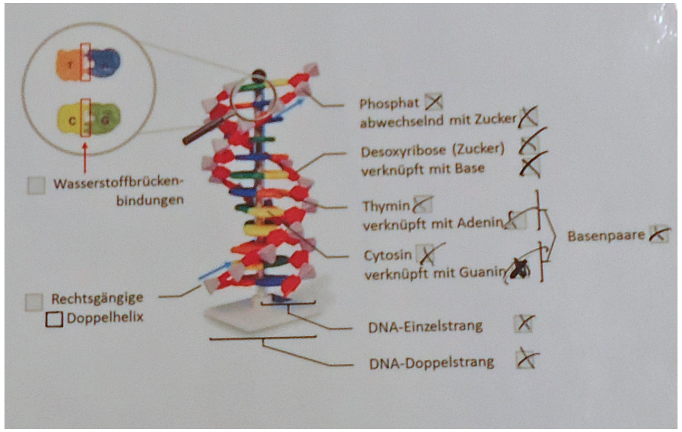
Both versions of the intervention began with a pre-lab phase (50 min) wherein students were familiarized with the lab equipment and relevant working techniques. Thereafter, the criminological relevance of DNA was introduced to contextualize the two main DNA-related experimental phases: DNA isolation from oral mucosal cells (60 min) and agarose gel electrophoresis (85 min). Both were connected to model-related phases (60 min). Students then tried to retrace Watson and Crick's research to solve the molecular puzzle of the DNA's structure.

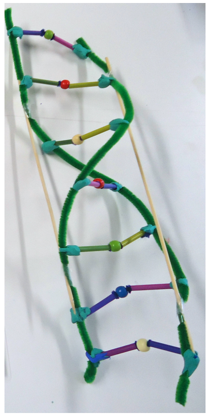
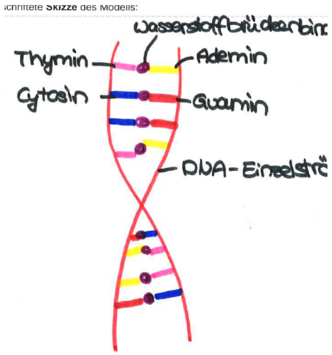
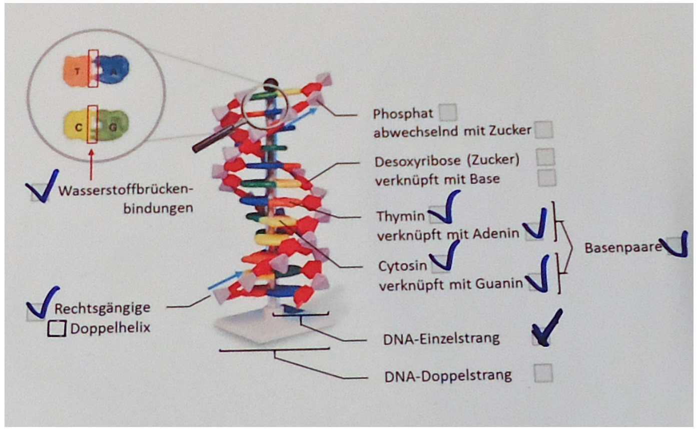
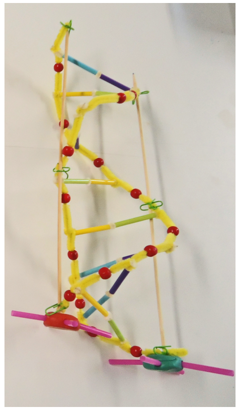
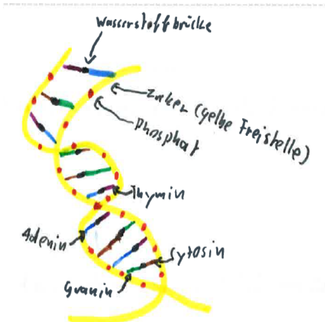
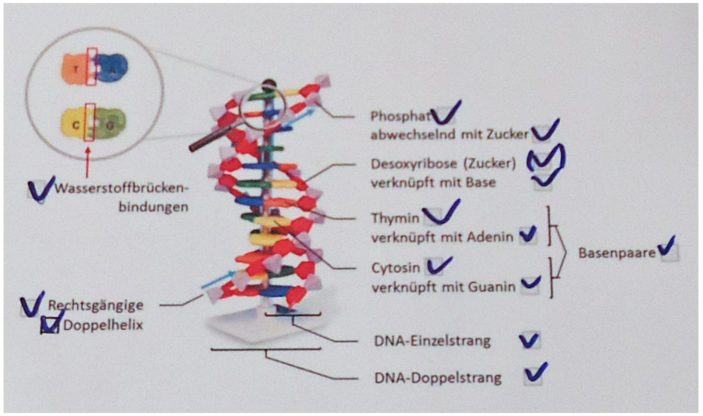
The modelling phases. These phases (Table 1) were key in providing a theoretical basis for experimental findings. Having read about the discovery of the structure of DNA (Usher 2013), students discussed and answered questions in their workbooks (e.g., "DNA's backbone: Label its components and describe their set up"; for details, see Electronic Supplementary Material [ESM] 1 as an online resource). The aim was to enable the students to internalize key aspects of the text which would later allow them to mentally model the DNA's structure. To this end, the text included references to all key components of the DNA (*e.g.* base pairing).

In the next stage, students transformed their mental DNA models into physical, hand-crafted DNA models (Table 1) using a DNA-modelling-kit containing crafting materials (*e.g.* colored beads and pipe cleaners; for three model examples, Table 2, 1st columns).

The model evaluation phases. These phases (Table 2, 2nd and 3rd column) were organized as a reciprocal self-evaluation of students' models. Modellers-1 and modellers-2 all evaluated their hand-crafted DNA model against their paper-and-pencil sketch of the model (evaluation-1). Modellers-2 then conducted an additional, comparison-based assessment of their hand-crafted DNA models, in which they were asked to draw a paper-and-pencil model with labels to identify the model's components (Table 2, 2nd column) and the elements of their hand-crafted models.

TABLE 2 Assessment of modellers-2' evaluation phases.

Phase			Assessment (model[a] / sketches[b] / self-evaluation sheet[c])
Modelling	Model evaluation		
Hand-crafted model	Evaluation-1: sketches	Evaluation-2: comparing and ticking the self-evaluation sheet	
			6 / 6 / 6

<table>
<tr>
<td></td>
<td>
</td>
<td>
</td>
<td>10 / 11 / 7</td>
</tr>
<tr>
<td></td>
<td>
</td>
<td>
</td>
<td>13 / 17 / 13</td>
</tr>
</table>

Notes: [a]score maximal 14 points; [b]score maximal 14 points; [c]score maximal the model score, see [a].

This was an opportunity to reflect on the process of building the model and on its informative value. In the meantime, ideas about the accuracy of their models could be exchanged. While modelling, students might have become aware of differences between their own and their classmates' models. Once they had completed a sketch of their models, students answered two open-ended questions on their worksheets: "Which features of the original DNA molecule are simplified in your model"; and "Explain why one might create different models of one biological original (in our case, the structure of the DNA)?" Thus, students were required to consider the scope and limitations of their models.

Modellers-2 also completed evaluation-2 (Table 1), wherein they were asked to compare and contrast their handmade models with a commercially available DNA model. A self-evaluation sheet (ESM 2) displaying the image of this scientific model served as the basis of their assessments. Each component included in both models was tagged to assist students' self-evaluations of their models.

In both versions, students' findings from the model phase were integrated into a final interpretation phase. Here, students discussed experimental results of the gel-electrophoresis, which they compared with previous hypotheses (Table 1).

Participants

Altogether, 296 ninth-graders (higher secondary school) participated in our study (girls 52.0%, boys 48.0%; $M_{\text{Class size}} = 22.8$, $SD = 6.2$; $M_{\text{Age}} = 14.6$, $SD = 0.8$). Six classes took part as modellers-1 ($n = 151$) and seven classes as modellers-2 ($n = 145$). Modellers-1 teamed up in 77 groups (75 2-person groups and two students working individually due to illness), modellers-2 in 73 groups (72 2-person groups and one student working individually due to illness). To avoid bias, we compared students' prior knowledge of biology with the respective biology grades. We found no significant difference (Mann-Whitney U test [MWU]: $Z = -1.144$, $p = .253$). Moreover, we compared individual students' prior in-class experience of modelling (3-item scale, adapted from Scharfenberg et al. 2007; Cronbach's Alpha .62) and did not discover a significant difference (MWU: $Z = -.859$, $p = .390$).

Participation was voluntary. Written parental consent was given prior to students' participation in our study, although the data collection was pseudo-anonymous and students could not be identified. The study was designed in accordance with the Declaration of Helsinki (2013), and the state ministry approved the questionnaires used.

Dependent variables

As dependent variables, we examined students' knowledge in a repeated measurement design: a pre-test (T0) two weeks before the intervention, a post-test (T1) after the module, and a retention test six weeks thereafter (T2). We examined students' sketches and their responses to the open questions from the evaluation-1 phase; for modellers-2, we additionally assessed the evaluation-2 phase. Throughout the entire intervention, students were unaware of any testing schedules.

Students' knowledge. We applied an ad-hoc knowledge test comprising 30 multiple-choice items: 12 items (examples, Table 3) assessed knowledge of the DNA-related phases (DNA relevance, hands-on isolation, and gel electrophoresis of DNA; Table 1) and 18 items (Table 3) analyzed knowledge related to the model phases (Table 1).

TABLE 3 Knowledge item examples related to the DNA-related phases (A) and the modelling phase (B).

Item relation	Item difficulty[a] (T0 / T1)	Item example[b]
DNA-related (A)[c]	63.7 / 85.5	A positively charged particle migrates through the electrical field … (Item 1) (a) between both poles (b) to the positive pole (c) to the negative pole* (d) not at all
	40.3 / 60.4	The DNA encodes the genetic information of … (Item 24) (a) all organisms* (b) great apes (c) all organisms but bacteria (d) vertebrates
	7.4 / 15.1	With the aid of gel electrophoresis, you get information about … (Item 6) (a) the molecular mass* (b) the number of bindings of a molecule (c) the components of a molecule (d) the atoms of a molecule
Modelling (B)[d]	48.9 / 82.7	Which of the following is not part of the DNA? (Item 2) (a) Adenine (b) Ribose* (c) Guanine (d) Deoxyribose
	24.8 / 64.7	In 1962, James Watson and Francis Crick were honoured with the Nobel prize for the discovery of … (Item 4) (a) of DNA located in cell nuclei (b) components that form the DNA (c) gel-electrophoresis (d) the DNA double helix structure*
	11.6 / 15.1	The analysis of a DNA-section revealed a proportion of guanine with 30 %. Following, the proportion of adenine is … (Item 26) (a)… not determinable. (b)…also 30%. (c) 70%. (d) 20%.*

Notes: [a]percentage of correct answers; [b]correct answer marked with asterisk *; [c]sum 12 items; [d]sum 18 items.

Content validity was given as the items were consistent with the state syllabus. Regarding construct validity, inter-item correlations below .20 (T0 = .08; T1 = .19; T2 = .18) confirmed that each item referred to different knowledge facets. Furthermore, the items' heterogeneity concerning complex constructs, such as cognitive achievement, emphasizes the given construct validity (Rost 2004). Cronbach's alpha values of .71 (T0), .64 (T1), and .70 (T2) indicate acceptable internal consistency when the value exceeds .70. According to Lienert & Raatz (1998) values between .50 and .70 allow for the differentiation of groups. Item difficulties (percentage of correct answers, Bortz & Döring 1995) ranged between 7% (high difficulty) and 88% (low difficulty). In the case of our intervention, the item difficulties reversed from T0 to T1 (Figure 1) and generally decreased.

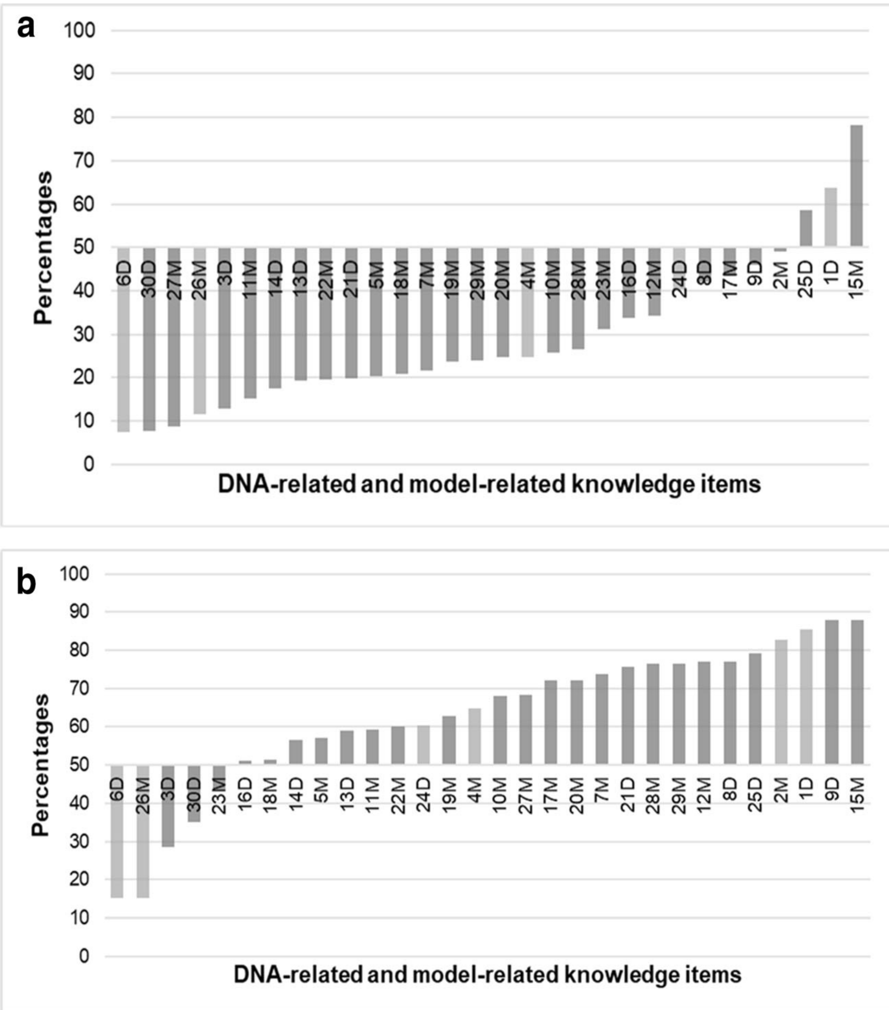


FIGURE 1a Repeated measurement design of both instructional variants in the outreach lab. **b** Item-difficulties for T0 and T1. DNA-related items shortened "D", model-related items shortened "M". Item examples shown in Table 3 highlighted in light grey (note the shift between the pre- and post-schedule)

We calculated the students' scores and analyzed these with regard to increases in knowledge (T1 minus T0), and retention rate (T2 minus T0). However, these different variables do not reflect actual knowledge growth. Thus, we calculated the actual learning success with respect to the maximal attainable score (30 correct answers): (T1 – T0) x (T1/30); and the persistent learning success (T2 – T0) x (T2/30) (Scharfenberg et al. 2007). Increased knowledge is, hence, weighted according to students' actual knowledge, making it possible to compare cognitive achievement despite some students exhibiting a huge increase in knowledge yet low final scores, and *vice versa*.

We also calculated correlations between biology grades and post-test (T1) scores for overall and model-related knowledge items using Spearman-Rho (Field 2012).

Evaluation-1 phase. In order to compare the evaluation variants for the evaluation-1 phase, we assessed students' model sketches (changed after Langheinrich & Bogner 2015; for definitions, examples, and frequencies, see ESM 3). We randomly selected 26 out of 150 drawings for a second scoring (17.3%). Cohen's kappa coefficient (Cohen 1968) scores of 0.88 and 0.82 for intra-rater and inter-rater reliability showed an "almost perfect" rating (Wolf 1997, p. 964). To avoid bias, sketches of both variants were compared and no significant difference could be identified (MWU: $Z = -.745$, $p = .456$).

Using content analysis (Bos & Tarnai 1999), we iteratively categorized the statements that students made in response to the open questions. For the first question 'Which features of the original DNA molecule are simplified in your model', four categories were employed: *level of DNA*, *level of substance*, *level of particles*, and *level of structure* (for definitions, examples, and frequencies, see ESM 4). We randomly selected 58 out of 384 statements for a second scoring (15.1%). We computed Cohen's kappa coefficient (Cohen 1968) scores of .98 and .78 for intra-rater and inter-rater reliability, which showed a "substantial" to "almost perfect" rating (Wolf 1997, p. 964). For the second question, 'Explain why one might create different models of one biological original (in our case, the structure of the DNA)?', we applied the adapted category system of Mierdel and Bogner (2019b) and identified five categories: *individuality of DNA*, *different interpretation, different model design, different focus,* and *different research state* (for definitions, examples, and frequencies, see ESM 5). We randomly selected 27 out of 150 statements for a second scoring (18.0%). We computed Cohen's kappa coefficient (Cohen, 1968) scores of .75 and .70 for intra-rater and inter-rater reliability, which showed a "substantial" rating (Wolf 1997, p. 964). To avoid bias, we compared the two evaluation variants in terms of category frequencies of responses to both open questions. We did not find any significant contingencies (adjusted Pearson's $C \leq .192$; $p \geq .065$; Pearson 1900).

TABLE 4 Assessment of the modellers'-2 evaluation-2 phase.

Analysis sector	Description (score)	Scores			
		Maximum	Students' self-evaluation sheet[a]	Assessment students' self-evaluation sheet[a]	Assessment students model[a]
Bases	- Four bases are indicated (2) - Base pairs correctly indicated (3) - Hydrogen bonds correctly indicated, differing in G/C and A/T (1)	6	5.1 (2.0/6.0)	4.3 (0/5.0)	5.0 (5.0/5.0)
Deoxyribose	- Deoxyribose indicated (1) - Deoxyribose linked to base (1)	2	1.3 (0/2.0)	1.0 (0/2.0)	1.5 (1.0/2.0)
Phosphate	- Phosphate indicated (1) - Phosphate and deoxyribose alternately arranged (1)	2	1.4 (0/2.0)	1.2 (0/2.0)	1.7 (1.0/2.0)
Primary structure	- Single strand visible (1) - Double strand visible (1)	2	1.6 (0.3/2.0)	1.6 (1.0/2.0)	2.0 (2.0/2.0)
Secondary structure	- Double helix visible (1) - Right-handed double helix visible (1)	2	1.1(0/2.0)	0.6 (0/1.8)	1.2 (0/2.0)
Sum		14	10.6 (6.0/13.0)	8.7 (2.5/11.0)	10.8 (8.0/13.0)

Note: [a]grouped median, 25th and 75th percentiles in brackets.

Evaluation-2 phase. A three-step approach was applied to evaluate the evaluation-2 phase (Table 4).

- Documentation of the students' self-evaluation: We counted each box that students had tagged on their self-evaluation sheet as one point (maximal score 14 points).
- Assessment of students' self-evaluation sheets: We analyzed the tagged boxes' conformity on the self-evaluation sheets using the respective models. Appropriate tags received one point each. If students tagged all their boxes correctly, they would reach the maximal core as was recorded on their self-evaluation sheets.
- Assessment of students' models: We independently assessed the models. Correct features each received one point whether or not they had been identified by the students (maximal score 14 points).

A comparison between the documented boxes and the assessment of the self-evaluation sheets enabled us to determine the extent to which students had correctly evaluated their models. Lower scores on the self-evaluation sheet indicate students' mistakes when assessing the quality of their models. However, a lower model score also indicates that a student may have documented model features that were not given. By contrast, higher model scores indicate model features that the student did not identify as such.

Statistical Analysis

We applied nonparametric methods due to an abnormal distribution of variables (Kolmogorov-Smirnov test (Lilliefors modification): partially $p < .001$), and, consequently, use boxplots to illustrate our results. Intra-group differences over the three test dates were analysed using Friedman test (F) in combination with a pairwise analysis from T0 to T1 and T2, and from T1 to T2, using Wilcoxon (W) signed-rank test. Mann-Whitney *U* tests (MWU) were used to evaluate inter-group differences. Due to multiple testing, we applied a Bonferroni correction (Field 2012). In the case of significant results, effect sizes *r* (Lipsey & Wilson 2001) were calculated with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. For correlation analyses, we applied Spearman's rank correlations and report Spearman's Rho values.

Results

We first provide an overview of our intra-group and inter-group analyses with regard to overall and model-related knowledge. This is followed by a detailed assessment of the evaluation-2 phase.

Intra-group Analyses of Cognitive Achievement

Intragroup analysis (F and W tests, Table 5) revealed significant changes for modellers-1 and modellers-2, in terms of both overall and model-related knowledge: They initially increased their knowledge at both levels, which then dropped between T1 and T2, but not below prior levels (T0). This suggests that students gain short-term and mid-term knowledge throughout the intervention (Table 6).

TABLE 5 Cognitive achievement of the students' sample as a whole and model understanding items only.

Items	Friedman test			Wilcoxon signed-rank test					
				T0 / T1		T0 / T2		T1 / T2	
	Chi-Square	*df*	*p*	*Z*	*p*	*Z*	*p*	*Z*	*p*
Model-related knowledge	275.65	2	< .001	-12.98	< .001	-8.64	< .001	-9.92	< .001
Overall knowledge	259.03	2	< .001	-12.95	< .001	-9.61	< .001	-9.08	< .001

Inter-group Analyses of Cognitive Achievement

To account for differences in students' prior knowledge (Table 6, superscript a), we calculated difference variables for short-term increases in knowledge and mid-term retention rates, and learning success variables to assess inter-group differences. These were the only variables taken into account. Based on sum scores, increases in knowledge (T1-T0) and retention rate (T2-T0) were calculated (Field 2012) for overall (30 items) and model-related (18 items) knowledge scores (Table 6).

TABLE 6 Dependent variables for both modellers-1 and modellers-2, analysed with regard to knowledge scores, difference variables and learning success. Over-all knowledge items and model-related items only were differentiated.

Dependent variable	**Knowledge**			
	Overall knowledge items (max. 30)		**Model-related knowledge items (max. 18)**	
	Modellers-1	**Modellers-2**	**Modellers-1**	**Modellers-2**
	***Mdn* (25th/75th *P*)**	***Mdn* (25th/75th *P*)**	***Mdn* (25th/75th *P*)**	***Mdn* (25th/75th *P*)**
T0[a]	10.3 (8.0/13.0)	6.3 (4.0/8.0)	5.7 (4.0/8.0)	3.8 (2.0/5.0)
T1[b]	20.6 (18.0/23.0)	19.4 (16.0/22.0)	13.9 (12.0/15.0)	12.8 (11.0/15.0)
T2[c]	16.9 (13.0/20.0)	17.2 (15.0/19.0)	10.9 (9.0/13.0)	10.5 (9.0/12.0)
Difference variables				
	Overall knowledge items (max. 30)[j]		**Model-related knowledge items (max. 18)[j]**	
	Modellers-1	Modellers-2	Modellers-1	Modellers-2
	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)
Increase in knowledge[d]	9.9 (7.0/12.3)	13.1 (9.0/16.0)	7.8 (6.0/10.0)	9.1 (7.0/11.0)
Retention rate[e]	6.9 (4.0/9.0)	10.4 (8.0/13.0)	5.4 (3.0/7.0)	6.6 (4.0/8.0)
Learning success				
	Overall knowledge items (max. 30)		**Model-related knowledge items (max. 18)[j]**	
	Modellers-1	Modellers-2	Modellers-1	Modellers-2
	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)
Actual learning success[f]	6.7 (4.2/8.6)	8.7 (5.1/11.9)[h]	3.4 (2.4/4.7)	3.9 (2.4/5.5)
Persistent learning success[g]	3.7 (1.8/5.9)	6.0 (3.7/7.9)[i]	1.9 (1.0/3.2)	2.3 (1.2/3.3)

Notes: [a]MWU-test over-all knowledge: $Z = -7.90$; $p < .001$; $r = .552$ and model-related items $Z = -5.95$; $p < .001$; $r = .439$; [b]MWU-test over-all knowledge: $Z = -2.26$; $p = .024$; $r = .185$ and model-related items $Z = -2.92$; $p = .004$; $r = .224$; [c]MWU-test over-all knowledge: $Z = -.271$; $p = .786$

and model-related items $Z = -1.59$; $p = .111$; [d]score: T1 - T0; MWU-test over-all knowledge: $Z = -4.99$; $p < .001$; $r = .386$ and model-related items $Z = -2.93$; $p = .003$; $r = .270$; [e]score: T2 – T0; MWU-test over-all-knowledge: $Z = -6.41$; $p < .001$; $r = .499$ and model-related items $Z = -2.82$; $p = .005$; $r = .304$; [f]actual learning success: (T1 – T0) x (T1/30) for overall knowledge and (T1 – T0) x (T1/18) for model related knowledge items; [g]persistent learning success: (T2 – T0) x (T2/30) for overall knowledge and (T2 – T0) x (T2/18) for model related knowledge items; [h]MWU-test over-all knowledge items: $Z = -3.36$; $p = .001$; $r = .293$; [i]MWU-test over-all knowledge items: $Z = -5.23$; $p < .001$; $r = .436$; [j]percentage share of correctly answered items: increase in knowledge overall modellers-1 and modellers-2 (33%, 44%) < model-related modellers-1 and modellers-2 (43%, 51%); retention rate overall modellers-1 and modellers-2 (23%, 35%) < model-related modellers-1 and modellers-2 (30%, 37%).

Overall Knowledge. Scores of the overall knowledge test, which included both DNA-related and model-related knowledge items showed differences in increases in knowledge and retention rates between modellers-1 and modellers-2. Modellers-2 scored significantly higher in increased knowledge and retention rate than modellers-1, with a medium-to-large effect size (Table 6, superscripts d/e).

As difference variables do not display the actual cognitive achievement, we analyzed learning success variables (see above). For short-term actual and mid-term persistent learning success, significant differences in overall knowledge were identified with medium-to-large effect sizes (Table 6; superscripts h/i). Compared to modellers-1, modellers-2 achieved higher scores in terms of actual and persistent learning success (Figure 2).

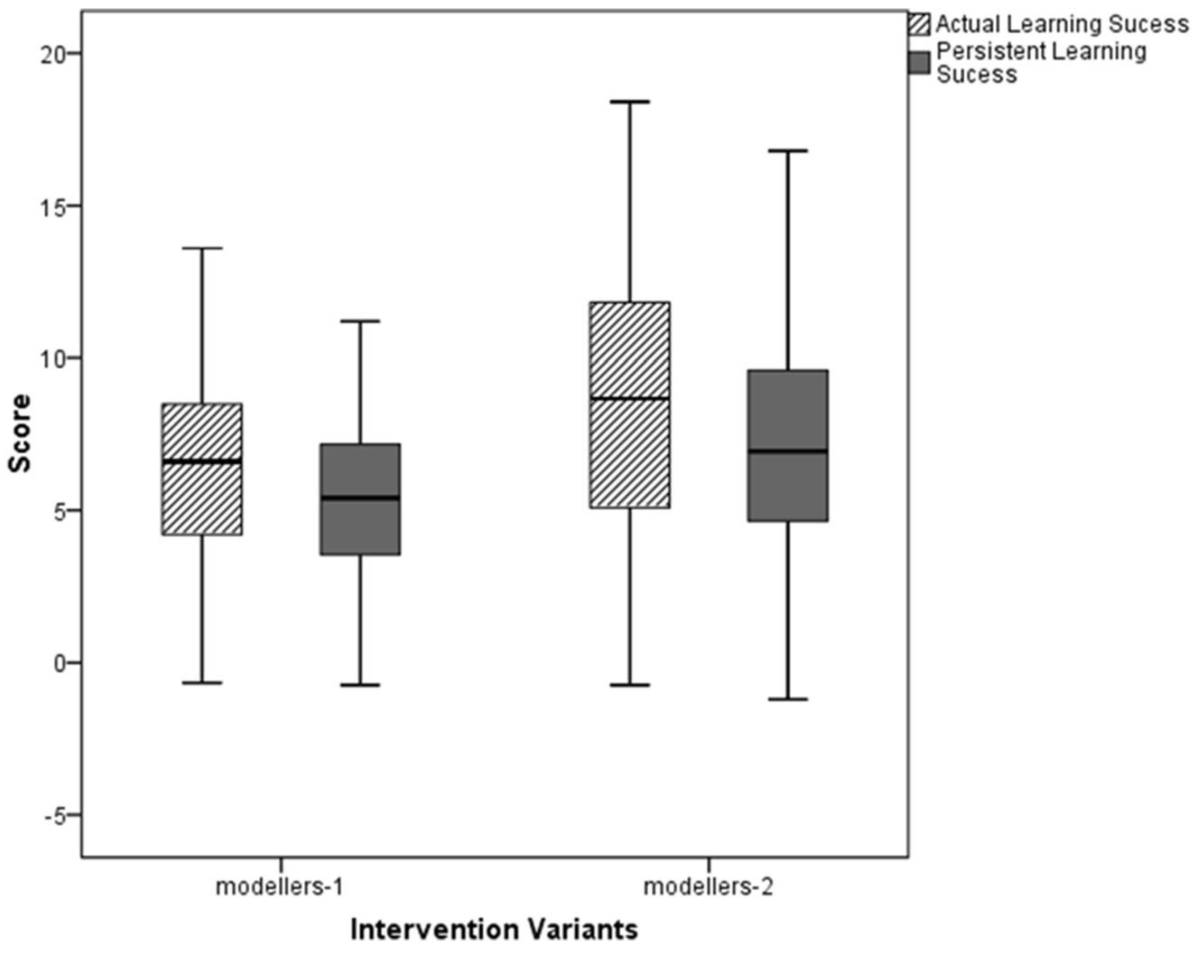


FIGURE 2 Changes in overall knowledge scores for both modellers-1 and modellers-2. All participating groups increased their actual learning success scores: (T1 – T0) x (T1/30). Persistent learning success scores: (T2 – T0) x (T2/30), however, dropped in comparison to actual learning success scores.

Model-related knowledge. Regarding scores based on 18, model-related knowledge items, modellers-2 achieved significantly higher increases in knowledge and retention rates with a medium effect size (Table 6, superscripts d/e). At the level of learning success, modellers-2 achieved higher scores (Figure 3). However, model-related cognitive achievement exceeded that of overall knowledge (Table 6, superscripts j).

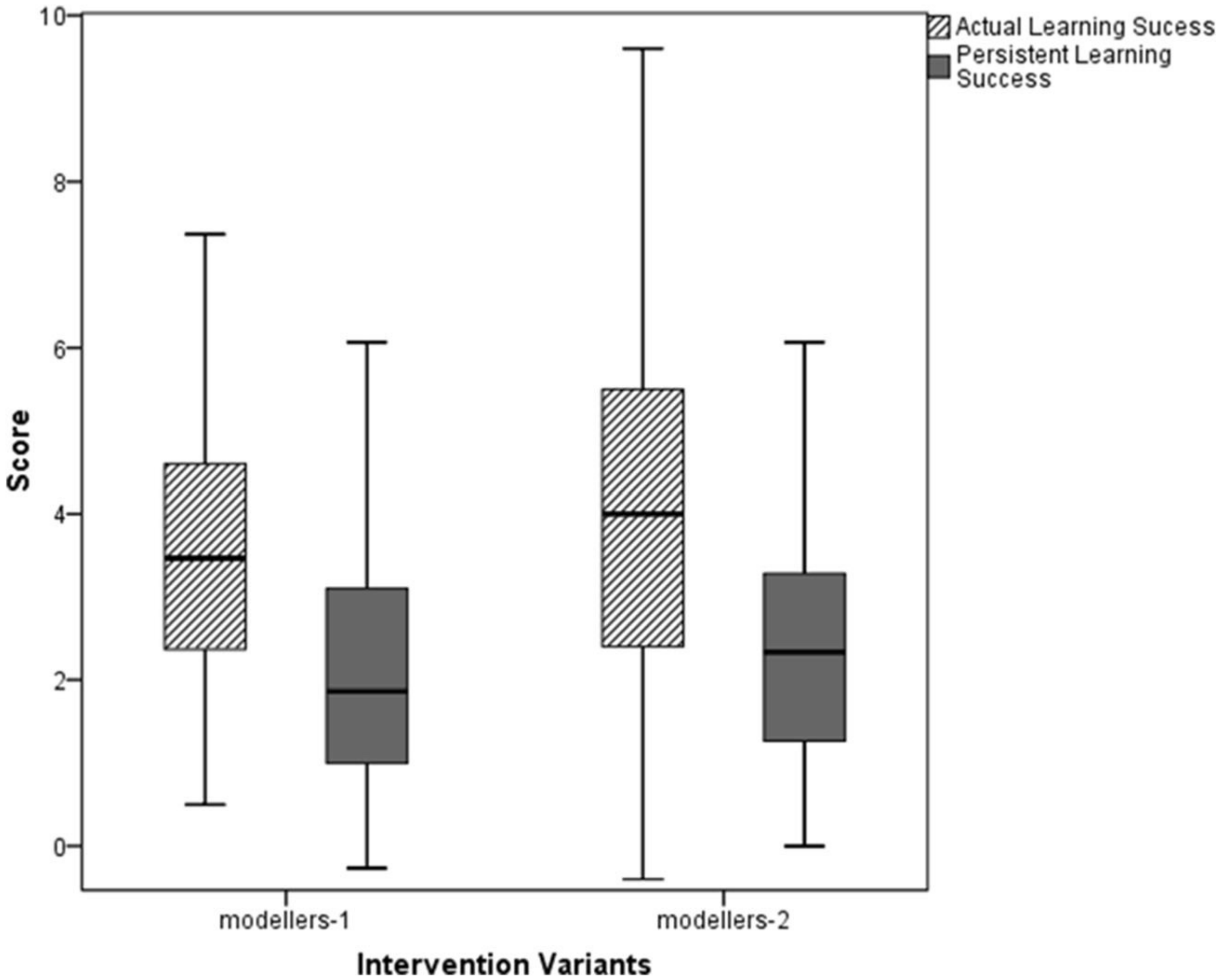


FIGURE 3 Changes in all 18model- related knowledge items for both modellers-1 and modellers-2. All participating groups increased their actual learning success scores: (T1 – T0) x (T1/18). Persistent learning success scores: (T2 – T0) x (T2/18), however, dropped in comparison to actual learning success scores.

Correlation biology grades. Biology grades and post-test (T1) scores displayed a weak negative correlation ($r_S = -.209$, $p < .001$). We obtained similar results by correlating students' biology grades and post-tests for model-related knowledge ($r_S = -.202$, $p = .001$). Splitting our evaluation variants, modellers-1 did not reveal any negative correlations ($p \geq .190$); modellers-2, however, displayed significant negative correlations of $r_S = -.235$ ($p = .006$) for overall and of r_S =-.196 (p = .026) for model-related knowledge. Thus, students with lower grades tend to achieve higher scores in overall and model-related knowledge post-tests than students with better grades. This was particularly evident in our evaluation-2 variant.

Assessment of Evaluation-2 Phase

Intra-group analyses of modellers-2 revealed the differences between students' self-evaluations and our assessment of their self-evaluation sheets and their model (Table 4; F: Chi-square 59.531, df = 2; p < .001). Pair-wise analysis revealed lower scores for their assessed self-evaluation sheets with a large effect (W: Z = -6.220, p < .001; r = .728). Thus, students identified as correct features that were not given in their model. This discrepancy was evident across all analyses sectors (Table 4; F: Chi-square ≥ 16.919, df = 2, p < .001, in each case; W: Z ≤ -2.715, p ≤ .007, in each case; r ≥ .318). In contrast – and also to a significant effect – some of the assessed models scored higher than the assessed self-evaluation sheets (W: Z = -5.804, p < .001; r = .684). Thereby, the students did not identify all of the correctly modelled features. This phenomenon was also evident across all analyses sectors (Table 4; W: Z ≤ -3.938, p ≤ .001, in each case; r ≥ .464).

Correlation analysis revealed small-to-medium correlations between the models' assessment scores, reflecting the models' quality, the actual learning success (r_S = .260, p = .003), and the persistent learning success (r_S = .251, p = .005). Further correlations between the above-mentioned variable and model-related actual learning success (r_S = .215, p = .015) and the model-related persistent learning success (r_S = .218, p = .014) were of particular importance. Thereby, models' assessment scores showed a small correlation with students' prior in-class experience of modelling (r_S = .217, p = .014).

Discussion

Our aim was to examine potential differences between participation in the two evaluation variants. The data suggest that students' short-term and mid-term DNA- and model-related knowledge was improved by the additional evaluation-2 phase.

Cognitive achievement

Overall knowledge. The effects of an additional model evaluation-2 phase on actual and persistent learning success indicate that this is a positive approach that supports effective learning. According to Ainsworth (2008), modelling affects three dimensions of learning which enable students to abstract, extend and relate their knowledge, to identify familiar concepts, and to appropriately display multiple aspects of the respective phenomenon. Altogether, this may lead to a deeper understanding of the subject (Oh and Oh 2001). In our previous study, the model evaluation-1 phase had already suggested that a one-day intervention combining model-related activities and hands-on experimentation could encourage learning in the abstract field of molecular genetics (Mierdel and Bogner 2019a). The effective integration of difficult scientific theory and working techniques into biology classrooms is, thus, a promising approach, as has been argued by Peel et al. (2019). Still, there was room for improvement which is why we conducted our present intervention with the additional model evaluation-2 phase. Compared to evaluation-1 variant (modellers-1), those participating in evaluation-2 (modellers-2) achieved higher scores regarding increases in knowledge and retention rate, and received higher actual and persistent learning success scores (Table 6). Students not only developed an initial sketch of their mental DNA model but also hand-crafted a three-dimensional model which they assessed in a comparison-based self-evaluation during which they worked in pairs to discuss and review their approach. In science, this type of evaluation is vital to assess the adequacy of previously-developed models and make efforts to improve upon them. In our study, these discussions required students to focus on providing consistent explanations and balanced assessments of their models (e.g., Passmore et al. 2009, Schwarz et al. 2009).

Model-related knowledge. Specific analysis of model-related knowledge items revealed a higher increase in knowledge and retention rate, for modellers-2 than for modellers-1. It also revealed higher actual and persistent learning success scores for modellers-2 (Table 6). Thus, our additional evaluation phase seems to impact model-based knowledge items. Schwarz et al. (2009) and Bryce et al. (2016) suggest a link between modelling practices in the classroom and students' learning success. Our measured cognitive achievement in model-related items (Table 6) might, then, be due to a "progression in knowledge and skills required for modelling, necessarily [entailing] progression in knowledge about the nature of models" (Gilbert and Justi 2016, p. 195). This possibility is also reflected in correlation analysis, with small-to-medium correlations between models' assessment scores and actual as well as persistent learning success. Thus, we can conclude that understanding of models is connected to knowledge about models (Peel et al. 2019).

Modelling also helped students to organize information about DNA as a model. It ultimately contributed to a deeper understanding of the learning content (e.g., Bryce et al. 2016; Grünkorn et al. 2014) and, in our case, supported students' understanding of DNA as a model.

Evaluation-2 Phase

Although actual and persistent learning success differed in both evaluation variants, there were also quantifiable differences in modellers-2' hand-crafted models (for example, Table 2). Some students correctly identified and labelled the DNA's different components, while some students only identified a few and others only modelled its basic structure. This result is in line with Howell et al. (2019) and Kim et al. (2015), who described students' difficulties in understanding DNA's structure-function relationships. Other studies about scientific modelling do not focus on DNA but, for instance, on the acoustic properties of materials (Hernández et al. 2015), natural selection and antibiotic resistance (Peel et al. 2019), or general classroom examples in biology, geography, and physics (Schwarz et al. 2009). It could also be discussed that students were, to a certain extent, unable to transfer their previously obtained factual knowledge into the new form of a model. Yet, although hardly any of the students had sufficient prior experience, they nonetheless found a way to excel – an achievement particularly notable among those who were considered, grade-wise, to be low-achievers (Bamberger & Davis 2011). Our own prior research suggests that most students find models (and other types of images: charts, graphs, diagrams, etc.) very difficult to engage with (personal research). Moreover, our results are in line with the findings of Quigley et al. (2017) who used the EcoSurvey tool to assess different kinds of models across different classrooms and correlated these with learning success. The authors suspected differences in modelling experience to be the underlying cause of success or failure. We also found a small correlation between the students' model quality and their prior experience in modelling at school. In our case, individual students stood out from the crowd if they either had a better understanding of models, processed new information more effectively, or read the text more carefully. Analysis of answers to DNA-model-related questions on the students' worksheet from model evaluation-1 (Table 1) supports this approach (ESM 4, ESM 5). Sample answers showed a broad understanding of DNA as a model; for instance, at particle level, that "the sugar and phosphate molecules are simplified (they usually consist of various atoms)" (ESM 4). This supports the claim that modelling is closely connected to students' academic performance (Quigley et al. 2017), as long as the modelling is not only about "doing school" but has scientific relevance (Schwarz et al. 2009, p.652).

Methodological Aspects

Firstly, as noted by Hernández et al. (2015, p. 257) several "recurring cycles of generation, evaluation and modification" would have been helpful to enhance model quality and deepen model understanding. Schwarz et al. (2009) suggest a four-step approach to successfully promote progressive understanding of models and scientific modelling. Yet the limited time available in our one-day outreach laboratory meant that this intervention was not suitable for such extensive modelling activities and we could not offer several cycles of improvement even though this might have positively influenced learning success scores (Louca et al. 2011). Such cycles would instead require regular in-class modelling and evaluation phases (e.g., José et al. 2015). To compensate for the lack of time, a more extensive pre-modelling phase prior to the outreach teaching unit could be included, during which students would be directly introduced to scientific modelling.

Second, our knowledge items about DNA and analytical methods in gene-technology laboratories only provide information about the development of students' factual knowledge. Thus, more open-ended, conceptual questions, such as those in our surveys, would give a deeper insight into student learning.

Conclusion

Although models are already used in biology lessons to encourage scientific reasoning, their effectiveness is rarely scrutinized (Werner et al. 2017). Yet, certain levels of complexity in experimentation and model design are required to adequately support scientific reasoning. Rinehart et al. (2016), therefore, have argued that all *cookbook* laboratories should be replaced with authentic, epistemic, scientific practices. Maintaining authenticity is, thereby, mandatory, as merely constructing models of scientific phenomena for the sake of modelling would miss the mark. Classroom activities that explicitly introduce students to the nature of models would be far more beneficial (José et al. 2015). Moreover, continuous in-class reflection and discussion are vital for retaining scientific authenticity and familiarizing students with real-life scientific practice (Acevedo 2008). Therefore, we consider our additional evaluation phase to be another valuable approach to integrating real scientific practices into science teaching. Using outreach laboratories, we demonstrated the impact of model-supported teaching on cognitive achievement. Based on our intervention variants, modellers-2 proved to be more effective than modellers-1 when focusing on DNA structure. Our intervention also confirms the effectiveness of research-based laboratory practice and active-learning protocols for cognitive achievement. Every student approaches new learning contents differently, and our gene-technology laboratory offers the required flexibility for differentiated teaching, addressing all types and speeds of learning (e.g., Mierdel and Bogner 2019a, b, Chen et al. 2016). Thus, teachers can apply our model evaluation and active-learning approach in the classroom and in other science subjects (e.g., in chemistry education for modelling protein structure (Torres & Correia 2007)). This will, of course, require science teachers to create new materials suitable for complex, inquiry-based lessons. Nonetheless, the modelling has the potential to encourage students to hypothesize, assess the accuracy of explanations, and identifying knowledge gaps, and is, thus, worth the effort (Svoboda and Passmore 2013). In future studies, extending the modelling phase to include several model-evaluation cycles, as suggested by Hernández et al. (2015), and assessing the impact on actual and persistent learning success would be of interest.

Funding Information

Open Access funding provided by Projekt DEAL.

Ethical Statement

1. Hereby, we consciously assure that for the manuscript "Self-evaluative scientific modelling in an outreach gene technology laboratory" the following is fulfilled:

a) This material is the authors' own original work, which has not been previously published elsewhere.
b) The paper is not currently being considered for publication elsewhere.
c) The paper reflects the authors' own research and analysis in a truthful and complete manner.
d) The paper properly credits the meaningful contributions of co-authors and co-researchers.
e) The results are appropriately placed in the context of prior and existing research.
f) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
g) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

2. Moreover, "all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Consent Statement

"Informed consent was obtained from all individual participants included in the study."

Open Access

References

Abedulwahet, M. & Nagy, Z.K. (2013). Applying Kolb's experiential learning cycle for laboratory education. *Journal of Engineering Education*, *98*, 283–294.

Acevedo, J. A. (2008). El estado actual de la naturaleza de la ciencia en la didáctica de lasciencias [The state of the art on nature of science in Science Education]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, *5*, 134–169.

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). London: Springer.

Ates, Ö. & Erylmaz A. (2011). Effectiveness of hands-on and minds-on activities on students' achievement and attitudes towards physics. *Asia-Pacific Forum on Science Learning and Teaching*, *12*, 1-23.

Bamberger, Y.M. & Davis E.A. (2011). Middle-school science students' scientific modelling performances across content areas and within a learning progression. *International Journal of Science Education, 35 (2),* 213-238.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation* [Research methods and evaluation] (2nd ed.) Springer, Berlin.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, *31*, 659–671.

Brody, M., Bangert, A., & Dillon, J. (2007). *Assessing Learning in Informal Science Contexts*. National Research Council for Science Learning in Informal Environments Committee. http://www7.nationalacademies.org/bose/Brody_Commissioned_Paper.pdf. Accessed 25 September 2019.

Bryce, C.M., et al. (2016). Exploring models in the biology classroom. *The American Biology Teacher*, *8*, 35-42.

Carpenter, S.L., et al. (2018). Models are a "metaphor in your brain": How potential and preservice teachers understand the science and engineering practice of modeling. *Science Education*, *119*, 275–286.

Chen, W., Shah, U., & Brechtelsbauer, C (2016). The discovery laboratory – A student-centred experiential learning practical: Part I – Overview. *Education for Chemical Engineers, 17,* 44-53.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.

Cook, T. D., & Campell, D. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.

Declaration of Helsinki, World Medical Association (2013). *Journal of the American Medical Association*, *310*, 2191-2194.

De Houwer, J., Barnes Holmes, D., & Moors, A. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*,202–209.

Domjan, M. (2015). *The principles of learning and behavior*. (7th ed.). Stamford, CT: Cengage Learning.

Field, A. (2012). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.

Franco C. & Colinvaux D. (2000). Grasping Mental Models. In: Gilbert J.K., Boulter C.J. (Eds.) *Developing models in science education* (pp. 93-118). Springer.

Gilbert, J.K., & Justi, R. (2016). *Modelling-based Teaching in Science Education*. Switzerland: Springer International Publishing.

Glowinski, I., & Bayrhuber, H. (2011). Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science. *International Journal of Environmental and Science Education*, *6*, 371–392.

Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, *36*, 1651–1684.

Haselgrove, M. (2016). *Learning: A very short introduction*. Oxford, UK: Oxford University Press.

Hernández, M.I., Couso, D., & Pintó, R. (2015). Analyzing students' learning progressions throughout a teaching sequence on acoustic properties of materials with a model-based inquiry approach. *Journal of Science Education Technology*, *24*, 356-377.

Hodson, D. (1998). *Teaching and Learning Science - a personalized approach*. Buckingham: Open University Press.

Howell, M.E., Booth, C.S., Sikich, S.M., Helikart, T., Roston, R.L., Couch, B.A., & van Dijk, K. (2019). Student understanding of DNA structure–function relationships improves from using 3D learning modules with dynamic 3D printed models. *Biochemistry and Molecular Biology Education, 47*, 303-317.

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Science Educator*, *17*, 10–17.

José, M.O., Del Mar Aragón, M., & Cuesta, J. (2015). The competence of modelling in learning chemical change: A study with secondary school students. *International Journal of Science and Mathematics Education*, *13*, 751-791.

Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, *24*, 369–387.

Kalyuga S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*, 402-410.

Kim et al. (2015). Cycles of exploration, reflection, and consolidation in model-based learning of genetics. *Journal of Science Education and Technology, 24,* 789-802.

Kind, V. (2015). Preservice science teachers' science teaching orientations and beliefs about science. *Science Education*, *100*, 122-152.

KMK (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Kovanović, V., et al. (2018). Understanding the relationship between technology-use and cognitive presence in MOOCs. *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge*, 582-583. New York: Association for Computing Machinery.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *Science and Children*, *49*, 10–13.

Lachman, S. J. (1997). Learning is a process: Toward an improved definition of learning. *Journal of Psychology*, *131*, 477–480.

Langheinrich, J., & Bogner, F. X. (2015). Student conceptions about the DNA structure within a hierarchical organizational level: Improvement by experiment- and computer-based outreach learn- ing. Biochemistry and Molecular Biology Education, 43(6), 393– 402.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Louca, L. T., Zacharia, Z. C., & Constantinou, C. P. (2011). In quest of productive modelling-based learning discourse in elementary school science. *Journal of Research in Science Teaching*, *48*, 919–951.

Lovey, B.R., & Riggs K.M. (2019). Flipping the laboratory: improving student engagement and learning outcomes in second year science courses. *International Journal of Science Education*, *41*, 64-79.

Mierdel, J., & Bogner, F.X. (2019a). Investigations of modellers and model viewers in an out-of-school gene technology laboratory. Research in Science Education, https://doi.org/10.1007/s11165- 019-09871-3.

Mierdel, J., & Bogner, F. X. (2019b). Comparing the Use of Two Different Model Approaches on Students' Understanding of DNA Models. Education Sciences, 9(115), 1–18.

Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C.D. (2006). Learning as a cultural process: Achieving equity through diversity. In R.K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 489-504). Cambridge: Cambridge University Press.

Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, *33*, 1109–1130.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education*, *96*, 21–47.

Orhan, T.Y., & Sahin, N. (2018). The impact of innovative teaching approaches on biotechnology knowledge and laboratory experiences of science teachers. *Education Sciences*, *8*, 213-237.

Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. *School Science and Mathematics*, *109*, 394–402.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, *50*, 157–175.

Peel, A., Zangori, L., Friedrichsen, P., Hayes, E., & Sadler, T. (2019). Students' model-based explanations about natural selection and antibiotic resistance through socio-scientific issues-based learning. *International Journal of Science Education, 41*, 510-532.

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. *European Scientific Journal*, *12*, 235–250.

Quigley, D., Ostwald, J., & Sumner, T. (2017). Scientific Modelling: Using learning analytics to examine student practices and classroom variation. *Proceedings of the Seventh*

International Learning Analytics and Knowledge Conference, 329-338. New York: Association for Computing Machinery.

Raviv, A., Cohen, S., & Aflalo, E. (2019). How should students learn in the school science laboratory? The benefits of cooperative learning. *Research in Science Education*, *49*, 331-345.

Rinehart, R., Duncan, R., Chinn, C., Atkins, T., & DiBenedetti, J. (2016). Critical design decisions for successful model-based inquiry in science classroom. *International Journal of Design for Learning*, *7*, 17-40.

Rost, J. (2004). *Lehrbuch Testtheorie–Testkonstruktion* [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

Rotbain, Y., Marbach-Ad, G., & Stavy, R. (2006). Effect of bead and illustrations models on high school students' achievement in molecular genetics. *Journal of Research in Science Teaching*, *43*, 500–529.

Saka, A., et al. (2006). A cross-age study of the understanding of three genetic concepts: how do they imagine the gene, DNA and chromosome? *Journal of Science Education and Technology, 15 (2),* 192-202.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. *Innovations in Education and Teaching International*, online first, 1–12.

Scharfenberg, F. J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus: Results of a teaching unit with authentic experiments. Biochemistry and Molecular Biology Education, 35(1), 28–39.

Scharfenberg, F. J., & Bogner, F. X. (2013). Teaching Gene Technology in an Outreach Lab: Students' Assigned Cognitive load clusters and the clusters' relationships to learner characteristics, laboratory vari- ables, and cognitive achievement. Research in Science Education, 43, 141–161.

Schwarz, C., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46,* 632–654.

Solomon, J. (1980).*Teaching children in the laboratory*. London: Crom Helm.

Svoboda, J., & Passmore, C. (2013). The strategies of modeling in biology education. *Science & Education*, *22*, 119–142.

Thorpe, W. H. (1943). Types of learning in insects and other arthropods. *British Journal of Psychology*, *33*, 220–234.

Torres, B.B. & Correia, P.R.M. (2007). Using denatured egg white as a macroscopic model for teaching protein structure and introducing protein synthesis for high school students. *Journal of Chemical Education, 84 (12),* 1941-1944.

Usher, S. (2013). *Letters of Note. Correspondence Deserving of a Wider Audience*. Edinburgh, Scotland: Canongate.

Werner, S., et al. (2017). Investigating how German Biology teachers use three-dimensional physical models in classroom instruction: A video study. *Research in Science Education*, *49*, 1–27.

Windschitl, M., & Thompson, J. (2006). Transcending simple forms of school science investigation: The impact of preservice instruction on teachers' understandings of model-based inquiry. *American Educational Research Journal*, *43*, 783–835.

Wolf, R. (1997). Rating scales. In J. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook* (pp. 958–965). Oxford, UK: Elsevier.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An Overview. *Theory Into Practice*, *41*, 64–70.

Electronic Supplemental Materials

Self-evaluative Scientific Modeling in an Outreach Gene Technology Laboratory

Tamara Roth[1]*, Franz-Josef Scharfenberg[1], Julia Mierdel[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de; ORCID ID: 0000-0002-9062-1489

Table of Contents

(Total pages 10)

Electronic Supplemental Material 1: Laboratory Manual

DNA-MODELING

After it had been proven by experiments that the DNA carries the genetic information, important questions about the structure of the DNA and the encryption of the genetic code were raised. At the beginning of the 1950s there was a race between several groups of researchers who wanted to answer these questions.

<u>**Tasks:**</u>

1) Read the **information text** about DNA-structure which you find in the modeling box!
2) **Underline** the chemical **components** that compose the **DNA**!
3) Answer the **following questions about DNA structure** with the help of the text!

(1) How many strands does the DNA consist of?

The DNA consists of two single strands.[1]

(2) Please tick as appropriate! The DNA-strands are …

- ☐ independent of each other.
- ☒ antiparallel.
- ☐ arranged offset to one another.
- ☐ identical.

(3) Explain the DNA-strands' cohesion!

The DNA single strands are linked to each other via hydrogen bonds between the individual base pairs.

(4) DNA's backbone: Label its components and describe their set up!

It consists of phosphate groups and a sugar called deoxyribose. These components are alternately arranged resulting in a phosphate-sugar-chain

(5) Which DNA backbone's compound are the bases linked to?

The single bases are firmly attached to the sugar components of the DNA backbone.

[1] Italicised sentences: potential student solutions.

(6) Name the bases and indicate possible base pairings!
The bases are adenine (A), thymine (T), guanine (G) and cytosine (C). Only two possible base pairings exist: (A) pairs with (T) and (G) pairs with (C) (complementary base pairing).

(7) Give the matching bases to the base sequence 'GATCTA'!
The base sequence of the complementary DNA strand is CTAGAT.

(8) Check! The DNA bases are located ...

- [x] inside of the DNA-molecule.
- [ ] outside of the DNA-molecule.

(9) The DNA forms a right-handed double-helix. Explain what it means!
If you look at the vertical single strands from above, you see that they are helically wound around each other in a clockwise direction.

(10) Describe how genetic information is encoded in the DNA!
The base sequence is a code: Certain sections form different genes which carry the basic information for the expression of different hereditary characteristics.

4) Slip into the role of the scientists Watson & Crick and **build a DNA-model** with the help of your previously answered questions! Work in pairs.
You will find a selection of materials in the box (glue, straws, scissors, pipe cleaners, colored beads and cardboards etc.).

Some clues:

- First consider which materials could represent the DNA's single components (several options are possible)!
- Decide yourself which material appears to be the right one for you and start modeling.
  Only build a short sequence of the giant molecule!
- Try to visualize as many characteristics of the DNA-structure as possible in your model!

5) Draw a **simplified sketch** of your model and label its components!
You will be given an extra sheet of paper for this purpose.

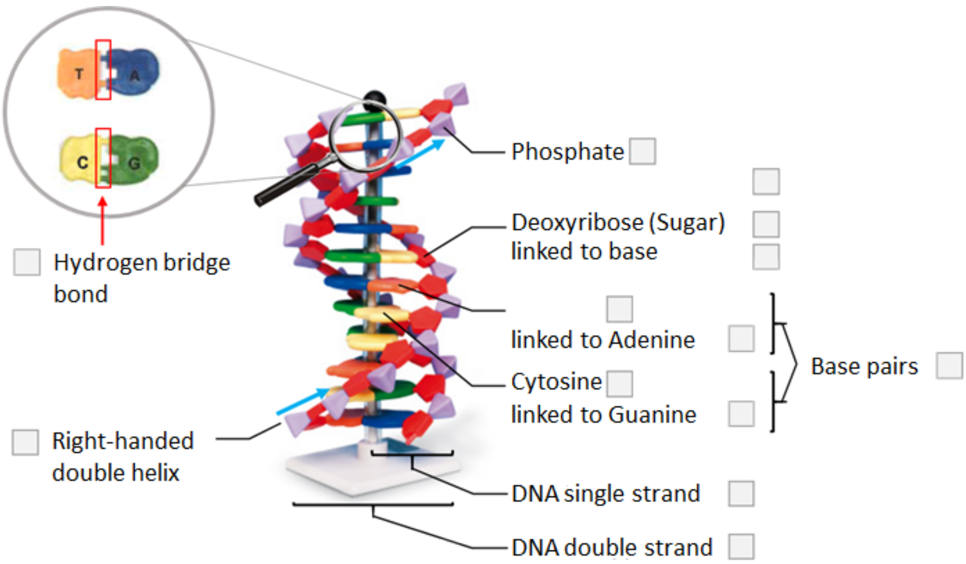
6) Finally, answer the two questions on the extra sheet!

Electronic Supplemental Material 2: Comparison of DNA Models Sheet

COMPARISON OF DNA MODELS

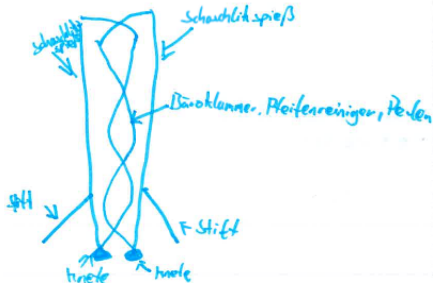
Tasks:

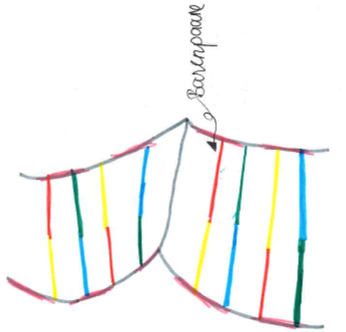
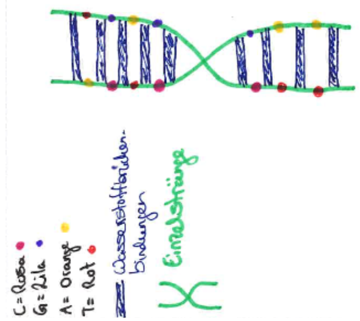
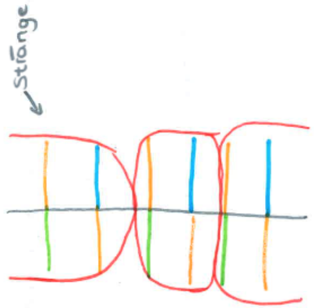
1. Carefully **look at** your DNA model and **compare** it to the school model which is handed out to you (similarities/differences)!
2. **Tick** the **appropriate boxes** in the figure below where components of your self-built model correspond with features of the school model.

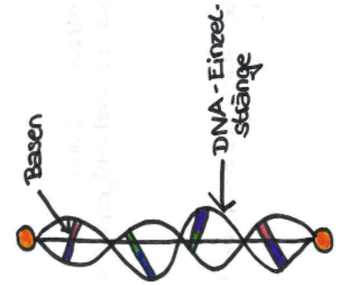
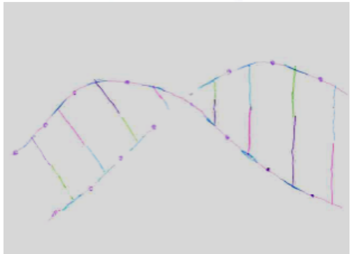
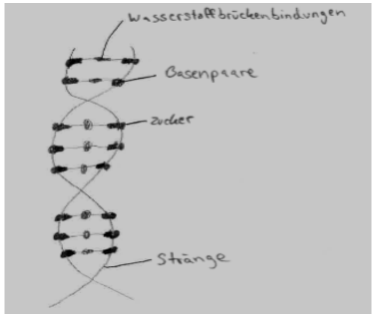


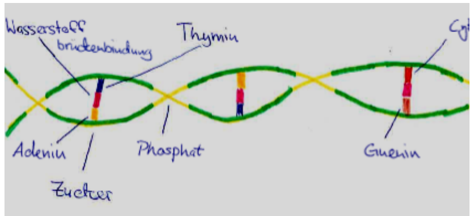
Electronic Supplemental Material 3: Assessment of Students’ Sketches

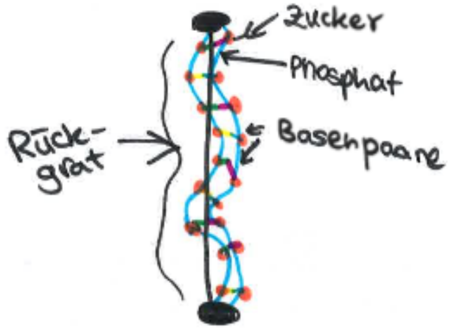
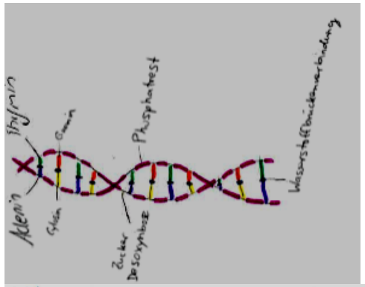
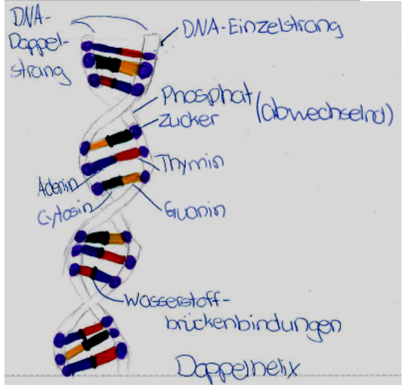
Assessment of students’ sketches based on their handcrafted models (modified after Langheinrich & Bogner 2015).

Analysis		Definition	Score	Example	Sketches’ scores *Md* (25^{th} / 75^{th} *P*)	
Sector	Subsec-tor				Modellers-1[a]	Modellers-2[b]
Bases	BA1	No bases drawn	0			
		Bases drawn	1	/		
		Bases drawn and labeled	2	/		
		Base pairs indicated	3	/		
		Base pairs indicated and labeled	4			

Sum BA1				3.2 (1.5/4.0)	3.4 (2.0/4.0)
BA2	Hydrogen bonds not indicated	0	 		
	Hydrogen bonds indicated	1	 		
	Hydrogen bonds correctly indicated	2	See above, example BA1-4		
Sum BA2				0 (0/0)	0.6 (0/2.0)
BA3	Bases indicated as not linked to the backbone	0	 		

		Bases indicated as linked to the backbone	1	See above, example BA1-4		
		Bases indicated and correctly linked to the backbone	2	See above, example BA1-3		
	Sum BA3				1.4 (1.0/2.0)	1.3 (1.0/2.0)
Deoxyri-bose	DE	Deoxyribose not indicated	0			
		Deoxyribose indicated	1			
		Deoxyribose indicated and labeled	2	See above, example BA1-3		
	Sum DE				1.4 (0/2.0)	1.2 (0/0.2)
Phosphate	PH	Phosphate not indicated	0			
		Phosphate indicated	1	See above, example DE1		
		Phosphate indicated and labeled	2	See above, example BA1-3		

	Sum PH				1.4 (0/2.0)	1.0 (0/2.0)
Primary structure	PR1	No primary structure visible	0			
		Single strand visible	1	Not determined		
		Double strand visible, composed of two single strands	2	See above, example BA1-3		
	Sum PR1				2.0 (2.0/2.0)	2.0 (2.0/2.0)
	PR2	No links between deoxyribose and phosphate	0	Not determined		
		Links between deoxyribose and phosphate indicated	1	See above, example BA1-3		
		Deoxyribose and indicated and alternatingly arranged as a polymer	2			
	Sum PR2				1.5 (0/2.0)	1.0 (0/2.0)
Secondary structure	SE	No secondary structure visible	0	See above, example PR1-0		

	Wrong secondary structure visible	1			
	Double helix visible	2			
	Right-handed double helix visible	3			
	Sum SE			2.1 (2.0/3.0)	2 (2.0/2.0)
Sum		19		13.3 (9.0/16.0)	12.0 (8.0/16.0)

Notes: [a]sum 77 sketches; [b]sum 73 sketches

Electronic Supplemental Material 4: Assessment of Students' Responses 1

Assessment of the students' responses to the question "Which features of the original DNA molecule are simplified in your model".

Category	Definition	Examples of students' answers	Frequencies	
			Modellers-1[a]	Modellers-2[b]
Level of	The student's answer mention a simplification of			
… DNA	… the DNA molecule's size or its length (e.g., by number of bases).	"Usually, the DNA molecule is longer (it has more phosphate and sugar components)"[c]. "Original DNA molecule has more base pairings (they are closer together)"[d].	35	29
… substances	… the DNA molecule due to the use of different materials or colors for the different components.	"The components are not illustrated chemically, but are colored for differentiation"[c]. "All DNA-Elements are represented by different colors"[d].	18	9
… particles	… the representation of the DNA molecule or its particles.	"The DNA molecules are simplified"[d]. "The [sugar and phosphate molecules] usually consist of various atoms"[d].	20	6
… structure	… the DNA structure in general or related to structural elements (e.g., base pairing, backbone, hydrogen bonds, or secondary structure).	"The whole DNA structure is simplified"[d]. "The base pairings are simplified"[c]. "The backbone is kept simple"[d]. "In our model the hydrogen bonds are simplified"[d]. "The double helix structure is shown in a simplified form as well"[d].	140	127

Notes: [a] sum 213; [b] sum 171; [c] modellers-1; [d] modellers-2

Electronic Supplemental Material 5: Assessment of Students’ Responses 2

Assessment of students’ responses to the question “Explain for what reason one may create different models of one biological original (in our case, the structure of the DNA)” (adapted from Mierdel & Bogner 2019).

Category	Definition	Examples from the students‘ answers	Frequencies	
	Students’ responses refers to		Modellers-1[a]	Modellers-2[b]
Individuality of DNA	… the individuality of the original DNA related to human beings.	„Every human being has another and an individual DNA”[d].	26	19
Different				
… interpretation	… different interpretations of the original DNA.	“Everyone has his or her own interpretation”[d].	14	12
… model design	... the different model design (e.g., different colors, materials, differing sizes or differing detail levels).	“Models can differ in their colors or sizes”[d]. “Models can look different, depending on the materials they are made of”[c]. „Modellers may model the DNA either roughly or just with all details“[c].	16	9
… focus	… different perspectives and variations of focus to be applied to the model (e.g., different sections, or states of the original).	“To highlight certain features, others have to simplified or ignored”[c]. “Every model has a different and specific purpose of illustration”[d].	10	14
… research state	… to new research findings about the original DNA structure.	“Models can differ because new findings are emerging constantly”[d].	1	1

Notes: [a] sum 77 (including 10 dyads with no answer); [b] sum 73 (including 18 dyads with no answer); [c] modeller-1; [d] modeller-2.

5.4 Teilarbeit B

Research in Science Education, DOI: 10.1007/s11165-021-10005-x, 2021

Testing creativity and personality to explore creative potentials in the science classroom

Tamara Roth[1*], Cathérine Conradty[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

Integrating creativity into science classes may pave the way to tapping complex scientific phenomena. Although not yet conclusively defined nor assessed using standardized measures, creativity is understood to support cognitive learning in formal and informal settings. However, the successful integration of creativity in educational modules depends on many factors. As our knowledge of how to identify these factors is still limited, teachers may have difficulties effectively monitoring and fostering creativity. Consequently, a valid means to measure creativity would help teachers to identify creativity and its influencing factors within the limited scope of science lessons. In the present study, we collected data from 538 Bavarian secondary school students (M±SD=16.96±2.99; 65.4%, female) focusing on personality and creativity measures. Comparable to previous studies, two subscales for creativity were applied: act, comprising conscious and adaptable cognitive processes, and flow, describing a creative mental state of full immersion. Since personality is understood to be linked to creativity, we used the Big Five scale with its shortened item battery to assess personality. We found that personal characteristics such as conscientiousness and flow, openness and agreeableness, and extraversion and neuroticism were significantly correlated. Anticipated gender and age differences were only evident when extreme groups were compared: age influenced act in younger male students and flow in older female students. Drawing on the literature and our results, we suggest pedagogical approaches to provide opportunities for creativity in science classrooms.

Keywords: Creativity. Personality. Science education. CPAC. BFI-10. Gender. Age

Introduction

The task of fostering creativity – along with other, related personality traits – is increasingly gaining attention in educational settings as a result of the Ministry of Education's focus on competence development (KMK, 2020). Creativity is understood to support the acquisition, transfer, and application of knowledge in schools and, hence, to contribute to an innovative and flourishing national economy (Lewis, 2008). Integrating creativity into science education thus promises a more sustainable approach to the provision and management of knowledge (Henriksen et al., 2018). However, authors such as Wyse and Ferrari (2015) have so far recommended that the integration of creativity should occur, primarily, in arts-related subjects. This excludes many other relevant subjects – for instance, science classes – which lack representation in training plans (Robinson, 2011). Yet, creativity is a critical aspect of science education, with factual knowledge alone no longer sufficient to foster the development of know-how and new scientific concepts (Brynjolfsson & McAfee, 2014). On the contrary, it is increasingly evident that the self-perpetuating system of "linearity, conformity, and standardization" does not advance "organic, adaptable, and diverse" societies (Robinson, 2011, p.4).

To quantitatively measure creativity in educational contexts, we applied the empirical tool Cognitive Processes Associated with Creativity (CPAC, Miller & Dumford, 2016) which has been significantly shortened by Conradty and Bogner (2018) who were, subsequently, the first to apply the tool in informal science learning modules (Conradty & Bogner, 2019). The empirical focus was primarily on the subscales act and flow (Conradty & Bogner 2018; Conradty & Bogner 2019). As creativity is understood to be specific to individuals and linked to personality factors (Kaufman et al., 2008; Barron & Harrington, 1981), we used the Big Five scale (Caprara et al., 1993) in its shortened version BFI-10 (Rammstedt & John, 2007) to explore previously reported interconnections with personality. Based on existing literature and our results, we suggest pedagogical approaches that create opportunities for creativity in science classrooms.

Theoretical Background

Attempts to define creativity have earned it a reputation as a particularly elusive construct (Corazza, 2016; Plucker and Renzulli 1999). Torrance (1988) suggested that "creativity defies a precise definition because it is largely unseen, nonverbal, and unconscious" (p. 43). As no standard definition exists, different elements related to individual perceptions of creativity are often added and later dismissed (Runco, 2019) in accord with individual perceptions of creativity. Runco and Jaeger (2012), for instance, first focused on "originality" and "effectiveness" as integral elements of the creative product but later reconsidered their decision and instead emphasized the less static creative process. For the present study, the definition of creativity for educational context proposed by Henriksen et al. (2018) seems to be most fitting: creativity is both "novel and effective, in addition to the subtler component of wholeness or context [...] which exists as a dynamic process emerging through a system of interactions" (p. 29).

Csikszentmihalyi (1988), for instance, extended the discourse around creativity beyond a mere definition with the claim that creativity is the outcome of three integral driving elements: social institutions (the environment that determines what is considered creative), a stable cultural domain (the social structures that help preserve and pass on what is creative), and the individual's innovative capabilities (the creative actions that may change the domain or environment). While, coincidentally, observing artists at work, Csikszentmihalyi also recognized how immersed they were in the process and how this influenced their creative power (Cseh, 2016). The word flow may best describe this state and has, thereby, become a key element of motivation and creativity studies (Piniel & Albert, 2020) and of relevance to the study reported here.

Creativity's Contemporary Relevance

The perceived importance of research into creativity reflects the increasing recognition of the need to support creative thinking in a competency-focused and innovative society (Corazza, 2016). On a daily basis, people are exposed to a flood of information, and efficiently transforming this information into know-how requires creativity (Chua, 2015). Perry-Smith (2014) designed a creativity-supporting environment that could also be transferred to science education settings. She proposed that individuals must be exposed to a variety of different frames – defined as "lenses through which individuals view a situation or problem, [or] a way for individuals to make sense of a problem" (Perry-Smith, 2014, p. 833) – to broaden their interpretative horizon. This exposure to different frames relies on meaningful social interaction and communication: both are key to promoting the type of diverse and resourceful working atmosphere which, by offering various frames, can nurture creativity. In return, creativity may foster innovation, which is important from an economic and social perspective (Perry-Smith, 2014). Innovation involves the development of effective, novel solutions to problems and, thus, entails creative problem-solving strategies. These strategies are particularly important in the fields of science and mathematics, meaning this form of expressed creativity is highly relevant for both science professions and science education (Aldous, 2007).

Creativity in Science Education

Current educational policies and attempts to standardize school systems tend to reduce creativity rather than to promote it (Kupers et al., 2018). Therefore, teachers have to find ways to encourage creativity, particularly in science classrooms, so that graduates can meet the demands of an innovative, globalized economy (Henriksen et al., 2018) and develop creative approaches to problems ranging in complexity (Aldous, 2007). Of course, creativity is crucial to areas beyond the sciences (Lucas, 2016), but, due to the scope of this paper, we focus mainly on the role of creativity in this discipline. Currently, few students associate creativity with science or see scientific knowledge as the product of creative processes, even though "creativity is inextricably linked to the nature of science itself" (Hadzigeorgiou et al., 2012). The integration of creativity in science education "should be rooted in and reflect aspects of creativity seen in scientific research", and educators should "devise a framework appropriate to children's needs and abilities" that allows for creative processes (Kind & Kind, 2007, p.3). In this approach to science education, students still acquire scientific knowledge, but do not generate new knowledge themselves. Hence in trying to stimulate scientific creativity in the classroom, the teacher must acknowledge the differences between the learner role of the student and the creativity required of professional scientists (Hodson, 1998; Kind & Kind, 2007). Richards and Cotterall (2016) offer guidelines to help teachers establish creativity in class. While collaborative learning approaches require various teaching techniques (Yager, 2005) and effective teacher training to sustain creative processes in the science classroom (Goodwin & Gotlib, 2004; Sawyer, 2012), some researchers suggest that enforced group work may diminish creative thinking (Csikszentmihalyi, 2000; Schmidt, 2011). Most importantly, extrinsic motivators – for example, the evaluation of current work, offers of specific rewards Lepper and Greene 1978, or situations that could be interpreted as relevant for exams – should be avoided. Although these are common school practices, they can have a detrimental effect on creative processes (Baer, 2010; Amabile, 1983). The specific rewards, in particular, may bear "hidden costs" for the development of creativity since they undermine intrinsic motivation (Baer, 2010, p.25).

Gender in Creative Science Education

Creative thinking in science subjects also appears to be influenced by gender. Okere and Ndeke (2012) identified considerable gender-related differences in scientific creativity in terms of flexibility, planning, and recognition of relations. Other researchers, however, have not found any significant gender-related differences regarding general creativity (Charyton & Snelbecker, 2007; Charyton et al., 2011). A possible reason for these diverging results could be male and female interest and performance in scientific subjects rather than in creativity itself. Trends in International Mathematics and Science Studies (TIMSS), which measure the basic understanding of math and science, support this theory since historically, male students have generally outperformed females in most science disciplines and expressed more interest in scientific topics (Neuschmidt et al., 2008; Goldman & Penner, 2016; Mejía-Rodríguez et al., 2020). Yet, in the most recent TIMSS 2019 report, they also found that girls may outperform boys in experiments where specific procedural steps were not included in the directions and many other scientific domains that involve reasoning and the application of knowledge (Mullis et al., 2020). If current understandings are correct that creativity is linked to intrinsic motivation (Csikszentmihalyi, 2000; Baer, 2010; Runco, 2014), genuine interest in the respective subject area may also indirectly influence scientific creative thinking and creative processes. Recent TIMSS reports (Mullis et al. 2020) also show that the gender gap is steadily decreasing (Neuschmidt et al., 2008; Mejía-Rodríguez et al., 2020), suggesting that the identified gender differences in creativity can be attributed to group effects (Baer, 2010; Charyton et al., 2011).

Establishing Creative Science Classrooms

Sawyer (2012) expects that collaborative learning approaches, in particular, will further contribute to gender balance in science subjects and thus in scientific creativity. The framework developed by Kaufman et al. (2008) for categorizing the development of an individual's creativity supports this assertion. Kaufman et al. divided creativity into four categories: (1) mini-c, which encompasses early and exploratory approaches to uncover individual creativity; (2) little-c, which develops with encouragement and several trials of mini-c creative processes; (3) pro-c, wherein the individual is "capable of working on problems, projects, and ideas that affect the field as a whole"; (4) and big-C, which is accomplished after many years of creative practice and is referred to as "eminent creativity" (Kaufman & Beghetto, 2009, p.6). Social interactions in the classroom may foster little-c creativity as classmates and teachers challenge creative scientific contributions made by individuals. This, in turn, can encourage mini-c creative thinking as individuals reflect on their own creative scientific interpretations.

However, the degree of structural freedom necessary to effectively foster creativity would require schools to fundamentally rethink classroom teaching (Cho et al. 2017). Since student learning in school science lessons is characterized by trial and error approaches and a lack of profound factual knowledge, teachers should create learning environments wherein extrinsic motivators are considerably reduced and which offer students the freedom to creatively explore scientific learning content individually or in groups to find an adequate solution to a scientific problem (Lepper & Greene, 1978). The structural freedom and open classroom may accordingly enable creative thinking both collectively and individually dependent on individual preferences (Amabile & Pratt, 2016; Beghetto, 2015). Yet, this necessary classroom transformation would require extensive teacher training and substantial investments beyond the capacity of many schools (Cho et al. 2017).

Measuring Creativity in Science Classrooms

Another major obstacle to promoting creativity in science education is the difficulty in measuring creativity and the resulting best practices for teaching. Although Hocevar and Bachelor (1989) identified over 100 applied measures, all followed different approaches. Piffer (2012) attributed this phenomenon to the lack of a standard definition of creativity, yet all measuring techniques were based on the four main categories of creativity – "process, product, person, and press" – each with a different focus on one category (Said-Metwaly et al., 2017, p.243). Accordingly, when selecting suitable measuring techniques, it is important to remember that, while there is no consensus on a single definition, there can be no single, authoritative measure for creativity.

At best, common creativity tests can grasp "aspects of [individual] creative potential" (Piffer, 2012 p.263). This, along with an often unnecessarily long test design, discourages teachers from identifying and fostering creative skills in the classroom. Thus, effectively integrating creativity into the science curriculum would require rapid testing and conclusive results (Kaufman et al., 2008). Teachers would also require sufficient tools to (1) understand test results, (2) draw links between the results and beneficial changes in science teaching and classroom setup, and (3) implement the findings in a way that effectively fosters creativity (Horng et al., 2005). That is, measuring changes in students' creativity and reading how to nurture this creativity help teachers to redesign their lessons, adapt the classroom environment to their students' various needs, and, finally, encourage individual scientific creativity (Südkamp et al., 2012; Gralewski & Karwowski, 2019).

Applied Empirical Instruments

Since creativity is understood to be linked to personality (Kaufman et al., 2008; Barron & Harrington, 1981), we decided to combine the shortened version of Miller's (2014) Cognitive Processes Associated with Creativity (CPAC) scale, which also best reflects the different dimensions of Henriksen et al.'s (2018) definition of creativity, with the Big Five questionnaire (Caprara 1993) in its ten-item version, commonly referred to as BFI-10 (Rammstedt & John, 2007), to explore students' creativity and their individual personality factors related to creative activities (Kaufman et al., 2008). In this context, we also show that each of the 25 items has already been successfully implemented in different science classroom settings. For teachers, this combination would, thus, provide two measuring techniques to reliably assess aspects of creative potential and personality (Conradty & Bogner, 2018; Conradty & Bogner, 2019).
The CPAC scale has proven particularly suitable for assessing six dimensions of creativity – "idea manipulation, idea generation, flow, imagery/sensory perception, incubation, and metaphorical/analogical thinking" – using a set of 28 items (Tsai, 2018, p.271). Since the use of a scale comprising 28 items would take more time than is likely to be available for conducting such tests in science classrooms, a shortened scale was suggested by Conradty and Bogner (2018), involving only 15 items. Although shorter than the original developed by Miller (2014), this revised scale contains all relevant dimensions of the longer variant and has already been validated in science outreach education (Mierdel & Bogner, 2019).

Previous studies have already revealed generalizable correlations between the positive personality traits of extraversion and openness and creativity (Furnham et al., 2013; Antinori et al., 2017; Hoseinifar et al., 2011). Openness has recently been revealed to affect both how curiously and creatively individuals explore the world and how they actually experience and perceive their environment from a neurocognitive point of view. Yet, this trait is still among the least well understood of all five (Antinori et al., 2017). However, it is not only personality factors regarded as positive that are correlated with creativity but also neuroticism (Batey et al., 2010), with its subscales volatility and withdrawal (Watrin et al., 2019). Neuroticism refers to individual predispositions towards negative feelings such as anxiety, depression, and anger, as well as certain responses to loss or frustration. Expressions of neuroticism differ between genders (Tackett & Lahey, 2017), yet the trait is both positively and negatively related to many different measures of creativity. Other traits, meanwhile, are tied to specific aspects of creativity: extraversion is linked to creative achievement and openness to self-rated creativity (Batey et al., 2010). For a long time, personality was considered stable and unchangeable. However, recent studies, such as Kitamura et al. (2015), revealed that certain high-order personality traits included in the Big Five could change in response to various influences. This has inspired the idea of designing educational interventions specifically aimed at fostering behaviours directly connected to creativity (Hoseinifar et al., 2011).

Comparable to creativity, the feasibility of assessing personality has been widely discussed for decades, particularly with regard to the number of factors required in such assessments. One widely endorsed outcome was the Big Five questionnaire (Caprara 1993), although Eysenck (1981) had previously proposed the three-factor model TFM measuring psychoticism, extraversion, and neuroticism, based on biological theories. Eysenck also effected a shift to higher-order (type) factors to analyse personality, rather than of primary (trait) factors (Wiseman & Bogner, 2003). Caprara's (1993) questionnaire, however, has its roots in the assumption that certain personality and behavioural patterns are present in creative individuals (Barron & Harrington, 1981) and, accordingly, distinguishes between personalities regarded as "normal", "extremes of the normal", and "abnormal", for instance, obsessive-compulsive (Furnham et al., 2013). Thus, the questionnaire breaks down the complex construct of personality into five essential features – neuroticism, friendliness or agreeableness, conscientiousness, extraversion, and openness/intellect – originally comprising over 100 items in its standardized version (Watrin et al., 2019). Many researchers (Carciofo et al., 2016; Balgiu, 2018) today, however, deploy the Big Five's shortened version with only 10 items, commonly referred to as BFI-10. The shortened scale is in no way inferior to the more detailed scale, especially in classroom settings, which has been proven by Rammstedt and John (2007), hence our decision to use the BFI-10.

Our study set out to answer three research questions. (i) How do the shortened versions of the CPAC and Big Five capture their respective constructs in classroom settings? (ii) How does personality – as described in the BFI-10 – correlate with act and flow in the CPAC question- naire? (iii) How do groups at either end of the age spectrum differ in gender-specific creativity and personality characteristics?

Materials and Methods

Sampling Procedure

Altogether, 538 Bavarian secondary school students from three schools with a focus on STEAM education participated in our study (girls 65.4%, boys 34.6%; M_{Age} = 16.96; SD = 2.99). They completed a questionnaire comprising 25 questions extracted from the BFI-10 and

CPAC scales during their biology classes. Regarding the construct validity of CPAC, inter-item correlations below 0.20 confirmed that each item referred to different creativity facets. Furthermore, the heterogeneity of the items with respect to complex constructs, such as creativity, emphasizes the given construct validity (Rost 2004). The Big Five consists of 10 items taken from the BFI-44, with two items for each Big Five domain (one reverse-scored): extraversion (items 6, 36); agreeableness (items 2, 22); conscientiousness (items 3, 23); neuroticism (items 9, 39); and openness (items 20, 41). Each item was assessed using a 5- point Likert scale ranging from “strongly disagree” (1) to “strongly agree” (5) (Rammstedt & John, 2007). The same Likert scale was used to assess the CPAC items. The CPAC comprised a total of 15 items representing the creativity domains of idea manipulation (7, 15), imagery/ sensory (1, 10), flow (2, 3, 9, 13, 14), metaphorical/analogical thinking (5), and idea generation (4, 6, 8, 11, 12) extracted from the original CPAC scale (Conradty & Bogner, 2018; Conradty & Bogner, 2019). The age and gender variables were collected by asking participating students for their age and gender, which was later categorized as a nominal variable.

Statistical Analyses

Subsequent statistical analyses were conducted using RStudio Team (2020) and IBM SPSS Statistics 26.0. Our data were not normally distributed following assessment with the Shapiro- Wilk test ($p < 0.001$). Despite our large sample size, we also refrained from normalizing the data as would be recommended by Bortz and Schuster (2010), since – due to Likert scaling – our measured values have no clearly interpretable relative distances. Moreover, our sample is largely heterogeneous in terms of age and gender distribution, which further encourages non- parametric analysis (Lomax, 1986).

We used RStudio Team (2020) for confirmatory factor analysis of the shortened CPAC to show that the predicted structure from our previous interventions (Conradty & Bogner, 2018; Mierdel & Bogner, 2019) was replicable (Thompson, 2004). Based on theory (Kaiser, 1970) and previous analysis (Conradty & Bogner, 2018), we assumed our sample would divide into two factors flow and act. To confirm the model’s adequacy, we calculated the comparative fit index (CFI = 0.837) and Tucker-Lewis index (TLI = 0.808) as well as the root mean square error of approximation (RMSEA < 0.050, $p = 0.492$). We also extracted factor scores for further calculations and factor loadings for a pattern matrix. Using a principal component analysis with subsequent oblique rotation – which reduces the data’s dimensionality while retaining its variation (Bro & Smilde, 2014) – we evaluated the Big Five test. In accordance with the Kaiser-Guttman criterion (Kaiser, 1970), it was divided into five factors with the Kaiser- Meyer-Olkin

(KMO = 0.54, χ^2 = 484.1) values being just about acceptable, indicating that conducting a factor analysis with our dataset was feasible.

To explore potential gender differences, particularly with regard to outlying group differences in age, we applied Mann-Whitney U (MWU) tests (Field, 2012). Before correlation analysis, however, we randomly selected 340 participants from the overall sample to avoid producing biased results due to differences in age and gender groups. In case of significant results, effect sizes r (Lipsey & Wilson, 2001) were calculated with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. The results were transferred from R to SPSS to create boxplots with sufficient graphic quality.

Results

Our results indicate that both CPAC and BFI-10 measure expected constructs, although not all observed variables adequately describe the latent constructs act and flow in the shortened CPAC scale. Moreover, gender differences only appear for outlying age groups and are otherwise not significant. Cronbach's α scored 0.675 for CPAC proving the scale's reliability to be acceptable (Lienert & Raatz, 1998). A subsequent principal component analysis with oblique rotation confirmed two factors based on eigenvalues <1.0, accounting for 60.03% of the total variance. All items displayed in the pattern matrix (Table 1) reached a total KMO score of 0.73 indicating reliable and distinct factors (Kaiser, 1970). Due to correlations between the components, however, we could calculate neither the sum of squared charges nor the total value.

TABLE 1 Principal component analysis with *CPAC* after oblique rotation (valid N = 538).

Pattern Matrix[a]		Components *factor analysis*	
		1	2
CPAC 08	If I get stuck on a problem, I try to take a different perspective of the situation.	,633	
CPAC 10	Imagining potential solutions to a problem leads to new insights.	,589	
CPAC 07	Thinking about more than one idea at the same time can lead to a new understanding.	,546	
CPAC 15	If I get stuck on a problem, I look for details that I normally would not notice.	,403	
CPAC 04	While working on something, I try to generate as many ideas as possible.	,419	
CPAC 01	Becoming physically involved in my work leads me to good solutions.	,291	
CPAC 05	If I get stuck on a problem, I look for clues in my surroundings.	,254	
CPAC 12	I get solutions to problems when my mind is relaxed.	,254	,241
CPAC 6	When I get stuck on a problem, a solution just comes to me when I set it aside.	,181	
CPAC 13	While working on something, I try to fully immerse myself in the experience.		,650
CPAC 02	When I am intensely working, I don’t like to stop.		,534
CPAC 14	I can completely lose track of time if I am intensely working.		,559
CPAC 03	If am intensely working, I am fully aware of “the big picture.”		,289
CPAC 09	While working on something I enjoy, the work feels automatic and effortless.	,221	,219

Extraction method: Principal Component Analysis.
Rotation method: Oblimin with Kaiser Normalisation.
The rotation converged into 2 iterations.

We conducted a confirmatory factor analysis to confirm the shortened scale and its two main factors – act and flow – from preceding studies, for instance, Conradty and Bogner (2018). We also retained the previous renaming of the two factors originally introduced by Miller and Dumford (2016). The resulting CFI = 0.837 and TLI = 0.808 are permissible since the RMSEA < 0.050 (p = 0.492) is good and also accounts for model fit. Chi-square (χ^2 = 164.65, df = 89, p < 0.001), however, indicates that the event occurs less than one time in a thousand and that data does not adequately fit the model (Phakiti, 2018). The resulting structural equation model (Fig. 1) indicates that some of the observed variables are better suited to describing the latent variables act and flow than other observed variables, as can be seen with respective factor loadings. Consequently, the short scale may need to be further modified to guarantee its reliability as a measuring instrument in science education (Fig. 1).

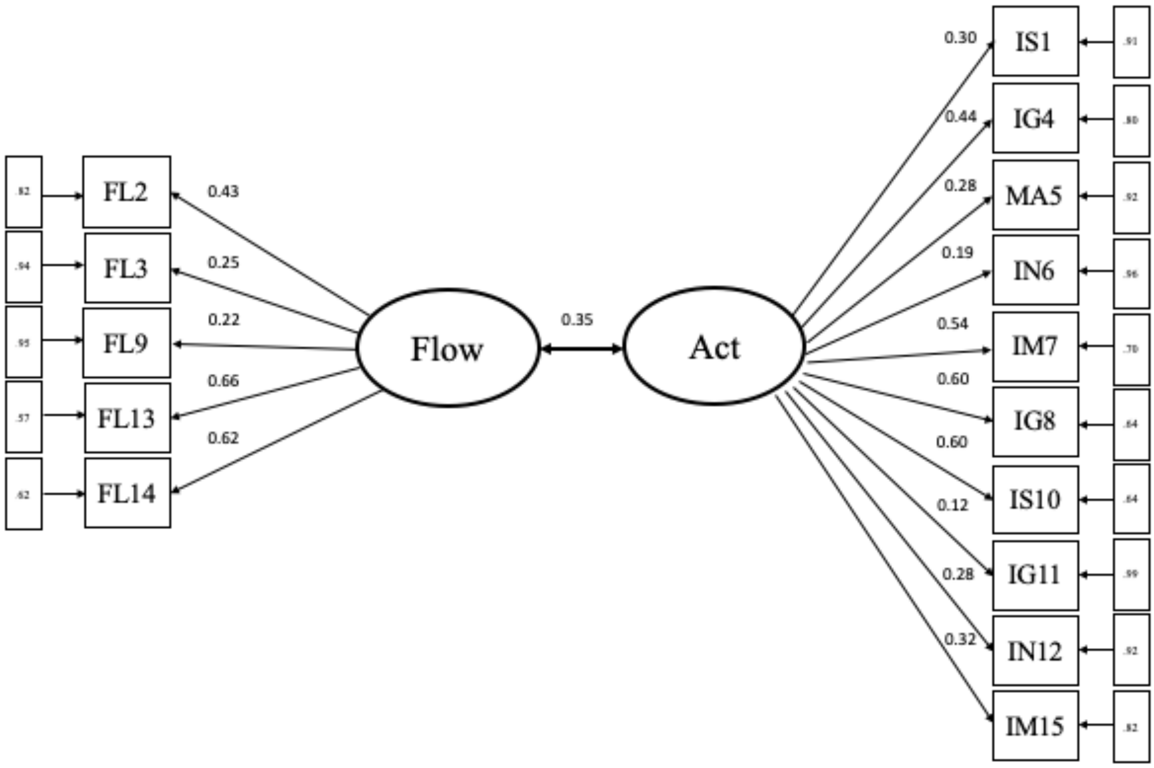


FIGURE 1 Structural equation model of CPAC after confirmatory factor analysis with two factor structure Act and Flow.

All factor loadings, apart from IG11 (p = 0.209) are significant with p ≤ 0.001. Error variances of the observed variables are indicated in square brackets coloured in light grey. Since specific variance and error variance are analytically, inversely related, increased specific variance and reliability lead to decreased error variances (Lomax, 1986).

For BFI-10, we completed a principal component analysis with oblique rotation as recom- mended in the literature (Caprara 1993). This confirmed five factors based on eigenval- ues <1.0. Due to correlations between the components, however, we could calculate neither the sum of squared charges nor the total value. Not all items as displayed in our pattern matrix showed loadings above the limit of 0.4 (Table 2) but still reached a total KMO value of 0.54, which barely indicates reliable and distinct factors (Kaiser, 1970) (Table 2). Bartlett's test sphe- ricity was $\chi^2 = 484.11$ (df = 45, $p < 0.001$). Thus, our data was just about suitable to replicate the five-factor structure of the BFI-10 and confirm the model.

TABLE 2 Principal component analysis with *BFI-10* after oblique rotation (valid N = 538).

Pattern Matrix[a]		Components *factor analysis*				
I see myself as someone who …		1 Extra 1+6	2 Consci 3+8	3 Neuro 4+9	4 Open 5+10	5 Friend 2+7
BF 6	… is outgoing, sociable	,961				
BF 1	… is reserved	,629				
BF 3	… tends to be lazy		,710			
BF 8	… does a thorough job		,685			
BF 4	… is relaxed, handles stress well			,512		
BF 9	… gets nervous easily			,785		
BF 10	… has an active imagination				,613	
BF 5	… has few artistic interests				,667	
BF 7	… tends to find fault with others					,653
BF 2	… is generally trusting					,324

Extraction method: Principal component analysis.
Rotation method: Oblimin with Kaiser Normalisation.
Total eigenvalue explained 84,56%

TABLE 3 Mann-Whitney-U-Test for gender effects after principal component analysis with *CPAC* and *BFI-10* (N = 538) for extreme-groups regarding age.

Age Percentile	*Mann-Whitney U-Test (MWU)*	Act	Flow	Extraversion	Conscientiousness	Neuroticism	Openness	Friendliness
	r	0.210	0.153	0.080	0.056	0.126	0.101	0.099
25	*Z*	-2.456	-1.790	-.924	-.657	-1.369	-.237	-1.153
	p	.014*	.073	.356	.511	.171	.237	.249
	r	0.089	0.199	0.027	0.164	0.063	0.094	0.079
75	*Z*	-.932	-2.068	-.273	-1.704	-.652	-.974	-.729
	p	.351	.039*	.758	.088	.514	.330	.466

After principal component analysis with CPAC and BFI-10, we applied the Mann-Whitney U test to assess the effects of gender on creativity and personality characteristics in outlying age groups. We discovered significant gender differences for the latent factor act in the lower age percentile. There, young boys reach higher scores for action-based creativity as compared to that of the girls. However, our results yielded significant gender differences for the latent factor flow in the upper age percentile (Table 3), suggesting that older girls achieve higher scores for flow experiences. Effect sizes ranged from low to intermediate for all factors analysed with regard to gender differences (Lipsey & Wilson, 2001) (Table 3).

As is displayed in Fig. 2, the most significant differences between female and male students emerged in outlying age groups for act and flow after calculation of gender effects with factor scores. Here, young boys achieve higher scores than young girls for act, while older girls reach higher scores than older boys for flow. Accordingly, gender only appears to influence creativity in outlying age groups and, thus, may not be significant for classroom design. It may be far more important to restructure science classrooms to encourage individual creative actions (Fig. 2).

We attempted to calculate age-related effects for BFI-10 using Kruskal-Wallis (Field, 2012). However, as the Big Five measures personality, which is regarded as stable, we could observe no immediate age-related effects (Watrin et al., 2019). Teachers should, therefore, not aim to influence the development of personality but try to provide learners with the freedom required to individually discover and enjoy their creative actions.

We also calculated correlations between CPAC and BFI-10 factors using Spearman-Rho (Field, 2012). We obtained significant positive correlations ($p < 0.001$) between conscientiousness and flow, with a correlation coefficient of 0.115 but with low effect sizes. No further significant correlations could be obtained between the scales, but flow and act had significant positive correlations ($p < 0.001$) of 0.453 within their scale with a medium to large effect size. Extraversion and neuroticism displayed a highly significant negative correlation ($p < 0.001$) with a correlation coefficient of −0.558 with a large effect size (Lipsey & Wilson, 2001). This confirms the predicted connections between personality and creativity, showing that individual preferences are involved in creative action.

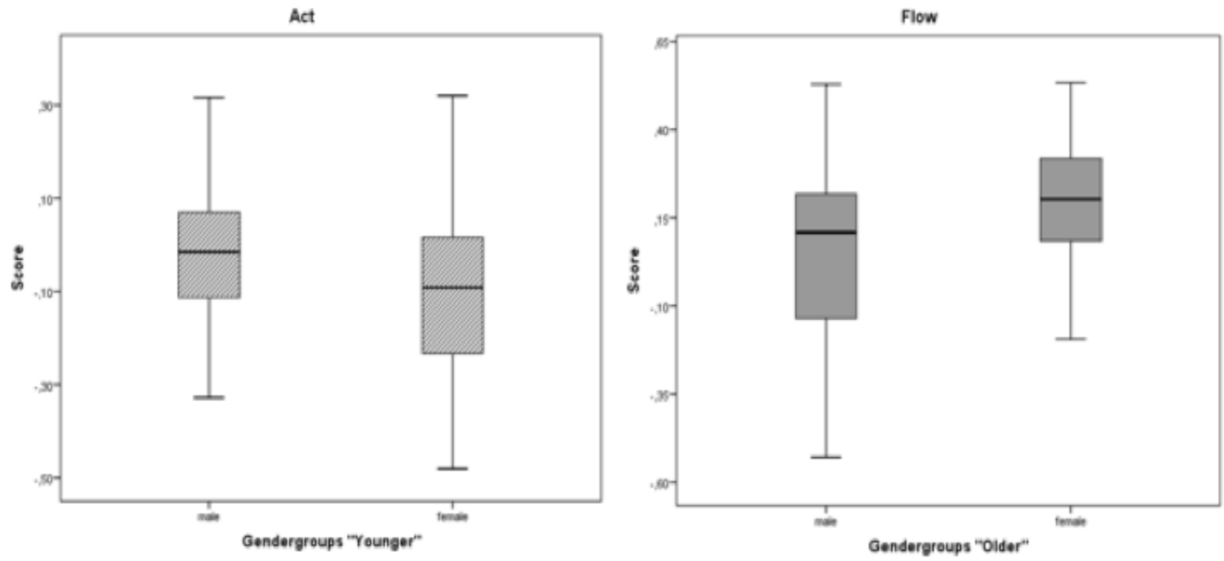


FIGURE 2 Gender differences for CPAC's creativity dimensions act and flow dependent on age groups.

Discussion

CPAC and Big Five Scale to Assess Creativity and Personality in Science Classrooms

CPAC to Assess Creativity in the Science Classroom We could confirm that the Big Five and CPAC scales in their shortened forms are both suitable for use in science classroom settings to assess personality and creativity, respectively. The shortened form of CPAC has already been used to help evaluate creativity across age groups and in different science teaching settings (Conradty & Bogner, 2018; Mierdel & Bogner, 2019). Our use of a confirmatory factor analysis to assess the shortened version of the test showed that some observable variables chosen from the original test are better suited to describing the latent construct than others. The variables FL3, MA5, IN6, FL9, and IN12 had particularly low factor loadings although they all produced significant descriptions of the latent construct. FL3 and IN6 were the weakest yet significant, observable variables. For FL9, significant cross-loadings additionally hamper a clear assign- ment to either flow or act. Without a clear reference, this observable variable could describe both constructs. IN6 is not a matter of assignment yet may be misunderstood. That is, the solution to the problem could come while the individual subconsciously continues working on it, or the problem may be solved if the individual attempts to find a solution after a certain time has passed. This ambiguity may have led to lower factor loadings in describing act. We should, thus, consider whether other observable variables from the original scale could replace those with low factor loadings or if an even shorter version might be suitable for implementation.

In view of the other observable variables, however, the shortened version of the CPAC may be able to capture act and flow as aspects of creativity in science classrooms. The scale is, thereby, able to successfully measure the learning effect of creativity skills (act) and emotional factors that contribute to enjoyment in learning (flow). Flow experiences are believed to be crucial in the evolution of creativity as they are rewarded with positive feelings (Csikszentmihalyi, 2000). To enable flow, special environments are required wherein the difficulty of tasks is balanced between demanding and easy. The perceived difficulty of tasks, however, varies individually. Furthermore, as students often completely immerse themselves in their tasks, feeling absolutely secure is another basic requirement for flow to occur (Csikszentmihalyi, 2015; Conradty & Bogner, 2018). This level of security may be accomplished in open learning environments that enable a high degree of self-regulation, irrespective of fixed lesson plans for students, and present teachers not as instructors but as mentors. In their new role as mentors, teachers may add to the students' perception of safety while encouraging students to try out individual learning and problem-solving approaches (Csikszentmihalyi, 2015). This can promote motivation (Conradty & Bogner, 2016), creativity (Conradty & Bogner, 2018), and even learning success (Thuneberg et al., 2018). In contrast, many rules and extrinsic motivators, such as exams or time pressure, impair creative actions (Csikszentmihalyi, 1975). Since some rules are, however, mandatory in educational processes, educational settings that still include preparation, incubation, eureka effects, evaluation, and elaboration as the five contributors to creative processes may also be of use (Csikszentmihalyi, 1996).

BFI-10 to Assess Personality in the Science Classroom This kind of open learning environ- ment also considers the needs of different personality types (Safwat et al., 2020), which could also influence creative actions (Kaufman et al., 2008). We measured personality using the BFI- 10 test, the short version of which is already established in most contexts, including science classrooms (Schumm, 2016; Carciofo et al., 2016). Yet, it is recommended that before using the scale, its feasibility should be assessed as, dependent on the culture, many variables can have different connotations and may, therefore, provide a distorted picture of personality distribution in the classroom (Carciofo et al., 2016). In our case, the BFI-10 has proven to be effective for measuring personality across all age groups in science classrooms after assessment with principal component analysis. In addition to substantial factor loadings for all five latent variables (extraversion, neuroticism, conscientiousness, friendliness, openness), we have also been able to indicate the strong negative correlation between the latent variables neuroticism and extraversion, as has been suggested by the arousal model and the Yerkes- Dodson law (Aguilar-Alonso 1996).

Correlations between CPAC and BFI-10

Our results also show connections between the CPAC's latent variable flow and the personality trait conscientiousness. A study assessing the impact of flow experience on job performance received comparable results for conscientiousness (Chui & Lee, 2012), confirming general flow theory. When students are goal-oriented and hardworking, their flow experience will strongly impact their performance (Demerouti, 2006) as well as their commitment and motivation (Cacioppe, 2017). Essential for flow experience, however, is an open and flexible environment that offers opportunities for individual development. Therefore, flow has mainly been assessed in the context of activities such as dancing, reading, online communication, and art (Lian et al., 2018). Practising these, however, also requires commitment as is inherent to conscientiousness (Chui & Lee, 2012). In science education, flow can be best experienced within the scope of hands-on experiments, which optimally involve students while giving them full control of their actions. This freedom, however, also requires a high degree of conscientiousness (Hoseinifar et al., 2011). Extensive note-taking in instruction-heavy classroom settings is, in contrast, deemed detrimental to flow experience. The perceived time limit of science lessons also prevents students from fully immersing themselves in their tasks (Boyer & Lamoreaux, 1997).

Many of the skills relevant to dancing and art – such as posing high-level problems and questions, making decisions, or combining ideas in new ways – also play a major role in science education and may contribute to fostering creative actions (Yager, 2005). Maor and Jost (2017), for instance, effectively mixed art and mathematics by combining forms, patterns, and numbers in endless variations. This stimulates active reflection and the testing of solutions (Sawyer, 2012). Their educational goal to stimulate acts is supposed to encourage students to creatively interact with the difficult subject of mathematics by simply exploring, observing, explaining, and proving its different facets (Maor & Jost, 2017). A collaborative, stimulating, and interactive environment is thereby crucial to experiencing flow (Miller & Dumford, 2016; Csikszentmihalyi, 2015).

Gender Differences Regarding CPAC and BFI-10 at either End of the Age Spectrum

In our study, the most eminent gender differences emerged for act and flow with regard to outlying age groups. There, young boys of the lowest age percentile score higher for act than girls in the same age group. In terms of best practice for the science classroom, action-based exercises for the promotion of creativity in younger male children can be developed based on our findings. That is, group work encouraging the lively social and creative exchange of ideas and the freedom to actively reflect and try out solutions are both indispensable (Sawyer, 2012). Explorative and hands-on teaching with many experiments and opportunities to discover and think creatively are recommended.

As adolescence proceeds, act-related creativity differences between the genders balance out. Instead, differences in flow experience come into play. Young women in the upper age percentile seem to have a significantly better flow experience than their male counterparts. A key reason for this is might be socialization processes and cultural influences during adolescence (Iiamura & Taku 2018). That is, from an early age, societal influences may lead males and females to develop different ways of coping and experiencing the world (Goodwin & Gotlib, 2004). Flow is, in fact, a mixture of different mental states and organizational requirements (Egbert, 2003). Since creativity is, however, described as linked to intrinsic motivation (Csikszentmihalyi, 2000; Baer, 2010; Runco, 2014), genuine interest in the respective subject area may also influence scientific creative processes and, consequently, act and flow. While the TIMSS 2019 report shows that the gender gap is steadily decreasing (Mullis et al., 2020; Mejía-Rodríguez et al., 2020), the different approach to combining science with art of our participating schools may have influenced results (Baer, 2010; Charyton et al., 2011).

Limitations

One limitation of our study is the self-reporting method of data collection which did not provide us with an “outside” or “independent” perspective on participants’ views. Self- reported data are vulnerable to inaccurate reporting as participants may represent themselves differently for a variety of conscious and unconscious reasons. Secondly, our design primarily relies on quantitative data, and is not the type of a multi-trait, multi-method design often regarded most suitable (Leutner et al., 2017). Finally, all data were collected in a single session, and longitudinal tracking of participants was not attempted.

Conclusion

The fact that creative thinking is based on normative cognitive processes emphasizes the need for close collaboration between cognition experts and educational instructors (Kröger 2015). In the context of science learning, productivity and freedom are important triggers of creativity. Gender differences in creative thinking appear more pronounced in certain age groups, and it is, thus, important that appropriate support is available to both genders. The inclusion of different personality traits must also be considered (Csíkszentmihályi et al. 2005). Creative cognitive processes are relevant in multiple educational disciplines – including science teaching – and not just in fields traditionally associated with creativity. Active reflection and the testing of solutions are, thereby, indispensable for science teaching that fosters and supports creativity (Sawyer, 2012). Using direct instruction and hands-on approaches (Boyer & Lamoreaux, 1997) to foster creativity could also be effective in developing scientific skills demanded by future employers (Brynjolfsson & McAfee, 2014). Thus, schools could benefit from integrating creative cognitive processes into curricular revisions for the improvement of science classes (Wyse & Ferrari, 2015). Providing open tasks that leave room for reflection on less restrictive time schedules is one feasible approach to encouraging creative cognitive processes (Yager, 2005). However, much more research is needed to identify best practices for fostering creativity in science education and to assess the extent to which creative cognitive processes are related to other positive outcomes of the school experience (Henriksen et al., 2018). Although the current study is exploratory, it reveals creativity to be an important player in the changing landscape of science education. After all, in a digitalizing world, creativity has the potential to become an indispensable soft skill to put an individual's scientific know-how into practice (Bruno & Canina, 2019).

Acknowledgements

Special thanks to the teachers and students involved in this study for their cooperation. Also, many thanks to the Horizon 2020 Framework Program and the German Research Foundation.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Ethics Approval

Hereby, we consciously assure that for the manuscript the following is fulfilled:

1) This material is the authors' own original work, which has not been previously published elsewhere.
2) The paper is not currently being considered for publication elsewhere.
3) The paper reflects the authors' own research and analysis in a truthful and complete manner.
4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
5) The results are appropriately placed in the context of prior and existing research.
6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
7) All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

Moreover, "all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Conflict of Interest

The authors declare no competing interests.

Open Access

References

Aguilar-Alonso, A. (1996). Personality and creativity. Personality and Individual Differences, 21(6), 959–969.

Aldous, C. R. (2007). Insights from history, cognitive psychology and neuroscience. International Education Journal, 8(2), 176–186.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. Journal of Personality and Social Psychology, 45, 357–376.

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. Research in Organizational Behavior, 36, 157–183.

Antinori, A., Carter, O. L., & Smillie, L. D. (2017). Seeing it both ways: Openness to experience and binocular rivalry suppression. Journal of Research in Personality, 68, 15–22.

Baer, J. (2010). Gender differences in the effects of anticipated evaluation on creativity. Creativity Research Journal, 10(1), 25–31.

Balgiu, B. A. (2018). The psychometric properties of the big five inventory-10 (BFI-10) including correlations with subjective and psychological well-being. Global Journal of Psychology Research, 8(2), 61–69.

Barron, F., & Harrington, D. (1981). Creativity, intelligence, and personality. Annual Review of Psychology, 32, 439–476.

Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. Learning and Individual Differences, 20(5), 532–535.

Beghetto, R. A. (2015). Teaching creative thinking in K12 schools. The Routledge International Handbook of Research on Teaching Thinking. New York: Routledge.

Bortz, J., & Schuster, C. (2010). Faktorenanalyse [Factor Analysis]. In: Bortz, J., & Schuster, C. (eds.). Statistik für Human- und Sozialwissenschaftler [Statistics for human and social scientists] (pp. 385–426). Berlin, Heidelberg, New York: Springer Verlag.

Boyer, S.E., & Lamoreaux, C.D. (1997). Flow theory as a construct for analyzing learning environments in a 7th- grade science classroom. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC document reproduction service no. ED409215).

Bro, R., & Smilde, A. K. (2014). Principal component analysis. Analytical Methods, 6, 2812–2831.

Bruno, C., & Canina, M. (2019). Creativity 4.0. Empowering creative process for digitally enhanced people. The Design Journal, 22(1), 2119–2131.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York, NY: WW Norton & Company.

Cacioppe, R. (2017). Integral mindflow: A process of mindfulness-in-flow to enhance individual and organization learning. The Learning Organization, 24(6), 408–417.

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The "big five questionnaire": A new questionnaire to assess the five-factor model. Personality and Individual Differences, 15(3), 281–288.

Carciofo, R., Yang, J., Song, N., Du, F., & Zhang, K. (2016). Psychometric evaluation of Chinese-language 44- item and 10-item big five personality inventories, including correlations with chronotype, mindfulness and mind wandering. PLoS One, 11(2), e0149963. https://doi.org/10.1371/journal.pone.0149963.

Charyton, C., & Snelbecker, G. E. (2007). General, artistic and scientific creativity attributes of engineering and music students. Creativity Research Journal, 19(2), 213–225.

Charyton, C., Basham, K. M., & Elliott, J. O. (2011). Examining gender with general creativity and preferences for creative persons in college students within the sciences and the arts. Journal of Creative Behavior, 42(3), 216–222.

Cho, H., Pemberton, C.L., & Ray, B. (2017). An exploration of the existence, value and importance of creativity education. Current Issue in Education, 20(1), Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/ view/1537.

Chua, R. Y. (2015). Innovating at cultural crossroads: How multicultural social networks promote idea flow and creativity. Journal of Management, 44(3), 1119–1146.

Conradty, C., & Bogner, F.X. (2016). Hypertext or Textbook: Effects on Motivation and Gain in Knowledge. Education Sciences, 6(3), 29:1–16.

Conradty, C., & Bogner, F.X. (2018). From STEM to STEAM: How to Monitor Creativity. Creativity Research Journal, 30(3), 233–240.

Conradty, C., & Bogner, F.X. (2019). From STEM to STEAM: Cracking the Code? How Creativity & Motivation Interacts with Inquiry-based Learning. Creativity Research Journal, 31(3), 284–295.

Corazza, G. E. (2016). Potential originality and effectiveness: The dynamic definition of creativity. Creativity Research Journal, 28(3), 258–267.

Cseh, G. M. (2016). Flow in creativity: A review of potential theoretical conflict. In Flow experience (pp. 79–94). Cham: Springer.

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety. San Francisco: Jossey-Bass.

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In R. Sternberg & J. E. Davidson (Eds.), Conceptions of giftedness (pp. 325–339). New York, NY: Cambridge University Press.

Csikszentmihalyi, M. (1996). Creativity: Flow and psychology of discovery and invention. New York: Harper Collins.

Csikszentmihalyi, M. (2000). Happiness, flow, and economic equality. American Psychologist, 55(10), 1163– 1164.

Csikszentmihalyi, M. (2015). Flow und Kreativität: Wie Sie Ihre Grenzen überwinden und das Unmögliche schaffen (flow and creativity: How to overcome your limits and achieve the impossible) (2nd ed.). Stuttgart: Klett-Cotta.

Demerouti, E. (2006). Job characteristics, flow, and performance: The moderating role of conscientiousness. Journal of Occupational Health Psychology, 11, 266–280.

Egbert, J. (2003). A study of flow theory in the foreign language classroom. Modern Language Journal, 87(4), 499–518.

Eysenck, H. J. (1981). A model for personality. New York: Springer.

Field, A. (2012). Discovering statistics using IBM SPSS statistics (4th ed.). London: SAGE.

Furnham, A., Hughes, D. J., & Marshall, E. (2013). Creativity, OCD, narcissism and the big five. Thinking Skills and Creativity, 10, 91–98.

Goldman, A. D., & Penner, A. M. (2016). Exploring international gender differences in mathematics self-concept. International Journal of Adolescence and Youth, 21(4), 403–418.

Goodwin, R. D., & Gotlib, I. H. (2004). Gender differences in depression: The role of personality factors. Psychiatry Research, 126(2), 135–142.

Gralewski, J., & Karwowski, M. (2019). Are teachers' ratings of students' creativity related to students' divergent thinking? A meta-analysis. Thinking skills and creativity, 33. https://doi.org/10.1016/j.tsc.2019.100583.

Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. Creative Education, 3(5), 603–611.

Henriksen, D., Mishra, P., & Fisser, P. (2018). Infusing creativity and technology in 21st century education: A systemic view for change. Educational Technology & Society, 19(3), 27–37.

Hocevar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In: Handbook of creativity (pp. 53–75). Springer US.

Hodson, D. (1998). Is this really what scientists do? Seeking a more authentic science in and beyond science education. In Wellington, J. (ed.) Practical work in school science. Which way now? (p. 93–108). London: Routledge.

Horng, J. S., Hong, J. C., ChanLin, L. J., Chang, S. H., & Chu, H. C. (2005). Creative teachers and creative teaching strategies. International Journal of Consumer Studies, 29(4), 352–358.

Hoseinifar, J., Siedkalan, M. M., Zirak, S. R., Nowrozi, M., Shaker, A., Meamar, E., & Ghaderi, E. (2011). An investigation of the relation between creativity and five factors of personality in students. Procedia-Social and Behavioural Sciences, 30, 2037–2041.

Kaiser, H. F. (1970). A second generation little jiffy. Psychometrika, 35(4), 401–415.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. Review of General Psychology, 13(1), 1–12.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). Essentials of creativity assessment (Vol. 53). John Wiley & Sons.

Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. Studies in Science Education, 43(1), 1–37.

Kitamura, S., Yasuno, F., Yamamoto, A., Kazui, H., Kudo, T., Matsuoka, K., Kiuchi, K., Kosaka, J., Nagatsuka, K., Iida, H., & Kishimoto, T. (2015). A structural model of age, grey matter volumes, education, and personality traits. Psychogeriatrics, 16(1), 46–53.

KMK (2020). Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife [educational standards in biology for the general university entrance qualification]. URL: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf.

Kröger, S. (2015). Neural correlates of creative thinking: Conceptual expansion processing: Dissertation to obtain the academic degree Doctor rerum naturalium. Gießen, Germany.

Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2018). Children's creativity: A theoretical framework and systematic review. Review of Educational Research, 89(1), 93–124.

Lepper, M. R., & Greene, D. (Eds.). (1978). The hidden costs of reward. Hillsdale: Erlbaum.

Leutner, F., Yearsley, A., Codreanu, S.-C., Borenstein, Y., & Ahmetoglu, G. (2017). From Likert scales to images: Validating a novel creativity measure with image-based response scales. Personality and Individual Differences, 106, 36–40.

Lewis, T. (2008). Creativity in technology education: Providing children with glimpses of their creative potential. International Journal of Technology and Design Education, 19(3), 255–268.

Lian, B., Kristiawan, M., & Fitriya, R. (2018). Giving creativity room to students through the friendly school's program. International Journal of Scientific &Technology Research,7(7), 1–7.

Lienert, G. A., & Raatz, U. (1998). Testaufbau und Testanalyse (test setup and test analysis). Grundlagen Psychologie. Weinheim, Germany: Beltz.

Iimura, S, & Taku, K. (2018). Gender Differences in Relationship Between Resilience and Big Five Personality Traits in Japanese Adolescents. Psychological Reports, 121(5), 920–931.

Lipsey, M. W., & Wilson, D. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage Publications.

Lomax, R. G. (1986). The effect of measurement error in structural equation modeling. The Journal of Experimental Education, 54(3), 157–162.

Lucas, B. (2016). A five-dimensional model of creativity and its assessment in schools. Applied Measurement in Education, 29(2), 278–290.

Maor, E., & Jost, E. (2017). Beautiful Geometry. Princeton: Princeton University Press.

Mejía-Rodríguez, A. M., Luyten, H., & Meelissen, M. R. M. (2020). Gender differences in mathematics self- concept across the world: An exploration of student and parent data of TIMSS 2015. International Journal of Science and Mathematics Education. https://doi.org/10.1007/s10763-020-10100-x.

Mierdel, J., & Bogner, F.X. (2019). Is creativity, hands-on modeling and cognitive learning gender-dependent?. Thinking Skills and Creativity, 31, 91–102.

Miller, A. L. (2014). A self-report measure of cognitive processes associated with creativity. Creativity Research Journal, 26(2), 203–218.

Miller, A. L., & Dumford, A. D. (2016). Creative cognitive processes in higher education. The Journal of Creative Behavior, 50(4), 282–293.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 international results in mathematics and science. In Retrieved from Boston College. TIMSS & PIRLS International Study: Center website https://timssandpirls.bc.edu/timss2019/international-results/.

Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995-2003). Studies in Educational Evaluation, 34, 56–72.

Okere, M. I. O., & Ndeke, G. C. W. (2012). Influence of gender and knowledge on secondar school students' scientific creativity skills Nakuru district, Kenya. European Journal of Educational Research, 1(4), 353– 366.

Perry-Smith, J. E. (2014). Social network ties beyond redundancy: An experimental investigation of the effect of knowledge content and tie strength on creativity. Journal of Applied Psychology, 99, 831–846.

Phakiti, A. (2018). Confirmatory factor analysis and structural equation modeling. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), The Palgrave handbook of applied linguistics research methodology. London: Palgrave Macmillan.

Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. Thinking Skills and Creativity, 7, 258–264.

Piniel, K., & Albert, Á. (2020). Motivation and flow. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), The Palgrave handbook of motivation for language learning (pp. 579–597). Cham: Palgrave Macmillan.

Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), Handbook of creativity (pp. 35–61). New York: Cambridge University Press.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. Journal of Research in Personality, 41, 203–212.

Richards, J., & Cotterall, S. (2016). Exploring creativity in language teaching. In R. H. Jones & J. C. Richards (Eds.), Creativity in language teaching (pp. 97–113). New York, NY: Routledge.

Robinson, K. (2011). Out of our minds: Learning to be creative. Oxford, UK: Capstone.

Rost, J. (2004). Lehrbuch Testtheorie–Testkonstruktion [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL: http://www. rstudio.com/ (accessed: 01.03.2021).

Runco, M. A. (2014). "Big C, little c" creativity as a false dichotomy: Reality is not categorical. Creativity Research Journal, 26(1), 131–132.

Runco, M. A. (2019). Creativity as a dynamic, personal, parsimonious process. In Dynamic perspectives on creativity (pp. 181–188). Cham, Switzerland: Springer. Advance online publication.

Runco, M.A., & Jaeger, G.J. (2012). The Standard Definition of Creativity. Creativity Research Journal, 24(1), 92–96.

Safwat, S., El Bolock, A., Alaa, M., Faltaous, S., Schneegass, S., & Abdennadher, S. (2020). The effect of student-lecturer cultural differences on engagement in learning environments - a pilot study. In F. De La Prieta et al. (Eds.), Highlights in practical applications of agents, multi-agent systems, and trustworthiness. Cham: Springer.

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. Creativity, 4(2), 238–275.

Sawyer, R. K. (2012). Explaining creativity: The science of human innovation. New York, NY: Oxford University Press.

Schmidt, A. L. (2011). Creativity in science: Tensions between perceptions and practice. Creative Education, 2, 435–445.

Schumm, M.F., & Bogner, F.X. (2016). Measuring adolescent science motivation. International Journal of Science Education, 38(3), 434–449.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achieve- ment: A meta-analysis. Journal of Educational Psychology, 104, 743–762.

Tackett, J. L., & Lahey, B. B. (2017). Neuroticism. In T. A. Widiger (Ed.), Oxford library of psychology. The Oxford handbook of the five factor model (p. 39–56). Oxford University Press.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington. DC: American Psychological Association.

Thuneberg, H.M., Salmi, H.S., & Bogner, F.X. (2018). How creativity, autonomy and visual reasoning contribute to cognitive learning in a STEAM hands-on inquiry-based math module. Thinking Skills and Creativity, 29, 153–160.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), The nature of creativity (pp. 43–75). New York, NY: Cambridge University Press.

Tsai, K.-C. (2018). Investigating the empirical links between creative and critical thinking. Psychology, Society, & Education, 11(3), 267–280.

Watrin, L., Geiger, M., Spengler, M., & Wilhelm, O. (2019). Forced-choice versus Likert responses on an occupational big five questionnaire. Journal of Individual Differences, 40(3), 134–148.

Wiseman, M., & Bogner, F.X (2003). A higher-order model of ecological values and its relationship to personality. Personality and Individual Differences, 34(5), 783–794.

Wyse, D., & Ferrari, A. (2015). Creativity and education: Comparing the national curricula of the states of the European Union and the United Kingdom. British Educational Research Journal, 41(1), 30–47.

Yager, R. E. (2005). A vision for what science education should be like for the twenty-five years of a new millennium. School Science and Mathematics, 100, 327–341.

5.5 Teilarbeit C

International Journal of Science Education, Revision Round 1

The Trade-Off Between STEM Knowledge Acquisition and Language Learning in Short-Scale CLIL Implementations

Tamara Roth[1*], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany
*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

Bilingual education could address the challenges introduced by an increasing internationalisation of the educational sector. Content and Language Integrated Learning (CLIL), in particular, may equip students with the necessary cultural and communicative skills to succeed in this changing academic environment. It is, however, not yet clear, how CLIL can effectively act in short-term educational contexts where full-term bilingual programs are not feasible. A one-day CLIL module in our outreach gene-technology lab is one such example. The assessment of our intervention with 252 high school students indicates that a CLIL module does not achieve the same learning success its German equivalent. Possible reasons may lie in the participants' limited prior experience with CLIL teaching. Since many students may not yet be fully fluent in the English language, they may face difficulties obtaining the same level of factual knowledge compared to the German treatment group. Even with additional language scaffolding material, full access to online dictionaries and crucial passages in the workbooks provided their native language, parallelising language learning and content learning has negatively influenced STEM knowledge acquisition. Thus, we argue that more attention should be paid to the inherent trade-off between language and content learning when carrying out short-term CLIL programs. In particular, we caution against using only content and language scaffolding to mediate this trade-off.

Keywords: Cognitive knowledge. Scientific bilingual learning. CLIL. Gene technology outreach learning. Science education.

Introduction

Excellent English language skills have become a basic requirement for most professions to effectively communicate internationally with people from various backgrounds (Sardegna et al. 2017). Despite undeniable relevance of this lingua franca (Oktaviani & Fauzan 2017) for effective participation in social and professional life, a high proportion of students lack English language competencies at the desired level (Rodenhauser & Preisfeld 2014). In response to these English language deficiencies, the European Commission (2004) has strongly recommended the introduction of Content and Language Integrated Learning (CLIL) to help students meet the requirements of an increasingly international society (Finkbeiner & Fehling 2006).

CLIL is understood to be an innovative approach to foreign language learning. While it retains the focus on content learning, the content is taught and discussed in both the native and foreign language (Eurydice 2005, Coyle et al. 2010). Unlike the Canadian model of immersion (Wode 1995), the foreign language is not just a medium to transfer knowledge but constitutes important didactic content. At German schools, incentives of the European Commission (2012) to adopt CLIL in regular classroom teaching, have resulted in an expansion of traditional CLIL subjects. Although CLIL was initially piloted primarily in social science, it has now been increasingly applied in science education, particularly in Biology (Bonnet 2012). As English is 'the undisputed language of science' (Meyerhöffer & Dreesmann 2019, p.1366), it is important for students in STEM classes to not just understand the scientific content in their native language but to also achieve scientific literacy in English (Lemke 1990, Klieme et al. 2010).

Studies that explored the influence of CLIL on science education have – so far – primarily focused on language attainment, i.e. higher communicative competencies, which only delivered few results for potential effects on science knowledge acquisition (e.g., Piesche et al. 2016, De Dios Martínez Agudo 2020). In particular, assessments of combinations of CLIL with practical experimentation in outreach (Biology) laboratories are scarce (e.g., Rodenhauser & Preisfeld 2014, Buse et al. 2018). Such informal approaches place particularly high demands on participating students, as both previously unknown scientific contents and specific scientific vocabulary are conveyed in both their native and a foreign language (e.g., Koch & Bünder 2006, Meyerhöffer & Dreesmann 2019).

CLIL is supposed to advance competencies required to be successful in an international academic environment (Grandinetti et al. 2013), we piloted a CLIL gene-technology outreach module, based on a previous non-CLIL outreach learning module. We retained the overall structure of the module and used the previous monolingual treatment group as comparison group

(Roth et al. 2020). Following cognitive load theory, we expect a general trade-off between science knowledge acquisition and language learning since both demand a high degree of mental capacities. Moreover, we conclude that scaffolding alone may not mediate this trade-off (e.g., Coyle 2007, Grandinetti et al. 2013). That is, short-scale CLIL modules should balance content and language learning, regarding both as equal contributors to mental effort, and consequently reduce content. In the following, we outline the relevant theoretical background including the basic principles of content and foreign language learning and how they can be successfully combined in CLIL education as well as practical experimentation, which has already proven its beneficial influence on content learning. All are vital to understand the design and procedure of our study.

Theoretical Background

Content Learning

Learning is often described as an adaptation of previous behaviours in response to experiences. The creation of meaning through experience reflects constructivist views of learning, in which knowledge acquisition is situational (Korthagen & Lagerwerf 1995). This view of learning presupposes a variety of distinct methods in science education to ensure curriculum-specific knowledge acquisition.

Learning, however, is more than just a generalisable outcome of experiences and one method does not equally take effect in all students. That is, experiences are highly individual and may be influenced by factors such as prior knowledge, or different approaches to learning (Lee et al. 2015). One content-specific method, thus, can impossibly lead to the exact same learning outcomes in heterogeneous student groups (Hodson 2014). Thus, inquiry learning has become an increasingly popular approach for effective science learning. It puts the students in a position where they can actively engage in learning method selection and learning environment design. Inquiry learning also adapts the regular scientific inquiry process at a student level, where students have to reason about information that goes beyond what they have been provided with to create knowledge (Vygotsky 1971).

In literature, inquiry learning is commonly perceived as a ‘bottom up’ approach that gives learners agency in the creation of knowledge through observation and experimentation, while the teacher acts as a guide (Rocard et al. 2007, Donaldson & Allen-Handy 2020). Of course, this does not necessarily mean that learners can freely decide on learning contents, since science curricula predefine a clear set of learning goals. Students can, however, choose suitable methods for individual learning. They are given the agency to shape their own classroom experience and create meaning from it in close collaboration with their classmates. (Renninger et al. 2018). The National Research Council (NRC) provided some essential features of how to design inquiry learning classrooms for science students, which include issues such as '(1) Learners are engaged by scientifically oriented questions. (2) Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions. (3) Learners formulate explanations from evidence to address scientifically oriented questions. (4) Learners evaluate their explanations considering alternative explanations, particularly those reflecting scientific understanding. [and] (5) Learners communicate and justify their proposed explanations.' (Minner et al. 2010).

English as a Foreign Language in Science Education

European language policies put up pressure to foster widespread inclusion of foreign languages in regular school curricula (European Commission 2012). In the spirit of cultural exchange and language diversity, students should speak at least two more foreign languages to receive 'Euro-qualification' (Finkbeiner & Fehling 2006, S.14). Since it is infeasible to include extra lessons for foreign language learning in the already tightly knit curriculum, CLIL was introduced as a means to meet the EU's language policy goals while avoiding cluttered lesson plans. It is also meant to effectively combine language learning and content learning, adding further motivational factors as compared to regular non-CLIL teaching (European Commission, 2004).

However, in current CLIL implementations in countries where English is a foreign language, lesson planning primarily focuses on content learning whereas the scientific language plays a subordinate role, in both the native and foreign language, (Rodenhauser & Preisfeld 2014). Only few acknowledge the intricate connection between native language, foreign language, and content (Sutton 1996, Yore & Treagust 2006). It is accordingly not feasible to only use the foreign language and to neglect the native language, since the creation of meaning and ultimately knowledge involves both (Poza 2016). This complex and mutually dependent relationship between language and content (Yore & Treagust 2006) was also addressed in Cummins' (1979) developmental interdependence hypothesis: It indicates that the first language has to exceed a certain threshold of proficiency to (1) enable higher-order thinking (2) to allow for adequate second language acquisition and (3) to facilitate content learning in the foreign language.

Moreover, Cummins (1979) distinguishes between everyday language and scientific language competencies. He labels them *Basic Interpersonal Communicative Skills* (BICS) and *Cognitive Academic Language Proficiency* (CALP) respectively. The latter are essential for understanding and effectively communicating scientific content, which is why they are of particular importance for CLIL. Especially in CLIL science classrooms, the use of CALP increases with the complexity of scientific learning while BICS becomes less important (Rodenhauser & Preisfeld 2014). CALP is also understood as easier to foster in older learners as compared to fluency and phonology in BICS according to the developmental interdependence hypothesis. To test CALP, cloze tests are recommended as an especially effective tool (Cummins 1979).

Yet, fluency of articulation, which is naturally ascribed to BICS (Cummins 1979) should also be fostered alongside CALP in CLIL science classrooms. Therefore, Buxton recommends to 'engage [students] in authentic communication through the use of hands-on tasks […] related to everyday experiences' (Buxton et al. 2008, p. 501). This is also what Gonzalez-Howard & McNeill (2016) advise, since the processing of information in a coherent and accessible manner stimulates CALP and BICS simultaneously (Nation & Newton 2008). Oga-Baldwin (2019) described this meaningful practice of the foreign language as engagement. Active engagement and argumentation in science discourse are regarded as main drivers of success in mastering scientific literacy in both the native and foreign language (Walker & Sampson 2013, Gonzalez-Howard & McNeill 2016, Oga-Baldwin 2019). Lam et al. (2012) substantiated this observation with their contextual model of engagement wherein 'engagement is the central mediator between the external world that students experience, their internal processes, and their degree of achievement' (Oga-Baldwin 2019, p.3).

In Germany, engagement with the foreign language in the form of CLIL learning is realised in either bilingual strands or singular CLIL modules. Singular CLIL modules do not have a set timeframe and often encompass only occasional implementations (Krechel 2013). Any meaningful engagement, however, appears to have a positive influence on second language learning (Bonnet 2004, Rodenhauser & Preisfeld 2014, Buse & Preisfeld 2018) whereas the influence of CLIL on scientific content learning is ambiguous (Haagen-Schützenhöfer et al. 2011, Piesche et al. 2016, Fernández-Sanjurjo et al. 2017).

One way to bridge the gap between the native language and the foreign language to create scientific literacy in both languages is scaffolding. Scaffolding is 'a type of teacher assistance that helps students learn new skills, concepts, or levels of comprehension of material that leads to the student successfully completing a specific learning activity with finite goals' (Maybin et al. 1992, p.188). Gottlieb (2016) identified four possible scaffolding categories:

linguistic, graphic, sensory, and interactive. In science subjects such as Biology, the graphic or visual scaffolding comes naturally. Visualisation already is an important technique to generate scientific knowledge and the uptake of therewith-associated scientific vocabulary in regular classroom settings. In CLIL, visualisation becomes even more essential, since it fosters understanding of the content and broadens scientific vocabulary as well as literacy (Fernández-Fontecha 2020).

Practical Experimentation

Hands-on experiments in classroom or outreach settings have already proven their positive effect on content learning (e.g., Solomon 1980, Kelly 2007, Mierdel and Bogner 2019). They provide the necessary basis for students to acquire knowledge and learn how to manage scientific equipment. Such first-hand experience is vital to understanding the nature of science, where experimental outcomes not always meet expectations (Loveys and Riggs 2019). Solomon (1980) described hands-on experimenting as key to increasing scientific knowledge. Scientific learning in the laboratory is, according to him, a natural and integral part of science teaching. Thus, different levels of scientific knowledge beyond content knowledge may be achieved through the combination of hands-on experimentation and scientific critical thinking. Such hands-on-minds-on approaches also foster skills as described in the national standards (KMK 2005), for instance the understanding of scientific processes. Crucial to this approach is an open and flexible learning environment, where critical thinking and self-evaluation skills may thrive. Hands-on experimentation, as an essential element of science discourse, may also contribute to practicing and assessing scientific knowledge obtained in the classroom (Kelly 2007). In this regard it offers an interesting approach for CLIL implementations. Especially with regard to trialling new scientific vocabulary in both the native and foreign language as well as to practicing scientific literacy, hands-on experimentation ensures equitable educational opportunities for all students (Nikula 2014, Evnitskaya & Morton 2011). That is, students who engage in experimentation while still learning a language that they effectively deploy to communicate their findings with classmates, feel less pressured and more confident (Honeycutt-Swanson et al. 2014).

In CLIL learning, this is also an effective means of scaffolding language learning, though interactive communication as such is also important when it comes to solely reassessing and self-evaluating results (Gottlieb 2016). A feasible method to encourage both foreign language acquisition and self-evaluation is reflective writing (Kovanović et al. 2018, Wilmes & Siry 2019) about particular experimental phases. With a focus on scientific content learning, this approach can encourage students to rethink and reassess certain steps in the experimental process (Mierdel and Bogner 2019; for details, see below). In this 'self-directive process', students 'transform their mental abilities into academic skills' (Zimmermann 2002, p. 65). Thereby, students autonomously verify or falsify hypotheses and modify preconceptions accordingly or develop new ideas (Schwarz et al. 2009). These competencies also reflect Kolb's (1984) categorisation of learners' abilities, which encompass abilities of concrete experience, reflective observation, abstract conceptualisation, active experimentation.

Objectives of the Study

The present study focuses on the following research questions:

1. How does a one-day CLIL science module influence the overall scientific academic achievement throughout the hands-on laboratory?
2. How does CLIL influence overall learning outcomes as compared to the monolingual German treatment group?

Thus, the specific objectives were three-fold:

- to assess students' overall scientific academic achievement throughout the laboratory
- to determine potential differences between monolingual German und CLIL treatment groups by comparing their overall scientific academic achievement
- to examine potential correlations between CLIL learner's performance and their Biology as well as English grades

Materials and Methods

Intervention Phases

Our study builds on the design of our previous intervention (Roth et al. 2020). That is, we retained the different experimental and modelling phases from our monolingual German module regarding structure and content. We only changed the language of instruction to English as well as all the materials provided to the students and added a separate vocabulary scaffolding exercise book. This should ensure overall treatment comparability (Table 1), while leveraging the burden of content and language learning in parallel. The following section is adapted from Roth et al. (2020) and describes the different phases of our one-day, hands-on bilingual module based on inquiry-based learning activities. The module was designed for ninth graders of Bavarian high schools and focused on the structure of DNA in combination with CLIL.

Throughout the module, classes remained intact and students worked in pairs. For each phase of the module, we provided vocabulary riddles in a language exercise book (vocabulary scaffolding) and provided key vocabulary in German and English in a laboratory manual (code-switching). The manual additionally contained visualizations of materials and graphic illustrations of experimental procedures in both its English and German version. Ms Roth, who has a background in English and Biology, functioned as guide throughout the module and provided demonstrations of key experimental steps ahead of experimentation. For theoretical input, she used interactive smartboard activities and a poster with clozes (Table 1).

TABLE 1 Quasi-experimental intervention design.

Intervention phases	Evaluation variants	
	Monolingual (non-CLIL) (*n* = 145)	Bilingual (CLIL) (*n* = 107)
Instruction and conversation language English	-	+
Vocabulary scaffolding material for all intervention phases	-	+
Pre-lab	+	+
DNA-related theoretical and experimental phases		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases		
Mentally modelling: text analysis	+	+
DNA modelling with craft materials	+	+
Model evaluation-1: Drawing a paper & pencil version of the crafted model and answering questions	+	+
Model evaluation-2: Comparing the crafted model with a scientific demonstration model	+	+
Interpretation	+	+

The content of the module matches requirements of the state's syllabus and follows national competency requirements (KMK 2005). The monolingual German module was conducted in 2018 and adapted to a CLIL intervention design for the current module in 2020 (Table 1). As classes from the 2018 module did not take part in the 2020 intervention and classes from 2020 were randomly assigned to intervention dates, German and CLIL treatment groups were treated as independent variable (Cook & Campell 1979) (Table 1).

Pre-lab phase. To familiarize students with the laboratory environment, since many students had only little prior experience in laboratory learning, we conducted an extensive pre-lab phase for both treatment groups. In this phase, the guide used interactive smartboard activities and demonstrations of how to use the equipment, to introduce students to different laboratory techniques and theoretical concepts connected to experimentation (e.g., Sarmouk et al. 2019). The CLIL treatment group received an additional short introduction language-specific objectives, for instance how to signal when instructions are too difficult to follow, where to find supporting vocabulary, and when to solve the vocabulary riddles in the vocabulary scaffolding exercise book. The guide was supported by junior research assistants proficient in English and Biology teaching. Accompanying teachers of the student classes either participated in the experiments themselves or observed their students' performance and involvement.

DNA-related theoretical and experimental phases. Each experimental phase was preceded by a DNA-related theoretical introduction. Starting with an as yet unsolved mystery murder in Munich, where the DNA of the assailant was obtained but did not match any of the suspects, (DNA relevance, Table 1), the guide used an interactive smart-board presentation to introduce the basic scientific properties of DNA and reactivate students' prior knowledge about cells. Students used their 'old' and 'new' knowledge to answer open questions in the laboratory manual, which also encouraged them to make predictions about the outcome of the subsequent experiments (Mierdel and Bogner 2019). After successful DNA extraction, students received an introduction to the scientific practice of gel-electrophoresis. The guide used a poster with clozes and predefined chunks of text as well as demonstrations to explain the basic functions of gel-electrophoresis. For the German treatment group, both the cloze and the text chunks were in German, while for the CLIL treatment group, the cloze was in English and the text chunks in German to provide all necessary information without overwhelming students through code-switching (Cheshire & Gardner-Chloros 1998). The alternation of theoretical and experimental phases should further ensure that students are not overburdened by too much new information.

Experimental Phases. We used an evidence-based, two-step approach (Roth et al. 2020) in our experimental phases. That is, students first answered questions in their individual laboratory manuals and thought about subsequent experimental procedures. Second, they worked in pairs to discuss every step before carrying out experiments that effectively combine hands-on and minds-on activities and require students to do more than simply follow instructions (Mierdel and Bogner 2019). For our hands-on approach, we provided sufficient content scaffolding material in our laboratory manual to enable independent experimentation and self-reliant protocolling of their observations, including open questions about theoretical details relevant to understanding the different experimental steps as well as questions about observations with respective space for answers (ESM 1). Illustrations further helped in carrying out the different experimental steps (ESM 2). Moreover, we offered open questions regarding components of the DNA that students had to extracted from a text about the discovery of DNA (Usher 2013) and respective craft materials for modelling (Roth et al. 2020). The laboratory manual was designed and adapted in the course of four laboratory evaluations (Goldschmidt and Bogner 2016, Langheinrich and Bogner 2016, Mierdel and Bogner 2019, Roth et al. 2020).

Model-related Phases. Both model-related phases followed the experimental phases after a short lunch break. As models are an integral part of scientific discourse to visualize and explain complex scientific phenomena, they are also increasingly important in science education (Krajcik & Merritt 2012). To help students understand the complex processes involved in modelling, we subdivided our model-related phases into a modelling phase, which involved mental modelling based on text analysis and modelling with craft materials, a model evaluation phase, which encompassed model evaluation-1 and model evaluation-2 (Table 1).

These different model-related phases were adapted from the four main stages of the Model of Modelling. That is, students first gathered information and experience about the scientific phenomenon of DNA. Based on an additional text about the discovery of DNA (Usher 2013), they formed a mental model. This mental model served as a foundation for the physical model crafted from different materials. In a last step, students evaluated limitations of their first model (Justi and Gilbert 2002). Mental modelling may, thereby, have been a key element to aggregate knowledge from experimentation and the text in a first simplified mental model (e.g., Roth et al. 2020, Franco & Colinvaux 2000).

After the modelling phases, students evaluated their model in a reciprocal self-evaluation throughout model evaluation-1 phase. That is, students critically reviewed their own handcrafted DNA models in a paper-and-pencil version. This is similar to the method of reflective writing and should encourage self-evaluation (Kovanović et al. 2018). Additional open-ended questions in the laboratory manual about the different components of DNA assisted in this endeavour (Roth et al. 2020). In model evaluation-2 phase, students consulted a commercially available DNA demonstration model and – based on this - assessed the quality of their hand-crafted DNA models.

Interpretation Phase. The interpretation phased started with the result of the model-related phases. The guide used an interactive smartboard presentation to go through the different model-related phases and revealed the original DNA model by Watson and Crick. This made students aware about the different forms of models to illustrate the structure of DNA but vary in their levels of detail. The guide also revealed the results of the gel-electrophoresis and again explained the meaning and formation of the different bars.

Language Scaffolding

To support content learning in a foreign language while fostering scientific literacy and language learning, we designed a language scaffolding exercise book (examples ESM3). It contained language specific riddles in the form of different word search puzzles or crossword puzzles in connection to an allocation of these words to English definitions. Students also had to match the words and their definitions with either given German translations or had to translate the words into German. At all times, students were allowed to use English-English and English-German dictionaries available in the laboratory. For each intervention phase, we had one specific language scaffolding exercise.

Moreover, we included short translations or explanations for scientifically crucial vocabulary in their laboratory manuals and added German equivalents for laboratory equipment to our presentation. To help students understand the complex procedure of gel electrophoresis, we had a poster with English descriptions and questions, but we used German answer tags and visuals to make the different process steps easier to follow. Oral explanations were provided in English.

Participants

Overall, 252 ninth graders of Bavarian high schools participated in our intervention (girls 52.4%, boys 47.6%; SDGender = 6.2; MAge = 14.6, SDAge = 0.7). Seven classes took part in our monolingual German intervention (n = 139) and eight classes in our CLIL intervention (n = 107). The German treatment group formed 70 student groups and the CLIL treatment group 53 student groups (52 2-person groups and one 3-person group). To account for potential influences of prior knowledge and school grades on student performance throughout the module, we compared students' T0 knowledge scores and Biology grades of both treatment groups.

We found no significant difference (Mann-Whitney U test [MWU]: Z = -.725, p = .468). For English language skills, we did not require any additional tests, as none of the students had considerable English language experiences outside of the classroom. Participating teachers primarily had a background in Biology and Chemistry with only one teacher who studies Biology and English for teaching.

Participation of schools and individual students was voluntary. We sent invitations to all high schools in a one hour driving range six month before the module started. The only prerequisite for participation was that teachers had not yet started their lessons on DNA. To comply with regulations of the Bavarian Ministry of Education and Cultural Affairs, we asked for written parental consent prior to students' participation in our study. The data collection was pseudo-anonymous and students could not be identified, which is in line with the Declaration of Helsinki (2013). The design of the module and the questionnaires were pre-approved by the Bavarian Ministry of Education and Cultural Affairs and received the reference number X.7-BO5106/149/10.

Variables

Students were asked to complete a knowledge test about scientific and structural characteristics of DNA at three different testing times: two weeks before participation (pre-test; T0), directly after the module (post-test; T1), and eight weeks after participation (retention-test; T2) in their native language. To retain objectivity, students did not know the exact purpose of the questionnaires that they completed.

Students' knowledge. To assess students content knowledge about scientific and structural characteristics of DNA, we have designed and developed an ad-hoc knowledge test with overall 30 multiple-choice items (examples, Table 2) in the course of three laboratory

TABLE 2 Knowledge item examples.

Item difficulty[a] (T0 / T1)		Item example[b]
Monolingual (non-CLIL)	Bilingual (CLIL)	
84.6/85.5	73.8/ 76.9	A positively charged particle migrates in an electrical field … (Item 1) (a) between the two poles (b) to the positive pole* (c) to the negative pole (d) not at all
52.9/28.5	8.4 / 15.0	What is wrong? The migration speed of a molecule within the electrical field depends on …? (Item 3) (a) the applied voltage* (b) the sample's density (c) the gel's density (d) its molecular size*
83.7/87.8	67.5 / 90.4	The molecular structure can be best compared to…? (Item 15) (a) a cardboard tube (b) a twisted rope ladder* (c) train tracks (d) a thread
55.6/35.1	5.4 / 23.7	What is wrong? DNA molecules are being made visible with gel-electrophoresis via… (Item 30) (a) blue coloured loading buffer (b) a dye that attaches to DNA molecules (c) a dye that glows under UV radiation (d) addition of dye to the gel*

Note: Correct answers are marked with an asterisk * evaluations (Langheinrich and Bogner 2016, Mierdel and Bogner 2019, Roth et al. 2020). Questions and respective multiple-choice answers were altered and randomly assigned after every testing time (T0, T1, T2) to prevent students from answering in systemic patterns.

The knowledge questionnaire fulfilled required quality criteria. Consistency with the state syllabus in the design of the module and questions provided content validity. Inter-item correlations below .20 (T0 = .08; T1 = .19; T2 = .18) indicated that the items were distinct and considered different areas of content knowledge. This alongside the complexity of the latent construct of cognitive achievement accounts for construct validity (Rost 2004). The reliability of the questionnaire was determined by calculations of Cronbach's alpha. Scores of .74 (T0), .76 (T1), and .78 (T2) exceeded the threshold of .70, which, according to Lienert & Raatz (1998), allows for the differentiation of groups. An additional calculation of item difficulties (percentage of correct answers, Bortz & Döring 1995) indicated a range between 5% (high difficulty) and 90% (low difficulty). Comparisons between the different testing times showed that the item difficulties considerably improved from T0 to T1 for the CLIL treatment group (Figure 1).

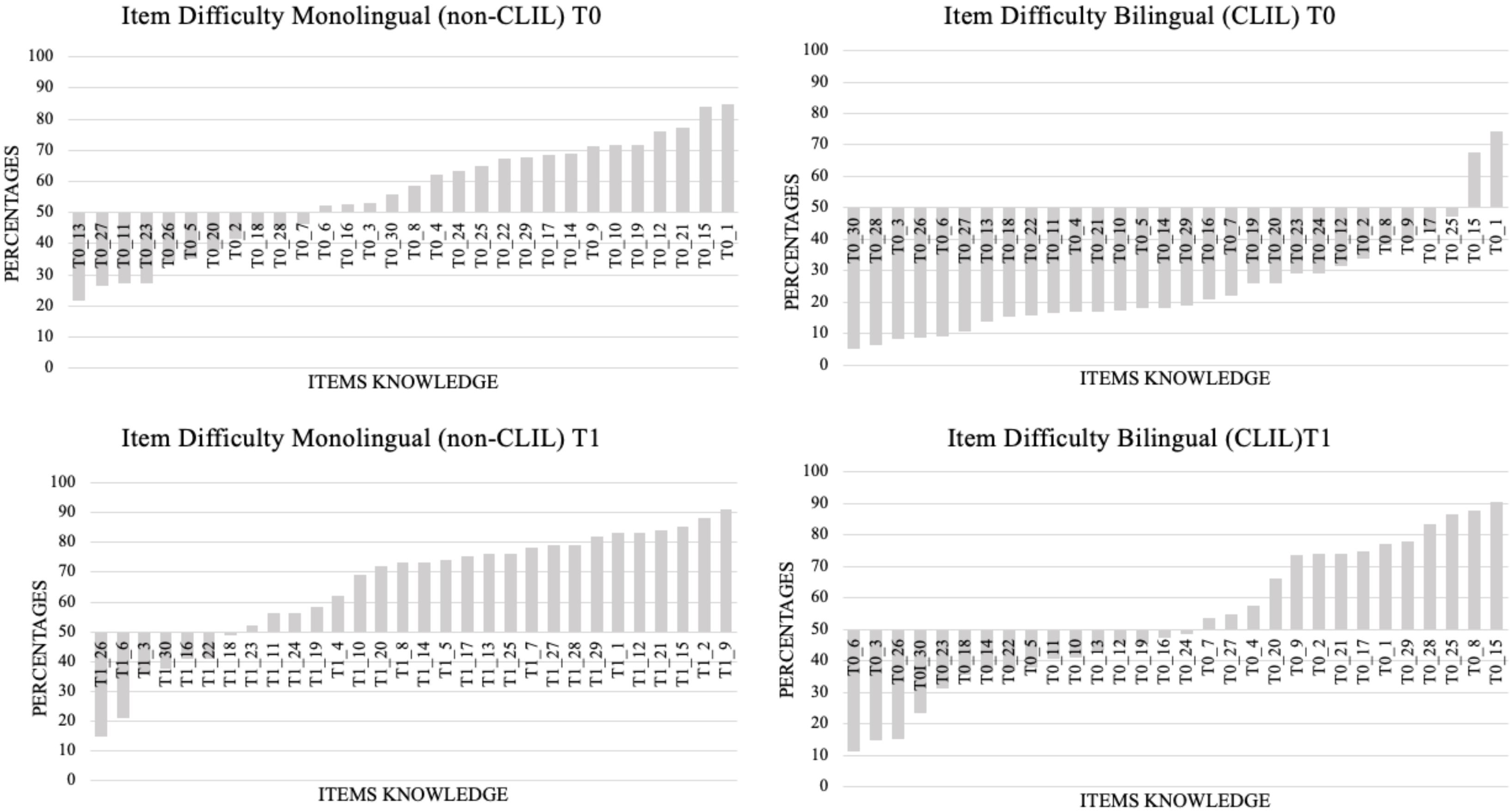


FIGURE 1 Item difficulties of monolingual and bilingual learners for knowledge items between T0 and T1.

We calculated the students' scores and analysed these with regard to increases in knowledge (T1 minus T0), and retention rate (T2 minus T0). Furthermore, we calculated the actual learning success with respect to the maximal attainable score (30 correct answers): (T1 – T0) x (T1/30); and the persistent learning success (T2 – T0) x (T2/30) (Scharfenberg et al. 2007). Increased knowledge is, hence, weighted according to students' actual knowledge, making it possible to compare cognitive achievement despite some students exhibiting a huge increase in knowledge yet low final scores, and *vice versa*. We also calculated correlations between Biology as well as English grades and post-test (T1) scores for knowledge items using Spearman-Rho (Field 2012).

Statistical Analysis. Statistical Analysis. We decided to use nonparametric tests to analyse our data, as first assessments showed a non-normal distribution of our variables (Kolmogorov-Smirnov test (Lilliefors modification): partially $p < .001$). Our assessment of subgroups that often reaches the threshold for Gaussian normal distribution encouraged our decision (Lomax 1986). We, therefore, also use boxplots visualise our results. To analyse intra-group differences of the three different test dates, applied the Friedman test (F) to illustrate general differences and Wilcoxon (W) signed-rank test to show concrete changes between T0 and T1, T1 and T2, and T0 and T2. For differences between the monolingual German and CLIL treatment group, we used the Mann-Whitney U tests (MWU). We decided to use a Bonferroni correction to eliminate barely significant results, which could have simply been a coincidence due to multiple testing (Field 2012). Should the results remain significant, we calculated effect sizes r (Lipsey & Wilson 2001) and categorised them into small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. Spearman's rank correlations served for correlation analyses and were reported as Spearman's Rho values.

Results

Intra-group Analyses of Cognitive Achievement

Intragroup analysis (F and W tests, Table 3) of students' overall knowledge of model-related and scientific knowledge of DNA indicated significant changes for monolingual and bilingual treatment groups: Their knowledge first increased and then dropped from T1 to T2, but never reached levels below T0. This suggests that both student groups were able to attain short-term and mid-term knowledge through participation in the intervention (Table 3).

TABLE 3 Cognitive achievement of monolingual (non-CLIL) and bilingual (CLIL) student sample.

Items	Friedman test			Wilcoxon signed-rank test					
				T0 / T1		T0 / T2		T1 / T2	
	Chi-Square	*df*	*p*	*Z*	*p*	*Z*	*P*	*Z*	*p*
Monolingual (non-CLIL)	161.62	2	< .001	-10.04	< .001	-8.49	< .001	-5.67	< .001
Bilingual (CLIL)	72.49	2	< .001	-7.69	< .001	-5.28	< .001	-3.42	< .001

Inter-group Analyses of Cognitive Achievement

We first calculated difference variables to determine students' short-term increase in knowledge and mid-term retention rates as well as to mitigate the influence of differences in students' knowledge at T0 (Table 4, notea). Our calculations for increases in knowledge (T1-T0) and retention rates (T2-T0) of overall 30 knowledge-test items were based on sum scores (Field 2012) (Table 4). Additional calculations aimed to determine learning success variables for actual learning success ((T1 – T0) x (T1/30)) and persistent learning success ((T2 – T0) x (T2/30)) (Scharfenberg et al. 2007). Both difference and learning success variables were then used to assess inter-group differences.

TABLE 4 Dependent variables for both monolinguals (non-CLIL) and bilinguals (CLIL), analyzed with regard to knowledge scores, difference variables and learning success.

Dependent variable	Knowledge	
	Knowledge items (max. 30)	
	Monolingual (non-CLIL)	Bilingual (CLIL)
	Mdn (25th/75th *P*)	*Mdn* (25th/75th *P*)
T0[a]	6.3 (4.0/8.0)	8.0 (5.0/13.0)
T1[b]	19.4 (16.0/22.0)	18.0 (14.0/25.0)
T2[c]	17.2 (15.0/19.0)	16.0 (11.0/19.0)
	Difference variables	
	Knowledge items (max. 30)[j]	
	Monolingual (non-CLIL)	Bilingual (CLIL)
	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)
Increase in knowledge[d]	13.1 (9.0/16.0)	9.0 (5.0/11.0)
Retention rate[e]	10.4 (8.0/13.0)	6.0 (3.0/11.0)
	Learning success	
	Overall knowledge items (max. 30)	
	Monolingual (non-CLIL)	Bilingual (CLIL)
	Mdn (25th/75th *P*)	Mdn (25th/75th *P*)
Actual learning success[f]	8.7 (5.1/11.9)[h]	4.4 (2.3/8.4)
Persistent learning success[g]	6.0 (3.7/7.9)[i]	8.5 (4.0/12.0)

Notes: a) MWU-test over-all knowledge: Z = -3.42; p = .001; r = .218. b) MWU-test over-all knowledge: Z = -3.31; p = .001; r = .210. c) MWU-test over-all knowledge: Z = -.1.51; p = 0.131; r =.096. d) score: T1 - T0; MWU-test over-all knowledge: Z = -6.02; p < .001; r = .383. e) score: T2 – T0; MWU-test over-all-knowledge: Z = -4.64; p < .001; r = .295; f) factual learning success: (T1 – T0) x (T1/30) for overall. g) persistent learning success: (T2 – T0) x (T2/30) for overall knowledge. h) MWU-test over-all knowledge items: Z = -2.49; p = .013; r = .159. i) MWU-test over-all knowledge items: Z = -1.51; p = 0.131; r = .096. j) percentage share of correctly answered items: increase in knowledge overall monolingual and bilingual (44%, 30%); retention rate in knowledge overall monolingual and bilingual (35%, 20%).

Overall Knowledge. Scores yielded from the assessment of the 30-item knowledge test displayed significant differences in increase in knowledge and retention rate between monolingual and bilingual learners. Monolingual learners scored significantly higher with a medium-to-large effect size in both difference variables than bilingual learners (Table 4, notes d/e).

While difference variables already give a good indication of the students' performance, they do not account for their cognitive achievement. That is why additional calculations of learning success variables were required. For short-term actual and mid-term persistent learning success, we identified significant differences with small effect sizes in overall knowledge were identified (Table 4; notes h/i). Compared to bilingual learners, monolingual learners achieved higher actual learning success scores in the 30-item knowledge test (Figure 2).

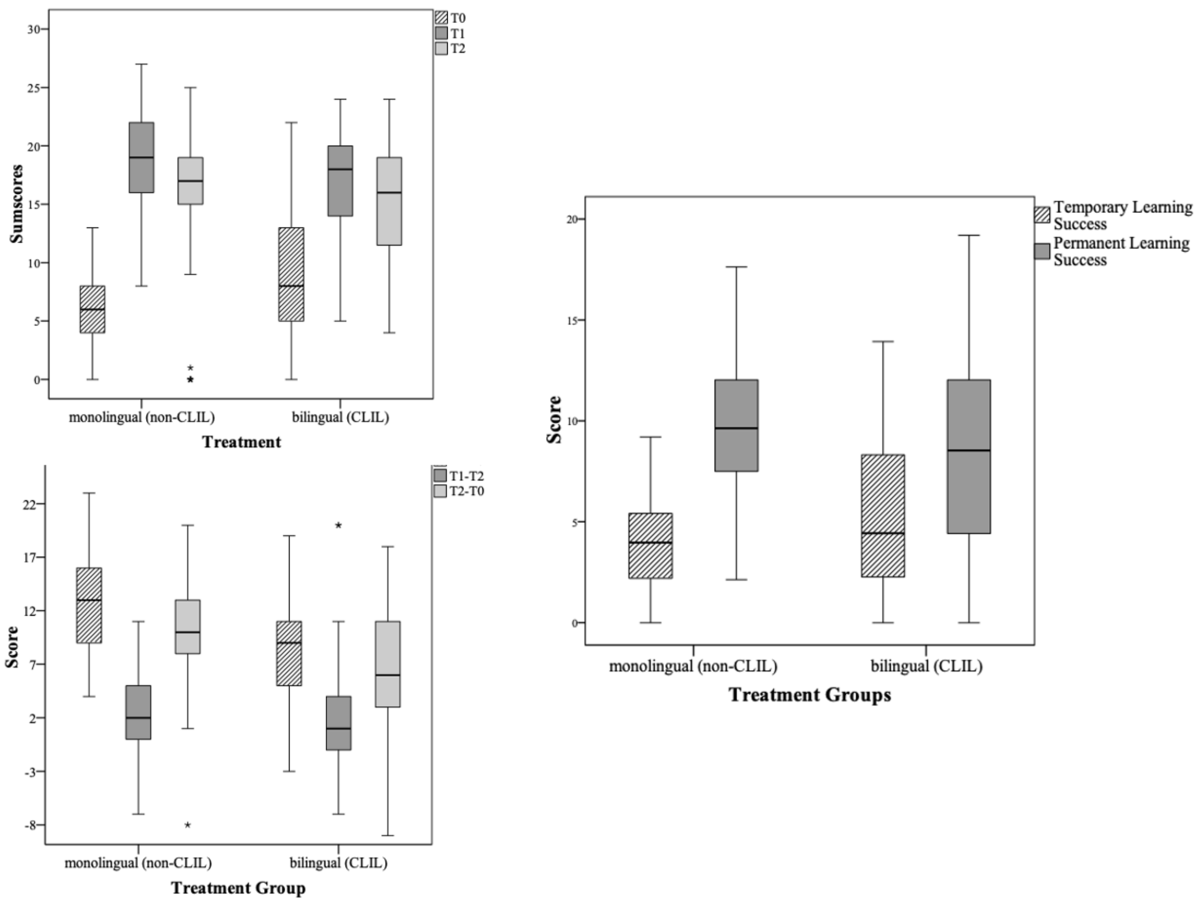


FIGURE 2 Differences between monolingual and bilingual learners with regard to cognitive achievements between testing times T0, T1, and T2 as well as increase in knowledge and retention rate as well as temporary and permanent learning success.

Correlation Biology grades. There were no correlations for bilingual learners (rS = -.049, p = 621) between Biology grades and post-test (T1) scores. Yet, monolingual learners yielded significant negative correlations (rS= -.171, p = .045) for difference and learning success variables of the 30-item knowledge test. Thus, students with lower grades appear to improve their difference and learning success scores in post-tests as compared to students with better grades. This is unfortunately not true for bilingual learners.

Discussion

Our one-day outreach module has a positive effect on scientific academic achievement regardless of CLIL or non-CLIL implementation method. That is, both treatment groups profited from the combination of hands-on tasks with minds-on activities. This informal hands-on-minds-on approach has been proven successful to convey fundamental scientific concepts of molecular biology in previous studies (e.g., Goldschmidt & Bogner 2016, Langheinrich & Bogner 2016, Mierdel & Bogner 2020, Roth et al. 2020). Since we refrained from deploying 'cookbook' procedures, we considered both cognitive and affective dimensions (Hofstein & Lunetta 2004). That is, besides confidently handling laboratory equipment, doing science away from the 'cookbook' entails the identification of problems, the generation of feasible hypothesis, the analysis and interpretation of data as well as the use of models to find possible explanations (Carmel et al. 2020).

All these combined could, however, if not regularly trained in class, go beyond individual mental capabilities and impair knowledge acquisition (e.g., Mierdel & Bogner 2020, Roth et al. 2020). We observed the impact of high mental load on scientific academic achievement in our previous findings, which are in line with cognitive load theory. Therein, the germane load is crucial for memorising and learning but is often impaired by high intrinsic and extraneous load (Sweller 2015). Although the non-CLIL comparison group achieved significantly higher scores in knowledge acquisition compared to students, who did not participate in any outreach module, the relative difference was not as high as expected (Mierdel & Bogner 2019).

Similar findings or even no differences at all were reported by e.g., Reynolds (1991). Possible explanations for these outcomes have been repeatedly provided with cognitive load theory, outlining learning difficulties for 'high information [situations] in which [the] working memory [may be] overloaded with incoming data' (Johnstone & Wham 1982).

Despite higher mental effort involved in learning scientific content in a foreign language, the discursive activity of negotiating meaning in the process of doing science by talking science helps students acquire knowledge (e.g., Evnitskaya & Morton 2011, Kress et al. 2001). Thus, although a potential overload of the individual working memory may have also influenced our findings, the combination of hands-on tasks with minds-on activities in CLIL learning (Glynn & Muth 1994) may have conveyed fundamental scientific concepts. Since our tests were conducted in the native language, concerns about the transfer of knowledge obtained primarily in the foreign language to the native language (e.g., Dalton-Puffer 2007, Grandinetti 2013) could also be dispensed with.

Nevertheless, short-term knowledge acquisition was lower for CLIL treatment groups, indicating that the mental processing of scientific content in a foreign language accompanied by hands-on activities may overstrain students' capabilities. Previous studies have already proven that outreach hands-on learning can be mentally strenuous (e.g., Scharfenberg & Bogner 2010, Meissner & Bogner 2012). Combining laboratory learning with scientific learning in a foreign language may further add to the mental load (Rodenhauser & Preisfeld 2014) and overstrain the working memory (Roussel et al. 2017). Our results confirm this tense relationship between language and content learning, given the lower results in our short-term knowledge acquisition for CLIL learners.

As is described in functional linguistics, language is 'an active partner in the constitution of reality' (Maxwell-Reed 2017, p.1750) and requires mental capabilities to create meaning from intake of information, which influences cognitive load. Although it would be beneficial for the mental load, language learning and learning of scientific contents through language are interdependent and cannot be separated (e.g., Maxwell-Reed 2017, Halliday 2003). This hypothesis is reinforced by general principles of scientific literacy. Scientific contents are understood and effectively communicated when the underlying scientific language is mastered. In short: language determines the content conveyed, and the content conveyed determines the language used (Sutton 1996, Yore & Treagust 2006). This relationship is reflected in Cummins' (1979) developmental interdependence hypothesis, which describes the native language as requirement to exceed a certain threshold of scientific literacy in order to facilitate content learning in the foreign language.

Thus, lacking capabilities of talking science in the native language may also affect content learning (Rodenhauser & Preisfeld 2014). In consequence, the mental overload in our case might originate in both, new contents and new vocabulary considering the levels of processing theory (Craik &Lockhart 1972). Thus, processing scientific contents in a foreign language while for the first time experimenting in an outreach hands-on lab seems to affect temporary learning success negatively. Piesche et al. (2016) even hypothesised that simultaneous learning of language and content affects mental capacities and diminishes knowledge acquisition.

Different rules, however, seem to apply to permanent learning success since difference variables display significant lower retention rates for CLIL compared to the monolingual group.

A possible explanation may be the cognitive processing of content on several levels for CLIL learners (Meyerhöffer & Dreesmann 2019): Although mentally much more strenuous, many passages require repeated and more precis reading in order to be understood, which is why a similar amount of knowledge may remain anchored in long-term memory (e.g., Rodenhauser & Preisfeld 2014, Marian & Fausey 2006). Piesche et al. (2016) or Fernández-Sanjurjo et al. (2017) provided a similar explanation for knowledge acquisition in both native language and CLIL learners. Authors such as Admiraal et al. (2006) or Haagen-Schützenhöfer et al. (2011), however, reported different explanations for differences in knowledge acquisition between CLIL and non-CLIL learners, making findings about CLIL learning highly ambiguous.

Since appropriate scaffolding is important to support bilingual learning, the amount and quality of scaffolding material may also impact content learning (e.g., Gottlieb 2016, Fernández-Fontecha et al. 2020). Even though we considered linguistic, graphic, and interactive scaffolds, knowledge acquisition did not achieve the expected results. Thus, ineffective language scaffolding can lead to fossilisation of students' language use and impairment of science sense-making (Buxton et al. 2019). To avoid this vicious circle, scaffolding material should ask students to reorganise and reconstruct knowledge on language and content level. Especially if the content is unknown and the content-cognitive demand high, the respective language-cognitive demand should be low (Grandinetti et al. 2013). Thus, scaffolding material should attempt to ease the informational burden by lowering the cognitive demand on either language or content level. This is easier said than done, which is why scaffolding approaches require constant systematic analysis and adaptation.

In line with indications of McGuiness (1999), we focused particularly on keeping the language-cognitive demand low by providing vocabulary quizzes as well as translations and explanations of difficult scientific terms. However, since genetics and laboratory procedures were unknown and new vocabulary was directly linked to new scientific methods, the language dimension could develop fast enough to have a moderating effect on content level. Especially in CLIL short-scale implementations, which are very densely packed and have high demands on scientific language proficiency, it takes more than regular scaffolding to guarantee sufficient knowledge acquisition (e.g., Coyle 2007, Grandinetti et al. 2013). Fernández-Fontecha et al. (2020) suggest multimodal scaffolding with, for instance images, to have to content presented in other semiotic forms than text.

Besides different kinds of scaffolding, it may also be feasible to learn key vocabulary before participating in short-scale CLIL implementations, such as our gene-technology lab. This could reduce mental load, since the basic scientific vocabulary is already established and only needs to be put into context (McGuiness 1999). Given the importance of science learning and scientific literacy in English for effective and successful participation in modern-day society, it is, accordingly, vital to foster and deepen the understanding of the complex relationship between language learning and content learning (European Commission 2004). It is vital to tackling the trade-off between STEM knowledge acquisition and language learning in short-scale bilingual implementations.

Limitations

First, our study built on a one-day outreach CLIL module with ninth graders with little prior experience in hands-on experimentation. Second, students are mostly monolingual and had only little real-life experience with the foreign language outside of classrooms. Third, a lack of commonly agreed standardised instruments for assessing CLIL content learning impairs adequate comparison (Dalton-Puffer 2011). Fourth, due to the context-dependency of CLIL learning, results cannot easily be extrapolated (Pérez-Cañado 2012). In consequence, generalising the success of CLIL learning requires acknowledgement of the diversity of possible CLIL implementations (Fernández-Sanjurjo et al. 2017). Further research in the context of short-scale implementations of CLIL is needed to get a better overview of options.

Conclusion

Previous studies primarily focused on language learning or content learning in long-scale CLIL modules (Meyerhöffer & Dreesmann 2019); our study contributes to understanding the relationship between language and science content learning in short-scale CLIL modules. Although the foreign language-learning group has proven less successful, CLIL outreach learning provides positive implications for long-term learning which is in line with Marian & Fausey (2006), since differences in short-term learning achievement equilibrated in the retention test. Although we cannot identify the causes of lower scores of the CLIL group, reducing content in later implementations might help (Martin 2015) as well as taking cognitive load into consideration (Sweller 2015).

Acknowledgements

Special thanks to the teachers and students involved in this study for their cooperation. Also, many thanks to the Horizon 2020 Framework Program and the German Research Foundation.

Funding

The research was supported by the Horizon 2020 Framework Program (grant number 665917) and the German Research Foundation (grant number LA 2159/8-6) as well as the University of Bayreuth.

Ethical Statement

Hereby, we, Tamara Roth and Franz X. Bogner, consciously assure that for the manuscript 'The trade-off between STEM knowledge acquisition and language learning in short-term CLIL implementations' generethical standards have been fulfilled and 'all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.'

Consent Statement

'Informed consent was obtained from all individual participants included in the study.'

References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12, 75–93. https://doi.org/10.1080/13803610500392160.

Bonnet, A. (2004). *Chemie im bilingualen Unterricht: Kompetenzerwerb durch Interaktion*. [Chemistry in bilingual education: competencies through interaction]. Opladen: Leske & Budrich.

Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal*, 1(4), 65–78.

Bortz, J., & Schuster, C. (2010). Faktorenanalyse [Factor Analysis]. In: Bortz, J., & Schuster, C. (eds.). Statistik für Human- und Sozialwissenschaftler [Statistics for human and social scientists] (pp. 385–426). Berlin, Heidelberg, New York: Springer Verlag.

Buse, M., Damerau, K. & Preisfeld, A. (2018). A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts. *International Journal of Environmental & Science Education*, 13(8), 647-660.

Buxton, C., Lee, O., & Santau, A. (2008). Promoting Science Among English Language Learners: Professional Development for Today's Culturally and Linguistically Diverse Classrooms. *Journal of Science Teacher Education*, 19 (5), 495-511.

Buxton et al. (2019). Understanding science and language connections: new approaches to assessment with bilingual learners. Research in Science Education, 49, 977-988. https://doi.org/10.1007/s10972-008-9103-x.

Carmel, J.H., et al. (2020). Helping students to „do science": characterizing scientific practices in general chemistry laboratory curricula. *Journal of Chemical Education*, 96(3), 423-434. https://doi.org/10.1021/acs.jchemed.8b00912.

Cheshire, J., & Gardner-Chloros, P. (1998). Code-switching and the sociolinguistic gender pattern. International Journal of the Sociology of Language, 129, 5–34.

Cook, T. D., & Campell, D. (1979). Quasi-experimentation. Design & analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543-562. https://doi.org/10.2167/beb459.0.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121-129.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated learning (CLIL) Classrooms*. Amsterdam: John Benjamins.

Dalton-Puffer, C. (2011). Content-and-language integrated learning: from practice to principles? *Annual Review of Applied Linguistics*, 31, 182-204. doi: 10.1017/S0267190511000092

Declaration of Helsinki, World Medical Association (2013). Journal of the American Medical Association, 310, 2191-2194.

De Houwer, J., Barnes Holmes, D., & Moors, A. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. Perspectives on Psychological Science, 6,202–209. https://doi.org/10.1177/1745691611400238.

De Dios Martínez Agudo, J. (2019). The impact of CLIL on English language competence in monolingual context: a longitudinal perspective, *The Language Learning Journal*, doi: 10.1080/09571736.2019.1610030.

Donaldson, J.P. & Allen-Handy, A. (2020). The nature and power of conceptualizations of learning. *Educational Psychology Review*, 32, 545-570. https://doi.org/10.1007/s10648-019-09503-2

Dunlosky, J., et al. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. https://doi.org/10.1177/1529100612453266.

Eurydice (Hg.) (2005). *Content and language integrated learning (CLIL) at school in Europe*. Brussels: Eurydice, European Unit.

European Commission (2004). *Promoting language learning and linguistic diversity. An action plan 2004-06.* Luxembourg, Office for Official Publications of the European Communities.

European Commission (2012). *FAQs on multilingualism and language learning*. Brussels. Retrieved from http://europa.eu/rapid/press-release_MEMO-12-703_en.pdf

Evnitskaya, N. & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language and Education*, 25(2), 109-127. https://doi.org/10.1080/09500782.2010.547199.

Fernández-Fontecha, A., O'Halloran, K.L., Wignell, P., & Tan. S. (2020). Scaffolding CLIL in the science classroom via visual thinking: a systematic functional multimodal approach. *Linguistics and Education*, 55, 1-10. https://doi.org/10.1016/j.linged.2019.100788.

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J.M. (2017). Analysing students' content-learning in science in CLIL vs. Non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, 22(6), 661-674. https://doi.org/10.1080/13670050.2017.1294142.

Field, A. (2012). *Discovering statistics using IBM SPSS statistics (4th ed.)*. London: SAGE.

Finkbeiner, Claudia & Fehling, Sylvia (2006): Investigating the role of awareness and multiple perspectives in intercultural education. In P. Ruggiano Schmidt & C. Finkbeiner: *ABC's of Cultural Understanding and Communication. Greenwich* (USA): Age Publishing.

Franco C. & Colinvaux D. (2000). Grasping Mental Models. In: Gilbert J.K., Boulter C.J. (Eds.) *Developing models in science education* (pp. 93-118). Springer.

Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: achieving scientific literacy. *Journal of Research in Science Teaching*, 31(9), 1057-1073. https://doi.org/10.1002/tea.3660310915.

Goldschmidt, M., & F.X. Bogner. (2016). Learning about genetic engineering in an outreach laboratory: Influence of motivation and gender on students' cognitive achievement. *International Journal of Science Education* 6(2): 166-187. https://doi.org/10.1080/21548455.2015.1031293.

Gonzalez-Howard, M. & McNeill, K.L. (2016). Learning in a community of practice: factors impacting English-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, 53(4), 527-553. https://doi.org/10.1002/tea.21310.

Gottlieb, M. (2016). *Assessing English language learners: Bridges for language proficiency to academic achievement*. Thousand Oaks: Corwin.

Grandinetti, M.m Langellotti, M., & Teresa Ting, Y.L. (2013). How CLIL can provide a pragmatic means to renovate science education – even in a sub-optimally bilingual context. *International Journal of Bilingual Education and Bilingualism*, 16(3), 354-374.

Haagen-Schützenhöfer, C., Mathelitsch, L., & Hopf, M. (2011). Fremdsprachiger Physikunterricht: Fremdsprachlicher Mehrwert auf Kosten fachlicher Leistungen? Auswirkungen fremdsprachenintegrierten Physikunterrichts auf fachliche Leistungen [Foreign language physics lessons: Added value for language skills at the cost of content achievement? Effects of content-language-integrated physics classes on content achievement]. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 223–260.

Halliday, M. A. K. (2003). Introduction: On the 'Architecture' of Human Language. In J.J. Webster (Ed.), *Collected Works of M.A.K. Halliday*, 3, 1–35. London: Continuum.

Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education*, 36(15), 2534-2553. https://doi.org/10.1080/09500693.2014.899722.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88, 28–54. https://doi.org/10.1002/sce.10106.

Honeycutt-Swanson, L., Bianchini, J. A., & Lee, J. S. (2014). Engaging in argument and communicating information: A case study of English language learners and their science teacher in an urban high school. *Journal of Research in Science Teaching*, 51(1), 31–64. https://doi.org/10.1002/tea.21124.

Johnstone, A.H. & Wham, A.J.B. (1979). A model for undergraduate practical work. Education in Chemistry, 16, 16–17.

Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, 24, 369–387. https://doi.org/10.1080/09500690110110142.

Kelly,G. J. (2007). Discourse in science classrooms. In S. K. Abell & N.G. Lederman (Eds.), *Handbook of research on science education* (pp. 443–469). Mahwah, NJ: Lawrence Erlbaum.

Klieme, E., et al. (Eds.) (2010). *PISA 2009: Bilanz nach einem Jahrzehnt* [PISA 2009: Review after a decade]. Münster: Waxmann.

KMK (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of Biology education for secondary school]. Munich: Luchterhand.

Koch, A., & Bünder, W. (2006). Fachbezogener Wissenserwerb im Bilingualen Naturwissenschaftlichen Anfangsunterricht [Content-related knowledge achievement in bilingual science lessons]. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 67–76.

Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Korthagen, F. & Lagerwerf, B. (1995). Levels in learning. *Journal of Research in Science Teaching*, 32(1), 1011-1038. https://doi.org/10.1002/tea.3660321004.

Kovanović, V., et al. (2018). *Understanding the relationship between technology-use and cognitive presence in MOOCs.* Proceedings of the Seventh International Conference on Learning Analytics and Knowledge, 582-583. New York: Association for Computing Machinery.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *Science and Children*, 49, 10–13.

Krechel, H.-L. (2013). Organisationsformen und Modelle in weiterführenden Schulen. [Organisation and models of bilingual education]. In W. Hallet & F. G. Königs (Eds.), *Handbuch Bilingualer Unterricht. Content and Language Integrated Learning* (pp.74–80). Seelze: Friedrich Verlag

Kress, G., C., Ogborn, J.J., & Tsatsarelis C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London: Continuum.

Lam, S. F.,Wong, B. P. H., Yang, H., & Liu, Y. (2012). Understanding student engagement with a contextual model. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 403-420). Boston, MA: Springer US.

Langheinrich, J., & F.X. Bogner. (2016). Computer-related self-concept: The impact on cognitive achievement. *Studies in Educational Evaluation*, 50, 46-52. https://doi.org/10.1016/j.stueduc.2016.06.003.

Lee, S., Kang, E., & Kim, H-B. (2015). Exploring the impact of students‘ learning approach on collaborative group modeling of blood circulation. Journal of Science Education Technology, 24, 234-255.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lomax, R.G. (1986). The effect of measurement error in structural equation modeling, The Journal of Experimental Education, 54 (3), 157-162.

Lovey, B.R., & Riggs K.M. (2019). Flipping the laboratory: improving student engagement and learning outcomes in second year science courses. *International Journal of Science Education*, 41, 64-79. https://doi.org/10.1080/09500693.2018.1533663.

Marian, V. & Fausey, C.M. (2006). Language-dependent memory in bilingual learning. *Applied Cognitive Psychology*, 20, 1025-1047. https://doi.org/10.1002/acp.1242.

Martin, M.A. (2015). Teacher education for content and language integrated learning: insights from a current European debate. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 18(3), 153-168.

Maxwell-Reed, C. (2017). Classroom discourse in bilingual secondary science: language as medium or language as dialectic? *International Journal of Bilingual Education and Bilingualsim*, 23(4), 499-512. https://doi.org/10.1080/13670050.2017.1377683.

Maybin, J., Mercer, N. & Stierer, B. (1992). Scaffolding: learning in the classroom. In K. Norman (Ed.), *Thinking Voices. The Work of the National Oracy Project* (pp. 186–95). London: Hodder & Stoughton.

McGuiness, C. (1999) *From Thinking Skills to Thinking Classrooms: A Review and Evaluation of Approaches for Developing Pupils' Thinking.* Research Report 115, DfEE: HMSO.

Meissner, B., & Bogner, F.X. (2012). Science teaching based on cognitive load theory: Engaged students, but cognitive deficiencies. *Studies in Educational Evaluation,* 38(4-5), 127-134). https://doi.org/10.1016/j.stueduc.2012.10.002.

Meyerhöffer, N. & Dreesmann, D.C. (2019). English-bilingual Biology for standard classes development, implementation and evaluation of an English-bilingual teaching unit in standard German high school classes. *International Journal of Science Education*, 41(10), 1366-1386. https://doi.org/10.1080/09500693.2019.1607620.

Mierdel, J., & F.X. Bogner. (2019). Comparing the use of two different model approaches on students' understanding of DNA models. Education Sciences 9(2), 115-133. https://doi.org/10.3390/educsci9020115.

Minner, D.D., Levy, A.J., & Century, J. (2010). Inquiry-based science instruction – what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research and Science Teaching,* 47(4), 474-496. https://doi.org/10.1002/tea.20347.

Nation, I. S. P., & Newton, J. (2008). *Teaching ESL/EFL listening and speaking*. London: Routledge.

Nikula, T. (2014). Hands-on tasks in CLIL science classrooms as sites for subject-specific language use and learning. *System*, 1-14. https://doi.org/10.1016/j.system.2015.04.003.

Oga-Baldwin, Q.W.L. (2019). Acting, thinking, feeling, making, collaborating: the engagement process in foreign language learning. *System*, 86, 1-10. https://doi.org/10.1016/j.system.2019.102128.

Oktaviani, A., & Fauzan, A. (2017). Teacher perceptions about the importance of English for young learners. *Linguistic, English Education and Art (LEEA) Journal*, 1(1), 1-15. https://doi.org/10.1016/j.system.2019.102128.

Pérez-Cañado, M. L. (2012). CLIL Research in Europe: Past, Present, and Future. International Journal of Bilingual Education and Bilingualism, 15(3), 315-341. https://doi.org/10.1080/13670050.2011.630064.

Piesche et al. (2016). CLIL for all? A randomised controlled field experiment with sixth grade students on the effects of content and language integrated science learning. *Learning and Instruction*, 44, 108–116. https://doi.org/10.1016/j.learninstruc.2016.04.001.

Poza, L.E. (2016). The language of ciencia: translanguaging and learning in a bilingual science classroom. *International Journal of Bilingual Education and Bilingualism,* DOI: 10.1080/13670050.2015.1125849

Renninger, K. A., Ren, Y., & Kern, H. (2018). Motivation, engagement, and interest: “In the end, it came down to you and how you think of the problem”. In C. E. Hmelo-Silver, S. R. Goldman, P. Reimann, & F. Fischer (Eds.), *International handbook of the learning sciences* (pp. 116–126). Milton: Routledge Ltd..

Reynolds, A.G. (1991). The cognitive consequences of bilingualism. In A.G. Reynolds (Ed.), Bilingualism, multiculturalism, and second language learning (pp. 145-182). Hillsdale NJ: Lawrence Erlbaum Associates.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Rocard report: “Science education now: A new pedagogy for the future of Europe”.* EU 22845, European Commission.

Rodenhauser, A. & Preisfeld, A. (2014). Bilingual (German-English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. *International Journal of Environmental and Science Education*, 10(1), 99-110.

Rost, J. (2004). Lehrbuch Testtheorie–Testkonstruktion [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

Roth, T., Scharfenberg, F-J., Mierdel, J., & F.X. Bogner. (2020). Self-evaluative scientific modeling in an outreach gene technology laboratory. *Journal of Science Education and Technology*, 29, 725-739. https://doi.org/0.1007/s10956-020-09848-2.

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: a cognitive load theory approach. *Learning and Instruction*, 52, 69-79. https://doi.org/10.1016/j.learninstruc.2017.04.007.

Sardegna, V. G., Lee, J., & Kusey, C. (2017). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. Language Learning. doi:10.1111/lang.12263.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. *Innovations in Education and Teaching International*, 1–12. https://doi.org/10.1080/14703297.2019.1604247.

Scharfenberg, F. J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus—Results of a teaching unit with authentic experiments. Biochemistry and Molecular Biology Education, 35(1), 28–39.

Scharfenberg, F-J., & Bogner, F.X. (2010). Instructional efficiency of changing cognitive load in an out-of-school laboratory. *International Journal of Science Education,* 32(6), 829-844. https://doi.org/10.1080/09500690902948862

Schwarz, C., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46, 632–654. https://doi.org/10.1002/tea.20311.

Solomon, J. (1980). *Teaching children in the laboratory*. London: Crom Helm.

Sutton, C. (1996). Beliefs About Science and Beliefs About Language. *International Journal of Science Education*, 18 (1), 1–18. https://doi.org/10.1080/0950069960180101.

Sweller, J. (2015). In academe, what is learned and how is it learned? *Current Directions in Psychological Science*, 24, 190-194. https://doi.org/10.1177/0963721415569570.

Usher, S. (2013). *Letters of Note. Correspondence Deserving of a Wider Audience.* Edinburgh, Scotland: Canongate.

Vygotsky, L. S. (1971). *The psychology of art (Scripta Technica, Inc., Trans.)*. Cambridge, MA/London: MIT Press

Walker, J.P. & Sampson, V. (2013). Learning to argue and arguing to learn – driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching*, 50(5), 561-596. https://doi.org/10.1002/tea.21082.

Wilmes, S.E.D. & Siry, C. (2019). Science notebooks as interactional spaces in a multilingual classroom: not just ideas on paper. *Journal of Research in Science Teaching*, 57(7), 999-1027. https://doi.org/10.1002/tea.21615.

Wode, Henning (1995). *Lernen in der Fremdsprache. Grundzüge von Immersion und bilingualem Unterricht* [Foreign language learning. Basic features of immersion and bilingual education]. Ismaning: Hueber (Forum Sprache).

Yore, L.D., & D.F. Treagust (2006). Current Realities and Future Possibilities: Language and Science Literacy - Empowering Research and Informing Instruction. *International Journal of Science Education*, 28 (2-3), 291–314.

https://doi.org/10.1080/09500690500336973.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An Overview. *Theory into Practice*, 41, 64–70. https://doi.org/10.1207/s15430421tip4102_2.

Electronic Supplemental Materials

The Trade-Off Between STEM Knowledge Acquisition and Language Learning in Short-Scale CLIL Implementations

Tamara Roth[1]*, Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

Correspondence: tamara.roth@uni-bayreuth.de; ORCID ID: 0000-0002-9062-1489

Table of Contents
(Total pages 3)

Electronic Supplemental Material 1: Laboratory Manual Questions

DNA-ISOLATION

FROM ORAL MUCOSAL CELLS

Materials:

- drinking cup
- distilled water
- 2 Pasteur pipettes (sterilized)
- snap-cap vial (Schnapdeckelglas)
- chronometer
- micropipettes
- graduated pipette (iced; 10ml)
- pipette pump (Pipettenheber)
- water bath (122°F)
- black placemat
- waterproof pens

Chemicals:

- lysis buffer (white) (Lysispuffer)
- mild detergent (green) (Feinwaschmittel)
- chilled Isopropyl alcohol (P)

Experimental procedure:

1) Briefly rinse your mouth twice with **distilled water (destilliertes Wasser)** from the drinking cup.

2) Then chew for 2 minutes (do not swallow!) so that **saliva (Speichel)** containing **ablated oral mucosal cells (abgetragene Mundschleimhautzellen)** forms.

3) Fill the **Pasteur pipette** with distilled water from the drinking cup up to the mark in front of the suction head (Saugkopf). Release the water into your mouth and **intensively swill (spülen) it around for 40 seconds** (as if rinsing with water after brushing your teeth).

4) Afterwards, the 'rinse water' should be spat **into** the **snap-cap vial**. Make sure you **label** your vial with your **group number** to avoid confusion.

5) Use a micropipette to add **2000 µl lysis buffer (white)** to the vial's contents. Then put on the snap-cap and carefully (!) tilt (kippen) it back and forth **five times** to mix it with the saliva.

Explain the role of the lysis buffer:

____

____

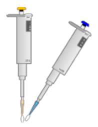
6) Use a second Pasteur pipette to add **5 drops of mild detergent (green)** to the solution in the snap-cap. Close the cap and carefully tilt it back and forth **once again**.

Explain the role of the mild detergent:

____

____

Electronic Supplemental Material 2: Laboratory Manual Visuals

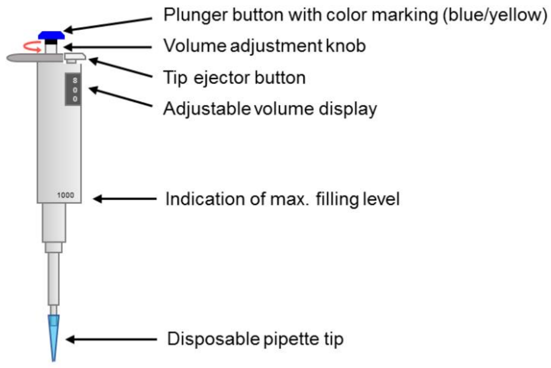


WORKING WITH MICROPIPETTES

Using **micropipettes** is one of the most important working techniques in gene technology labs. The handling of such instruments is necessary for several experimentation steps today. Named after its manufacturer, micropipettes are also called 'Eppendorf pipettes'.

With these special laboratory instruments, it is possible to pipette very low volumes of 2 µl to 1000 µl, depending on the selected micropipette and the previously adjusted volume.

Elements of a micropipette:

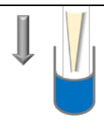
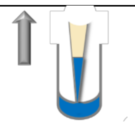
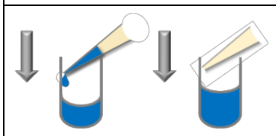


Please note:

Micropipettes are very expensive, sensitive instruments and must be handled carefully!

- **Micropipettes may only be used with a pipette tip attached!**
- **Always keep the tip of the pipettes downwards!**
- **Never use the same pipette tip for different substances!**

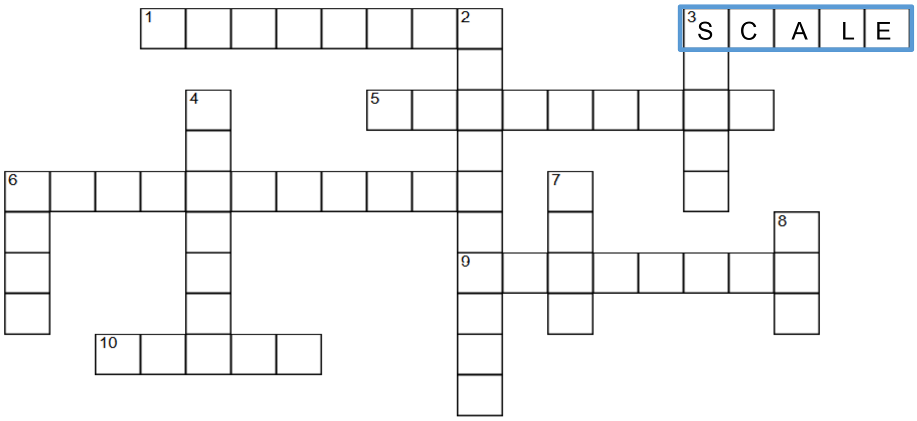
How to handle a micropipette:

1		Turn the ***volume adjustment knob*** **(Volumenregler)** to select the desired amount of liquid. The volume is given in µl (said **'microliter'**) and read from top to bottom. *e.g., 81,2 µl are indicated in the picture on the left* Now put on a suitable ***pipette tip*** **(Pippetenspitze)**.
2		Press the plunger button (Druckknopf) down to the ***first pressure point*** **(erster Druckpunkt)** and hold, then immerse (eintauchen) the tip 0.5 cm in the liquid.
3		Slowly release (loslassen) the plunger button. The medium is sucked in (einsaugen).
4		To eject (auswefen) the medium from the pipette tip, press the plunger button down to the first pressure point again, then down to the ***second pressure point*** **(zweiter Druckpunkt)** so that the tip is completely empty. Always place droplets on the vessel wall (Gefäßwand).

Electronic Supplemental Material 3: Language Exercise Book

DNA GEL ELECTROPHORESIS

Find the matching words to the definitions and put them in the crossword!



ACROSS

1 a small piece of equipment made of two narrow strips of metal joined at one end. It is used to pull out hairs or to pick up small objects by pressing the two strips of metal together with the fingers

3 a set of numbers, amounts, etc., used to measure or compare the level of something (example)

5 a place for storing liquid, especially a natural or artificial lake providing water for a city

6 to change from being a liquid or gas to a solid form, or to make something do this

9 to make something less pure

10 a special container that keeps drinks hot or cold

DOWN

2 a liquid in which small pieces of solid are contained, but not dissolved

3 to move quickly with a twisting, circular movement, or to make something do this

4 a cooking utensil with a wide, flat blade that is not sharp, used especially for lifting food out of pans

6 a flat piece of plastic, wood, or metal with a thin row of long, narrow parts along one side, used to tidy and arrange your hair

7 a hollow container into which you pour a soft or liquid substance

8 to change the colour of something using a special liquid

TWEEZERS	SUSPENSION
DYE	COMB
SPATULA	SCALE
CONTAMINATE	SWIRL
RESERVOIR	FLASK
SOLIDIFY	MOLD

5.6 Teilarbeit D

Interactive Learning Environments, Submitted

Content and Language Integrated Scientific Modelling: A Novel Approach to Model Learning

Tamara Roth[1*], Franz-Josef Scharfenberg[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

The relevance of English language competencies in authentic, discipline-specific contexts at school is increasingly acknowledged outside of English-speaking countries. Since any understanding of complex scientific problems requires the combination of scientific literacy with other competencies (such as scientific modeling), the appropriate application of Content and Language Integrated Learning (CLIL) is of great importance. The present study focuses on an established, hands-on outreach genetic education module on DNA structure, which it extends with a bilingual adaption to examine the influence of non-CLIL and CLIL learning on students' scientific modeling skills and model understanding. When comparing non-CLIL learners (n = 149) and CLIL learners (n = 316), the former received higher scores in the assessment of model-related self-evaluation sheets and built better models. We also found that non-CLIL learners achieved better temporary model-related knowledge scores and, for model evaluation, were more reflective in determining similarities and differences between their hand-crafted model and a commercial DNA school model. However, CLIL learners performed better in comparing their model sketches with their hand-crafted models. They also used different approaches to develop models and conceptualize integral components of models, as reflected in their advanced model understanding. That is, CLIL learners no longer perceive models as exact replicas of the original but as visualizations of scientific evidence that may change in response to new findings. We conclude that CLIL influences modeling qualities on different levels, by fostering model competencies and, in particular, model understanding.

Keywords: CLIL learning. Classroom modeling. Evaluating models. Changing nature of models. Gene technology outreach learning. Science education.

Introduction

English has become the internationally acknowledged lingua franca of science, used to communicate scientific developments and discourses. To prepare for life after school and to partake in the global science community, students need to increase their English language competencies beyond levels commonly achieved in school language lessons (Rodenhauser & Preisfeld 2014). In response to a lack of English language proficiency, the European Commission (2004) has emphasized the need for Content and Language Integrated Learning (CLIL) to prepare young students for the demands of a globalized society. At German schools, CLIL subject teaching is increasingly applied in science education. Thereby, students are encouraged to understand scientific content in English at the same time as achieving an appropriate level of scientific literacy (Lemke 1990, Klieme et al. 2010).

Definitions of scientific literacy commonly fit into two complementary groups: (1) understanding scientific content and working scientifically, enabling one's development as a future scientist, and (2) understanding science and using scientific information to make informed decisions as a conscientious citizen (Ke et al. 2020). Scientific literacy requires knowledge of relevant vocabulary, which can be used to explain and argue ideas (Norris & Phillips 2003), and an understanding of scientific practices that enable and encourage in-depth analyses of observed phenomena (Passmore & Svoboda 2012; Mendonça & Justi 2013). One such procedure is scientific modeling (Ke et al., 2020), which helps break down complex phenomena in the search for possible explanations and encourages scientific discourse (Lee et al. 2015).

Modeling has received considerable attention in science education classrooms. As "an explanatory system that represents objects or phenomena via discourse, writing, behavior, and drawing" (Lee et al. 2015, p.234), a model engages – and, potentially, fosters – various scientific competencies. This, however, only hold true "if students understand the nature and the purpose of models in science, as well as comprehend how models are constructed" (Sins et al. 2009, p.1206). The fact that their use encourages students to actively discuss scientific findings and ideas makes models a particularly important tool for achieving scientific literacy. In combination with CLIL learning, scientific modeling may advance many competencies required to be successful in an international academic environment (Grandinetti et al. 2013), such as scientific literacy in more than one language, in-depth understanding of scientific phenomena, and the capability to effectively communicate this understanding (Passmore & Svoboda 2012, Mendonça & Justi 2013, Ke et al. 2020).

To explore the effect of CLIL learning on model-related content knowledge, model understanding, and the ability of students to develop and evaluate a model, we piloted a CLIL genetics outreach education module structurally identical to a previously conducted non-CLIL outreach learning module (Roth et al. 2020). The module lasted for one day, which authentically reproduces the overall time that teachers in the classroom have to foster model understanding – irrespective of CLIL or non-CLIL (Bingimlas 2009). Focusing on the recently hypothesized relationship between scientific modeling and scientific literacy (Ke et al. 2020), we particularly explore the influence of this relationship on model understanding for non-CLIL and CLIL learners. Specifically, we examine the changing nature of models (CNM) as a subscale of model competency (Treagust et al. 2002). While the model-of-approach of our module only covers the most basic level of model learning, even at this level, students often do not display model understanding. Prior studies by, for instance, Treagust et al. (2002) and Krell et al. (2012), indicated that many students understand the multiplicity of models but still perceive them as an exact replica of the scientific phenomenon. Such weak epistemological understanding may – if not improved – prevent students from developing feasible scientific models and related content knowledge or understanding (Schwarz & White 2005, Sins et al. 2009). We also investigate model-related knowledge scores, yet do so with the understanding that both modeling and language learning require a high degree of mental capacity, which could result in a mental trade-off (Grandinetti et al. 2013). However, the increasing adoption of CLIL outside of English-speaking countries requires the examination of potential positive and negative outcomes of this combination. The same may be true for the ability to develop and evaluate models. To explore these possible mutual influences, we first outline the relevant theoretical background, including the basic principles of CLIL learning and model learning, and the connection between model learning and scientific language. This information is important in that it will enable readers to better understand the design and procedure of our study.

Theoretical Background

CLIL Learning

"Reading, writing, and oral communication are critical literacy practices for participation in a global society. In the context of scientific inquiry, literacy practices support learners by enabling them to grapple with ideas, share their thoughts, enrich understanding and solve problems" (Krajcik and Sutherland 2010, p. 456). The central importance of written and oral language competencies makes scientific literacy in English indispensable for those aiming to understand international scientific discourse and actively contribute to society via engagements with science (Yore & Treagust 2006). This means that efforts to improve scientific literacy in English need to be increased at school. European language policies support this endeavor by encouraging English language learning in regular school curricula outside of English language learning classes via CLIL (European Commission 2012). Used effectively, CLIL combines language learning and content learning so that neither proficiency in English nor the respective subject area is a prerequisite for successful learning (European Commission 2004, Stoddart et al. 2002).

CLIL "encompasses any activity in which a foreign language is used as a tool in the learning of a non-language subject in which both language and subject have a joint role" (Marsh 2002, p. 58). Combining these predominantly separate aspects of learning contrasts with traditional instructional approaches, which see language proficiency as an essential requirement of effective content learning (Cummins 1981, Stoddart et al. 2002). Yet, although CLIL learning, by definition, requires content and language to be equally integrated into lesson planning, the scientific language often acts only as a mediator (Rodenhauser & Preisfeld 2014). Few acknowledge the intricate connection between language and content (Meskill & Oliveira 2019). The one-dimensional view of language and content learning fails to acknowledge that content and its contextualization induce an act of language, while language gives meaning to content. Such contextualization gives relevance to the act of language, leading to an overall better understanding of the relationship between science learning, students' experiences, and language use (Krashen 2013, Tolbert et al. 2019). Authentic settings, wherein "visual cues, concrete objects, and hands-on activities" (Stoddart et al. 2002, p. 666) can, thus, provide the necessary basis for natural language development (Meskill & Oliveira 2019).

Accordingly, enriching such settings with CLIL encourages scientific literacy in both the native language and English (via the regular practice) in scientific contexts (Gonzalez-Howard & McNeill 2016, Tolbert et al. 2019). The exploration of scientific phenomena helps students process scientific content in a coherent and accessible manner, and the ensuing academic scientific communication and discourse encourage critical reflection (Nation & Newton 2008). The engagement in and discussion of scientific issues is commonly regarded as a primary driver of success in mastering scientific literacy (Gonzalez-Howard & McNeill 2016). Examples from practice (Lam et al. 2012) confirm that "the creation of immersive environments where students participate in discourse patterns that mirror those seen in the scientific community" has a positive effect on scientific literacy and content learning (Quarderer & McDermott 2020, p. 3). Such environments are often created through inquiry learning, which combines hands-on activities with experiments and the exploration of a scientific phenomenon (Hampton & Rodriguez 2001). Thereby, students engage in meaningful scientific practice, such as describing observed phenomena, formulating hypotheses, assessing results, and reflecting on findings, which demands the use of a discipline-specific register (Stoddart et al. 2002). To foster such practice in science classrooms, Tolbert et al. (2019) suggested four dimensions of instruction, including contextualization, scientific reasoning, scientific discourse and scientific literacy. This way, language is inherently contextualized – no longer a mere mediator - and helps students to structure and communicate acquired information (Lemke 1990, Hampton & Rodriguez 2001).

Since CLIL learning requires considerable mental capabilities to manage content and language teaching simultaneously, effective scaffolding is necessary to decrease the cognitive load (Grandinetti et al. 2013, Tolbert et al. 2019). Scaffolding is "a type of teacher assistance that helps students learn new skills, concepts, or levels of comprehension of material that leads to the student successfully completing a specific learning activity with finite goals" (Maybin et al. 1992, p.188). Such scaffolding activities are particularly important for the discipline-specific register of the foreign language, as the "language provides learners with meaningful cues that help them interpret the content being communicated" (Stoddart et al. 2002, p.666). Gottlieb (2016) describes four possible scaffolding dimensions: linguistic, graphic, sensory, and interactive. In science subjects such as biology, graphic or visual scaffolding involving diagrams, graphs, or charts is a commonly used tool (Buxton et al., 2019). Other forms of scaffolding that may come in handy for CLIL science teaching include linguistic scaffolding, which can provide definitions of key terms, language frames, or bridging and prompting to foster scientific literacy (Unsal et al. 2018, France 2019).

In Germany, CLIL learning is realized in either bilingual strands or singular CLIL modules. Whereas singular CLIL modules have varying timeframes and may be used occasionally, bilingual strands involve regular CLIL lessons. Regardless of time and frequency, any meaningful engagement with language and content positively influences English scientific literacy (Rodenhauser & Preisfeld 2014, Buse & Preisfeld 2018).

Model Learning

One increasingly popular inquiry method for fostering scientific literacy is scientific modeling (Akerson 2009, Ke et al. 2020). Models are commonly used to visualize or explain phenomena (Krajcik & Merritt 2012). The process skills of scientific modeling include revealing underlying mechanisms, showing causal links, raising questions, and testing multiple hypotheses (Akerson 2009). Whenever a prediction proves incorrect or new evidence emerges, a model can be adapted and refined accordingly (Passmore et al. 2009).

This adaptability of models to different theoretical perspectives is described as the changing nature of models (CNM) in the SUMS questionnaire (Students' Understanding of Models in Science), which was designed to gain insights into students' understanding of models (Treagust et al. 2002). It is an essential aspect of modeling practice since most students believe models to be exact replicas (EM) "of reality that embody different spatio-temporal perspectives [in contrast to] constructed representations that may embody different theoretical perspectives" (Grosslight et al. 1991, p. 799). A study by Krell et al. (2012) confirmed these findings. Treagust et al. (2002) initially hypothesized that the analysis of more abstract concepts would automatically lead to a more abstract perception of scientific models, but their results indicated that this does not hold true. Other studies, such as Gobert et al. (2011), observed a subject-dependency of model understanding and suspected it to be "based on the models to which they have been exposed within these respective domains (p.673). Also, Schwarz et al. (2002) alongside Sins et al. (2009) and Louca and Zacharia (2012) questioned the development of epistemic knowledge about models and decontextualized understanding of models. In classroom teaching, scientific models often lack contextualization and are only introduced superficially. Thus, students do not usually understand the value of scientific models for explaining phenomena, appreciate the adaptability of models to emerging evidence, or recognize differences between models and explained phenomena (Krajcik & Merritt 2012, Ke et al. 2020).

Models are regarded as the cornerstones of every scientific discipline, with their combination of theory with modeled processes and functions making models invaluable for student learning (Passmore et al. 2009). Natural phenomena are often difficult to observe, yet models can provide an authentic alternative experience, are usually easier to understand, and do not require exhaustive preparatory work (Carpenter et al. 2019). The benefits of models are apparent in, for example, the finding that students gain better content knowledge scores when applying three-dimensional DNA models rather than two-dimensional DNA models (Rotbain et al. 2006, Saka et al. 2006). The benefits of models can, however, only be leveraged if students understand the nature of models, their inherent function, as well as the process of developing them (Gilbert & Justi 2002, Sins et al. 2009).

The positive relationship between modeling and content knowledge may lie in the application of in-depth strategies inherent to deep learning, such as cognitive reasoning through metacognition (Sins et al. 2009, Mendonça & Justi 2013). Reflective thinking to evaluate possible learning outcomes and the generation of explanations for observed phenomena are particularly salient for deep learning strategies (Lee et al. 2015). Such strategies require communicative action between students to convey their thoughts and build valid arguments that explain the observed phenomenon (Passmore & Svoboda 2012, Ke et al. 2020). The use of modeling thus provides students with an experience that mirrors those of scientists, who also have to evaluate, modify, and argue in support of their models when presenting them to their peers (Mendonça & Justi 2013, Ke & Schwarz 2020). Students, thereby, "do not just describe empirical experiences [when they engage in modeling in science lessons] as they would with experiments and observations; they can also reason, explain, and communicate phenomena or systems using empirical experiences as evidence" (Lee et al. 2015, p. 236).

Objectives of the Study

The present study focuses on the following research questions:

1. How does a one-day CLIL science module influence students' knowledge of DNA as a model throughout the hands-on laboratory?
2. How does CLIL influence students' general understanding of models and modelling as compared to the monolingual German treatment group?
3. How does CLIL influence students' ability to build meaningful models as compared to the monolingual German treatment group?
4. How does CLIL influence students' ability to evaluate models – specifically, the two implemented evaluation phases (evaluation-1 and evaluation-2) – as compared to the monolingual German treatment group?

Thus, the specific objectives were four-fold:

- to assess students' content knowledge regarding DNA as a model throughout the laboratory
- to identify potential differences between monolingual German and CLIL treatment groups by comparing their general understanding of models
- to identify potential differences between monolingual German and CLIL treatment groups by comparing their ability to build meaningful models
- to identify potential differences between monolingual German and CLIL treatment groups by comparing their ability to evaluate models in model evaluation-1 and model evaluation-2

Materials and Methods

Intervention Phases

Our study integrated CLIL into our gene-technology module's design but otherwise followed the format of our previous interventions (Mierdel & Bogner 2019, Roth et al. 2020) to maintain overall treatment comparability. Specifically, we retained our four intervention phases, but extended the module’s time by three hours to enable the processing of language learning tasks. For each of the intervention phases, we provided linguistic and graphic scaffolding material in a separate exercise book and laboratory manual throughout our laboratory to reduce difficulties in understanding (Table 1). The following section is adapted from Roth et al. (2020).

TABLE 1 Quasi-experimental intervention design (adapted from Roth et al. 2020).

Intervention phases	Evaluation variants	
	Monolingual (non-CLIL) (*n* = 145)	Bilingual (CLIL) (*n* = 107)
Instruction and conversation language English	-	+
Vocabulary scaffolding material for all intervention phases	-	+
Pre-lab (~1.5 hours)	+	+
DNA-related theoretical and experimental phases (~3 hours)		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases (~2 hours)		
Mental modeling: text analysis	+	+
DNA modeling with craft materials	+	+
Model evaluation-1: Drawing a paper & pencil version of the crafted model and answering questions	+	+
Model evaluation-2: Comparing the crafted model with a scientific demonstration model	+	+
Interpretation (~0.5 hours)	+	+

Our one-day, hands-on bilingual module offered inquiry-based learning activities focused on the structure of DNA and bilingual CLIL learning adapted to the capabilities of ninth graders. Students worked in pairs to complete their tasks, with technical guidance provided in a laboratory manual and language scaffolding exercises to foster scientific vocabulary. The module's content is in line with the state's syllabus and follows national competency requirements (KMK 2005). While CLIL may add another layer of complication to the already demanding goal of developing students' modeling competence, such practice is very authentic with the increasing adoption of CLIL outside of English-speaking countries. We conducted two versions of the intervention, which differed only in their language: monolingual German for the non-CLIL and bilingual English for the CLIL variant (Table 1). Usually, only quasi-experimental designs are feasible for use with students in intact class groups (Cook & Campell 1979). Thus, we randomly assigned classes to the respective language variants and treated non-CLIL and CLIL treatment groups as an independent variable (Table 1).

Pre-lab Phase. Both treatment groups began with a pre-lab phase. Due to a lack of prior laboratory experience, the pre-lab phase was rather extensive. We introduced students to different laboratory techniques and theoretical concepts connected to experimentation (e.g., Sarmouk et al. 2019). That is, students learned how to handle laboratory equipment – such as micropipettes and centrifuges – necessary for experimentation. The teacher's role was limited to providing guidance and explaining relevant theoretical concepts via presentations and graphic illustrations on a poster. Practical concepts of experimentation were conveyed using demonstrations. Both theoretical and practical exercises were vital for conducting the subsequent experiments. This phase should also prevent students from feeling overwhelmed in response to the experimental situation (Mierdel & Bogner 2019).

DNA-Related Theoretical and Experimental Phases. We outlined the practical relevance of the subsequent experiments (DNA relevance, Table 1) and introduced the scientific concepts of DNA isolation and gel electrophoresis. We asked students to reactivate their knowledge of cells and to use this information for hypothesizing about the experiments' potential outcomes (Mierdel & Bogner 2019, Roth et al. 2020). Our decision to provide information in a series of theoretical phases ensured that students were not overwhelmed by the amount of information and we able to focus on the forthcoming experimental phases. This sequential structure was essential for explaining the difficult concept of gel electrophoresis, which involves chemical knowledge about the contents of DNA and requires abstract thinking to imagine the processes involved.

For the experimental phases, we used an evidence-based, two-step approach (Mierdel & Bogner 2019, Roth et al. 2020). That is, students first answered questions in their laboratory manuals and considered subsequent experimental procedures. Secondly, they worked in pairs to discuss each step before carrying out experiments that effectively combine hands-on and minds-on activities, requiring them to do more than simply follow instructions (Mierdel & Bogner 2019, Roth et al. 2020). For our hands-on approach, we provided sufficient scaffolding material in our laboratory manual to enable independent experimentation and self-reliant protocolling of their observations, as well as the materials necessary for modeling (Mierdel & Bogner 2019, Roth et al. 2020). The laboratory manual was designed and adapted over the course of four laboratory evaluations (Goldschmidt & Bogner 2016, Langheinrich & Bogner 2016, Mierdel & Bogner 2019, Roth et al. 2020)

Model-Related Phases. Both model-related phases directly followed and built on the experimental, DNA-related phases. As models are vital for visualizing and explaining phenomena in science and science education (Krajcik & Merritt 2012), we subdivided our model-related phases into a mental modeling phase involving text analysis and either the construction and discussion of a purely mental model or a rough sketch, a modeling phase involving craft materials, a model evaluation-1 phase, and a model evaluation-2 phase (Table 1).

We based our model-related phases on the four main stages of the Model of Modelling (Gilbert & Justi 2002, p. 370 ff.): (1) “experience[s] of the phenomenon being modeled”, (2) “forming a mental model”, (3) “decision … about the mode of representation in which it is to be expressed”, and (4) “testing … scope and limitations of the model”. Mental modeling was, thereby, the key to providing a theoretical basis for experimental findings. In our module, a text about the discovery of the DNA’s structure (Usher 2013) provided fundamental knowledge about the necessary components of the DNA. Students could then apply this knowledge to their simplified mental models, which either remained imaginary or were roughly sketched on paper (e.g., Mierdel & Bogner 2019, Franco & Colinvaux 2000). Although model learning requires primarily the testing of hypotheses or the explanation of phenomena, such practice often goes beyond classroom capabilities (Bingimlas 2009). Our ‘model-of-approach’, which focused on conveying CNM instead of EM, may not be popular (Gouvea & Passmore 2017), but followed the curriculum and laid the foundation for ‘model-for-approaches’ by partially developing a DNA model from experimentation with DNA and Crick’s letter to his son about the discovery of DNA as well as students’ critical reflection on their model (ISB 2019).

That is, after determining the DNA’s representation and building the hand-crafted model from the text, we conducted a model evaluation-1 phase as a reciprocal self-evaluation. As combining sketching and hand-crafting models proved to be important (e.g., Prabha 2016), students evaluated their hand-crafted DNA model based on a detailed paper-and-pencil version. Another effective method for encouraging self-evaluation is reflective writing (Kovanović et al. 2018). Open-ended questions about model-related components encourage students to rethink and reassess certain steps and decisions in developing the mental model into its physical counterpart (Roth et al. 2020). In evaluation-2 phase, students assessed their hand-crafted DNA models using a comparison-based self-evaluation with a commercially available DNA demonstration model.

Interpretation Phase. Here, students compared their hypothesis about the outcome of their experiments with images of the gel electrophoresis. They also discussed their models in class and compared these to the molecular DNA model by Watson & Crick.

Language Scaffolding

To support content learning in a foreign language while fostering scientific literacy and language learning, we designed a language scaffolding exercise book. The book contained language-specific riddles, such as word search puzzles and crossword puzzles, allocating words to English definitions. Students were also asked to match the words and their definitions with German translations or to translate the words into German. At all times, students were allowed to use English-English and English-German dictionaries available in the laboratory. For each phase of the intervention, we had one specific language scaffolding exercise (Electronic Supplemental Material [ESM] 1).

We also included short translations or explanations of scientifically essential vocabulary in the students' laboratory manuals and added German words for laboratory equipment to our presentation. To help students understand the complicated procedure of gel electrophoresis, we provided a poster with English descriptions and questions. Still, we used German answer tags and visuals to make the different process steps easier to follow. Oral explanations were provided in English.

Participants

Altogether, 465 ninth graders (higher secondary school) participated in our study (girls 50.8%, boys 49.2%; $M_{\text{class size}} = 21.9$, $SD = 4.6$; $M_{\text{age}} = 14.7$, $SD = 0.7$). Seven classes took part in our non-CLIL intervention ($n = 149$) and 14 in our CLIL intervention ($n = 316$). Non-CLIL students teamed up in 69 two-person groups (and three three-person groups, two students working individually due to illness) and CLIL students in 151 two-person groups (and four three-person groups, two students working individually due to illness).

Participation was voluntary. Written parental consent was given before students participated in our study, although the data collection was pseudo-anonymous, and students could not be identified. The study was designed in accordance with the Declaration of Helsinki (2013), and the state ministry approved the questionnaires used.

Dependent variables

As dependent variables, we examined students' model-related content knowledge scores and model understanding in a repeated measurement design: a pre-test (T0) two weeks before the intervention, a post-test (T1) directly after the module, and a retention test six weeks later (T2). We examined students' sketches and their responses to the open questions as part of the evaluation-1 phase, as well as students' models and their self-evaluation sheets as part of the evaluation-2 phase. Throughout the entire intervention, students were unaware of any testing.

Students' Content Knowledge regarding the structure of DNA. We applied an ad-hoc knowledge test comprising eighteen multiple-choice questions about content knowledge related to the different model phases regarding the structure of DNA. Students had to select one correct answer out of four possible answers. To avoid response patterns, we randomly rearranged questions and possible answers for each testing time. Item examples can be found in Table 2 (complete questionnaire in ESM 2). Content validity was a given as the items were consistent with the state syllabus. Likewise, construct validity was confirmed based on the items' heterogeneity in relation to complex constructs, such as building up content knowledge and understanding of scientific concepts (Rost 2004). Cronbach's alpha values of .70 (T0), .71 (T1), and .69 (T2) indicate acceptable internal consistency. According to Lienert & Raatz (1998), values between .50 and .70 allow for the differentiation of groups.

Students' Model Understanding. Following Mierdel & Bogner (2019), we deployed a shortened version of the SUMS questionnaire (Students' Understanding of Models, Treagust et al. 2002) focusing on the subscales of models as *exact replicas* (ER) and the *changing nature of models* (CNM). Students had to answer general, decontextualized questions about models (Krell et al. 2014) using a five-point Likert-Scale ranging from strongly disagree (1) to strongly agree (5). Item examples can be found in Table 2 (complete questionnaire in ESM 3). ER and CNM best address the two dimensions that form the focus of our modeling phases. As ER is a conceptual error, low scores are desirable. CNM, on the other hand, shows true model understanding, meaning that high scores are considered beneficial (Treagust et al. 2002). Cronbach's alpha values above .7 (T0 = .81, T1 = .76, T2 = .70) indicate acceptable internal consistency (Lienert & Raatz 1998). The SUMS questionnaire uses a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

TABLE 2 Item examples to model-related content knowledge and model understanding.

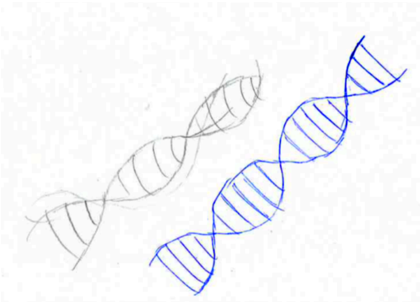
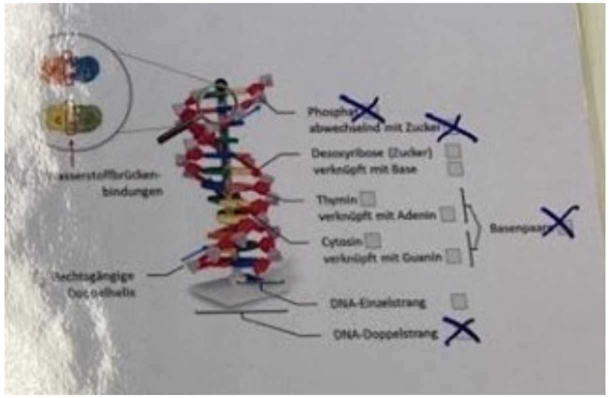
Item relation	Item example
Model-related knowledge[a]	The proportion of sugar to phosphate in the DNA-molecule is … (a) 2:1 (b) 1:2 (c) 3:1 (d) 1:1* The DNA bases are located … (a) at the outside of the DNA molecule and bound to phosphate. (b) at the inside of the DNA molecule and bound to phosphate. (c) at the inside of the DNA molecule bound to sugar.* (d) at the outside of the DNA molecule and bound to sugar.
Model understanding[b]	Models can change if new theories or evidence prove otherwise. (1) strongly disagree (2) disagree (3) neither agree nor disagree (4) agree (5) strongly agree A model has to be close to the real thing. (1) strongly disagree (2) disagree (3) neither agree nor disagree (4) agree (5) strongly agree

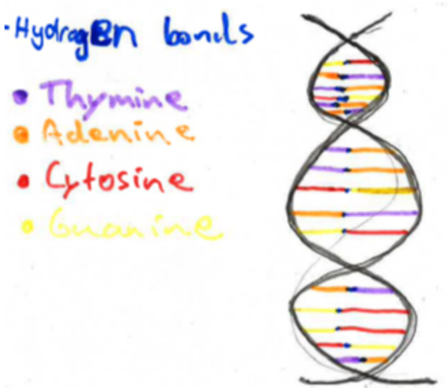
Notes: [a]Sum 18 items; [b]Sum 14 items; correct answers are marked with an asterisk *

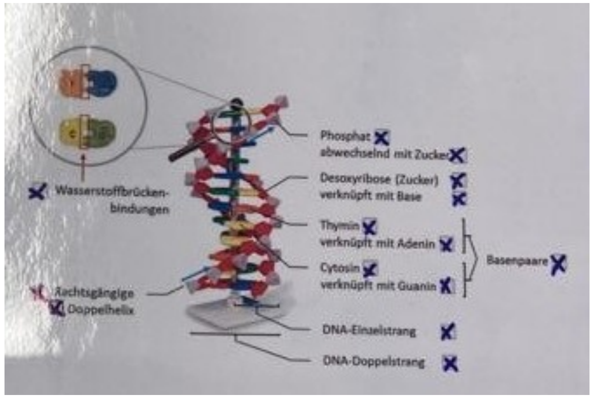
Evaluation-1 Phase. To compare the effects of both treatment approaches on the evaluation-1 phase, we assessed students' model sketches (changed after Langheinrich & Bogner 2015; for definitions, examples, and frequencies, see ESM 3). We then randomly selected 38 out of 231 drawings for a second scoring (16.4%). Cohen's kappa coefficient (Cohen 1968) scores of 0.90 and 0.85 for intra-rater and inter-rater reliability indicated an "almost perfect" rating (Wolf 1997, p. 964).

Using content analysis (Bos & Tarnai 1999), we iteratively categorized the statements made by students in response to the open questions. For the first question, 'Which features of the original DNA molecule are simplified in your model', we assigned each answer to one of four categories: level of DNA, level of substance, level of particles, and level of structure (for definitions, examples, and frequencies, see ESM 4). We then randomly selected 76 out of 474 statements for a second scoring (16.0%). We computed Cohen's kappa coefficient (Cohen 1968) scores of .97 and .84 for intra-rater and inter-rater reliability, which indicated an "almost perfect" rating (Wolf 1997, p. 964). For the second question, 'Explain why one might create different models of one biological original (in our case, the structure of the DNA)?', we applied the adapted category system developed by Langheinrich & Bogner (2015) and Mierdel & Bogner (2019) and identified five categories: individuality of DNA, different interpretation, different model design, different focus, and different research state (for definitions, examples, and frequencies, see ESM 5). We then randomly selected 61 out of 416 statements for a second scoring (14.7%). We computed Cohen's kappa coefficient (Cohen, 1968) scores of .89 and .72 for intra-rater and inter-rater reliability, which showed a "substantial" to "almost perfect" rating (Wolf 1997, p. 964).

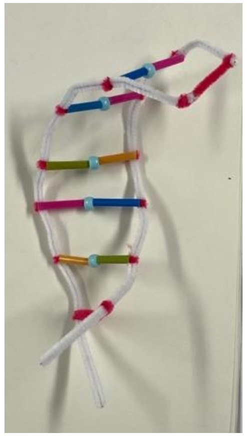
TABLE 3 Assessment of students' evaluation phases.

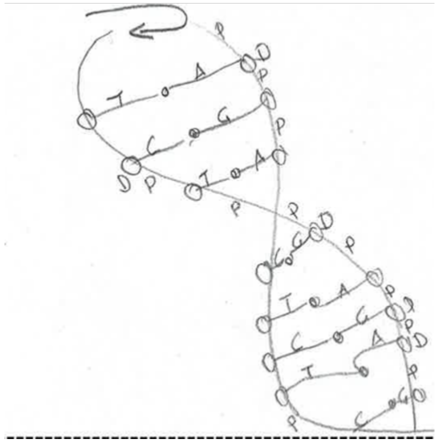
Phase			Assessment (model[a] / sketches[b] / self-evaluation sheet[c])
Modelling		Model evaluation	
Hand-crafted model	Evaluation-1: sketches	Evaluation-2: comparing and ticking the self-evaluation sheet	
			3[a] / 6[b] / 2[c]

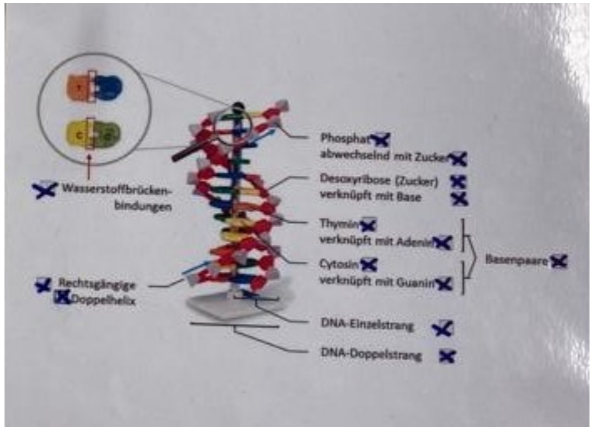




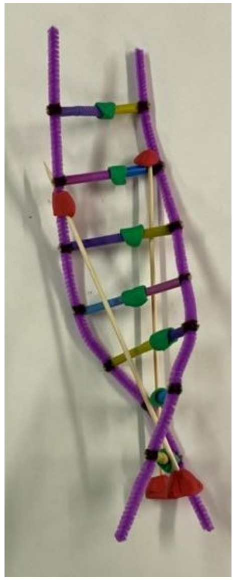
8[a]/ 11[b] / 8[c]

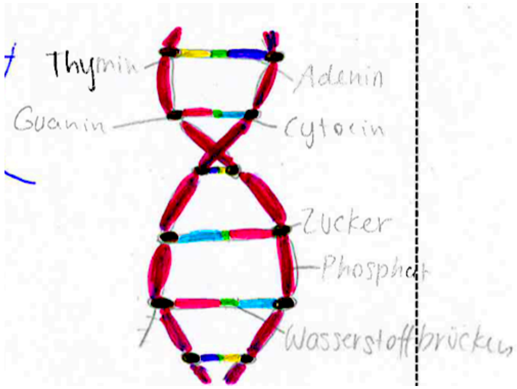


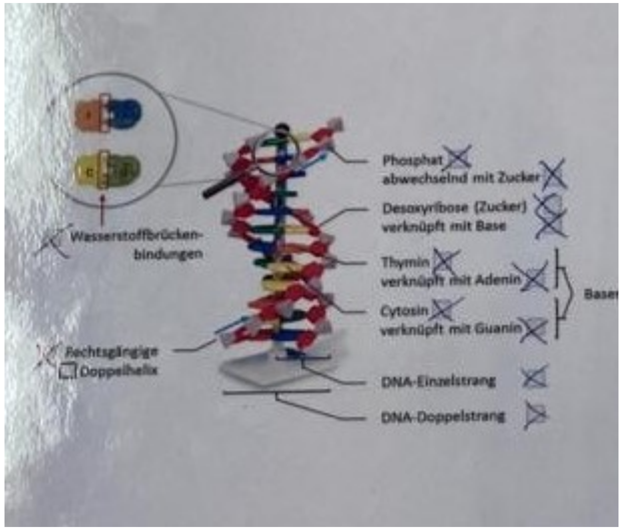




11[a] / 19[b] / 11[c]







13[a] / 18[b] /13[b]

Notes: [a]score maximal 14 points; [b]score maximal 19 points; [c]score maximal the model score, see [a].

Evaluation-2 Phase. A three-step approach was applied to evaluate evaluation-2 phase (Table 3).

- Documentation of the students' self-evaluation: We counted each box that students had ticked on their self-evaluation sheet as one point (maximal score 14 points, for description, see Table 3).
- Assessment of students' self-evaluation sheets: We analyzed the conformity of ticked boxes on self-evaluation sheets using the respective models. Where ticks matched the models, students received one point each. The maximum score could be attained by checking all boxes of the self-evaluation sheet with model features.
- Assessment of students' models: We independently assessed the models. Correct features received one point each, whether or not they had been identified (maximal score 14 points).

Both assessors randomly selected 30 self-evaluation sheets, from a total of 231, and 30 models, from a total of 231, for a second scoring (13.0%, either). Cohen's kappa coefficient (Cohen 1968) scores of 0.90 and 0.89 for intra-rater and of 0.77 and 0.79 for inter-rater reliability showed "substantial" to "almost perfect" ratings (Wolf 1997, p. 964).

Comparing documented boxes and assessing the self-evaluation sheets helped us determine the degree to which students were correctly evaluating the models. Lower assessment scores from students' self-evaluation sheets indicate that students had performed poorly in their evaluations of models, yet these scores also imply that students may have documented model features that were not given. High model scores may, in contrast, reveal that students did not correctly identify existent model features.

Comparison of Students' Evaluation Phases

To compare students' evaluation-1 and evaluation-2 phases, we scored their sketches, with a maximum possible score of 14 points, by adding scores of their self-evaluation sheets and our assessment of their models. When compared to our model scores, scores of the recoded sketches indicate the quality of their evaluation-1 phase, while scores of their self-evaluation sheets indicate the quality of their evaluation-2 phase.

Statistical Analysis

Statistical tests were conducted using SPSS Statistics 27. We applied nonparametric methods due to the non-normal distribution of variables (Kolmogorov-Smirnov test (Lilliefors modification): partially $p < .001$), and, consequently, use boxplots to illustrate our results. Intra-group differences over the three test dates were analyzed using the Friedman test (F) combined with a pairwise analysis from T0 to T1 and T2, and from T1 to T2, using the Wilcoxon (W) signed-rank test. Mann-Whitney U tests (MWU) were used to evaluate inter-group differences. To account for the use of multiple testing methods, we applied a Bonferroni correction and decreased the Alpha level to .017, respectively (Field 2012). In case of significant results, effect sizes *r* (Lipsey & Wilson 2001) were calculated with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. For contingency analyses, we calculated the adjusted Pearson's contingency coefficient (*C*; Pearson 1900). Since Pearson's *C* is a member of the *r* effect size family (e.g., Ellis 2010), we also treat it as an effect size score. We used a principal component analysis with subsequent varimax rotation, which reduces the data's dimensionality while retaining its variation (Bro & Smilde 2014), to evaluate the shortened SUMS scale. Following the Kaiser-Guttman-Criterion (Kaiser 1970), it divided into two factors: the Kaiser-Meyer-Olkin (KMO = .80, $\chi 2 = 599.2$) values being *acceptable* to *good*, indicating that conducting a factor analysis with our dataset was feasible.

Results

We first provide an overview of our intra-group and inter-group analyses with regard to model-related knowledge and model understanding. This overview is followed by a detailed assessment of the evaluation-1 and evaluation-2 phases.

Intra-group Analyses of Model Related Knowledge

Intragroup analysis revealed changes for both non-CLIL and CLIL learners for model-related knowledge scores (F: $\chi^2_{\text{non-CLIL/CLIL}}$ (2, n = 149/316) = 202.94/114.71, $p < .001$). With each approach, knowledge initially increased before dropping between T1 and T2, but not below levels of T0 (Figure 1; $W_{T0/T1}$: non-CLIL/CLIL Z = -10.07/-8.64; $p < .001$; $W_{T0/T2}$: non-CLIL/CLIL Z = -9.77/-7.43, $p < .001$; $W_{T1/T2}$: non-CLIL/CLIL Z = -8.06/-2.62, p < .001/ =

.009). This suggests that students gained short-term and mid-term model-related knowledge throughout the intervention.

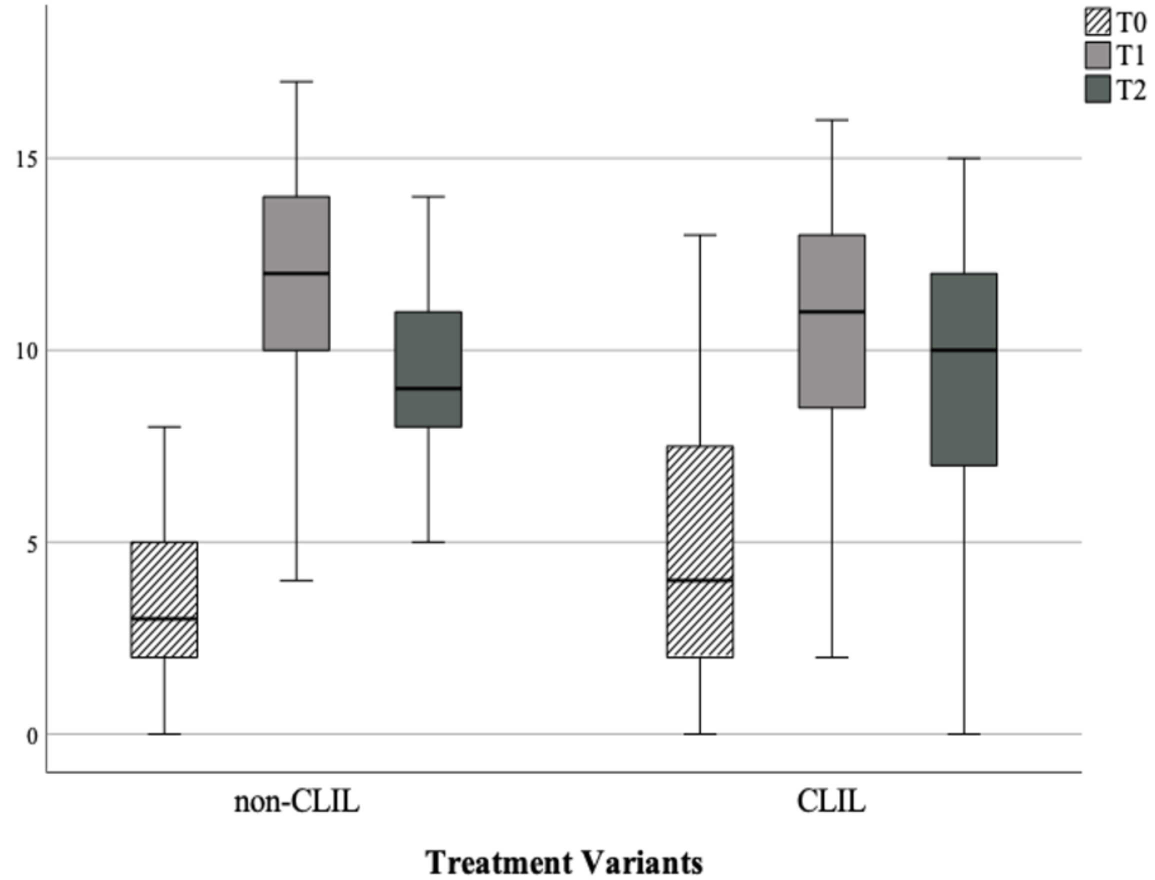


FIGURE 1 Differences between non-CLIL and CLIL learners for model-related knowledge acquisition at the three different testing times.

Inter-group Analyses of Model Related Knowledge

Non-CLIL and CLIL students started with similar model-related knowledge scores (T0, Figure 1; MWU: Z = -2.27; p = .023). After the intervention, with a small-to-medium effect, non-CLIL students scored higher than CLIL students in model related knowledge (T1, Figure 1; MWU: Z = -2.47; p = .013; r = .218). In the follow-up test, the model related knowledge scores of both treatment variants decreased to similar levels (T2, Figure 1; Z = -2.09; p = .036). Yet, participants in both groups increased their understanding of the DNA as a model based on the knowledge obtained throughout the different experimental steps and modeling activities.

Factor Analysis Model Understanding. Principal component analysis (PCA) of seven items from the SUMS questionnaire, involving varimax rotation, resulted in two factors based on eigenvalues >1.0. The Kaiser–Meyer–Olkin measure verified the sampling adequacy (KMO

= .796), which is well above the acceptable lower limit of .5 (Field 2012). Bartlett's test of sphericity ($\chi^2 = 599.224$, $p < .001$) indicated that correlations between items were sufficient for performing a PCA. Examination of the Kaiser–Guttman criterion yielded empirical justification for the retention of two factors, which explained 63.69% of the total variance.

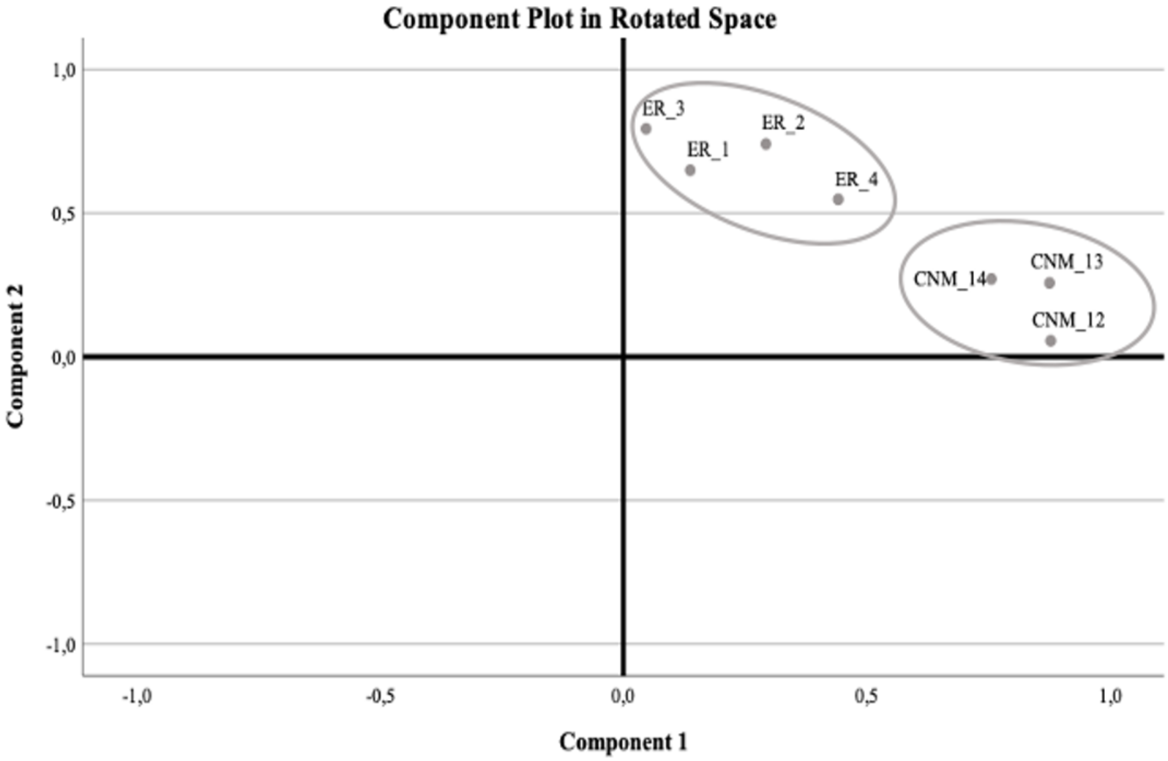


FIGURE 2 Component plot in rotated space after principal component analysis with clearly distinct subscales models as exact replicas (ER) and changing nature of models (CNM).

The component plot in rotated space (Figure 2) supported our two-factor solution and confirmed the original subscales. Factor loadings under .35 were suppressed. We could, however, observe cross-loadings for ER_4 (Table 4). The percentage of variance explained by models as exact replicas (ER) was 34.46%, and the variance explained by the changing nature of models was 29.23% (CNM).

TABLE 4 Factor loadings for *models as exact replicas* (ER) and *changing nature of models* (CNM) in the rotated component matrix.

		1	**2**
ER_1	A model should be an exact replica.	.650	
ER_2	A model needs to be close to the real thing.	.741	
ER_3	A model needs to be close to the real thing by being very exact so that nobody can disprove it.	.794	
ER_4	Everything about a model should be able to tell what it represents.	.548	.441
CNM_12	A model can change if there are new theories or evidence proves otherwise.		.878
CNM_13	A model can change if there are new findings.		.876
CNM_14	A model can change if there are changes in data or belief.		.756

Inter-group Analysis Model Understanding. To account for differences in students' understanding of models with regard to ER and CNM between non-CLIL and CLIL learners, we calculated mean scores for all testing times. Starting at similar levels (T0, MWU: $Z = -1.750$, $p = .80$), we found differences for T1 (MWU: $Z = -7.845$, $p < .001$, $r = .513$) and T2 (MWU: $Z = -8.339$, $p < .001$, $r = .543$) regarding CNM (Figure 3, left part) and ER (Figure 3, right part) between non-CLIL and CLIL learners with large effect sizes (Lipsey & Wilson 2001).

Intragroup analysis revealed changes for non-CLIL and CLIL learners for model understanding (F: $\chi^2_{non\text{-}CLIL/CLIL}$ (5, n = /149316) = 152.41/125.94, $p < .001$). Non-CLIL learners changed their notions of models from ER to CNM after participation (T1) but returned to ER after another six weeks (T2). Among CLIL learners, in contrast, perception of models as ER was reinforced by participation (T1) but changed to CNM after a six week reflection phase (T2) (Figure 3; $W_{T0/T1}$: non-CLIL/CLIL ER $Z = -4.74/-4.71$; $p < .001$; $W_{T1/T2}$: non-CLIL/CLIL ER $Z = -7.90/-7.06$, $p < .001$; $W_{T0/T2}$: non-CLIL/CLIL ER $Z = -2.96/-4.54$, $p = .003/$ $p < .001$; $W_{T0/T1}$: non-CLIL/CLIL CNM $Z = -2.42/-1.65$; $p = .015/$ p = .098; $W_{T1/T2}$: non-CLIL/CLIL CNM $Z = -6.60/-5.39$, $p < .001$; W_{T0T2}: CLIL/CLIL CNM $Z = 2.91/-4.46$, $p = .004/$ p < .001).

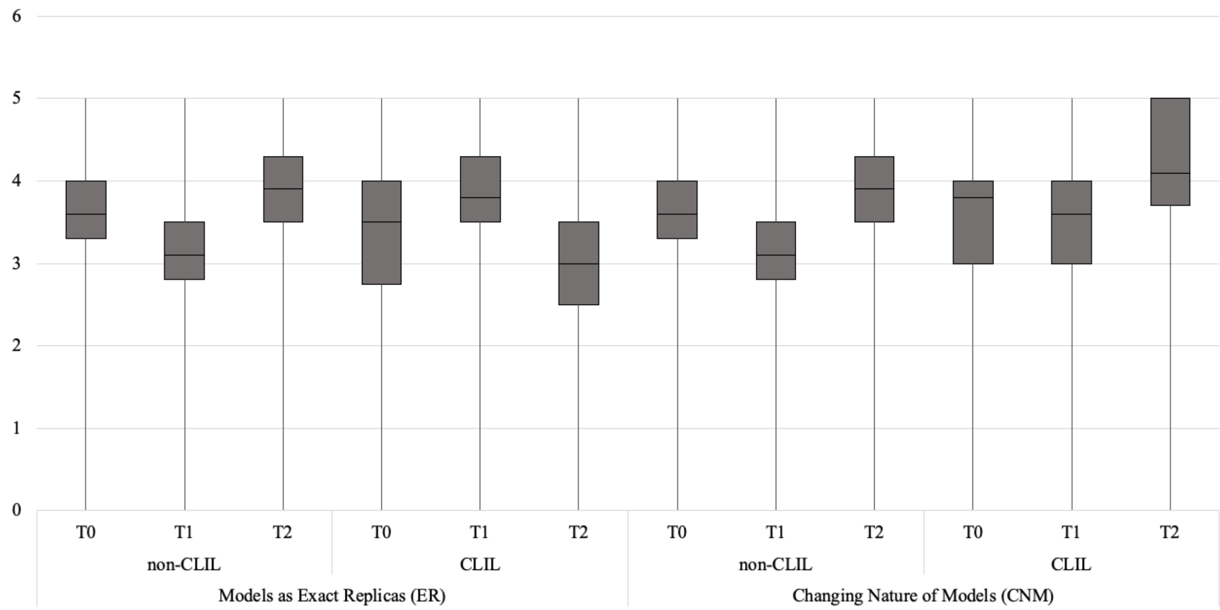


FIGURE 3 Differences between non-CLIL and CLIL learners in 'models as exact replicas' and 'changing nature of models' across testing points.

Assessment of Evaluation-1 Phase

We compared sketches of both language variants and identified significantly better results with a medium effect size (MWU: $Z = -6.574$, $p < .001$, $r = 0.304$; for details, see ESM2) for the non-CLIL approach.

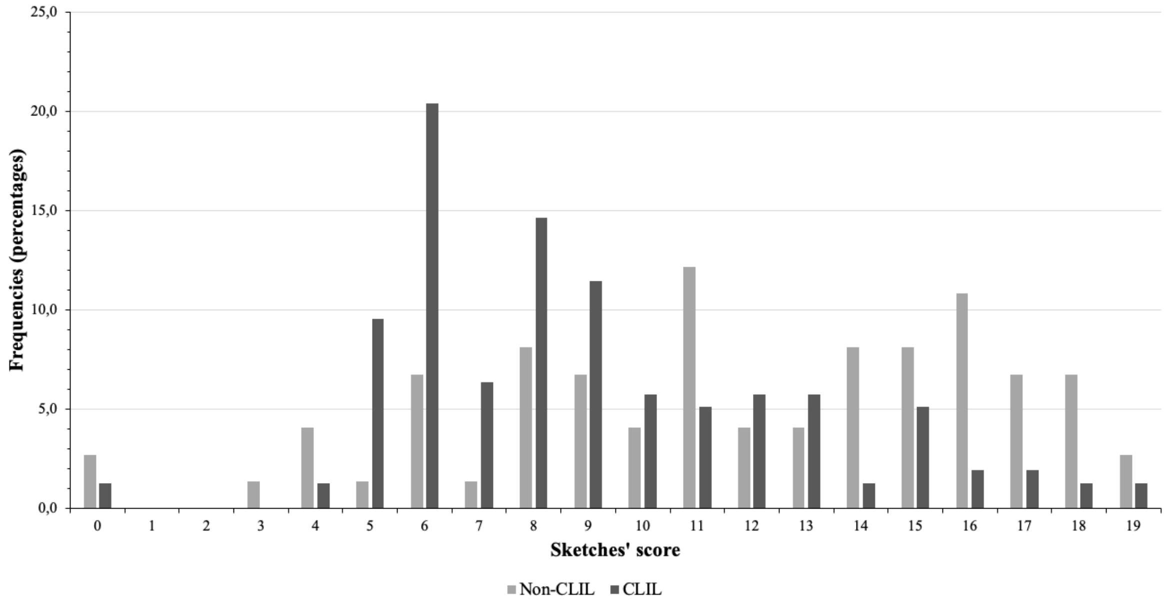


FIGURE 4 Category frequencies after analyses of students' responses to the open questions for both non-CLIL and CLIL learners.

Comparing the two treatment groups in terms of category frequencies and students' responses to the open questions, we found differences for both questions ($C = .426$ and $C = .445$, $p < .001$ in both cases). In their answers to the question 'Which features of the original DNA molecule are simplified in your model', CLIL learners focused almost entirely on the structural level category ($C = .335$, $p < .001$) while non-CLIL learners also pointed to the categories DNA and particle levels ($C = .427/.211$, $p < .001$).

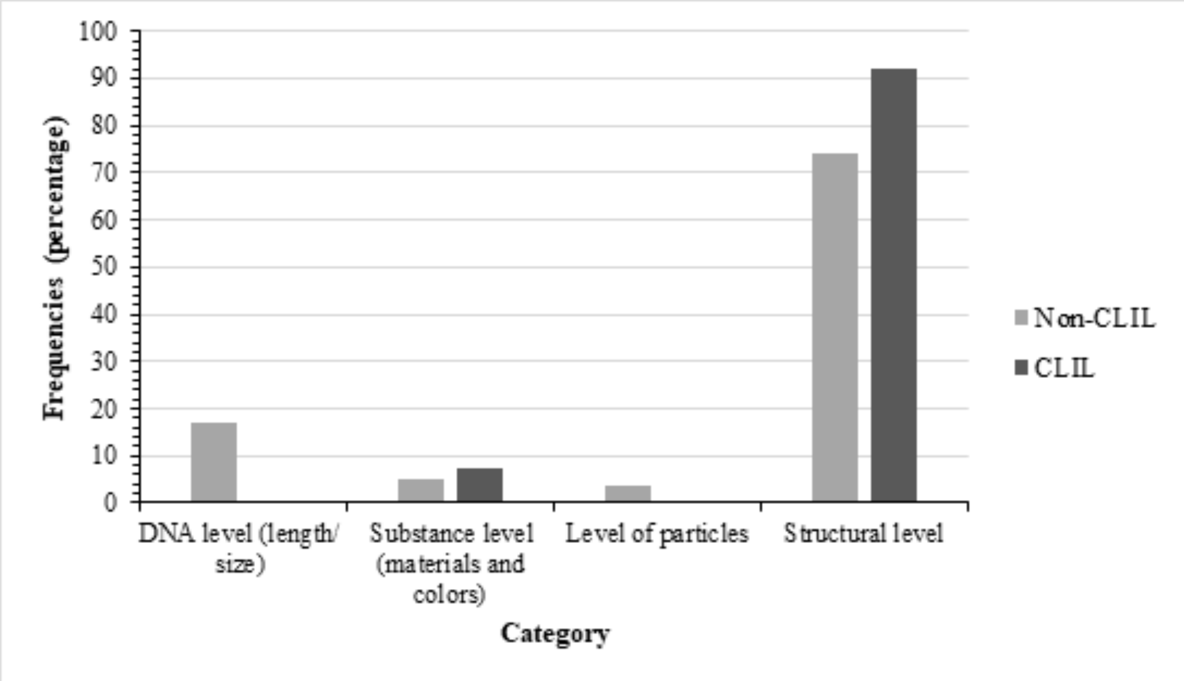


FIGURE 5 Category frequencies of the open question 'Which features of the original DNA molecule are simplified in your model' for non-CLIL and CLIL learners.

In their answers to the question 'Explain why one might create different models of one biological original (in our case, the structure of the DNA)', non-CLIL learners commonly pointed to the category *Individuality of DNA* ($C = .372$, $p < .001$) while CLIL learners focused instead on different research states as explanations for different DNA models ($C = .322$, $p < .001$). The instructor also observed that CLIL learners re-read texts more often and discussed their models more intensely than non-CLIL learners. That is, they often consulted with other groups, engaged the instructor more frequently, and used own spare sheets to brainstorm about the construction of their model.

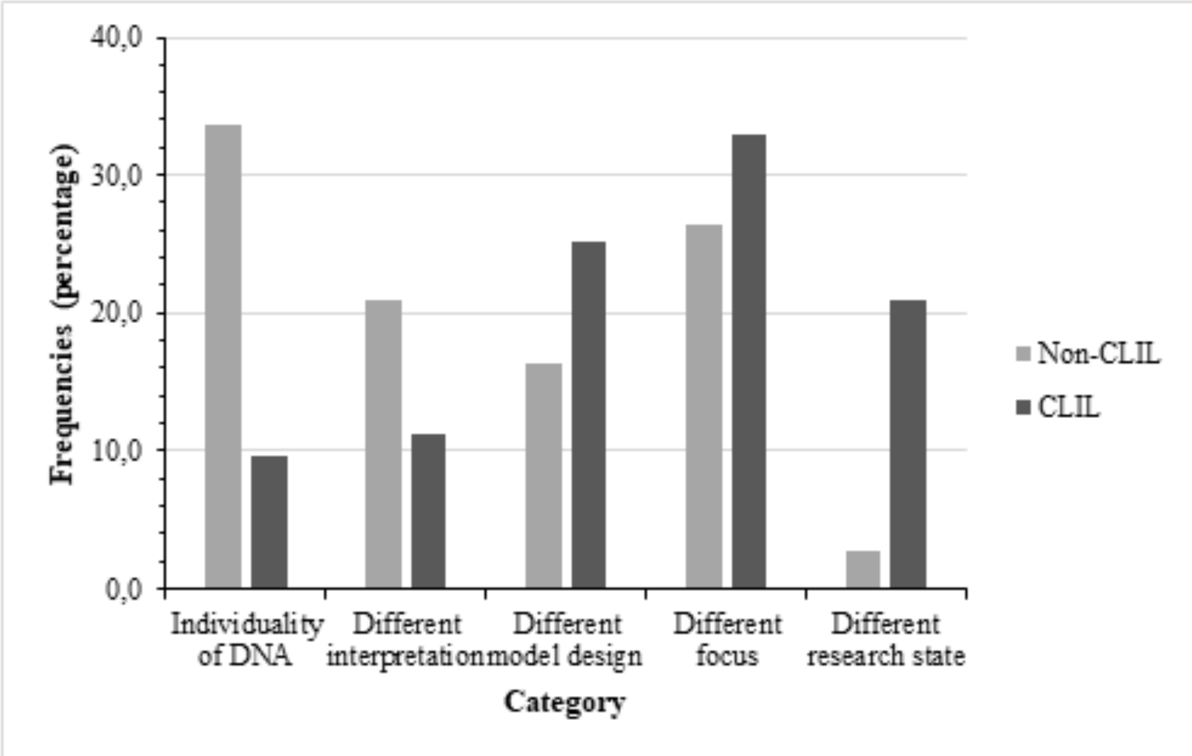


FIGURE 6 Category frequencies of the open question 'Explain why one might create different models of one biological original (in our case, the structure of the DNA)' for non-CLIL and CLIL learners.

Assessment of Evaluation-2 Phase

We compared students' self-evaluation sheets, our assessment of their sheets, and our assessment of their models. For non-CLIL learners, we identified higher scores both for the assessment of students' self-evaluation sheets (MWU: $Z = -3.711$, $p < .001$; $r = .171$, small-to-medium effect) and for the assessment of their models (MWU: $Z = -8.576$, $p < .001$; $r = .396$, medium-to-large effect). Students' self-evaluation sheets did not differ significantly (MWU: $Z = -0.893$, $p = .372$; see Figure 7 and for details Table 5).

Intra-group analyses of both treatment variants indicated similar differences between students' self-evaluations and our assessment of their self-evaluation sheets and assessment of their model (Table 5; F: $\chi^2_{\text{non-CLIL/CLIL}}$ (2, $n = 149/316$) ≥ 121.68, $p < .001$). The pairwise analysis revealed lower scores for our assessment of their self-evaluation sheets compared to their own self-evaluated scores, with large effects (W: $Z \leq -8.818$, $p < .001$; $r \geq .722$). Thus, students identified features as correct that were not given in their model. This discrepancy was evident across all analyses of all sections in both treatment variants (except for the section *primary* *structure*; Table 5; F: $\chi^2_{\text{non-CLIL/CLIL}}(2, n = 149/316) \geq 16.92, p < .001$); W: $Z \leq -2.762, p \leq .006$, $r \geq .326$ medium-to-large effects)

TABLE 5 Assessment of the evaluation-2 phases in the non-CLIL and CLIL approaches.

Analysis sector	Description (score)	Scores						
		Maximal	Students' self-evaluation sheet[a]		Assessment students' self-evaluation sheet[a]		Assessment students model[a]	
			Non-CLIL[b]	CLIL[c]	Non-CLIL[b]	CLIL[c]	Non-CLIL[b]	CLIL[c]
Bases	- Four bases are indicated (2) - Base pairs correctly indicated (3) - Hydrogen bonds correctly indicated, differing in G/C and A/T (1)	6	5.1 (2.0/6.0)	3.9 (1.0/6.0)	4.3 (0/5.0)	0.8 (0/5.0)	5.0 (5.0/5.0)	1.1 (0/5.0)
Deoxyribose	- Deoxyribose indicated (1) - Deoxyribose linked to base (1)	2	1.3 (0/2.0)	1.1 (0/2.0)	1.0 (0/2.0)	0.5 (0/1.0)	1.5 (1.0/2.0)	0.8 (0/2.0)
Phosphate	- Phosphate indicated (1) - Phosphate and deoxyribose alternately arranged (1)	2	1.4 (0/2.0)	1.4 (0/2.0)	1.2 (0/2.0)	0.6 (0/1.0)	1.7 (1.0/2.0)	0.7 (0/2.0)

Primary structure	- Single strand visible (1) - Double strand visible (1)							
		2	1.6 (0.3/2.0)	1.6 (1.0/2.0)	1.6 (1.0/2.0)	1.6 (1.0/2.0)	2.0 (2.0/2.0)	2.0 (2.0/2.0)
Secondary structure	- Double helix visible (1) - Right-handed double helix visible (1)							
		2	1.1 (0/2.0)	1.5 (1.0/2.0)	0.6 (0/1.8)	0.9 (0/1.0)	1.2 (0/2.0)	1.0 (0/2.0)
Sum		14	10.6 (6.0/13.0)	9.1 (5.0/13.0)	8.7 (2.5/11.0)	5.0 (3.0/9.0)	10.8 (8.0/13.0)	6.6 (4.0/10.0)

Note: [a]grouped medians, 25^{th} and 75^{th} percentiles in brackets; [b]monolingual approach; [c]bilingual approach.

In contrast, the assessed models scored higher than the assessed self-evaluation sheets (W: $Z \leq -8.280$, $p < .001$; $r \geq .504$, medium-to-large effect). That is, students did not identify all the correctly modeled features. This phenomenon was also evident across all analyses of sectors in both treatment variants (except for the sector bases and secondary structures in the CLIL variant; Table 5; F: $\chi^2_{\text{non-CLIL/CLIL}}$ (2, n = 149/316) $\geq$ 16.92, $p < .001$); W: $Z \leq -3.191$, $p \leq .001$, in each case; $r \geq .180$, small-to-medium effects).

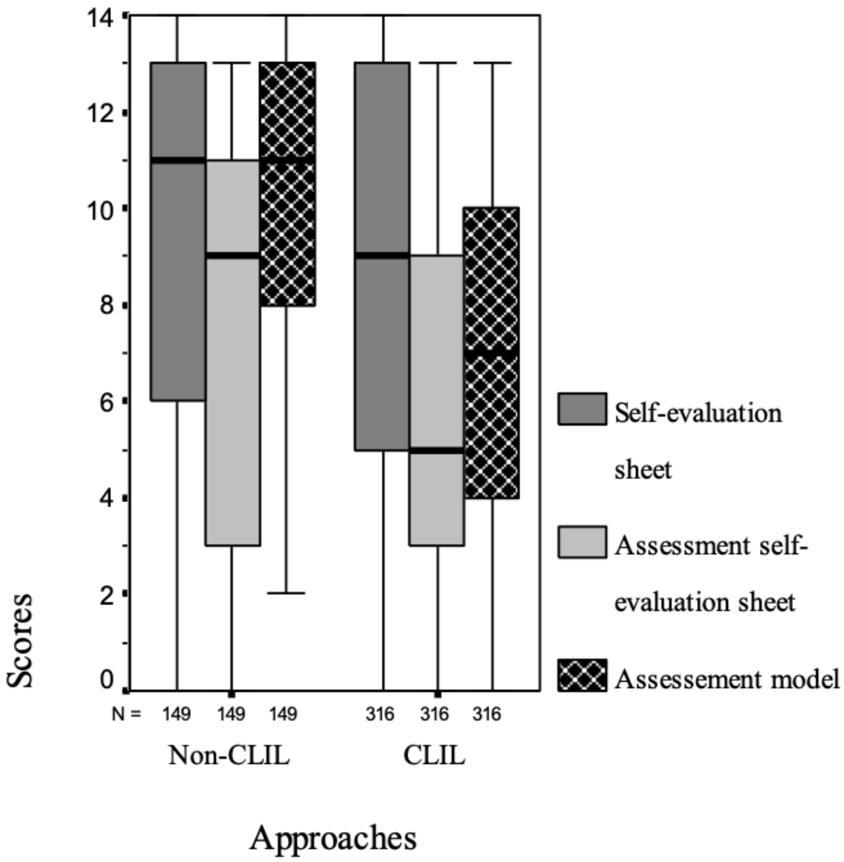


FIGURE 7 Differences between non-CLIL and CLIL learners in scores for self-evaluation sheets, subsequent model assessment and reassessment of the students' self-evaluation sheets.

Comparison of Evaluation-1 and Evaluation-2 Phases

In both language variants, we compared the scores of recorded sketches, which represent the quality of evaluation-1 phase, and the scores of students' self-evaluation sheets, which reflect the quality of their evaluation-2 phase, with our model scores.

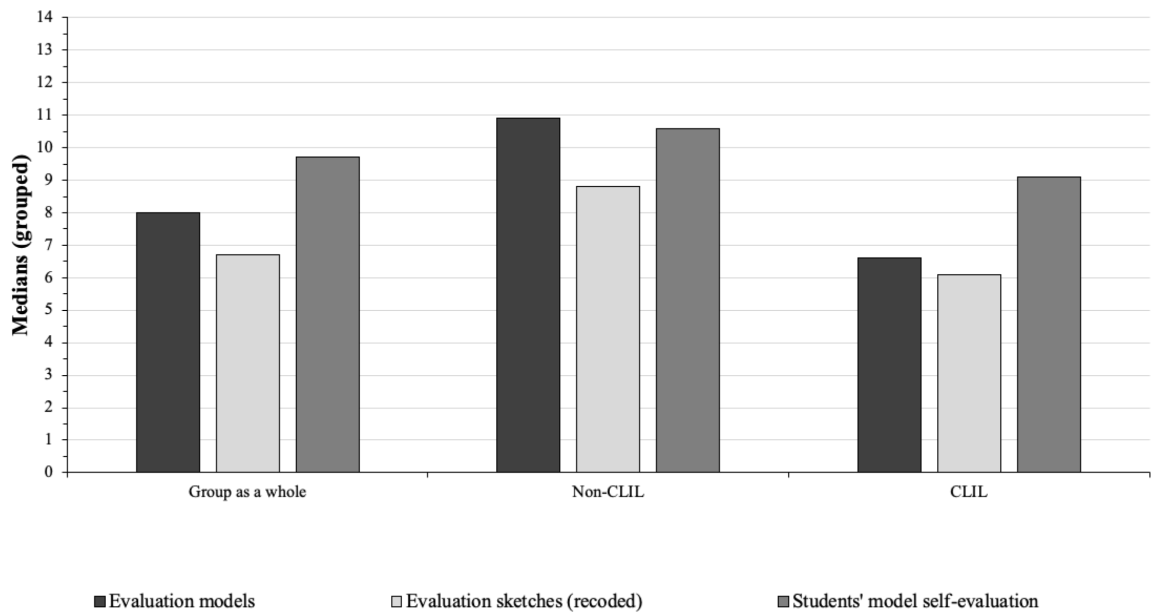


FIGURE 8 Differences between non-CLIL and CLIL learners for model evaluation, recoded evaluation of students' sketches and students' model self-evaluation.

Non-CLIL Learners. Intra-group analyses of non-CLIL students revealed differences between our model scores and the scores for students' sketches (recoded) and the students' self-evaluation sheets (F: $\chi^2_{\text{non-CLIL}}$ (2, n = 149) = 10.79, p < .001). Pair-wise analysis revealed lower scores for their sketches (recoded) compared to the model scores (W: Z = -4.051, p < .001; r = .332, medium effect), but not for their self-evaluation sheets (W: Z = -0.824, p = .410). That is, monolingual taught non-CLIL students performed better in evaluation-2 phase than in evaluation-1 phase.

CLIL Learners. Intra-group analyses of CLIL learners revealed differences between our model scores, the scores of students' sketches (recoded), and their self-evaluation sheets (F: χ^2_{CLIL} (2, n = 316) = 109.28, p < .001). Pair-wise analysis displayed similar scores for students' sketches (recoded) as compared to the model scores (W: Z = -0.352, p = .725), but higher scores for their self-evaluation sheets (W: $Z = -8.100$, $p < .001$, $r = 0.456$, medium-to-large effect). That is, CLIL learners outperformed non-CLIL learners in evaluation-1 phase. However, CLIL learners also identified features as correct that were not given in their model.

Discussion

Our one-day outreach module had a positive effect on model-related cognitive achievement, model-understanding, modeling, and model evaluation, regardless of non-CLIL or CLIL implementation. Our engagement with students' discipline-specific scientific practices involving hands-on and minds-on activities has led to a deeper understanding of the represented scientific content and models in science (Schwarz et al. 2009), which may also have contributed to scientific literacy (Ke et al. 2020). The different experimental steps for isolating and visualizing the students' DNA from oral mucosal cells have provided the scientific content basis for effectively discussing the observed phenomenon and constructing and evaluating the ensuing models (Lemke 1990, Berland & Reiser 2009).

Cognitive Achievement

The influence that our CLIL science module had on content knowledge and understanding regarding DNA as a model throughout the hands-on laboratory was in keeping with our previous findings (Roth et al. 2020). That is, non-CLIL learners outperformed CLIL learners with regard to short-term knowledge acquisition. We propose that the increased mental load from both content and language learning may have influenced students' performance. That is, the mental processing of scientific content in a foreign language, accompanied by hands-on experimentation and modeling and model evaluation activities, may have overstrained students' mental capabilities. Our previous studies of non-CLIL outreach learning have already indicated that combined experimentation, modeling, and model evaluation can be strenuous (Scharfenberg & Bogner 2013, Mierdel & Bogner 2019, Roth et al. 2020). The addition of language learning to such mentally demanding learning situations may further increase the mental load and overstrain the working memory (Rodenhauser & Preisfeld 2014). Yet, it mirrors an authentic setting in science classrooms in connecting with the increasing adoption of CLIL outside of English-speaking countries.

Although the performance of the CLIL learners was poorer as compared to non-CLIL learners, there was a significant knowledge acquisition that could be sustained both temporarily and permanently. While the comparably higher mental effort involved in learning scientific content in a foreign language may have influenced overall performance (Rodenhauser & Preisfeld 2014, Sweller 2015), the discursive activities – as observed and encouraged by the instructor - involved in the construction and negotiation of an appropriate model led to in-depth learning (Krell et al. 2015, Ke et al. 2020, Ke & Schwarz 2020) of information relevant to the design of the DNA model. Thus, the combination of hands-on tasks with minds-on activities in CLIL learning (Glynn & Muth 1994) may have conveyed model-related knowledge, while the negotiation of meaning involved in the process of talking about science could have encouraged retention of this knowledge (Evnitskaya & Morton 2011). That is, the various scaffolding exercises, which had to be solved in group-work, may have encouraged more and deeper discussion of the DNA-related experiments and ensuing construction of a three-dimensional DNA model.

Model Understanding

The influence that our CLIL science module had on overall model-understanding, compared to the non-CLIL approach, supports our assumption that the incorporation of scientific discourse into model negotiation and model construction, induced by more extensive scaffolding exercises, encourages in-depth learning. In particular, students displayed a significant increase in their awareness for CNM and a respective decrease in the misconception of EM after participation in our CLIL science module. This counteracts the widespread yet erroneous notion among students that models are exact copies of real-life phenomena that only differ across time and do not change in response to new evidence (Grosslight et al. 1991, Ke & Schwarz 2020). In fact, students are largely unaware of the importance of evidence and the constant negotiation of a models' validity by peer scientists (Treagust et al. 2002, Mendonça & Justi 2013, Ke et al. 2020). Since the findings of our non-CLIL approach reflect the hypothesized lack of model understanding – students still perceived models as ER after participation - our CLIL science module appears to have compensated for this lack.

One key aspect that may have influenced model understanding is discursive action: The construction, articulation, and vindication of scientific arguments help "students […] develop a deeper understanding of content" (Passmore & Svoboda 2012, p. 1537). This also includes the analysis and interpretation of data and the re-evaluation of their model, if required (Ke et al. 2020). As such, critical reflection of models and the therewith associated re-construction of certain elements may have contributed to CNM (Schwarz et al. 2009). The increased discursive action may have also led to the development of epistemological understanding of modelling processes, hence encouraging deeper cognitive processing of their material (Sins et al. 2009).

Yet, discourse and model evaluation were elements of our previous non-CLIL laboratory. Gilbert & Justi (2002) assume that "scientists always mentally rehearse the design [of their model before] actual empirical testing takes place [and, if the outcomes are not feasible,] attempt[s] will have to be made to modify it and to re-enter the cycle" (p.372). Accordingly, repeated engagement with both the content and the resulting model is essential, firstly, to develop the model and, secondly, if necessary, to improve it (Schwarz et al. 2009, Passmore & Svoboda 2012, Lee et al. 2015). This also entails a fundamental understanding of the elements necessary to include in the model based on their purpose (Grosslight et al. 1991, Sins et al. 2009), such as the base pairs and phosphate-sugar-strands in the DNA model. Since our participating students were not yet fully fluent in English, and only at the beginning of their constantly developing scientific literacy, they had to re-read, re-think, and extensively discuss their mental models before they proceeded with construction. This led to several cycles of mental rehearsal (Oh & Oh 2011, Gilbert & Justi 2002, Khan 2011), in the process of which the students could have slowly become aware of the changing nature of models, dependent on the information (Treagust et al. 2002) they extracted from their manuals and the text about the discovery of DNA (Usher 2013).

This is, of course, only one possible explanation, but it is supported by theories about the development of scientific literacy and associated deep learning. Often, models are simply regarded as representations of a phenomenon and not as a "sense-making tool for considering how and why patterns and mechanisms of phenomena occur" (Ke & Schwarz 2020, p. 2). Finding answers to the 'how' and 'why' moves modeling away from purely content-based learning and focuses more on the practice of science. This encourages students to engage in proposing scientific hypotheses and finding scientific arguments to justify these hypotheses. Not only does this stimulate in-depth exploration of the subject at hand, which encourages the development of scientific literacy (Quarderer & McDermott 2020, France 2019), but it also promotes deep learning (Lee et al. 2015). In particular, the transformational nature of modeling is supposed to trigger deep learning. That is, "deep-level processing" of the given information is required (Case & Gunstone 2010, p. 461), which, it seems, our non-CLIL approach did not achieve to the same degree as our CLIL module.

For deep-level processing, students also have to understand what they are reasoning about. Since the language of science differs from the language used in everyday settings (Wang & Chen 2014), scientific literacy in the respective science discipline is indispensable (Prain 2004). This not only includes vocabulary and hands-on activities but also, and in particular, minds-on activities. Such minds-on activities include the explanation and abstraction of scientific content (Glynn & Muth 1994). Although minds-on activities were also included in our non-CLIL module in the form of open questions in the laboratory manuals encouraging students to summarize observations, we further developed these activities in our CLIL module. As students were still learning English as their second language, we provided additional scaffolding material in the form of a language workbook and made sure the students had understood the content of our laboratory by encouraging them to summarize the procedures in their own words. We also extended this method to the modeling phases, which is why the purpose of a model was probably more clearly understood than in the non-CLIL module. That is, the "verbal representation" of a model may be at least as important as its physical manifestation in explaining and justifying the proposed model (Campbell & Fazio 2020, p. 2302). Yet, in classroom teaching, such representation of mental models, including the critiquing, revising, and enriching of mental models is often neglected. Specifically, the involvement of teachers to create models from scientific discourse appears to be a key element to model understanding and model construction (Khan et al. 2011).

Model Evaluation

The influence that our CLIL science module had on overall model evaluation-1 and model evaluation-2 as compared to the non-CLIL treatment group mirrors our findings on model-related cognitive achievement. Non-CLIL learners received higher scores for their model sketches and hand-crafted models and outperformed CLIL learners in model evaluation-2. The better overall performance of non-CLIL learners in model construction, in their sketches and in evaluation-2 may – much like content knowledge and understanding of the scientific concept – be attributed to the increased mental load from both content and language learning for the CLIL learners (Sweller 2015). As has previously been outlined, the practice of modeling requires an understanding of the content and an extensive period of negotiation (Gilbert & Justi 2002, Passmore & Svoboda 2012). The lack of language proficiency in English may, thus, have influenced students' performance and contributed to an overload of mental demands from the combination of content and language learning (Johnstone & Wham 1982).

In both non-CLIL and CLIL treatment variants, some students correctly identified and labelled the DNA's different components, some students identified only a limited number, and some only modelled the DNA's basic structure. This result is in line with the findings of previous studies (Howell et al. 2019), wherein the researchers described difficulties in students' understanding of DNA's structure-function relationships. We also reported such findings in a comparative study of non-CLIL outreach modules (Roth et al. 2020).

Analysis of answers to DNA-model-related questions on the students' worksheet of model evaluation-1 emphasize the increased model understanding of CLIL learners. Sample answers showed that CLIL learners focused, in particular, on the 'structural level of DNA' and 'state of research' as determinants of how such models are being developed. This demonstrates a broad understanding of DNA as a model and the realization that models are not exact representations of observed phenomena but may change dependent on emerging evidence (Grosslight et al. 1991, Ke & Schwarz 2020). This awareness is in stark contrast to typical notions among students of the nature of models (Schwarz et al. 2009, Krell et al. 2014, Krell et al. 2015). Similar to overall model understanding, the students' lack of fluency in English and their developing scientific literacy may have led to more intense re-reading, re-thinking, and extensive discussion of their mental models before they proceeded with their construction. This could have induced several cycles of mental rehearsal (Oh & Oh 2001, Gilbert & Justi 2002), which expedited the awareness of CNM (Treagust et al. 2002). Findings from evaluation-1, wherein students were asked to draw sketches of their hand-crafted models, supports this theory. Although the models were certainly not as exact as those built by non-CLIL learners, CLIL learners designed sketches that were far more accurate and labeled components of their models accordingly. This task demands a deeper understanding of models than simply comparing hand-crafted models to a commercial DNA model and ticking features on a pre-designed list, as was the case in evaluation-2 where non-CLIL learners outperformed CLIL learners. Hence, CLIL learners not only understood the importance of CNM but also knew the function and place of components that they crafted in their models.

Limitations

Firstly, our study involved ninth-graders, who had little to no prior experience in hands-on experimentation, modeling, and model evaluation, which was evident from field notes taken after questions throughout the pre-lab phase. Secondly, the students had little real-life experience of English outside English language lessons, as we learned from conversations with students and teachers. Thirdly, due to Bonferroni corrections, comparisons of T0 and T2 knowledge scores were slightly above the reduced threshold of significance. Therefore, we cannot exclude a potential beta error. However, an additional comparison of different variables only fortified the higher short-term achievement of non-CLIL learners. A lack of commonly agreed standardized instruments for assessing CLIL content learning, which, in our case, also extends to modeling and model evaluation, may impair adequate comparison (Dalton-Puffer 2011). Fourthly, due to the context-dependency of CLIL learning, results cannot easily be extrapolated (Pérez-Cañado 2012). As a consequence, generalizing about the success of CLIL learning requires an acknowledgement of the diversity of possible CLIL implementations. Further research in the context of short-term implementations of CLIL in combination with model-related learning is required to pinpoint key challenges and consider possible means to increase its success. That is, the implementation of long-term modules as investigated by, for instance Meyerhöffer and Dreesmann (2019) and confirmed by Admiraal et al. (2006) or Haagen-Schützenhöfer et al. (2011), show that the positive outcomes of our module, such as model understanding, could be enhanced while negative outcomes related to CLIL instruction, such as content knowledge, could be leveled out. Yet, other short-term CLIL modules by, for instance, Rodenhauser & Preisfeld (2014) did not produce significant differences between CLIL and non-CLIL participants. For further implementations, we should consider reducing the extraneous load (Chandler & Sweller 1991), by providing more scaffolding materials and spreading the module over two consecutive days. This way, and in line with Meyerhöffer and Dreesmann (2019) and Craik and Lockhart's (1972) levels of processing theory, CLIL students should succeed and even outperform non-CLIL students also regarding content knowledge, and model building as well as evaluation. Moreover, the inclusion of qualitative discourse analysis in further developments of the module could shed more light on the importance of talking about science in the creation of knowledge and model understanding. Moreover, the inclusion of qualitative discourse analysis in further developments of the module could shed more light on the importance of talking about science in the creation of knowledge and model understanding.

Conclusion

Our study furthers understanding of the relationship between CLIL learning and modeling, which encompasses aspects of language learning and scientific literacy (e.g., Prain 2004), content learning and language learning (e.g., Stoddart et al. 2002, Gonzalez-Howard & McNeill 2016), modeling and content learning (e.g., Schwarz et al. 2009, Lee et al. 2015), as well as modeling and scientific literacy (e.g., Quarderer & McDermott 2020, Ke et al. 2020). While most previous studies have focused, primarily, on only one of these combinations, our module encompasses aspects of them all. Moreover, our study explores the potential of short-term CLIL modules, rather than the long-term CLIL modules more commonly explored by other researchers (Meyerhöffer & Dreesmann 2019).

Although the CLIL treatment group received overall lower scores than non-CLIL learners, CLIL outreach learning holds the potential to improve model understanding. As the development of model understanding is rather cumbersome, and stimulating environments are difficult to identify (Glynn & Muth 1994, Schwarz et al. 2009), our CLIL module could provide a possible approach by combining hands-on laboratory experiments with language learning and associated minds-on activities. Although we cannot identify the reasons for lower scores among the CLIL group, we would – in line with cognitive load theory (Sweller 2015) – encourage the reduction of content in later implementations.

Acknowledgements

Special thanks to the teachers and students involved in this study for their cooperation. Also, many thanks to the Horizon 2020 Framework Program and the German Research Foundation, who funded this publication, as well as the University of Bayreuth.

Ethical Statement

Our study was in line with the requirements of the Declaration of Helsinki and the study as well as questionnaires were approved by the Bavarian State Ministry of Education and Cultural Affairs, Science and the Arts (Reference IV.7-BO5106/149/18)

Conflict of Interest Statement

For the article “Content and Language Integrated Scientific Modelling: A Novel Approach to In-depth Model Learning”, we certify that we have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. Educational Research and Evaluation, 12, 75–93.

Akerson, V.L., Townsend, J.S., Donnelly, L.A., Hanson, D.L., Tira, P., & White, O. (2009). Scientific modeling for inquiry teachers network (SMIT'N): the influence on elementary teachers' views of nature of science, inquiry, and modeling. Journal of Science Teachers Education, 20, 21-40.

Berland, L.K., & Reiser, B.J. (2009). Making sense of argumentation and explanation. Science Education, 93(1), 26–55.

Bingimlas, K.A. (2009). Integration of ICT in teaching and learning environments: A review of the literature. Eurasia Journal of Mathematics, Science & Technology Education, 5(3), 235-245.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. International Journal of Educational Research, 31, 659–671.

Bro, R., & Smilde, A.K. (2014). Principal component analysis. Analytical Methods, 6, 2812-2831.

Buse, M., Damerau, K. & Preisfeld, A. (2018). A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts. International Journal of Environmental & Science Education, 13(8), 647-660.

Buxton, C., Harman, R., Cardozo-Gaibisso, L., Jiang, L., Bui, K., & Allexsaht-Snider, M. (2019). Understanding science and language connections: new approaches to assessment with bilingual learners. Research in Science Education, 49, 977-988.

Campbell, T., & Fazio, X. (2020). Epistemic frames as an analytical framework for understanding the representation of scientific activity in a modeling-based learning unit. Research in Science Education, 50, 2283-2304.

Carpenter, S.L., et al. (2019). Models are a "metaphor in your brain": How potential and preservice teachers understand the science and engineering practice of modeling. School Science and Mathematics, 119, 275–286.

Case, J., & Gunstone, R. (2002). Metacognitive development as a shift in approach to learning: an in-depth study. Studies in Higher Education, 27(4), 459–470.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and Instruction, 8(4), 293-332.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213–220.

Cook, T. D., & Campell, D. (1979). Quasi-experimentation. Design & analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671-684.

Cummins, J. (1981). Bilingualism and minority-language children. Toronto: OISE Press.

Dalton-Puffer, C. (2007). Discourse in Content and Language Integrated learning (CLIL) Classrooms. Amsterdam: John Benjamins.

Declaration of Helsinki, World Medical Association (2013). Journal of the American Medical Association, 310, 2191-2194.

Ellis, P. D. (2010). The essential guide to effect sizes statistical power, meta-analysis, and the interpretation of research results. Cambridge: Cambridge University Press

European Commission (Hrsg.) (2004): Promoting language learning and linguistic diversity. An action plan 2004-06. Luxembourg, Office for Official Publications of the European Communities. European Commission (2012). FAQs on multilingualism and language learning. Brussels. Retrieved from http://europa.eu/rapid/press-release_MEMO-12-703_en.pdf.

European Commission (2012). FAQs on multilingualism and language learning. Brussels. Retrieved from http://europa.eu/rapid/press-release_MEMO-12-703_en.pdf

Evnitskaya, N. & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. Language and Education, 25(2), 109-127.

Field, A. (2012). Discovering statistics using IBM SPSS statistics (4th ed.). London: SAGE.

France, A. (2019). Teachers using dialogue to support science learning in the primary classroom. Research in Science Education. https://doi.org/10.1007/s11165-019-09863-3.

Franco C. & Colinvaux D. (2000). Grasping Mental Models. In: Gilbert J.K., Boulter C.J. (Eds.) Developing models in science education (pp. 93-118). Springer.

Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: achieving scientific literacy. Journal of Research in Science Teaching, 31(9), 1057-1073.

Gobert, J., O'Dwyer, L., Horwitz, P., Buckley, B., Levy, S., & Wilensky, U. (2011). Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry. International Journal of Science Education, 33, 653–684.

Goldschmidt, M., & Bogner, F.X. (2016). Learning about genetic engineering in an outreach laboratory: Influence of motivation and gender on students' cognitive achievement. International Journal of Science Education, 6(2), 166-187.

Gonzalez-Howard, M. & McNeill, K.L. (2016). Learning in a community of practice: factors impacting English-learning students' engagement in scientific argumentation. Journal of Research in Science Teaching, 53(4), 527-553.

Gottlieb, M. (2016). Assessing English language learners: Bridges for language proficiency to academic achievement. Thousand Oaks: Corwin.

Grandinetti, M.M., Langellotti, M., & Teresa Ting, Y.L. (2013). How CLIL can provide a pragmatic means to renovate science education – even in a sub-optimally bilingual context. International Journal of Bilingual Education and Bilingualism, 16(3), 354-374.

Grosslight, L., Unger, C., & Jay, E. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. Journal of Research in Science Teaching, 28(9), 799–822.

Haagen-Schützenhöfer, C., Mathelitsch, L., & Hopf, M. (2011). Fremdsprachiger Physikunterricht: Fremdsprachlicher Mehrwert auf Kosten fachlicher Leistungen? Auswirkungen fremdsprache- nintegrierten Physikunterrichts auf fachliche Leistungen [Foreign language physics lessons: Added value for language skills at the cost of content achievement? Effects of content-language-integrated physics classes on content achievement]. Zeitschrift für Didaktik der Naturwissenschaften, 17, 223–260.

Howell, M.E., Booth, C.S., Sikich, S.M., Helikart, T., Roston, R.L., Couch, B.A., & van Dijk, K. (2019). Student understanding of DNA structure–function relationships improves from using 3D learning modules with dynamic 3D printed models. Biochemistry and Molecular Biology Education, 47, 303-317.

Johnstone, A.H. & Wham, A.J.B. (1979). A model for undergraduate practical work. Education in Chemistry, 16, 16–17.

Gilbert, J. K., & Justi, R. S. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. International Journal of Science Education, 24, 369–387.

Gouvea, J., & Passmore, C. (2017). 'Models of' versus 'Models for'. Towards an agent-based conception of modeling the science classroom. Science and Education, 26(1-2), 49-63.

Hampton, E., & Rodriguez, R. (2001). Inquiry science in bilingual classrooms. Bilingual Research Journal, 24(4), 461-478.

ISB (2019). Ergänzende Informationen zum LehrplanPLUS [Supplementary information on the curriculum]. URL: https://www.lehrplanplus.bayern.de/sixcms/media.php/71/B8_10_Info_Modelle_20190916.pdf (accessed 4 May 2021).

Kaiser, H.F (1970). A second generation little jiffy. Psychometrika, 35, 401–415.

Ke, L., & Schwarz, C. (2020). Using epistemic considerations in teaching: Fostering students' meaningful engagement in scientific modeling. In A. Upmeier zu Belzen, D. Krüger, & J. Van Driel (Eds.), Towards a Competence-based View on Models and Modeling in Science Education (pp. 181–199). Springer International Publishing.

Ke, L., Zangori, L., Sadler, T.D., & Friedrichsen, P.J. (2020). Integrating scientific modeling and socio-scientific reasoning to promote scientific literacy. In Warden A. Powell (Ed.), Socioscientific issues-based instruction for scientific literacy development (pp. 31-56). Hershey: IGI Global.

Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education,* 22, 535-560.

Klieme, E., et al. (Eds.) (2010). PISA 2009: Bilanz nach einem Jahrzehnt [PISA 2009: Review after a decade]. Münster: Waxmann.

KMK (2005). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

Kovanović, V., et al. (2018). Understanding the relationship between technology-use and cognitive presence in MOOCs. Proceedings of the Seventh International Conference on Learning Analytics and Knowledge, 582-583. New York: Association for Computing Machinery.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? Science and Children, 49, 10–13.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. Science, 328 (5977), 456–459.

Krashen, S. (2013). The case for non-targeted, comprehensible input. Journal of Bilingual Education Research & Instruction, 15(1), 102-110.

Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2012). Students' understanding of the purpose of models in different biological contexts. International Journal of Biology Education, 2(2), 1–34.

Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2014). Students‘ levels of understanding models and modelling in Biology: global or aspect-dependent? Research in Science Education, 44, 109-132.

Krell, M., Reinisch, B., & Krüger, D. (2015). Analyzing students' understanding of models and modeling referring to the disciplines Biology, Chemistry, and Physics. Research in Science Education, 45, 367-393.

Lam, S. F.,Wong, B. P. H., Yang, H., & Liu, Y. (2012). Understanding student engagement with a contextual model. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), Handbook of research on student engagement (pp. 403-420). Boston, MA: Springer US.

Langheinrich, J., & Bogner, F. X. (2015). Student conceptions about the DNA structure within a hierarchical organizational level: Improvement by experiment- and computer-based outreach learn- ing. Biochemistry and Molecular Biology Education, 43(6), 393– 402.

Langheinrich, J., & Bogner, F.X. (2016). Computer-related self-concept: The impact on cognitive achievement. Studies in Educational Evaluation, 50, 46-52.

Lee, S., Kang, E., & Kim, H-B. (2015). Exploring the impact of students‘ learning approach on collaborative group modeling of blood circulation. Journal of Science Education Technology, 24, 234-255.

Lemke, J. L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Lienert, G. A., & Raatz, U. (1998). Testaufbau und Testanalyse [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lipsey, M. W., & Wilson, D. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage Publications.

Louca, L., & Zacharia, Z. (2012). Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. Educational Review, 64, 471–492.

Marsh, D. (Ed.). (2002). CLIL/EMILE—the European dimension: Actions, trends and foresight. European Commission: Potential Public Services Contract DG EAC.

Maybin, J., Mercer, N. & Stierer, B. (1992). Scaffolding: learning in the classroom. In Kate Norman (Ed.), Thinking Voices. The Work of the National Oracy Project (pp.186–95). London: Hodder & Stoughton.

Mendonça, P.C.C., & Justi R. (2013). The relationships between modelling and argumentation from the perspective of the model of modelling diagram. International Journal of Science Education, 35(14), 2407–2434.

Meskill, C., & Oliveira, A.W. (2019). Meeting the challenges of English learners by pairing science and language educators. Research in Science Education, 49, 1025-1040.

Meyerhöffer, N. & Dreesmann, D.C. (2019). English-bilingual Biology for standard classes development, implementation and evaluation of an English-bilingual teaching unit in standard German high school classes. International Journal of Science Education, 41(10), 1366-1386.

Mierdel, J., & F.X. Bogner. (2019). Comparing the use of two different model approaches on students' understanding of DNA models. Education Sciences 9(2), 115-133. https://doi.org/10.3390/educsci9020115.

Nation, I. S. P., & Newton, J. (2008). Teaching ESL/EFL listening and speaking. London: Routledge.

Norris, S.P., & Phillips, L.M. (2003). How literacy in its fundamental sense is central to scientific literacy. Science Education, 87(2), 224-240.

Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: An overview. International Journal of Science Education, 33, 1109–1130.

Passmore, C., Stewart, J., & Cartier, J. (2009). Model-based inquiry and school science: Creating connections. School Science and Mathematics, 109, 394–402.

Passmore C., & Svoboda J. (2012). Exploring opportunities for argumentation in modelling classrooms. International Journal of Science Education, 34(10), 1535–1554.

Pearson, K. (1900). On the theory of contingency and its relation to association and normal correlation. London, UK: Dulau.

Pérez-Cañado, M. L. (2012). CLIL Research in Europe: Past, Present, and Future. International Journal of Bilingual Education and Bilingualism, 15(3), 315-341.

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. European Scientific Journal, 12, 235–250.

Prain, V. (2004). The role of language in science learning and literacy. In Writing and learning in the science classroom (pp. 33–45). Dordrecht: Springer.

Quarderer, N.A., & McDermott, M.A. (2020). Examining science teacher reflections on argument-based inquiry through a critical discourse lens. Research in Science Education, 50, 2483-2504.

Rodenhauser, A. & Preisfeld, A. (2014). Bilingual (German-English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. International Journal of Environmental and Science Education, 10(1), 99-110.

Rost, J. (2004). Lehrbuch Testtheorie–Testkonstruktion [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

Rotbain, Y., Marbach-Ad, G., & Stavy, R. (2006). Effect of bead and illustrations models on high school students' achievement in molecular genetics. Journal of Research in Science Teaching, 43, 500–529.

Roth, T., Scharfenberg, F-J., Mierdel, J., & F.X. Bogner. (2020). Self-evaluative scientific modeling in an outreach gene technology laboratory. Journal of Science Education and Technology, 29, 725-739. https://doi.org/0.1007/s10956-020-09848-2.

Saka, A., et al. (2006). A cross-age study of the understanding of three genetic concepts: how do they imagine the gene, DNA and chromosome? Journal of Science Education and Technology, 15 (2), 192-202.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. Innovations in Education and Teaching International, online first, 1–12.

Scharfenberg, F. J., & Bogner, F. X. (2013). Teaching Gene Technology in an Outreach Lab: Students' Assigned Cognitive load clusters and the clusters' relationships to learner characteristics, laboratory vari- ables, and cognitive achievement. Research in Science Education, 43, 141–161.

Schwarz, C., & White, B. (2005). Metamodeling knowledge: developing students' understanding of scientific modeling. Cognition and Instruction, 23, 165–205.

Schwarz, C., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. Journal of Research in Science Teaching, 46, 632–654.

Sins, P., Savelsbergh, E., Van Joolingen, W., & Van Hout-Wolters, B. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. International Journal of Science Education, 31, 1205–1229.

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. Journal of Research in Science Teaching, 39(8), 664-687.

Sweller, J. (2015). In academe, what is learned and how is it learned? Current Directions in Psychological Science, 24, 190-194.

Tolbert, S., Knox, C., & Salinas, I. (2019). Framing, adapting, and applying: Learning to contextualize science activity in multilingual science classrooms. Research in Science Education, 49, 1069-1085.

Treagust, D.F., Chittleborough, G.D., & Mamiala, T.L (2002). Students' understanding of the role of scientific models in learning science. International Journal of Science Education, 24, 357–368.

Unsal, Z., Jakobson, B., Molander, B.-O., & Wickman, P.-O. (2018). Language use in a multilingual class: A study of the relation between bilingual students' languages and their meaning-making in science. Research in Science Education, 48(5), 1027–1048.

Usher, S. (2013). Letters of Note. Correspondence Deserving of a Wider Audience. Edinburgh, Scotland: Canongate.

Wang, J-R., & Chen, S-F. (2014). Exploring mediating effect of metacognitive awareness on comprehension of science tests through structural equation modeling analysis. Journal of Research in Science Teaching, 50(2), 175-191.

Wolf, R. (1997). Rating scales. In J. Keeves (Ed.), Educational research, methodology and measurement: an international handbook (pp. 958–965). Oxford, UK: Elsevier.

Yore, L.D., & Treagust, D.F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. International Journal of Science Education, 28(2–3), 291–314.

Electronic Supplemental Materials

Content and Language Integrated Scientific Modelling: A Novel Approach to Model Learning

Tamara Roth[1]*, Franz-Josef Scharfenberg[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

Correspondence: tamara.roth@uni-bayreuth.de; ORCID ID: 0000-0002-9062-1489

Table of Contents

(Total pages 5)

Electronic Supplemental Material 1a: Language Exercise Book

DNA-MODELLING

Find the matching words to the definitions and German translations in your word-cloud and write them down!

English	Definition	Deutsch
Example: *clockwise*	**moving around in the same direction as the hands of a clock**	**rechtsgängig; im Uhrzeigersinn**
	not parallel	antiparallel
	the force causing molecules of the same substance to stick together	Zusammenhalt
	a single thin piece of thread, wire, hair, etc.	Stränge
	a chemical element; a gas that is the lightest of all the elements; it combines with oxygen to form water.	Wasserstoff
	strong connection	Bildung
	things that are different but together form a useful or attractive combination of skills, qualities or physical features	einander ergänzend

Electronic Supplemental Material 1b: Knowledge Questionnaire

A. Knowledge questionnaire (adapted and extended from Authors 2016, Authors 2020). Correct answers are written in italics.

1	**The genetic information is encoded in the DNA by …**
☐	the coils of the DNA strand.
☐	the fusion of egg and sperm cells during fertilization.
☐	the formation of different chromosomes.
☐	*the sequence of the single bases.*

2	**The proportion of sugar to phosphate in the DNA-molecule is …**
☐	2:1
☐	1:2
☐	3:1
☐	*1:1*

3	**The abbreviation 'DNA' stands for …**
☐	Oxyribonucleic acid.
☐	*Deoxyribonucleic acid.*
☐	Deoxynucleic acid.
☐	Dideoxyribonucleic acid.

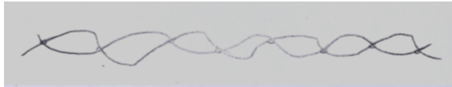
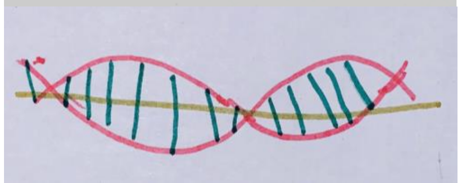
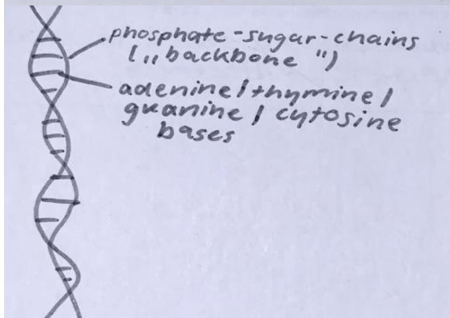
4	**A section of DNA that provides the basic information for building a particular characteristic is called …**
☐	genome.
☐	plasmid.
☐	chromosome.
☐	*gene.*

5	**The DNA bases are located …**
☐	at the outside of the DNA molecule and bound to phosphate.
☐	at the inside of the DNA molecule and bound to phosphate.
☐	*at the inside of the DNA molecule bound to sugar.*
☐	at the outside of the DNA molecule and bound to sugar.

6	**Which of the following components is <u>not</u> included in the DNA?**
☐	adenine
☐	*ribose*
☐	guanine
☐	deoxyribose

Electronic Supplemental Material 2: Assessment of Students' Sketches

Assessment of students' sketches based on their handcrafted models (modified after Authors 2015).

Analysis		Definition	Score	Example	Sketches' scores *Md* (25^{th} /75^{th} *P*)	
Sector	Subsector				Non-CLIL students[a]	CLIL students[b]
Bases	BA1	No bases drawn	0			
		Bases drawn	1			
		Bases drawn and labeled	2			
		Base pairs indicated	3			

Electronic Supplemental Material 3: Assessment of Students' Responses 1

Assessment of the students' responses to the question "Which features of the original DNA molecule are simplified in your model"

Category	Definition	Examples of students' answers	Frequencies	
			Non-CLIL students[a]	CLIL students[b]
Level of	The student's answer mentions a simplification of			
… DNA	… the DNA molecule's size or its length (e.g., by number of bases).	"The size of the model is different."[c] "Original DNA molecule has more base pairings (they are closer together)."[d]	29	2
… substances	… the DNA molecule due to the use of different materials or colors for the different components.	"The shapes of the chemicals have been simplified."[c] "All DNA-Elements are represented by different colors."[d]	9	22
… particles	… the representation of the DNA molecule or its particles.	"The DNA molecules are simplified."[d] "The [sugar and phosphate molecules] usually consist of various atoms."[d]	6	0
… structure	… the DNA structure in general or related to structural elements (e.g., base pairing, backbone, hydrogen bonds, or secondary structure).	"The whole DNA structure is simplified."[d] "Thymine and cytosine are simplified."[c] "The backbone is kept simple."[d] "In our model the hydrogen bonds are simplified."[d] "The double helix structure is shown in a simplified form as well."[d]	127	279
No adequate answer	The student's answer does not refer to the DNA molecule.	"The proteins have been simplified."[c]	40	156

Notes: [a] sum 171, monolingual; [b] sum 303, bilingual; [c] CLIL student; [d] Non-CLIL student.

Electronic Supplemental Material 4: Assessment of Students' Responses 2

Assessment of students' responses to the question "Explain for what reason one may create different models of one biological original (in our case, the structure of the DNA)" (adapted from Authors, 2019b).

Category	Definition	Examples from the students' answers	Frequencies	
			Non-CLIL students[a]	CLIL students[b]
	Students' responses refer to			
Individuality of DNA	… the individuality of the original DNA related to human beings.	„Every human being has another and an individual DNA."[c]	37	20
Different				
… interpretation	… different interpretations of the original DNA.	"Everyone has his or her own interpretation."[c]	23	23
… model design	... the different model design (e.g., different colors, materials, differing sizes or differing detail levels).	"Models can differ in their colors or sizes."[c] "Models can be of a different type."[d] „Some models are more detailed."[d]	18	52
… focus	… different perspectives and variations of focus to be applied to the model (e.g., different sections, or states of the original).	"In order to highlight different aspects."[d] "Every model has a different and specific purpose of illustration."[c]	29	68
… research state	… to new research findings about the original DNA structure.	"Models can differ because new findings are emerging constantly."[c]	3	43

Notes: [a] monolingual, sum 146 (including 36 with no or inadequate answer); [b] bilingual, sum 357 (including 151 with no or inadequate answer); [c] monolingual; [d] bilingual.

5.7 Teilarbeit E

The Language Learning Journal, Revision Round 1

Influence of COVID-19 Restrictions on Language Learning in a CLIL Outreach Genetics Laboratory

Tamara Roth[1*], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

Bilingual learning may prepare students for successful academic careers in contexts where English is not the lingua franca. One approach to familiarizing students from non-English speaking countries with the necessary cultural and communicative skills is Content and Language Integrated Learning (CLIL). Although useful in various educational contexts, only a few best practices relating to short-term variants are available. A one-day CLIL module in our outreach genetics lab is one such example wherein both scientific language learning and scientific literacy can be effectively fostered. The apparent success of this approach may lie in a systematic exposure to discourse activities combined with hands-on laboratory experimentation. Additional language scaffolding material, dictionaries, and code-switching may also positively influence the under-standing and uptake of scientific vocabulary. Thus, the short-scale CLIL program effectively combines science learning with scientific discourse and fosters English scientific language learning and scientific literacy. Yet, plummeting retention scores in the Cloze test, which deter-mines linguistic knowledge, were unexpected. We view these to a “Corona effect” since retention test data have been collected during the first nation-wide lockdown. Accordingly, we argue that language learning is largely dependent on communicative practice and that more attention should be paid to discourse activities in the classroom.

Keywords: CLIL, gene technology outreach learning, science education, scientific literacy, science communication, COVID-19 pandemic

Introduction

Fluency in English is indispensable for understanding and contributing to scientific discourse (Meyerhöffer & Dreesmann 2019). Particularly in international contexts, the spoken and written language of science is English (Oktaviani & Fauzan 2017). Still, many non-native students lack the English language competencies necessary for full engagement (Rodenhauser & Preisfeld 2014, Meskill & Oliveira 2019). The European Commission (2004) responded to this need for fluency in English by issuing Content and Language Integrated Learning (CLIL) recommendations. These recommendations emphasize English communication competencies in various subject areas and provide approaches tailoring these subjects toward improving English scientific language proficiency (Admiraal et al. 2006, Meyerhöffer & Dreesmann 2019). These recommendations have led to increased CLIL subject teaching at German schools, which has gained particular traction in science subjects (Bonnet 2012).

Classroom settings that provide the required freedom to combine both hands-on and minds-on activities relevant to the development of scientific literacy are scarce (Glynn & Muth 1994, Gonzalez-Howard & McNeill 2016, Fazio & Gallagher 2019). Still, practical CLIL-based experimentation – for instance, in outreach Biology laboratories – would offer such an environment (e.g., Rodenhauser & Preisfeld 2014, Buse et al. 2018, Roth et al. 2020). For language learning, in particular, the combination of everyday language and scaffolding with scientific language provides the necessary foundation for students to explore scientific content in a meaningful way while also encouraging scientific literacy and foreign language learning (Schoerning 2014, Ryoo 2015).

CLIL is intended to advance English language competencies to a level required for successful careers in an international academic environment (Admiraal et al. 2006, Dalton-Puffer 2007). A modified non-CLIL outreach learning module can offer a sufficient empirical baseline against which the effects of CLIL can be monitored (Mierdel & Bogner 2020, Roth et al. 2020). In line with previous studies that found CLIL to have a positive influence on English language competencies (Lorenzo et al. 2010, Dalton-Puffer 2007), increases in English language learning and scientific literacy can be expected. One key factor in improvement appears to be regular practice of the foreign language and scientific terminology (Thürmann 2008). Below, we outline the relevant theoretical background including the basic principles of foreign language learning and scientific literacy, how this learning and literacy can be achieved using content and language integrated laboratory learning, and how the COVID-19 pandemic has impacted education. All are vital to understanding the design, procedure, and results of our study.

Theoretical Background

Scientific Literacy

Existing science curricula mainly aim to develop scientific literacy by understanding and critically reflecting scientific findings (KMK 2005). In CLIL, science education still heavily relies on content learning, with language often regarded as a mere medium of instruction (Stoddart et al. 2002, Pérez Vidal & Roquet 2015). However, the importance of language in understanding, conceptualizing, reflecting, and critically discussing scientific information (Roth 2014, Meskill & Oliveira 2019) needs further consideration. Not least, because the development of scientific literacy goes beyond scientific knowledge. It also requires students to develop the reading and writing abilities necessary for evaluating and distributing scientific findings (Glynn & Muth 1994, p. 1058).

Broadly speaking, there are two key dimensions to commonly discussed definitions of scientific literacy: The first refers to the academic relevance of scientific literacy, focusing primarily on understanding scientific content and resultant scientific practices. The second goes beyond academia by encouraging students to act responsibly and in an informed way as citizens of a global society (Ke et al. 2020). Krajcik & Sutherland (2010) add to these two dimensions five key principles of scientific literacy which contribute to effective science teaching. "(1) linking new ideas to prior knowledge and experiences, (2) anchoring learning in questions that are meaningful in the lives of students, (3) connecting multiple representations, (4) providing opportunities for students to use science ideas, and (5) supporting students' engagement with the discourses of science" (p. 456 f.).

When it comes to foster scientific literacy, simply reading or producing scientific texts is not sufficient. Rather, it is students' engagement with the intricate relationship between scientific language, scientific content, and scientific practice - in critical reflection and iterative discourse – that fosters the development of scientific literacy (Moje et al. 2001, Fazio & Gallagher 2019). Thus, discipline-specific scientific content and respective literacy practices combined and integrated into a stimulating learning environment may promote scientific understanding and language (Bradbury 2014, Fazio & Gallagher 2018). The intense exercise of both language and scientific practice and the transformation of the obtained knowledge into different representational forms are understood to increase scientific literacy (Glynn & Muth 1994, Fazio & Gallagher 2019, Quarderer & McDermott 2020).

English Language Learning in Science Education

The European Commission strongly advocates for broader inclusion of foreign languages, and subject teaching in foreign languages, into regular school curricula (European Commission 2012). While language classes primarily train students to communicate at the level of everyday language, CLIL fosters expression in discipline-specific contexts using the required technical and specialized language (Fehling 2008, Meskill & Oliveira 2019). The increased use of the foreign language and the associated additional speaking and listening time is also expected to improve general foreign language competencies (Shaw et al. 2010, Fernández-Fontecha et al. 2020).

As curricula are usually already densely packed, adding extra lessons for foreign language learning is often not practical. CLIL was established as an alternative means of achieving the desired level of foreign language competencies without increasing students' workload (European Commission, 2004). Thus, the objective of CLIL is to integrate language learning and content learning, teaching students in a learning environment that better reflects the students' future employment contexts than does regular teaching (Meyerhöffer & Dreesmann 2019). Since CLIL does not involve evaluations of the learners' linguistic performance, the individual learner's linguistic repertoire in classroom discourse might benefit from expansion and interactive feedback.

Language awareness is essential in this regard. Language awareness describes "a person's sensitivity to and conscious awareness of the nature of language and its role in human life" as well as the "awareness that learners have of language, independently of conscious reflection on language" (Little 1997, p. 93). Such awareness is indispensable for both the language learner and language teacher since it contributes to effective language learning (Lo 2017, Lindahl 2020). For the learner, language awareness can be enabling and empowering and support a more self-determined language learning experience. By contrast, the instrumental function of language awareness is crucial for the language teaching process (Little 1997, Lindahl 2020). For teachers using CLIL, language awareness encompasses three integral dimensions: "(1) awareness of how language construes content; (2) awareness of how such language may impose difficulties on [foreign language] learners and how scaffolding can be provided; and (3) awareness of [foreign language] learning theories and pedagogy" (Lo 2017, p. 2).

In CLIL lesson planning for science subjects such as Biology, language is often approached as a mere mediator and means of conveying information (Rodenhauser & Preisfeld 2014). This naïve approach ignores the intricate connections between native language, foreign language, and content (Sutton 1996, Yore & Treagust 2006). It is not sensible to simply replace the L1 (first language) with the L2 (second language) as students require both to create meaning and, ultimately, knowledge (Poza 2016). One approach to integrating the L1, scientific content, and the L2 is scaffolding as "a type of teacher assistance that helps students learn new skills, concepts, or levels of comprehension of material" (Maybin et al. 1992, p.188). Gottlieb (2016) describes four types of scaffolding: linguistic, graphic, sensory, and interactive. In science subjects, graphic or visual scaffolding is easy to implement. Visualization is commonly used in non-CLIL classes to help students understand complex content and internalize the relevant scientific vocabulary. In CLIL classroom settings, visualization becomes even more essential as it also helps students to learn the required scientific vocabulary in the L2 and improves L2 literacy (Fernández-Fontecha 2020). Another way to improve scientific vocabulary and L2 literacy is repeated practice of the foreign language in realistic settings (Gonzalez-Howard & McNeill 2016). This exercise helps students process information in a coherent and accessible manner and stimulates argumentation using the learned scientific vocabulary (Nation & Newton 2008). This type of active engagement in scientific discourse is commonly identified as a key success factor for scientific literacy in L2 and L1 (Walker & Sampson 2013, Oga-Baldwin 2019).

In Germany, CLIL science learning is provided in either bilingual strands or singular CLIL modules. Singular CLIL modules do not have a set timeframe and are often used occasionally (Diehr 2012). Even such occasional use appears to positively impact second language learning (Bonnet 2004, Rodenhauser & Preisfeld 2014, Buse et al. 2018). However, the effect of CLIL on scientific content learning is ambiguous (Piesche et al. 2016, Fernández-Sanjurjo et al. 2017).

Impact of COVID-19 on Education

COVID-19 has forced a temporary closure of all schools across the nation and a shift to distance learning. Subsequent adaptations to new learning methods were primarily based on online services designed to share learning material and control learning progress (Hoenig & Wenz 2020). Yet, most learning activities have been redirected to self-study (Huber & Helm 2020), making students and parents responsible for learning success (König et al. 2020). Especially for language learning, where constant training is key to acquiring a certain proficiency level (Thürmann 2008), conversion to online learning and self-study was not expected to produce satisfactory results (Gao & Zhang 2020). Without the opportunity to prepare students, the role of teachers (Huber & Helm 2020), "changed from the traditional knowledge imparter and the classroom activity organizer to the resource integrator and the supervisor for students' autonomous learning in online teaching mode" (Gao & Zhang 2020, p. 6). While this may be a long-term goal for teaching in general (Buck et al. 2015, Nikula 2015), the rapid change under COVID-19 restrictions has put pressure on students, parents, and teachers alike to retain the previous level of education while faced with substantial limitations (Hoening & Wenz 2020, Huber & Helm 2020).

Objectives of the Study

The present study focuses on the following research questions:

3. How does a one-day CLIL science module influence overall language learning and scientific literacy in the hands-on laboratory?
4. How did the COVID-19 pandemic influence overall student performance concerning language competencies?

Thus, the specific objectives were three-fold:

- to assess students' overall language learning achievement in the laboratory
- to examine students' development of scientific literacy in the laboratory
- to determine possible effects of the COVID-19 pandemic on the retention test by comparing the students' overall language competencies

Materials and Methods

Intervention Phases

Our study involved a modified version of a previous intervention, in which the four different intervention phases were maintained (Table 1) to ensure overall comparability. To integrate CLIL, linguistic scaffolding material was added into a language exercise book. Graphic scaffolding material was also available throughout the four phases, as were visualization elements to improve understanding, which could be found in the laboratory manual (Table 1).

TABLE 1 Quasi-experimental intervention design.

Intervention phases	Evaluation variants	
	Monolingual (non-CLIL) (*n* = 145)	Bilingual (CLIL) (*n* = 107)
Instruction and conversation language English	-	+
Vocabulary scaffolding material for all intervention phases	-	+
Pre-lab	+	+
DNA-related theoretical and experimental phases		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases		
Mental modelling: text analysis	+	+
DNA modelling with craft materials	+	+
Model evaluation-1: Drawing a paper & pencil version of the crafted model and answering questions	+	+
Model evaluation-2: Comparing the crafted model with a scientific demonstration model	+	+
Interpretation	+	+

Our bilingual module lasted one full school day. It emphasized hands-on and inquiry-based learning about the structure of DNA, and bilingual conversation in English and German. The lab was adapted to the capabilities of ninth graders, where required, to improve learning outcomes. We asked students to work in pairs, using technical guidance provided in a laboratory manual to complete language scaffolding exercises designed to enhance the learning of scientific vocabulary. We selected the content of the student lab from the state's syllabus and followed national competency requirements (KMK 2005). The following chapter is summarized according to Roth et al. (2020).

Pre-lab phase. Both treatment groups – that is, the non-CLIL group in our previous treatment and the CLIL group – began with a pre-lab phase. Since most students had no prior experience with hands-on laboratory experimentation, the pre-lab phase was comparatively long. We used this phase to introduce students to laboratory techniques and theoretical concepts connected to experimentation. Specifically, we used hands-on exercises to teach students to handle the required equipment, such as micropipettes and centrifuges. The students' teachers acted as external guides for these exercises. We also introduced theoretical concepts via presentations and graphic illustrations on a poster. For practical concepts involved in experimentation, we used demonstrations. Both theory and practical exercises were vital for conducting subsequent experiments. The pre-lab phase was also important in that it could reduce the likelihood of students feeling overwhelmed in the three subsequent phases (e.g., Kalyuga 2009).

DNA-related theoretical and experimental phases. We began the second phase by introducing real-life background to the subsequent experiments (DNA relevance, Table 1), the underlying scientific concepts of DNA isolation, and gel electrophoresis. We then split the remainder of the second phase into a series of subphases, each with theoretical and experimental parts. By doing so, we reduced the risk of students becoming overwhelmed by too much new information and allowed them to focus on the forthcoming experimental phase. This precaution was essential for explaining the difficult concept of gel electrophoresis, which involves chemical knowledge about the contents of DNA and abstract thinking to imagine the processes involved.

DNA-related experimental phases. We used an evidence-based, two-step approach (Mierdel & Bogner 2019) in our experimental phases. We firstly asked students to answer questions in their laboratory manuals and think about subsequent experimental procedures. Secondly, we asked them to work in pairs to discuss every step before carrying out experiments that effectively combined hands-on and minds-on activities and required students to do more than simply follow instructions (Roth et al. 2020). For our hands-on approach, we provided sufficient scaffolding and modelling material in our laboratory manual to enable students to experiment and record their observations independently (Mierdel & Bogner 2019). Specifically, we ensured that the laboratory manual included open questions about theoretical details relevant to understanding the different experimental steps, and questions about the students' observations. Moreover, we provided indications of relevant materials for experimentations and illustrations to help the students carry out the different experimental steps. That is, we "offer authentic and direct practice with designing and interpreting controlled experiments" (Schwichow et al. 2016, p. 980). Our laboratory manual drew strongly on the manual we had designed and fine-tuned in the course of four previous laboratory evaluations (Goldschmidt & Bogner 2016, Langheinrich & Bogner 2016, Mierdel & Bogner 2020, Roth et al. 2020)

Model-related Phases. We conducted the model-related phases directly after the experimental, DNA-related phases. As models are vital to visualizing and explaining phenomena in science and science education (Krajcik & Merritt 2012), we subdivided our model-related phases into four sub-phases: a mental modelling phase that involved text analysis and a mental model or a rough sketch, a modelling phase with craft materials, a model evaluation-1 phase, and a model evaluation-2 phase (Table 1).

We used a text about the discovery of DNA's structure (Usher 2013) to familiarize students with the basic components of DNA and its structure. The students could then use this knowledge to construct a simple but accurate mental model (e.g., Mierdel & Bogner 2019, Roth et al. 2020). We then asked the students to handcraft a model of DNA using pre-selected crafting material. Once they had finished the crafting exercise, we organized two model evaluation phases that we designed as reciprocal self-evaluations. For the first phase (model evaluation-1), we asked students to draw a detailed and labeled paper and pencil version of the crafted model and answer open-ended questions to encourage self-evaluation. We used the drawing element as combined handcrafting and sketching exercises have been shown to improve learning outcomes (e.g., Prabha 2016). Students evaluated their handcrafted DNA model with the help of their paper-and-pencil versions. We used open-ended questions because reflective writing is understood to be an effective method for encouraging self-evaluation (Roth 2014, Meskill & Oliveira 2019). Our questions about model-related components encouraged students to rethink and reassess their steps and decisions in translating their mental model into its physical counterpart (Roth et al. 2020). For the second evaluation phase (model evaluation-2), we provided students with a commercially available DNA demonstration model to enable comparison-based self-evaluation.

Interpretation Phase. In this phase, we asked students to compare their hypotheses about the outcome of their experiments with images of the gel electrophoresis. We also asked them to discuss their individual models in class and compare these to the molecular DNA model by Watson & Crick.

Language Scaffolding

We designed a language scaffolding exercise book to support integrated content and foreign language learning. We included language-specific riddles such as word search puzzles and crossword puzzles that required students to match scientific terms with English definitions. Moreover, we asked students to match the terms and their definitions with given German translations or translate the terms into German. At all times, we allowed students to use English-English and English-German dictionaries, which we provided. We used one specific language scaffolding exercise for each intervention phase.

The laboratory manuals and the presentation for our theoretical phases also included short translations or explanations of important scientific terms. To help students understand the complicated procedure of gel electrophoresis, we designed a poster with English descriptions and questions. Answers to key questions were provided in German and accompanied by visuals to simplify the content. If necessary, we made use of code-switching for the explanation of important experimental steps.

Participants

Our subjects were 179 ninth graders (of upper stratification grammar schools) (girls 48.5%, boys 51.5%; $SD_{Gender} = 6.2$; $M_{\text{Age}} = 14.5$, $SD_{Age} = 0.6$). Students teamed up in 88 groups (84 2-person groups and four 3-person groups). Participation was voluntary, and we asked for written parental consent before the students' arrival. Data collection was pseudo-anonymous, and we identified the students using only pseudonymous identifiers. We designed the laboratory in line with the Declaration of Helsinki (2013), and our questionnaires were approved by the relevant German state ministry.

Variables

Students' language learning. In a series of repeat measurements, students completed a pre-test (T0) two weeks before the intervention, a post-test (T1) after they had participated in the bilingual laboratory, and a retention-test six weeks after the post-test (T2). To avoid individual preparation effects (e.g., Bogner 1995), students were not made aware of this testing schedule. For each of the three tests (T0, T1, T2), we also randomly assigned the questions and respective multiple-choice answers to avoid systemic response patterns. The Cloze test we conducted comprised 14 multiple-choice items (examples, Table 2).

TABLE 2 Cloze test item examples.

Item difficulty (T0 / T1)	Item example
CLIL learners	
84.6/85.5	Examples _____ organelles are, inter alia, … (Item 7) (a) for (b) from (c) pro (d) of*
52.9/28.5	[…] the nucleus, _____, and vacuoles of different sizes. (Item 8) (a) membrane (b) mitochondria* (c) blood cells (d) liver
83.7/87.8	_____ name shows that the relationship between …? (Item 3) (a) Her (b) Those (c) This* (d) That
55.6/35.1	[…] cells is similar to the relationship ____an organism and its organs … (Item 6) (a) inter (b) amid (c) concerned (d) between*

Note: Correct answers are marked with an asterisk *

We ensured content validity by using items consistent with the state syllabus. The heterogeneity of items used to describe complex constructs, such as language learning, confirms construct validity (Rost 2004). Cronbach's alpha scores of .61 (T0), .63 (T1), and .66 (T2) indicate acceptable internal consistency and allow for the differentiation of groups Lienert & Raatz (1998). Item difficulties (percentage of correct answers, Bortz & Döring 1995) ranged between 20% (greater difficulty) and 73% (lower difficulty). In our CLIL intervention, the item difficulties slightly improved from T0 to T1 (Figure 1).

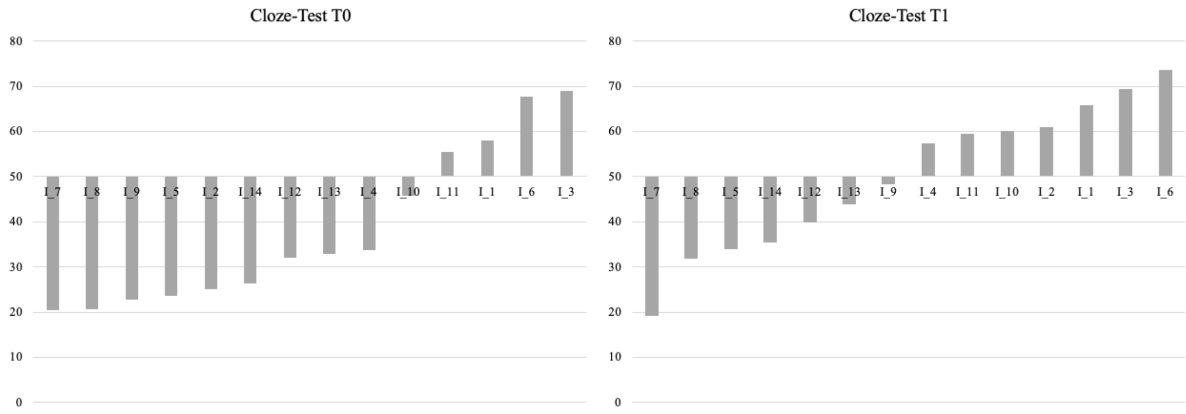


FIGURE 1 Item difficulties for the Cloze test at two different testing times (T0 and T1)

We calculated the students' test scores and analyzed these for increases in language knowledge (T1 minus T0) and retention rate (T2 minus T0). Moreover, we calculated the actual language learning success in relation to the maximum attainable score (14 correct answers): (T1 – T0) x (T1/14); and the continued language learning success (T2 – T0) x (T2/30) (Scharfenberg et al. 2007). We weighted increased language according to students' individual language knowledge, making it possible to compare language learning achievement despite some students exhibiting a considerable increase in language knowledge yet low final scores, and *vice versa*. We also calculated correlations between English grades and post-test (T1) scores for knowledge items using Spearman-Rho (Field 2012).

Analyses of open questions. We used content analysis (Bos & Tarnai 1999) to iteratively categorize the statements students made in response to the open questions. We randomly selected 52 out of 179 student answers (example Table 3) for a second scoring (30%). We computed Cohen's kappa coefficient (Cohen 1968) scores of .87 and .69 for intra-rater and inter-rater reliability, which showed a "substantial" to "almost perfect" rating (Wolf 1997, p. 964).

TABLE 3 Example question and respective statements of different content qualities.

Question: Explain the role of the lysis buffer!					
School number	Group	Answer	Language score	Content score	Overall score
22	1	Destroys the membrane of the cell and core	2	4	6
	2	Destroyed cell membrane and nuclear membrane	2	4	6
	3	It destroys the cell membrane	2	3	5
	4	Destroys cell- and nucleus membrane	2	4	6
	5	It should destroy the cell membrane and the nucleus	2	4	6
	6	It destroys the cell and the nucleus	2	3	5
	7	The use of lysis buffer is to destroy the membrane of the cell	2	3	5
	8	The lysis buffer destroys the membrane of the DNA	2	2	4
	9	It destroys the cell membrane	2	3	5
	10	Deytroyes the cell membrane	2	3	5

We used a three-step approach to evaluate the open questions (Table 3).

- Assessment of students' language use: We assigned two points if the students used English in their answers, one point if they answered in German, and zero points if they did not answer the question. We did not punish spelling errors or grammatical inconsistencies as long as the statement was legible.
- Assessment of students' content: We independently assessed the content. We gave one point for correct or partially correct responses. Maximum scores varied across questions between two and six points.
- Assessment of students' overall score: We added students' scores for language use and content assessment

These enabled us to determine the extent to which students had correctly understood the question and answered accordingly.

Statistical Analysis. We applied nonparametric methods due to the variable's non-normal distribution (Kolmogorov-Smirnov test (Lilliefors modification): partially $p < .001$), and, consequently, use boxplots to illustrate our results. We used the Friedman test (F) to calculated intra-group differences in the results of the Cloze test over the three test dates and conducted pairwise analysis from T0 to T1 and T2, and from T1 to T2, using the Wilcoxon (W) signed-rank test. We used Mann-Whitney *U* tests (MWU) to evaluate inter-group differences between male and female students. To account for multiple testing, we applied a Bonferroni correction (Field 2012). In the case of significant results, we calculated effect sizes *r* (Lipsey & Wilson 2001) with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. For correlation analyses, we applied Spearman's rank correlations and report Spearman's Rho values.

Results

We firstly provide an overview of our detailed assessment of the students' scientific literacy by analyzing language, content, and overall performance in responses to the open questions in our laboratory manual. Secondly, we present an intra- and inter-group analysis of linguistic language competencies evident in the results of our Cloze test.

Analyses of open questions. We compared answers to questions (such as those shown in Table 3) and calculated independent scores for language (Chronbach's $\alpha = .86$), content (Chronbach's $\alpha = .74$), and overall performance (Chronbach's $\alpha = .83$). Thereafter, we computed sum scores for all three categories and compared these scores with regard to gender. We did not identify any significant gender differences for our CLIL approach in relation to language

(MWU: Z = -.253, p = .800), content (MWU: Z = -1.101, p = .271), or overall (MWU: Z = -.406, p = .685) performance (Figure 2).

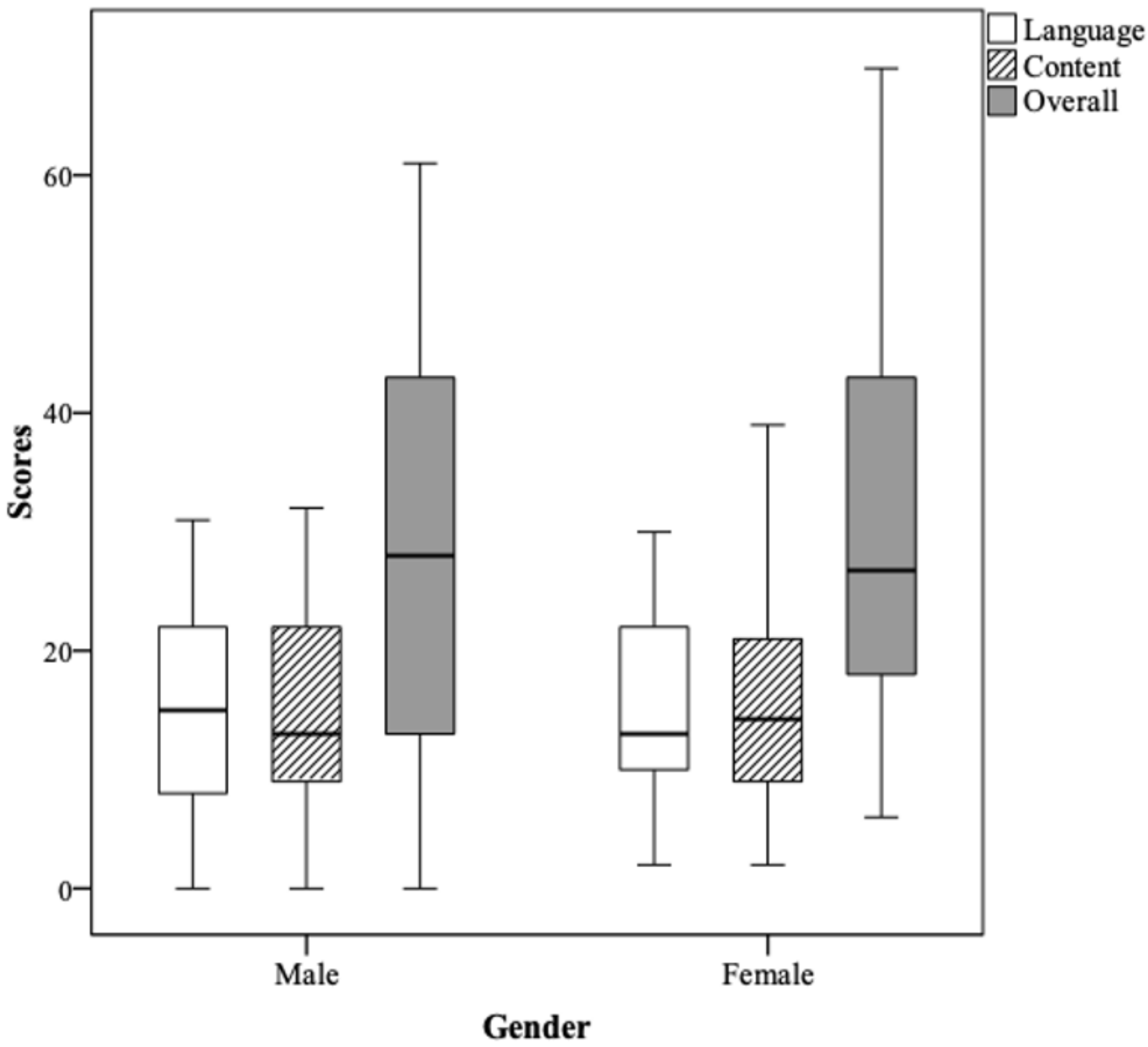


FIGURE 2 No significant gender differences between language (Mdnmale = 15.0, 25th = 8.0, 75th = 22.0; Mdnfemale = 13.2, 25th = 10.0, 75th = 22.0), content (Mdnmale = 12.8, 25th = 8.0, 75th = 22.5; Mdnfemale = 14.3, 25th = 9.0, 75th = 21.0), and overall performance (Mdnmale = 27.8, 25th = 13.0, 75th = 44.0; Mdnfemale = 26.9, 25th = 18.0, 75th = 43.0) of students in answering open questions in the laboratory manual.

Intra-group Analyses of Language Learning

Intragroup analysis (Friedman and Wilcoxon tests) revealed significant changes for CLIL learners when it came to linguistic language knowledge scores: The students initially increased their knowledge from T0 to T1, which then dropped between T1 and T2, this time below T0 levels (Figure 3, left).

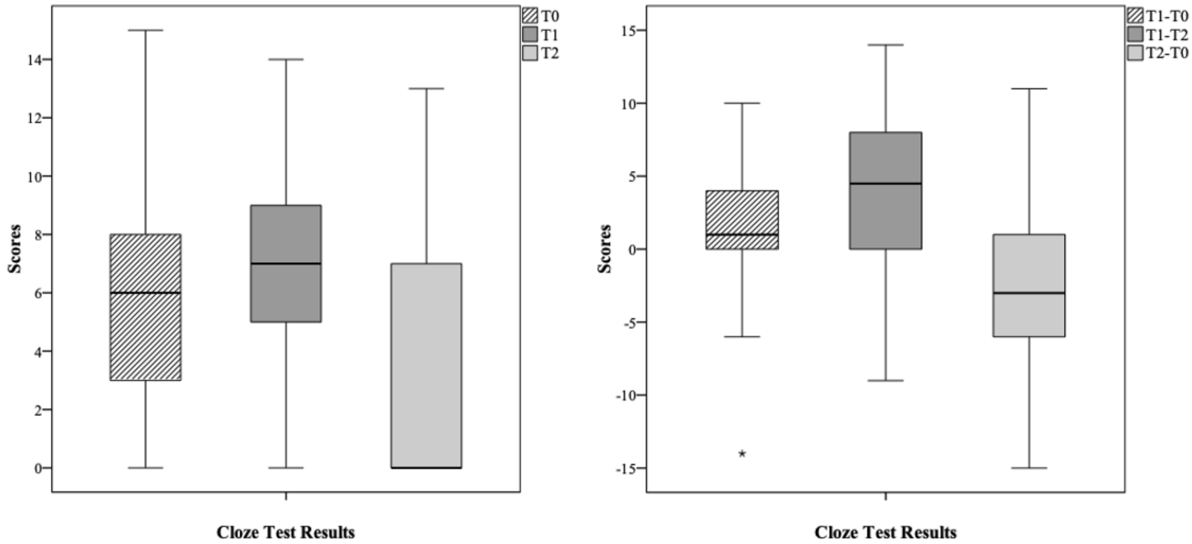


FIGURE 3 Differences for language-related knowledge acquisition at the three different testing times.

Our Friedman test for CLIL learners was significant (*p* < .001, *df* = 2, χ^2= 29.99), confirming differences between the three testing times (Field 2012). Wilcoxon signed-rank test explicated these differences for T0 to T1 (*p* < .001, *Z*= -4.968), for T0 to T2 (*p* < .001, *Z* = -4.405), and for T1 to T2 (*p* = .668, *Z* = -.429). This suggests that students increase their knowledge (*p* < .001, *Z* = -4.405) and gain short-term linguistic language knowledge (*p* < .001, *Z* = -3.242) throughout the intervention.

Inter-group Analyses of Linguistic Language Knowledge

We calculated difference variables for short-term increases in knowledge, mid-term retention rates, and learning success variables to assess inter-group differences and account for differences in students' prior knowledge. Based on sum scores (MWU_{T0}: *p* = .039, *Z* = -2.067; MWU_{T1}: *p* = .087, *Z* = -1.710; MWU_{T2}: *p* = .982, *Z* = -.022) increases in knowledge (T1-T0) and retention rate (T2-T0) were calculated (Field 2012) for linguistic knowledge scores.

Scores showed no differences in increases in linguistic knowledge (MWU: *p* = .648, *Z* = -.457) and retention rates (*p* = .082, *Z* = -1.742) between male and female learners. As difference variables do not display actual cognitive achievement, we analyzed learning success variables. For short-term actual (*p* = .173, *Z* = -1.363) and mid-term persistent (*p* = .244, *Z* = -1.164) learning success, no significant differences in linguistic knowledge were identified.

Discussion

A one-day CLIL outreach module can positively affect linguistic knowledge by allowing access to scientific content through language. This effect holds, regardless of gender. The same is true for a deeper understanding of the represented scientific content and underlying scientific language. All findings are in line with the extant literature (e.g., Glynn & Muth 1994). The combination of minds-on and hands-on activities may be the key to increase scientific literacy (Ke et al. 2020). On such a basis, hands-on scientific content in combination with scientific language scaffolding prepares the linguistic basis for effectively discussing observed phenomena and for constructing the ensuing models (Lemke 1990; Berland & Reiser 2009).

Linguistic Knowledge

The Cloze test revealed a shift from only four out of fourteen correctly answered items (over 50%) in the pre-test to over seven out of fourteen in the post-test, indicating that CLIL had a significant influence. This is quite impressive for a short-scale CLIL implementation, especially in the context of outreach, hands-on learning (e.g., Rodenhauser & Preisfeld 2014, Buse et al. 2018). Such effects on language learning are usually only reported for long-term interventions, giving hope for the success of short yet intense CLIL phases (e.g., Lorenzo et al. 2010, Dalton-Puffer 2007).

After an initial increase, linguistic knowledge dropped six to eight weeks after participation. Surprisingly, these subsequent scores for linguistic knowledge were lower than those observed prior to participation. This is rather uncommon, especially after intense training (Thürmann 2008), as second language learning is generally perceived to initially develop as "context embedded", becoming increasingly "decontextualized" over the course of several years (e.g., Rosenthal 1993, p. 438). Moreover, school learning of the biological disciplines of biotechnology and genetics involves many English terms (Meyerhöffer & Dreesmann 2019) which should, thus, be familiar to students (e.g., transcription, translation). Consequently, cognitive overload may be ruled out as an explanation for the observed drop in linguistic knowledge (Sweller 2015, Ryoo 2015). Thus, individual language competencies may have been affected by the severe lockdown measures introduced in response to the COVID-19 pandemic on the last day of the CLIL module phase. As all schools across Germany switched to distance learning for several weeks (König et al. 2020, Hoenig & Wenz 2020) students often had no permanent access to computers or internet connection, which impaired access to instructional materials, including those relevant for language learning and training. Teachers were also not able to verify that students effectively engaged with their tasks and materials (Fraillon et al. 2019).

Effective language learning requires continual practice to reduce errors, perpetuate knowledge, and improve literacy (Thürmann 2008, Pérez-Vidal & Roquet 2015). If the students had not practiced the foreign language for weeks, low scores for subject-specific vocabulary in the foreign language and subject knowledge would not be surprising (Thürmann 2008). A related study reported a cut in learning time from 7.4 to 3.6 hours a day during distance learning (Wößmann et al. 2020).

Switching from controlled classroom environments to online-supported self-learning placed students in an unfamiliar situation (König et al. 2020). Yet, attempts to increase instructional time devoted to meaningful learning and "prevent student misbehavior caused by boredom, overload, disorientation, [or] difficulties in understanding […]" have been made (Huber & Helm 2020, p. 244). Despite these efforts, clear disadvantages associated with online language learning remain (Gao & Zhang 2020). Thus, the drop in linguistic knowledge scores may indicate that distance learning cannot effectively replace face-to-face teaching when it comes to language learning.

Scientific Literacy

The laboratory manual questions were answered just by 48% of our participants (while just 2% provided answers in the national language of German). This result is consistent with Cheshire & Gardner-Chloros (1998), who noted that the language used is not necessarily related to understanding. In CLIL learning, for instance, students are sometimes even encouraged to answer in German rather than not responding if they feel they lack the required vocabulary (KMK 2013).

Since learning English and processing scientific content simultaneously can be cognitively demanding, sufficient support is needed to avoid mental overload (Ryoo 2015, Sweller 2015). Code-switching – as we did with oral instructions – or written answers should maintain word flow and may allow for single words or whole sentences in the native language to be used in argumentation (Cheshire & Gardner-Chloros 1998). Additional scaffolding can also help to reduce cognitive load (Coyle 2007, Grandinetti et al. 2013) by spreading it across the four possible dimensions (linguistic, graphic, sensory, and interactive) (Gottlieb 2016). We used all of these methods to help students develop content and linguistic knowledge. Since students have to understand both the scientific content and the underlying scientific vocabulary to increase scientific literacy, it is essential to include these in equal parts (Unsal et al. 2018, France 2019).

Scientific literacy primarily requires minds-on-hands-on activities and vocabulary to help students explain and abstract scientific content (Glynn & Muth 1994, France 2019). Our module contained both activities: (1) students had to understand and conduct experiments independently, using information from the prelab phase and course manual; (2) students had to transform their knowledge from the laboratory into different forms of representation; (3) students had to effectively communicate the ideas they generated. Respective discipline-specific terminology and rhetoric were, thereby, relevant (Norris & Phillips 2003, Prain 2004) as scientific language significantly differs from everyday language (Wang & Chen 2014). This was particularly useful for the minds-on activities, wherein we asked students to summarize observations gathered from their hands-on activities. To avoid mental overload, we used appropriate scaffolding material and encouraged students to orally summarize the procedures in their own words (Sweller 2015).

These time-consuming actions may be a possible explanation for scientific literacy and foreign language learning scores in the CLIL module: most students answered the open questions correctly but had only time to respond to a selection of questions (Glynn & Muth 1994, Meyerhöffer & Dreesmann 2019). Finding answers to the 'how' and 'why' requires time to think – since students propose scientific hypotheses and arguments – while stimulating in-depth exploration of the topic. It also encourages the development of scientific terminology, scientific content, and, finally, scientific literacy (Quarderer & McDermott 2020, France 2019).

Gender Differences

The recent PISA study has revealed similar scores in scientific literacy for both genders (Reiss et al. 2018). In general, this lack of gender differences is in line with our results since, on average, both males and females performed equally well when answering the open questions in our laboratory manual. Even the Cloze test did not reveal any significant gender differences. While such results are encouraging, many other studies still observe substantial gender differences when it comes to language learning (e.g., Andreou et al. 2005, van der Slik et al. 2015). That is, specific social and classroom factors are understood to positively influence language learning in female students but may show adverse effects in males. At worst, this results in a significant reduction in intrinsic motivation and self-confidence among boys (Mady & Seiling 2017). To avoid discrimination, teaching strategies and learning environments that equally engage both genders with the foreign language must be created (Bećirović 2017). Yet, gender differences are not only an issue with language but also science learning. Girls, in particular, show a much lower self-confidence despite higher scores in science subjects and are less likely to pursue a STEM career (Delaney & Devereux 2019). Thus, literature also suggests a stimulating and motivating environment that equally engages both genders (Mady & Seiling 2017, Mierdel & Bogner 2019, Conradty & Bogner 2019). Since our CLIL outreach module effectively combines language and science learning activities in a care-free albeit demanding hands-on-minds-on environment, males and females both seem to be motivated, with neither gender favored nor disadvantaged as compared to the other. This may have led to similar overall scores for both language and science learning.

Limitations

Firstly, our ninth graders had limited prior experience with hands-on experimentation and with English use outside of language lessons. Secondly, the Cloze test and the evaluation of manuals only give indications of language learning and scientific literacy as there is still a lack of commonly-agreed standardized instruments (Norris & Philips 2003, Rodenhauser & Preisfeld 2014). Thirdly, due to the context-dependency of CLIL learning, results cannot easily be extrapolated (Pérez-Cañado 2012), which also applies to the drop in linguistic knowledge scores for T2 as a potential result of COVID-19 lockdown policies (Thürmann 2008, König et al. 2020, Hoenig & Wenz 2020). As a consequence, generalizing the success of our CLIL module would require more research regarding short-scale implementations of CLIL in combination with language learning and scientific literacy (Fernández-Sanjurjo et al. 2017).

Conclusion

The success of short-term CLIL modules to foster language learning and scientific literacy relies heavily on intensive hands-on and minds-on activities. Moreover, the retention of linguistic knowledge relies on frequent training (Thürmann 2008), which COVID-19 lockdown restrictions prevented (Gao & Zhang 2020). So far, educational studies of the pandemic have mainly explored the possible general consequences (König et al. 2020, Hoenig & Wenz 2020, Wößmann et al. 2020). Nevertheless, as previous studies of language learning and scientific literacy (e.g., Norris & Phillips 2003, Prain 2004) have also predominantly focused on long-scale CLIL modules (Meyerhöffer & Dreesmann 2019), our study provides relevant information to the scientific community

Although only about 50% of our students answered our laboratory manual questions, our study still holds positive implications for language learning and scientific literacy. The Cloze test also indicated that CLIL learning has a positive impact on linguistic understanding and language learning. The development of language learning and scientific literacy is rather cumbersome and requires stimulating environments (e.g., Norris & Phillips 2003, Prain 2004), and CLIL offers an approach combining hands-on laboratory experiments with language learning via the much-discussed minds-on activities (Rodenhauser & Preisfeld 2014, Buse et al. 2018). Although we cannot pinpoint the reasons why some students did not answer the laboratory manual questions, in line with cognitive load theory (Sweller 2015) we would encourage a content reduction in later implementations.

Acknowledgements

This study was supported by University of Bayreuth in the APC funding program ‘Open Access Publishing’ and the ‘Qualitätsoffensive Lehrerbildung’ Program funded by the German Federal Ministry of Education and Research. Special thanks to the teachers and students involved in this study for their cooperation.

Ethical Statement

a) Hereby, we, Tamara Roth and Franz X. Bogner, consciously assure that for the manuscript “Content and Language Integrated Scientific Modelling: A Novel Approach to In-depth Model Understanding” the following is fulfilled:

1) This material is the authors’ own original work, which has not been previously published elsewhere.
2) The paper is not currently being considered for publication elsewhere.
3) The paper reflects the authors’ own research and analysis in a truthful and complete manner.
4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
5) The results are appropriately placed in the context of prior and existing research.
6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
7) All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

b) Moreover, “all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.”

Consent Statement

“Informed consent was obtained from all individual participants included in the study.”

References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12, 75–93.

Andreou, G., Vlachos, F., & Andreou, E. (2005). Affecting factors in second language learning. *Journal of Psycholinguistic Research*, 34, 429-438.

Bećirović, S. (2017). The relationship between gender, motivation and achievement in learning English as a foreign language. *European Journal of Contemporary Education,* 6(2), 210-220.

Berland, L.K., & Reiser, B.J. (2009). Making sense of argumentation and explanation. *Science Education*, 93(1), 26–55.

Bogner, F.X. (1995). Eine bilinguale Unterrichtseinheit zum Thema Glas (mit konsequenter Umsetzung des Erkundungsprinzips vor Ort). *Chemkon*, 2(1), 9-16.

Bonnet, A. (2004). *Chemie im bilingualen Unterricht: Kompetenzerwerb durch Interaktion*. [Chemistry in bilingual education: competencies through interaction]. Opladen: Leske & Budrich.

Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal*, 1(4), 65–78.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation* [Research methods and evaluation]. Berlin: Springer.

Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, *31*, 659–671.

Bradbury, L. (2014). Linking science and language arts: a review of the literature which compares integrated versus non-integrated approaches. *Journal of Science Teacher Education*, 25(4), 465–488.

Buck, G., Mast, C., Ehlers, N., & Franklin, E. (2005). Preparing teachers to create a mainstream science classroom conducive to the needs of English-language learners: a feminist action research project. *Journal of Research in Science Teaching,* 42 (9), 1013-1031.

Buse, M., Damerau, K. & Preisfeld, A. (2018). A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts. *International Journal of Environmental & Science Education*, 13(8), 647-660.

Cheshire, J., & Gardner-Chloros, P. (1998). Code-switching and the sociolinguistic gender pattern. *International Journal of the Sociology of Language,* 129, 5–34.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.

Conradty, C., & Bogner, F.X. (2019). From STEM to STEAM: Cracking the code? How creativity & motivation interacts with inquiry-based learning. *Creativity Research Journal*, 31(3), 284-295.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543-562.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated learning (CLIL) Classrooms*. Amsterdam: John Benjamins.

Declaration of Helsinki, World Medical Association (2013). *Journal of the American Medical Association*, 310, 2191-2194.

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. Economics of Education Review, 72, 219–238.

Diehr, B. (2012). What´s in a name? Terminologische, typologische und programmatische Überlegungen zum Verhältnis der Sprachen im Bilingualen Unterricht. [What's in a name? Reflections on terminology, typology and programme regarding the relation of languages in bilingual education]. In B. Diehr & L. Schmelter (Eds.), Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven (pp.17-36). Frankfurt a.M.: Peter Lang.

European Commission (2004): *Promoting language learning and linguistic diversity. An action plan 2004-06*. Luxembourg, European Communities.

European Commission (2012). *FAQs on multilingualism and language learning*. Brussels. URL: http://europa.eu/rapid/press-release_MEMO-12-703_en.pdf (accessed: 2 March 2021)

Fazio, X., & Gallagher, T.L. (2018). Bridging professional teacher knowledge for science and literary integration via design-based research. *Teacher Development*, 22(2), 267-280

Fernández-Fontecha, A., O'Halloran, K.L., Wignell, P., & Tan, S. (2020). Scaffolding CLIL in the science classroom via visual thinking: A systemic functional multimodal approach. *Linguistics and Education,* doi: 10.1016/j.linged.2019.100788.

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J.M. (2017). Analysing students' content-learning in science in CLIL vs. Non-CLIL programmes: empirical evidence from Spain. International *Journal of Bilingual Education and Bilingualism*, 22(6), 661-674.

Field, A. (2012). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.

Fraillon, J., Ainley, J, Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: the IEA international computer and information literacy study 2018 international report*. New York: Springer.

France, A. (2019). Teachers using dialogue to support science learning in the primary classroom. *Research in Science Education*. https://doi.org/10.1007/s11165-019-09863-3.

Gao, L.X., & Zhang, L.J. (2020). Teacher learning in difficult times: examining foreign language teachers' cognitions about online teaching to tide over COVID-19. *Frontiers in Psychology,* doi: 10.3389/fpsyg.2020.549653.

Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: achieving scientific literacy. *Journal of Research in Science Teaching*, 31(9), 1057-1073.

Goldschmidt, M., & Bogner, F.X. (2016). Learning about genetic engineering in an outreach laboratory: Influence of motivation and gender on students' cognitive achievement. *International Journal of Science Education*, 6(2), 166-187.

Gonzalez-Howard, M. & McNeill, K.L. (2016). Learning in a community of practice: factors impacting English-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, 53(4), 527-553.

Gottlieb, M. (2016). *Assessing English language learners: Bridges for language proficiency to academic achievement*. Thousand Oaks: Corwin.

Grandinetti, M.M., Langellotti, M., & Teresa Ting, Y.L. (2013). How CLIL can provide a pragmatic means to renovate science education – even in a sub-optimally bilingual context. *International Journal of Bilingual Education and Bilingualism*, 16(3), 354-374.

Hoening, K., & Wenz, S.E. (2020). Education, health behavior, and working conditions during the pandemic: evidence from a German sample. *European Societies,* doi: 10.1080/14616696.2020.1824004.

Huber, S.G., & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises - reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability,* 32, 237-270.

Kalyuga S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction*, *19*, 402-410.

Ke, L., Zangori, L., Sadler, T.D., & Friedrichsen, P.J. (2020). Integrating scientific modeling and socio-scientific reasoning to promote scientific literacy. In Warden A. Powell (Ed.), *Socioscientific issues-based instruction for scientific literacy development* (pp. 31-56). Hershey: IGI Global.

KMK (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

KMK (2013). *Konzepte für den bilingualen Unterricht - Erfahrrungsbericht und Vorschläge zur Weiterentwicklung* [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany – concepts for bilingual teaching – progress report and suggestions for further development]. Munich: Luchterhand.

König, J., Jäger-Biela, D.J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education*, 43(4), 608-622

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *Science and Children*, *49*, 10–13.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science*, 328 (5977), 456–459.

Langheinrich, J., & Bogner, F.X. (2016). Computer-related self-concept: The impact on cognitive achievement. Studies in Educational Evaluation, 50, 46-52.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lindahl, K. (2020). Connecting ideology and awareness: critical multilingual awareness in CLIL contexts. *English Teaching & Learning,* 44, 211-228.

Lipsey, M. W., & Wilson, D. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage Publications.

Little, D. (1997). Language awareness and the autonomous language learner. *Language Awareness*, 6(2-3), 93-104.

Lo, Y.Y. (2017). Development of the beliefs and language awareness of content subject teachers in CLIL: does professional development help?. *International Journal of Bilingual Education and Bilingualism*, doi: 10.1080/13670050.2017.1318821.

Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning (CLIL) in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31, 418–442.

Mady, C., & Seiling, A. (2017). The coupling of second language learning motivation and achievement according to gender. *Theory and Practice in Language Studies,* 7(12), 1149-1159.

Maybin, J., Mercer, N. & Stierer, B. (1992). Scaffolding: learning in the classroom. In Kate Norman (Ed.), *Thinking Voices. The Work of the National Oracy Project* (pp.186–95). London: Hodder & Stoughton.

Meskill, C., & Oliveira, A.W. (2019). Meeting the challenges of English learners by pairing science and language educators. *Research in Science Education*, 49, 1025-1040.

Meyerhöffer, N. & Dreesmann, D.C. (2019). English-bilingual Biology for standard classes development, implementation and evaluation of an English-bilingual teaching unit in standard German high school classes. *International Journal of Science Education*, 41(10), 1366-1386.

Mierdel, J., & Bogner, F.X. (2019). Comparing the use of two different model approaches on students' understanding of DNA models. *Education Sciences*, 9(2), 115-133.

Mierdel, J., & Bogner, F.X. (2020). Simply InGEN(E)ious! How creative DNA modeling can enrich classic hands-on experimentation. *Journal of Microbiology and Biology Education*, 21(2), 21-45.

Moje, E.B., Collazo, T., Carrillo, R., & Marx, R.W. (2001). "Maestro, what is 'quality'?": language, literacy, and discourse in project-based science. *Journal of Research in Science Teaching,* 38(4), 469-498.

Nation, I. S. P., & Newton, J. (2008). *Teaching ESL/EFL listening and speaking*. London: Routledge.

Ngai, P. B.-Y. (2002). Bilingual education for all: A benefits model for small towns. Bilingual Research Journal, 26(2), 269-294.

Nikula, T. (2014). Hands-on tasks in CLIL science classrooms as sites for subject-specific language use and learning. *System*, 1-14.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224–240.

Oga-Baldwin, Q.W.L. (2019). Acting, thinking, feeling, making, collaborating: the engagement process in foreign language learning. *System*, 86, 1-10.

Oktaviani, A., & Fauzan, A. (2017). Teachers perceptions about the importance of English for young learners. *Linguistic, English Education and Art (LEEA) Journal*, 1(1), 1-15.

Pérez Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: an analysis measuring relative gains. *System.* doi: 10.1016/j.system.2015.05.004

Piesche et al. (2016). CLIL for all? A randomized controlled field experiment with sixth grade students on the effects of content and language integrated science learning. *Learning and Instruction*, 44, 108–116.

Poza, L.E. (2016). The language of ciencia: translanguaging and learning in a bilingual science classroom. *International Journal of Bilingual Education and Bilingualism*, dio: 10.1080/13670050.2015.1125849

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. *European Scientific Journal*, *12*, 235–250.

Prain, V. (2004). The role of language in science learning and literacy. In *Writing and learning in the science classroom* (pp. 33–45). Dordrecht: Springer.

Quarderer, N.A., & McDermott, M.A. (2020). Examining science teacher reflections on argument-based inquiry through a critical discourse lens. *Research in Science Education*, 50, 2483-2504.

Reiss, K., Weiß, M., & Klieme, E. (2018). PISA 2018 - Grundbildung im internationalen Vergleich [PISA 2018 – International comparison of basic education]. In Olaf Köller (Ed.) *OECD Programme for International Student Assessment (PISA).*

Rodenhauser, A. & Preisfeld, A. (2014). Bilingual (German-English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. *International Journal of Environmental and Science Education*, 10(1), 99-110.

Rost, J. (2004). *Lehrbuch Testtheorie–Testkonstruktion* [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

Roth, T., Scharfenberg, F-J., Mierdel, J., & F.X. Bogner. (2020). Self-evaluative scientific modeling in an outreach gene technology laboratory. Journal of Science Education and Technology, 29, 725-739. https://doi.org/0.1007/s10956-020-09848-2.

Roth, W-M. (2014). Science language wanted alive: through the dialectical/dialogical lens of Vygotsky and the Bakhtin Circle. *Journal of Research in Science Teaching,* 51(8), 1049-1083.

Rosenthal, J.W. (1993). Theory and practice: science for undergraduates of limited English proficiency. *Journal of Science Education and Technology,* 2(2), 435-443.

Ryoo, K. (2015). Teaching science through the language of students in technology-enhanced instruction. *Journal of Science Education Technology,* 24, 29-42.

Scharfenberg, Franz-Josef (2005): *Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse.* [Experiment-based biology education for genetics outreach learning: emprical research regarding acceptance, knowledge acquisition, and interest] (am Beispiel des Demonstrationslabors Bio- /Gentechnik der Universität Bayreuth mit Schülern aus dem Biologie-Leistungskurs des Gymnasiums). Dissertation. Bayreuth.

Schoerning, E. (2014). The effect of plain-English vocabulary on student achievement and classroom culture in college science instruction. *International Journal of Science and Mathematics Education*, 12, 1-21.

Schwichow, M., Zimmerman, C., Croker, S., Härtig, H. (2016). What students learn from hands-on activities. *Journal of Research in Science Education*, 53(7), 980-1002.

Shaw, J.M., Bunch, G.C., & Geaney, E.R. (2010). Analyzing language demands facing English learners on science performance assessments: the SALD framework. *Journal of Research in Science Teaching,* 47(8), 909-928.

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integration inquiry science and language development for English language learners. *Journal of Research in Science Teaching,* 39(8), 664-687.

Sutton, C. (1996). Beliefs About Science and Beliefs About Language. *International Journal of Science Education*, 18 (1), 1–18.

Sweller, J. (2015). In academe, what is learned and how is it learned? *Current Directions in Psychological Science*, 24, 190-194.

Thürmann, E. (2008): Eine eigenständige Methodik für den bilingualen Sachfachunterricht? [Methodology for bilingual learning] In: Gerhard Bach und Susanne Niemeier (Hg.): *Bilingualer Unterricht. Grundlagen, Methoden, Praxis, Perspektiven.* Frankfurt am Main: Peter Lang, S. 71–89.

Unsal, Z., Jakobson, B., Molander, B.-O., & Wickman, P.-O. (2018). Language use in a multilingual class: A study of the relation between bilingual students' languages and their meaning-making in science. *Research in Science Education*, 48(5), 1027–1048.

Usher, S. (2013). *Letters of Note. Correspondence Deserving of a Wider Audience*. Edinburgh, Scotland: Canongate.

van der Slik, F.W.P., van Hout, R.W.N.M., & Schepens, J.J. (2015). The gender gap in second language acquisition: gender differences in the acquisition of dutch among immigrants from 88 countries with 49 mother tongues. *PLOS ONE*, 10(11), doi.org/10.1371/journal.pone.0142056.

Walker, J.P. & Sampson, V. (2013). Learning to argue and arguing to learn – driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *Journal of Research in Science Teaching*, 50(5), 561-596.

Wang, J-R., & Chen, S-F. (2014). Exploring mediating effect of metacognitive awareness on comprehension of science tests through structural equation modeling analysis. *Journal of Research in Science Teaching,* 50(2), 175-191.

Wößmann, L., Freundl, V.Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2020). Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen? [Education in the corona crisis: how have school children spent the period of school closures, and what educational measures do Germans support]. *Ifo Schnelldienst*, 73 (9), S. 25–39.

Wolf, R. (1997). Rating scales. In J. Keeves (Ed.), *Educational research, methodology and measurement: an international handbook* (pp. 958–965). Oxford, UK: Elsevier.

Yore, L.D., & Treagust, D.F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, 28(2–3), 291–314.

5.8 Teilarbeit F

Studies in Educational Evaluation, Revision Round 1

The Relevance of School Self-Concept and Creativity for CLIL Outreach Learning

Tamara Roth[1*], Cathérine Conradty[1], Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de

Abstract

Content and Language Integrated Learning (CLIL) combined with experimental outreach learning activities may affect the inherently correlated constructs creativity and self-concept. Evaluation of our one-day CLIL genetics lab confirms this correlation for both Biology ($p < .001$, $r = .739$) and English ($p < .001$, $r = .802$) with model-fit parameters of CFI = .943 and TLI = .936 for creativity and self-concept in Biology and CFI = .943 and TLI = .936 for creativity and self-concept in English. Participation did not produce any gender differences except for the latent variable Social of the self-concept scale. Moreover, self-concept correlates with short- and long-term cognitive achievement and long-term linguistic knowledge, confirming the mutually reinforcing relationship between self-concept and academic achievement. For Biology and English grades, negative correlations with self-concept indicate effective support for low achievers. In consequence, an increased use of short-term CLIL interventions may enable creativity and a healthy self-concept.

Keywords: Creativity. School Self-Concept. Science and Language Learning. Science Education. Hands-on Outreach Learning. Gender.

Introduction

Integrating creativity into science education promises to make the provision and management of knowledge more sustainable (Henriksen et al. 2018). Creativity is now understood to support the acquisition, transfer, and application of knowledge at schools and, hence, to contribute to an innovative national economy (Lewis 2008). Distinct from, yet connected to, creativity (e.g., Sisk 1972, Bournelli et al. 2009) is self-concept, which also contributes significantly to the development of individual academic capabilities (e.g., Jansen et al. 2015, Hoferichter et al. 2018). A more positive self-concept in a specific subject is linked to better skills, and vice versa (Marsh et al. 2015). This two-way relationship is also a key factor in the development of scientific literacy (Wang et al. 2008).

To be scientifically literate, a student must be able to read, communicate, and critically discuss scientific evidence and their own hypothesis (Glynn & Muth 1994). This goes beyond traditional knowledge acquisition and, in combination with creativity, may foster the development of know-how and new scientific concepts (Brynjolfsson & McAfee 2014). Since it is increasingly evident that the self-perpetuating system of "linearity, conformity, and standardization" does not advance ‘organic, adaptable and diverse’ societies (Robinson 2011, p.4), learning environments are required that allow for open, individual, and self-directed exploration of scientific content (Richards & Cotterall 2016, Ke et al. 2020).

Such informal approaches can be realized as Content and Language Integrated Learning (CLIL) modules, wherein both previously unknown scientific contents and specific scientific vocabulary are conveyed in the students’ native and the foreign language (e.g. Rodenhauser & Preisfeld 2014, Buse et al. 2018). Yet, such learning places particularly high demands on participating students, often beyond individual mental capabilities (e.g. Meyerhöffer & Dreesmann 2019).

To explore the proposed effect of creativity and self-concept on linguistic and scientific knowledge in CLIL environments, we piloted a genetics outreach education module structurally identical to a previously conducted German module (Roth et al. 2020). Following the hypothesized relationship between self-concept and creativity (e.g., Sisk 1972, Mawang et al. 2018), we first explored their interaction using the School Self-Concept (SESSKO, Schöne et al. 2002) questionnaire and the Cognitive Processes Associated with Creativity (CPAC, Miller & Dumford 2016) scale. We also investigated correlations between school self-concept for Biology and English and respective grades, as well as cognitive achievement scores. Self-concept is already understood to be mutually influenced by school achievement (e.g. Bakadorova &

Raufelder 2020) but may differ with regard to gender, independent of academic performance (Jansen et al. 2014, Mullis et al. 2020). Consequently, we also explored gender differences for the latent variables Criterial, Individual, Social, and Absolute of the SESSKO scale. To assess the scales' interaction with each other and academic achievement, in addition to possible gender differences, we first outline the relevant theoretical background, including the basic principles of school self-concept, creativity in science education and CLIL learning. These are important for understanding the design and process of our study.

Theoretical Background

School Self-Concept

Self-concept is understood to influence achievement and performance in academic contexts (Marsh & Martin 2011, Jansen et al. 2015). Students' reflections about their capabilities and the potential to further develop their skills in the classroom may either positively or negatively impact their motivation to retain, expand, and demostrate their skills. Thus, it is essential to create situations where students can develop a positive albeit reflected self-concept (Patall et al. 2014, Bakadorova & Raufelder 2020) since it is more likely to support good academic achievement. Both are understood to be mutually dependent and affirmative (Arens et al. 2016): Students with positive self-concept beliefs are more likely to show behaviors that induce academic success, while the academic success they achieve validates their positive self-concept beliefs (Hoferichter et al. 2018). Yet, what students perceive as 'good' or 'bad' performance often depends on their own standards (Jansen et al. 2015).

The TIMSS study(Trends in International Mathematics and Science Study) indicated a "set of individual beliefs about cognitive abilities in achievement-related situations" (Mullis et al. 2020, Bakadorova & Raufelder 2020, p. 2) to be an influence on career decisions (Mejía-Rodríguez et al. 2020). That is, regardless of high achievement scores in science or mathematics (Goldman & Penner 2016), students may decide against STEM-careers due to a negative self-concept. This is paricularly decisive for the gender-gap in STEM professions (Mejía-Rodríguez et al. 2020) and requires the immediate identification of influencing factors to help create self-concept enhancing learning environemnts for both genders (Hardy 2014, Mawang et al. 2018). Unfornunately, such reasoning is still largely overlooked in lesson planning (Patall et al. 2014).

The same applies to the interrelationship between creativity and self-concept. According to the literature, creativity positively influences self-concept, especially among gifted students

(Fleith et al. 2002). Yet, few acknowledge creativity as an influencing factor of scientific achievement although "creativity is inextricably linked to the nature of science itself" (Hadzigeorgiou et al. 2012). Teachers, in particular, often focus on logical thinking and uptake of scientific knowledge as a means to increase self-concept, rather than recognizing creativity creativity as a key factor (Fleith et al. 2002).

Creativity in Science Education

The integration of creativity into science education "should be rooted in and reflect aspects of creativity seen in scientific research", which is why educators should "devise a framework " that allows for creative processes (Kind & Kind 2007, p.3). Yet, research indicates that school learning does not necessarily meet the requirements to foster students' creative potential (Runco et al. 2017) even though creativity is also considered a possible influencing factor for academic performance (Gadja 2017). Hence, Richards & Cotterall (2016) offer guidelines to help teachers establish creativity in the classroom. Environments that encourage cognitive flexibility - i.e. the ability "to generate novel or unusual ideas and switch between unrelated concepts in a broad manner" - support creative processes (Filippetti & Krumm 2020, p.4). This can include individual and collaborative learning approaches with a high degree of flexibility and frequent open discussions (Miller 2014). By no means should learning approaches be predefined, even if, for instance, group work might help to develop creativity (Lian et al. 2018).

Creativity cannot be observed outside of the social environment (Csikszentmihalyi 1988). This is reinforced by the four categories of creativity: (1) mini-c, which encompasses early and exploratory approaches to uncover individual creativity, (2) little-c, which develops with encouragement and several trials of mini-c creative processes, (3) Pro-c, wherein the individual is "capable of working on problems, projects, and ideas that affect the field as a whole", (4) and Big-C, which is accomplished after many years of creative practice and is referred to as "eminent creativity" (Kaufman & Beghetto 2009, p.6).

Despite the relevance of social embeddedness for creative processes, it is essential to trigger intrinsic motivation and not force the creative process via extrinsic factors. Although the latter are commonly employed in schools – i.e., in the form of exams, regular evaluations, and rewards for good performance - they can have a detrimental effect on creative processes (Baer 2010, Amabile 1983), not least, because intrinsic motivation is understood as directly linked to creativity and school achievement (Gajda 2016). In a creativity-enhancing learning environment - whether in individual or group work - extrinsic motivators should be reduced to provide students with space to creatively engage with science learning content (Amabile & Pratt 2016, Beghetto 2015).

One aspect of creativity, which is particularly relevant to bilingual learning, is language. Language may express culture and is, consequently, involved in shaping creativity via the cultural domain (Csikszentmihalyi 1988). Research about creativity in bilingual contexts has confirmed this notion and identified a higher degree of divergent thinking in bilingual students. A more flexible approach and an ambiguity-tolerant mind is understood to be the main reason for these results (Fleith et al. 2002).

CLIL in Science Education

Since curricula are already densely packed, CLIL was established to enable regular subject teaching while intensely fostering foreign language learning, all without increasing students' overall workload (European Commission, 2004). The additional use of the foreign language as a medium to communicate subject-specific content is expected to improve foreign language competences (Shaw et al. 2010, Fernández-Fontecha et al. 2020). Instead of everyday communication, CLIL focuses on discipline-specific terminology (Meskill & Oliveira 2019). The combination of language and content learning creates authentic settings wherein students complete tasks comparable to those they will encounter at work (Liu et al. 2014, Meyerhöffer & Dreesmann 2019). Students thereby improve their critical thinking skills and receive important insights into written and oral scientific discourse (Liu et al. 2014, Gonález-Howard & McNeill 2016). Such exercisea also help students process information in a coherent and accessible manner that stimulates argumentation using the acquired scientific vocabulary (Nation & Newton 2008). This active engagement in scientific discourse is commonly identified as essential for scientific literacy foreign and the native languages (Walker & Sampson 2013, Oga-Baldwin 2019).

Since the learners' linguistic performance is not evaluated in CLIL instruction, it is important to identify their linguistic repertoire in classroom discourse and improve it through interactive feedback. With regard to foreign language learning, the connection between the native language, foreign language, and content is crucial to enable sensible language learning that both creates factual knowledge and meaning (Sutton 1996, Yore & Treagust 2006, Poza 2016). One approach to bridging the gap between language learning and scientific content is scaffolding. Scaffolding is understood to be a special kind of teacher assistance to help students acquire knowledge, improve understanding, and develop skills (Myhill & Warren 2005). There are four common types of scaffolding - linguistic, graphic, sensory, and interactive - all of which may play a role in science education (Gottlieb 2016). Visualization is among the most important for helping students understand complex content and internalize the relevant scientific vocabulary. For CLIL, visualization supports the uptake of new scientific terminology in the native and foreign language and fosters scientific literacy (Fernández-Fontecha 2020).

Objectives of the Study

The present study focuses on the following research questions:

1. How do school self-concept and creativity interact in a one-day CLIL hands-on science module?
2. How does the school self-concept influence CLIL learning in a one-day CLIL hands-on science module?
3. How does gender influence school self-concept in a one-day CLIL hands-on science module?

Thus, the specific objectives were five-fold:

- to assess students' school self-concept regarding the four-factor structure of *SESSKO* in Biology and *SESSKO* in English
- to assess students' creativity regarding the two-factor structure of the respective *SESSKO* questionnaires and the shortened *CPAC* scale
- to identify potential correlations between *SESSKO* in Biology and *SESSKO* in English and the shortened *CPAC* scale
- to identify potential correlations for *SESSKO* in Biology and *SESSKO* in English with pre-, post-, and retention tests of knowledge tests for Biology and English
- to identify potential gender differences regarding *SESSKO* in Biology and *SESSKO* in English

Materials and Methods

Intervention Phases

We adapted the design of previous, science modules (e.g., Langheinrich & Bogner 2016, Mierdel & Bogner 2020, Roth et al. 2020) to the requirements of a CLIL module while retaining the overall structure for treatment comparability. That is, all four intervention phases included specific linguistic and graphic scaffolding material in a separate exercise book and laboratory manual to reduce difficulties in understanding (Table 1). Our one-day inquiry-based learning module on the structure of DNA was expanded using CLIL and adapted to the requirements of ninth graders. The following section is adapted from Roth et al. (2020).

TABLE 1 Quasi-experimental intervention design.

Intervention phases	Evaluation variants	
	Monolingual (non-CLIL) (*n* = 145)	Bilingual (CLIL) (*n* = 107)
English as the language of instruction and conversation/interaction	-	+
Vocabulary scaffolding material for all intervention phases	-	+
Pre-lab	+	+
DNA-related theoretical and experimental phases		
DNA relevance	+	+
Hands-on isolation of DNA	+	+
Gel electrophoresis of DNA	+	+
Model-related phases		
Mental modelling: text analysis	+	+
DNA modelling with craft materials	+	+
Model evaluation-1: Drawing a paper & pencil version of the crafted model and answering questions	+	+
Model evaluation-2: Comparing the crafted model with a scientific demonstration model	+	+
Interpretation	+	+

Throughout the module, students worked in pairs (or groups of three) to complete different experimental steps, which were detailed in a laboratory manual. Additional language scaffolding provided relevant scientific vocabulary. The module's content is in line with the state's syllabus and adheres to national competency requirements (KMK 2005). Classes were randomly assigned to either our previous science module or the current CLIL module, following a quasi-experimental design (Cook & Campell 1979).

Pre-lab phase. Students began with an extensive pre-lab phase to familiarize themselves with laboratory requirements such as different laboratory techniques and theoretical concepts connected to experimentation (e.g., Sarmouk et al. 2019). The teacher only acted as a guide, conveying necessary theoretical concepts and demonstrating practical concepts relevant to experimentation. Further instructions and explanations could be retrieved from the laboratory manual. In doing so, the pre-lab phase prevented students from feeling overwhelmed in response to the experimental situation (e.g., Kalyuga 2009).

DNA-related theoretical and experimental phases. Building on the students' prior knowledge of cells, we introduced the practical relevance of subsequent experiments (DNA relevance, Table 1) and familiarized the students with the concepts of DNA isolation and gel electrophoresis. To ensure that students were not overwhelmed by the flood of information, we employed a sequential approach, alternating between theoretical and practical phases. This allowed the students to focus on the experiments at hand.

Experimental Phases. We used an evidence-based, two-step approach (Mierdel & Bogner 2020, Roth et al. 2020) in our experimental phases: Firstly, students answered questions in their laboratory manuals related to subsequent experimental procedures. Secondly, they worked in small teams to discuss and then take experimental steps, combining both hands-on and minds-on activities (Rinehart et al. 2016). Since the teacher only provided guidance where required, scaffolding material in our laboratory manual allowed for independent experimentation and self-reliant protocolling of their observations (Mierdel & Bogner 2020, Roth et al. 2020).

Model-related Phases. The experiments were followed by a two-part modelling phase, which comprised modelling and model evaluation. We decided to use modelling since models are essential to explain and visualize phenomena in science and science education (Krajcik & Merritt 2012). For modelling, we first encouraged a mental modelling phase after text analysis, where students either discussed their mental model or used a rough sketch for visualization. Thereafter, students used craft materials to construct the mental models they had discussed.

We based our model-related phases on the four main stages of the Model of Modelling (Gilbert & Justi 2002, p. 370 ff.): (1) "experience[s] of the phenomenon being modeled", (2) "forming a mental model", (3) "decision … about the mode of representation in which it is to be expressed", and (4) "testing … scope and limitations of the model". Accordingly, mental modelling provided the theoretical basis after students had obtained fundamental knowledge about the structure of DNA from experimentation and a text (Usher 2013). Students could then apply this knowledge to their simplified mental models and use it to build a hand-crafted DNA model (e.g., Conradty & Bogner 2019, Franco & Colinvaux 2000).

After modelling, the models were evaluated following Gilbert & Justi (2002). The evaluation was split into two parts and consisted of evaluation-1and evaluation-2 phases. Evaluation-1 phase was organized as a reciprocal self-evaluation, whereby students evaluated their self-built model using a detailed and labelled sketch. This combined the benefits both of hand-crafting and sketching models (e.g., Prabha 2016) as well as the method of reflexive writing to encourage self-evaluation (Ke et al. 2020). Open-ended questions about model-related components further encouraged students to consider about possible adjustments to the mental and physical models (Mierdel & Bogner 2020, Roth et al. 2020). In evaluation-2 phase, students assessed their hand-crafted DNA models using a comparison-based self-evaluation involving a commercially available DNA demonstration model.

Interpretation Phase. Here, students compared their hypothesis about the outcome of their experiments with images of the gel electrophoresis. They also discussed their models in class and compared these to the molecular DNA model by Watson & Crick.

Language Scaffolding

To support content learning in a foreign language while fostering scientific literacy and language learning, we designed a language scaffolding exercise book. It contained language-specific riddles such as word search puzzles or crossword puzzles to allocate words to English definitions. Students also had to match the words and their definitions with given German translations or translate the words into German. At all times, students were allowed to use English-English and English-German dictionaries available in the laboratory. For each intervention phase, we had one specific language scaffolding exercise.

Moreover, we included short translations or explanations of essential scientific vocabulary in their laboratory manuals and added German equivalents for laboratory equipment to our presentation. To help students understand the complicated procedure of gel electrophoresis, we provided a poster with English descriptions and questions. Code-switching was allowed to make process steps easier to follow and prevent obstruction of word flow. Oral explanations were primarily provided in English.

Participants

Overall, 252 ninth graders (from upper stratification grammar schools) participated in our student laboratory (girls 48.5%, boys 51.5%; SD_{Gender} = 6.2; M_{Age} = 14.5, SD_{Age} = 0.6). Students teamed up in 88 groups (84 2-person groups and four 3-person groups). Participation was voluntary, and we asked for written parental consent before the students' arrival. Data collection was pseudo-anonymous, and we only ever referred to the students using pseudonymous identifiers. We designed the laboratory in line with the Declaration of Helsinki (2013), and we had relevant German state ministry approved our questionnaires.

Variables

Students completed a pre-test (T0) two weeks before the intervention, asking them about their school self-concept and creativity. Throughout the entire intervention, students were unaware of any testing schedules.

Students' School Self-Concept. We applied the *SESSKO* (Skalen zur Erfassung des schulischen Selbstkonzepts [Scales to capture school self-concept]) questionnaire comprising 22 five-point Likert-scaled items that belong to one of four latent variables *Criterial*, *Individual*, *Social*, and *Absolute*. Content validity was given as reference norms could be provided by factor analysis (Schöne et al. 2002). Cronbach's alpha scores of .82 for self-concept in Biology and .91 for self-concept in English indicate acceptable internal consistency with values exceeding .70. According to Lienert & Raatz (1998), values between .50 and .70 allow for the differentiation of groups.

Students' Creativity. Miller's (2014) *Cognitive Processes Associated with Creativity* (*CPAC*) scale has proven particularly suitable for assessing six dimensions of creativity – "idea manipulation, idea generation, flow, imagery/sensory perception, incubation, and metaphorical/analogical thinking" – using a set of 28 items (Tsai 2018, p.271). Since using a scale comprising 28 items would take more time than is likely to be available for conducting such tests in science classrooms, a shortened scale was suggested by Conradty & Bogner (2018), involving only 15 items. Although shorter than the original scale developed by Miller (2014), this revised scale contains all relevant dimensions of the more extensive version and has already been validated in science outreach education (Conradty & Bogner 2019).

Students' Knowledge in English and Biology. In a repeated measurement design, students completed a pre-test (T0) two weeks before the intervention, a post-test (T1) after the module, and a retention-test six weeks thereafter (T2). Throughout the entire intervention, students were unaware of the testing schedule. To assess knowledge in Biology, we used an ad-hoc knowledge test comprising 30 multiple-choice items. For each round of tests (T0, T1, T2), the questions and respective multiple-choice answers were randomly assigned to avoid systemic response patterns. Content validity was given as the items were consistent with the state syllabus. Regarding construct validity, inter-item correlations below .20 (T0 = .08; T1 = .19; T2 = .18) confirmed that each item referred to different knowledge facets.

For knowledge in English, we applied a Cloze test with 14 items with the same repeated measurement design used to assess knowledge of Biology. The heterogeneity of items to describe complex constructs, such as language learning, confirms construct validity (Rost 2004). Cronbach's alpha scores of .61 (T0), .63 (T1), and .66 (T2) indicate acceptable internal consistency and allow for the differentiation of groups (Lienert & Raatz 1998).

Statistical Analysis. Subsequent statistical analyses were conducted using AMOS (IBM Corp. 2020) and IBM SPSS Statistics 26.0. We used AMOS (IBM Corp. 2020) for confirmatory factor analysis of the shortened *CPAC* to show that the predicted structure from our previous interventions (Conradty & Bogner 2018, Conradty & Bogner 2019) was replicable (Thompson 2004). Based on theory (Kaiser 1970) and previous analysis (Conradty & Bogner 2018) we assumed our sample would divide into two factors *Flow* and *Act.* We also used AMOS (IBM Corp. 2020) with a confirmatory factor analysis of *SESSKO* in Biology and in English, which, as predicted, showed the four-factor structure postulated by Schöne et al (2002). Our data were not-normally distributed following assessment with the Shapiro-Wilk test ($p < .001$). Despite our large sample size, we also refrained from normalizing the data as would be recommended by Bortz & Schuster (2010), since – due to Likert-scaling – our measured values have no clearly interpretable relative distances. Mann-Whitney *U* tests (MWU) were deployed to evaluate inter-group differences regarding gender. To account for multiple testing, we applied a Bonferroni correction $p < .017$ (Field 2012). In the case of significant results, effect sizes r (Lipsey & Wilson 2001) were calculated with small (> 0.1), medium (> 0.3), and large (> 0.5) effect sizes. For correlation analyses, we applied Spearman's rank correlations and reported Spearman's Rho values.

Results

Our results indicate that creativity, self-concept in Biology and in English measured expected constructs and significantly correlated with on another. We did not observe any significant gender differences for the latent variables of our respective scales but did observe *Social* for self-concept in Biology. Cronbach α scored .684 for creativity, .818 for self-concept in Biology and .91 in English, confirming that the internal consistency of the scales was acceptable and allowed for the differentiation of groups (Lienert & Raatz 1998). A subsequent confirmatory factor analysis with oblique rotation confirmed two factors with high factor loadings for creativity and four factors with high factor loadings for self-concept in Biology and in English. The model fit with CFI = .943 and TLI = .936 for creativity and self-concept in Biology is excellent and indicates reliable and distinct factors, which is supported by $RMSEA$ < .050 (p = .110) and $CMIN/DF$ < 5.00 (Schermelleh-Engel et al. 2003). The resulting structural equation model (Figure 1) indicates that all observable variables, other than *A3* in the subscale *Absolute,* adequately describe their latent variables, as can be seen with respective factor loadings

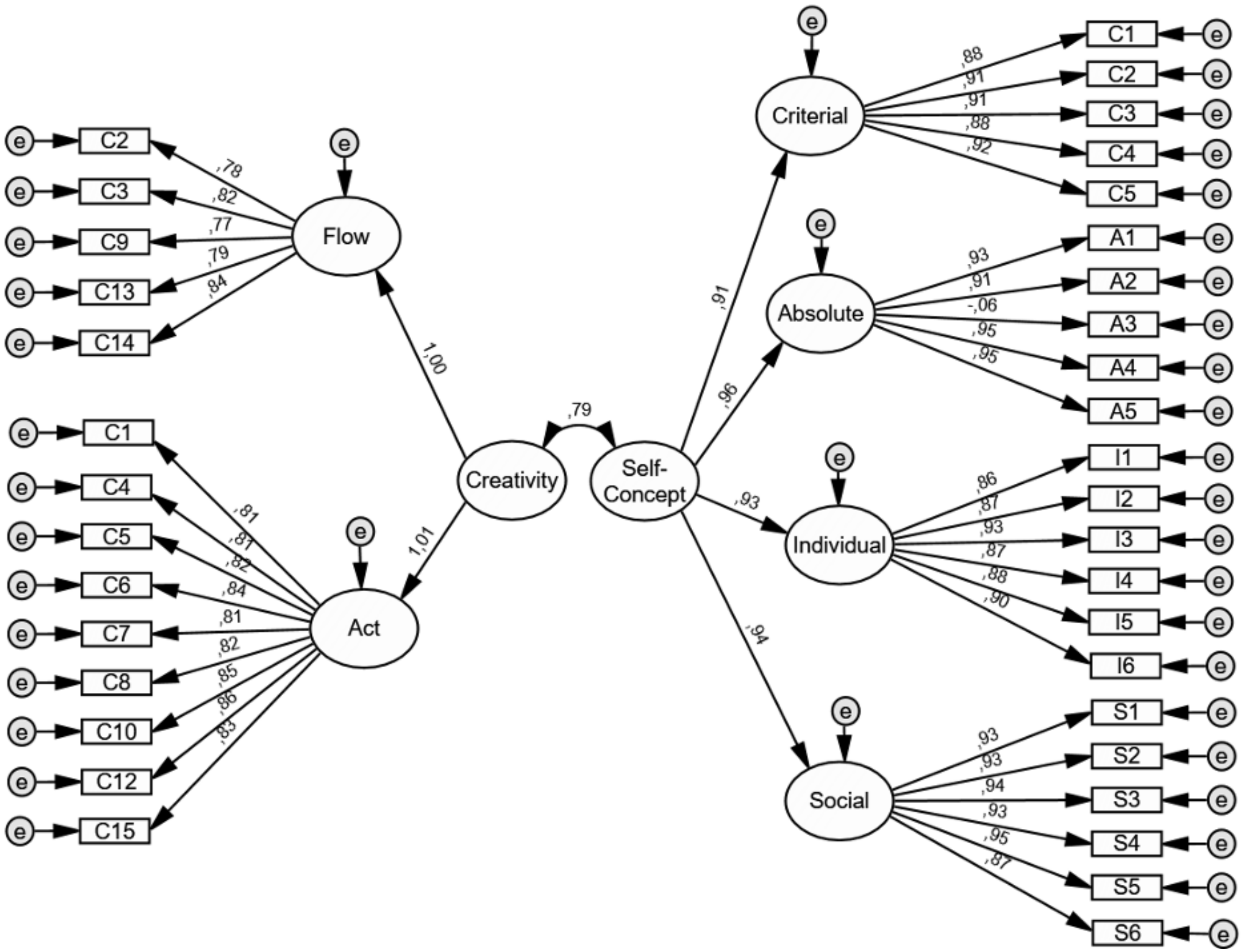


FIGURE 1 Structural Equation Model for creativity testing with creativity and self-concept in Biology after Confirmatory Factor Analysis with oblique rotation and correlation analysis between the two constructs.

For self-concept in English, the model fit with *CFI* = .943 and *TLI* = .936 as well as *RMSEA* < .050 (p = .006) and *CMIN/DF* < 5.00 (Schermelleh-Engel et al. 2003) also show reliable and distinct factors. The resulting structural equation model (Figure 2) indicates that all observable variables apart from *A3* in the subscale *Absolute* adequately describe their latent variables, as can be seen with respective factor loadings. However, in both cases, chi-square (creativity + self-concept in Biology; χ^2= 1425.59, *df*= 587, p<.001; creativity + self-concept in English; χ^2= 1513.42, *df*= 587, p<.001) implies that the event occurs less than one time in a thousand and that data does not adequately fit the model (Phakiti 2018). Yet, considering the other parameters that indicate a very good model fit (Schermelleh-Engel et al. 2003), χ^2 can be neglected, since it is also very sensitive to sample size and is no longer the basis for acceptance or rejection of a model (Schermelleh-Engel et al. 2003, Vandenberg 2006). This approach is supported by the results of power for test of close fit (MacCallum et al. 2006) with λ = .84 for creativity and self-concept in Biology and λ = .96 for creativity and self-concept in English, indicating good approximate fit with values above λ = .80 (Cohen 1992). Error variances of the observed variables are represented by circles shortened with "e" and coloured in light grey. Since specific variance and error variance are, analytically speaking, inversely related, increased specific variance and reliability lead to decreased error variances (Lomax 1986).

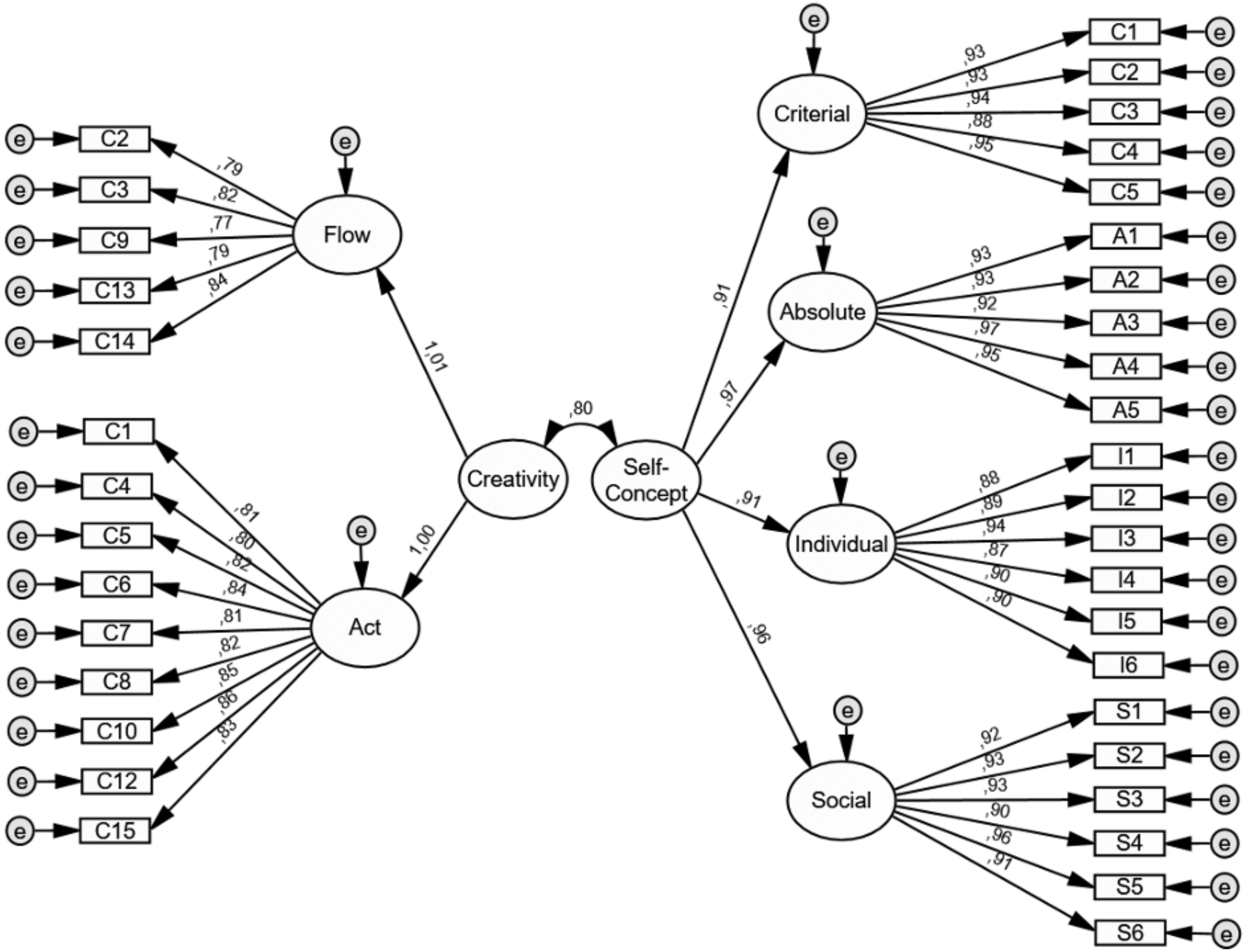


FIGURE 2 Structural Equation Model for creativity testing with creativity and self-concept in English after Confirmatory Factor Analysis with oblique rotation and correlation analysis between the two constructs.

To describe *CPAC's* latent variables *Act* and *Flow*, we retained the renaming of the two factors originally introduced by Miller & Dumford (2016). The same applies to the four latent variables *Criterial, Individual, Social,* and *Absolute* of the *SESSKO* scale (Schöne et al. 2002).

After confirmatory factor analysis with creativity, self-concept in Biology, and in English, we correlated the respective self-concept scales with creativity. This correlation, using Spearman-Rho (Field 2012), displayed significant correlations ($p < .001$) for both self-concept in Biology and creativity, with $r = .739$, and self-concept in English and creativity, with $r = .802$. Individual correlations between the different latent variables of self-concepts and observable variables of creativity revealed significant correlations (Table 2) with effect sizes ranging from intermediate to large effects (Lipsey & Wilson 2001). This confirms previously postulated connections between self-concept and creativity, showing that individual perception of achievement is involved in creative action.

TABLE 2 Detailed correlation analyses between the observable variables of creativity and self-concept in Biology and self-concept in English, respectively, as well as correlations between Biology and/or English grades and self-concept in Biology and self-concept in English, respectively for T1. The latent factors of the creativity scale are shortened as follows: idea manipulation (IM), imagery/sensory (IS), flow (FL), metaphorical/analogical thinking (MA), idea Generation (IG), and incubation (IN).

	Biology				**English**			
	Criterial	Individual	Social	Absolute	Criterial	Individual	Social	Absolute
IS1	.410**	.364**	.415**					
FL3	.313**	.305**	.243*	.268*		.253*		.268*
IG4	.308**	.362**	.452**	.302**		.302**		.302**
MA5		.241*	.296*	.322**		.352**	.272*	.322**
IN6	.290*	.374**	.244*	.324**	.243*	.308**	.230*	.324**
IM7	.327**	.284*	.316**	.310**	.226*	.297*	.240*	.310**
IG8				.358**	.333**	.457**	.262*	.358**
FL9				.247*		.277*		.247*
IS10	.273*	.281*	.223*	.312**	.248*	.326**	.281*	.312**
IN12				.231*				.231*
FL13	.289*		.358**	.307**			.233*	.307**
FL14	.316**	.333**	.354**	.232*		.326**		.232*
IM15	.308**	.402**	.371**	.298*		.297*	.235*	.298*
Grade Bio.	-.447**	-.237*	-.284*					
Grade Engl.				-.337**	-.408**		-.350**	-.337**

** = $p < .001$
* = $p < .017$

Moreover, we discovered significant negative correlations using Spearman-Rho (Field 2012) with intermediate to large effect sizes (Lipsey & Wilson 2001) between self-concept in English and English grades as well as between self-concept in Biology and Biology and English grades (Table 2). That is, those with lower grades in the respective subjects received higher self-concept scores in English and Biology after participation. We also identified significant intermediate correlations (Lipsey & Wilson 2001) between:

- post-tests of the Biology knowledge test and the respective self-concept in Biology (*Criterial:* $r = .193$, $p = .037$; *Individual*: $r = .330$, $p < .001$; *Social*: $r = .256$, $p = .005$; *Absolute*: $r = .277$, $p = .002$) and
- retention-test of the English Cloze test and self-concept in English (*Criterial:* $r = .296$, $p = .001$; *Individual*: $r = .290$, $p = .002$; *Social*: $r = .273$, $p = .003$; *Absolute*: $r = .247$, $p = .007$).

We also applied the Mann-Whitney-*U* test to assess the possible effect of gender on school self-concept. We discovered no significant gender differences for all latent variables of self-concept but the latent variable *Social* for Biology (Figure 3). There, male students reached higher scores, as compared to female students, regarding various perceptions of their Biology competency in comparison to classmates with a high effect size (MWU; $Z = -2.788$, $p = .005$, $r = .697$; Lipsey & Wilson 2001) while grades ($MWU_{Biology}$; $Z = -.654$, $p = .513$; $MWU_{English}$; $Z = -.046$, $p = .963$) and testing scores did not differ.

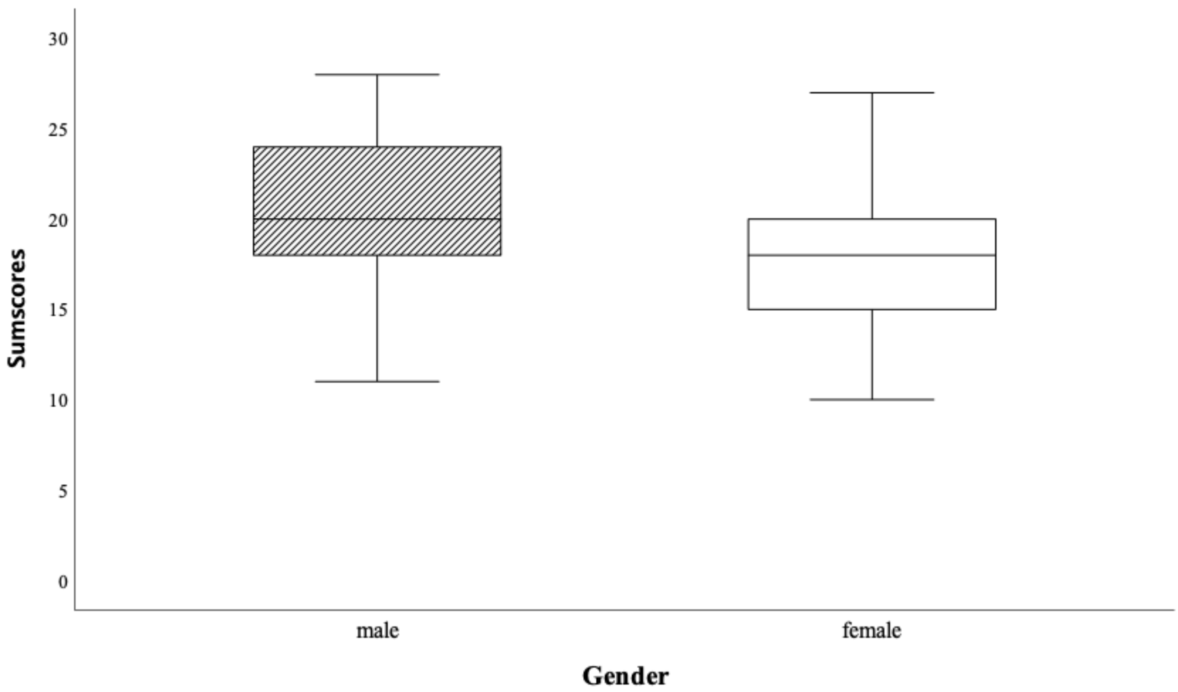


FIGURE 3 Gender differences for the latent variable Social of self-concept in Biology after MWU testing.

Discussion

A one-day outreach CLIL module affects the self-concept of low achievers in Biology and English regardless of gender. This promising finding is in line with the existing literature (e.g. Patall et al. 2014), as does the finding that, for both genders, self-concept strongly correlates with creativity, which has previously been presented by authors such as Sisk (1972) or Mawang et al. (2018). The absence of overarching gender differences is encouraging, although girls still perceive their subject-specific achievement as lower when compared to classmates (e.g., Mullis et al. 2020, Mejía-Rodríguez et al. 2020). Inducing such comparisons, as does the latent variable *Social,* has already been proven detrimental. Ideally, self-worth and self-concept would never depend on anything else but the appreciation of individual capabilities (Hoferichter et al. 2018, Bournelli 2009).

Relationship between school self-concept and creativity in CLIL modules

Creativity matters for CLIL outreach learning: Flow experiences are reportedly linked with positive feelings (Csikszentmihalyi 2000). To enable Flow, adequate environments are required wherein the difficulty of tasks is either balanced between demanding and easy or provides students with a spectrum of difficulty levels to choose from (Lian et al. 2018). For instance, modelling the DNA structure from a text (Usher 2013, Mierdel & Bogner 2020, Roth et al. 2020), asks for creative solutions and helps develop a positive self-concept (Bournelli et al. 2009) as students perceive themselves as competent agents capable of finding solutions. This strengthens subject-specific or competence-specific self-efficacy and, hence, self-concepts (Justo 2008). Our module required teams to develop both scientific and language problem-solving strategies, creating a learning environment that supports both creativity (Conradty & Bogner 2019) and self-concept in Biology and in English. The extensive scaffolding necessary to successfully conduct the CLIL module may have also created the relevant basis to flexibly increase or decrease the difficulty of tasks dependent on individual preferences (e.g., Grandinetti 2013, Gottlieb 2016).

Feeling secure is considered a basic requirement for both *Flow* and self-concept (Conradty & Bogner 2018, Justo 2008). Security supports complete immersion in tasks (Csikszentmihalyi 2000) and facilitates the development of solutions without fear of assessment. We, therefore, advised our teachers to abandon "the role of information transmitters" and to "take on the role of learning environment creators" wherein students could work on both their strengths and weaknesses (Justo 2008, p. 45). Having teachers as guides may increase the students' perception of safety and encourage individual learning and problem-solving strategies (Csikszentmihalyi 2000). This can promote motivation (Marsh et al. 2015, Conradty et al. 2016), creativity (Conradty & Bogner 2018, Conradty & Bogner 2019), and even learning success (Conradty et al. 2020, Mierdel et al. 2019, Roth et al. 2020). Therefore, the learning process in our module was initiated and controlled by students, while allowing for a high degree of self-regulation irrespective of fixed lesson plans (Justo 2008, Marsh et al. 2015) However, outreach learning still is an exception to daily school routines, which focuses primarily on preparation, incubation, evaluation, and elaboration (Csikszentmihalyi 1988, Justo 2008, Bournelli et al. 2009).

Influence of school self-concept on CLIL

A second major result reveals the interrelation of Biology and English cognitive achievement scores with respective self-concept scores. The mutual and affirmative influence between self-concept and achievement is in line with relevant literature (Jansen et al. 2015, Arens et al. 2016, Wang et al. 2008, Mejía-Rodríguez et al. 2020): A positive self-concept in a subject encourages performance and higher goals of achievement (Patall et al. 2014). This may, in turn, lead to more self-confidence and security the respective subject (Justo 2008). Yet, the role of peers and teachers is also crucial for the development of self-concept. Dependent on others' expectations of and confidence in a student's individual abilities, their self-concept can be either strengthened or weakened, influencing academic performance. Szumski & Karwowski (2019) have named this phenomenon after the Victorian play by George Bernard Shaw 'Pygmalion Effect'. Consequently, it is important to provide affirmative and encouraging learning environments (Blöte 1995, Burnett 2003) such as outreach learning featuring practical experimentation (Rodenhauser & Preisfeld 2014, Buse et al. 2018, Roth et al. 2020). Extensive scaffolding also allowed for successful completion of the experiments while accounting for different levels of student performance. This prevented undesired mental workloads, which could have negatively influenced self-concept (Grandinetti et al. 2013).

Studies examining self-concept in CLIL have been scarce, with a focus primarily on elementary or university level education. In contrast to our findings, self-concept often is reported to be quite low for CLIL learners, especially in comparison to non-CLIL learners (Roiha & Mäntylä 2019). The increased demand for language competency required to communicate scientific knowledge could have prevented students from feeling competent in the target language (Doiz et al. 2014). Since our outreach module provided extensive linguistic scaffolding to reduce cognitive overload (Coyle 2007, Grandinetti et al. 2013) and allowed for code-switching (Königs 2013), we hoped that students could experience learning in a foreign language without feeling deficient. In general, the combination of science and language learning in a realistic and hands-on setting may boost students' confidence in using the foreign language (Buse et al. 2018). In line with a study by Coyle (2007), students find this learning situation much more motivating. The increased motivation and the positive feedback (Roiha & Mäntylä 2019) could also be achieved from successfully completing our outreach module. Thus, the reciprocal nature of achievement and self-concept could also have contributed to a more positive self-concept and academic achievement (Jansen et al. 2015, Arens et al. 2016).

Gender effects in school self-concept

In a third step, our study produced gender differences for the latent variable, *Social,* relating to self-concept in Biology. Although such differences have been widely discussed in the past, empirical studies are still scarce (Wang et al. 2008). Yet, differences regarding the self-concepts in science subjects are already known. Girls often perceive themselves as less competitive than boys, despite equal performance (Mullis et al. 2020, Mejía-Rodríguez et al. 2020). Although a reverse distribution of self-concept was expected for English learning, we discovered no significant differences between girls and boys (e.g. Andreou et al. 2005, van der

Slik et al. 2015). This could indicate that our CLIL module reinforces the English self-concept in such a way that - even in comparison with others - both genders are equally supported.

Yet, our findings suggest that the latent variable *Social* does not help to promote a positive self-concept. Students learn through self-reflection, stimulating questions in the subscales *Criterial, Individual,* and *Absolute* to constructively assess their own abilities constructively. It is only the variable *Social* that asks them to compare their achievements with those of other students (Hoferichter et al. 2018). Students already perceive themselves as differently talented for "domains such as: scholastic competence, athletic competence, peer acceptance, physical appearance, and conduct or behavior", particularly in comparison with peers (Bournelli et al. 2009, p. 105). Depending on who students compare themselves to, this can have positive or negative consequences for their self-concept: If high-achievers compare themselves with other high-achievers it may be beneficial and provides the necessary stimulus to continue improving. If low-achievers or moderately gifted students - who are much more common - compare themselves possibly better students, it can be detrimental (Hoferichter et al. 2018). The demand for comparison can further be expected to encourage violent communication. Comparison with others is understood to be a form of (de)valuation. According to Rosenberg (2012), such behavior provokes and consolidates conflict.

Limitations

Firstly, the self-reporting method of data collection for *SESSKO* and *CPAC* may not have provided us with an 'outside' or 'independent' perspective on participants' views. Self-reported data are often understood to be vulnerable to inaccurate reporting as participants may represent themselves differently for various conscious and unconscious reasons. Secondly, our design primarily relied on quantitative data and is not the type of multi-trait, multi-method design, which some authors recommend (Leutner et al. 2017). Thirdly, due to the context-dependency of CLIL learning, results cannot easily be extrapolated (Pérez-Cañado 2012). As a consequence, generalizing learnings from CLIL modules requires acknowledgement of the diversity of other possible implementations (Fernández-Sanjurjo et al. 2017).

Conclusion

Understanding the relationship between creativity and school self-concept may help to influence learning in Biology and English on the basis of a one-day CLIL hands-on science module. Few previous studies have investigated this relationship with CLIL (e.g. Fleith et al.

2002). Our study revealed positive effects in both English and Biology in terms of self-concept and cognitive achievement, indicating that a hands-on minds-on approach with different experimental and modelling tasks provides a cognitively stimulating, self-reflective, and creativity-endorsing environment (Rodenhauser & Preisfeld 2014, Bakadorova & Raufelder 2020, Ke et al. 2020). Nevertheless, a longitudinal study assuring such beneficial effects is still necessay.

A second major conclusion emanates from the exploration of creativity, school self-concept, and achievement in short-time CLIL modules (e.g. Fleith et al. 2002, Meyerhöffer & Dreesmann 2019). So far, the majority of studies have focused on long-term CLIL approaches in combination with hands-on minds-on tasks may have stronger influences on content and language learning, the enhancement of school self-concept is also difficult in these contexts (Hattie & Marsh 1996, Fleith et al. 2002). Yet, the gender differences observed for the subcale *Social,* which we and numerous TIMSS studies have identified (e.g., Mullis et al. 2020, Mejía-Rodríguez et al. 2020), indicate that further in-depth research to improve gender equality in science learning is required.

Acknowledgements

First of all, we thank all the students as well as their teachers, who participated in our study. We are also very grateful for the support of our junior assistants throughout the module phase.

Ethical statement

Hereby, we consciously assure that for the manuscript the following is fulfilled:

1) This material is the authors' own original work, which has not been previously published elsewhere.

2) The paper is not currently being considered for publication elsewhere.

3) The paper reflects the authors' own research and analysis in a truthful and complete manner.

4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

5) The results are appropriately placed in the context of prior and existing research.

6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

7) All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

Moreover, "all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Consent statement

Informed consent was obtained from all individual participants included in the study.

Conflict of Interest or Disclosure Statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. Journal of Personality and Social Psychology, 45, 357-376.

Amabile, T.M., & Pratt, M.G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. Research in Organizational Behavior, 36, 157-183.

Andreou, G., Vlachos, F., & Andreou, E. (2005). Affecting factors in second language learning. Journal of Psycholinguistic Research, 34, 429-438.

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., and vom Hofe, R. (2016). Math self-concept, grades, and achievement test scores: long-term reciprocal effects across five waves and three achievement tracks. Perspectives on Psychological Science 1, 133–163.

Baer, J. (2010). Gender differences in the effects of anticipated evaluation on creativity. Creativity Research Journal, 10(1), 25-31.

Bakadorova, O., & Raufelder, D. (2020): The relationship of school self-concept, goal orientations and achievement during adolescence. Self and Identity, DOI: 10.1080/15298868.2019.1581082.

Bećirović, S. (2017). The relationship between gender, motivation and achievement in learning English as a foreign language. European Journal of Contemporary Education, 6(2), 210-220.

Beghetto, R. A. (2015). Teaching creative thinking in K12 schools. The Routledge International Handbook of Research on Teaching Thinking. New York: Routledge.

Blöte, A.W. (1995). Students' self-concept in relation to perceived differential teacher treatment. Learning and Instruction, 5, 221-236.

Bournelli, P., Makri, A., & Mylonas, K. (2009). Motor creativity and self-concept. Creativity Research Journal, 21(1), 104-110.

Bortz, J., & Döring, N. (1995). Forschungsmethoden und Evaluation [Research methods and evaluation]. Berlin: Springer.

Bortz, J., & Schuster, C. (2010). Faktorenanalyse [Factor Analysis]. In: Bortz, J., & Schuster, C. (eds.). Statistik für Human- und Sozialwissenschaftler [Statistics for human and social scientists] (pp. 385-426). Berlin, Heidelberg, New York: Springer Verlag.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York, NY: WW Norton & Company.

Burnett, P.C. (2003). The impact of teacher feedback on student self-talk and self-concept in reading and mathematics. The Journal of Classroom Interaction, 38(1), 11-16.

Buse, M., Damerau, K. & Preisfeld, A. (2018). A Scientific Out-of-School Programme on Neurobiology Employing CLIL. Its Impact on the Cognitive Acquisition and Experimentation-Related Ability Self-Concepts. International Journal of Environmental & Science Education, 13(8), 647-660.

Cohen, J. (1992). Statistical power analysis. Current Directions in Psychological Science, 1(3), 98–101

Conradty, C., & Bogner, F.X. (2016). Hypertest or textbook: Effects on motivation and gain in knowledge. Education Sciences, 6(3), 29, 1-16.

Conradty, C., & Bogner, F.X. (2018). From STEM to STEAM: How to monitor creativity. Creativity Research Journal, 30(3), 233-240.

Conradty, C., & Bogner, F.X. (2019). How Creativity in STEAM Modules Intervenes with Self-Efficacy and Motivation. Education Sciences, 10(3), 1-15.

Conradty, C., Sotiriou, S.A., & Bogner, F.X. (2020). From STEM to STEAM: Cracking the code? How creativity & motivation interacts with inquiry-based learning. Creativity Research Journal, 31(3), 284-295.

Cook, T. D., & Campell, D. (1979). Quasi-experimentation. Design & analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

Coyle, D. (2007). Listening to learners: an investigation into 'successful learning' across CLIL contexts. International Journal of Bilingual Education and Bilingualism, 16(3), 244-266:

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In R. Sternberg & J. E. Davidson (Eds.), Conceptions of giftedness (pp. 325–339). New York, NY: Cambridge University Press.

Csikszentmihalyi, M. (2000). Happiness, flow, and economic equality. American Psychologist, 55(10), 1163–1164.

Declaration of Helsinki, World Medical Association (2013). Journal of the American Medical Association, 310, 2191-2194.

Doiz, A., Lasagabaster, D., & Sierra, J.M. (2014). CLIL and motivation: the effect of individual and contextual variables. The Language Learning Journal, 42(2), 209-224.

European Commission (Hrsg.) (2004): Promoting language learning and linguistic diversity. An action plan 2004-06. Luxembourg, Office for Official Publications of the European Communities.

Fehling, S. (2008). Language Awareness und bilingualer Unterricht. Eine komparative Studie [Language awareness and bilingual education. A comparative study]. Frankfurt am Main: Internationaler Verlag der Wissenschaften.

Fernández-Fontecha, A., O'Halloran, K.L., Wignell, P., & Tan, S. (2020). Scaffolding CLIL in the science classroom via visual thinking: A systemic functional multimodal approach. Linguistics and Education, DOI: 10.1016/j.linged.2019.100788.

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J.M. (2017). Analysing students' content-learning in science in CLIL vs. Non-CLIL programmes: empirical evidence from Spain. International Journal of Bilingual Education and Bilingualism, 22(6), 661-674.

Field, A. (2012). Discovering statistics using IBM SPSS statistics (4th ed.). London: SAGE.

Filippetti, V.A., & Krumm, G. (2020). A hierarchical model of cognitive flexibility in children: Extending the relationship between flexibility, creativity and academic achievement. Child Neuropsychology, DOI: 10.1080/09297049.2019.1711034

Fleith, D. S., Renzulli, J. S., & Westberg, K. L. (2002). Effects of a creativity training program on divergent thinking abilities and self-concept in monolingual and bilingual classrooms. Creativity Research Journal, 14, 373-386.

Franco C. & Colinvaux D. (2000). Grasping Mental Models. In: Gilbert J.K., Boulter C.J. (Eds.) Developing models in science education (pp. 93-118). Springer.

Gadja, A. (2017). The relationship and moderators of school achievement and creativity at different educational stages. Thinking Skills and Creativity, 24, 250-267.

Gilbert, J. K. & Justi, R. S. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. International Journal of Science Education, 24, 369–387.Glynn, S.M. & Muth, K.D. (1994). Reading and writing to learn science: achieving scientific literacy. Journal of Research in Science Teaching, 31(9), 1057-1073.

Goldman, A. D., & Penner, A. M. (2016). Exploring international gender differences in mathematics self- concept. International Journal of Adolescence and Youth, 21(4), 403–418.

Gonzalez-Howard, M. & McNeill, K.L. (2016). Learning in a community of practice: factors impacting English-learning students' engagement in scientific argumentation. Journal of Research in Science Teaching, 53(4), 527-553.

Gottlieb, M. (2016). Assessing English language learners: Bridges for language proficiency to academic achievement. Thousand Oaks: Corwin.

Grandinetti, M.M., Langellotti, M., & Teresa Ting, Y.L. (2013). How CLIL can provide a pragmatic means to renovate science education – even in a sub-optimally bilingual context. International Journal of Bilingual Education and Bilingualism, 16(3), 354-374.

Hardy, G. (2014). Academic self-concept: Modeling and measuring for science. Research in Science Education, 44, 549-579.

Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. Creative Education, 3(5), 603-611.

Hattie, J., & Marsh, H.W. (1996). The relationship between research and training: A meta-analysis. Review of Educational Research, 66(4), 507-542.

Henriksen, D., Mishra, P., & Fisser, P. (2018). Infusing Creativity and Technology in 21st Century Education: A Systemic View for Change. Educational Technology & Society, 19(3), 27–37.

Hoferichter, F., Lätsch, A., Lazarides, R., & Raufelder, D. (2018). The big-fish-little-pond effect on the four facets of academic self-concept. Frontiers in Psychology, 9(1247), 1-11.

IBM Corp. (2020). IBM SPSS Amos for Windows (Version 26). IBM Corp. Armonk, New York.

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. Learning and Individual Differences, 30, 11-21.

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. Contemporary Educational Psychology, 41, 13-24.

Justo, C.F. (2008). Creative relaxation, motor creativity, self-concept in a sample of children from early childhood education. Electronic Journal of Research in Educational Psychology, 14(6), 29-50.

Kaiser, H. F. (1970). A second generation little jiffy. Psychometrika, 35(4), 401–415.

Kalyuga S. (2009). Knowledge elaboration: A cognitive load perspective. Learning and Instruction, 19, 402-410.

Kaufman, J.C., & Beghetto, R.A. (2009). Beyond big and little: the four C model of creativity. Review of General Psychology, 13(1), 1-12.

Ke, L., Zangori, L., Sadler, T.D., & Friedrichsen, P.J. (2020). Integrating scientific modeling and socio-scientific reasoning to promote scientific literacy. In Warden A. Powell (Ed.), Socioscientific issues-based instruction for scientific literacy development (pp. 31-56). Hershey: IGI Global.

Kind, P.M., & Kind, V. (2007). Creativity in science education: perspectives and challenges for developing school science. Studies in Science Education, 43(1), 1-37.

KMK (2005). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss [Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany - standards of biology education for secondary school]. Munich: Luchterhand.

König, J., Jäger-Biela, D.J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. European Journal of Teacher Education, 43(4), 608-622

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? Science and Children, 49, 10–13.

Langheinrich, J., & Bogner, F.X. (2016). Computer-related self-concept: The impact on cognitive achievement. Studies in Educational Evaluation, 50, 46-52.

Leutner, F.; Yearsley, A.; Codreanu, S.-C.; Borenstein, Y.; Ahmetoglu, G. (2017). From Likert scales to images: Validating a novel creativity measure with image-based response scales. Personality and Individual Differences, 106, 36–40.

Lewis, T. (2008). Creativity in technology education: Providing children with glimpses of their creative potential. International Journal of Technology and Design Education, 19(3), 255-268.

Lian, B., Kristiawan, M., & Fitriya, R. (2018). Giving creativity room to students through the friendly school's program. International Journal of Scientific and Technology Research, 7(7), 1-7.

Lienert, G. A., & Raatz, U. (1998). Testaufbau und Testanalyse [Test setup and test analysis] (6th ed.). Psychologie Verlags Union: Weinheim.

Lipsey, M. W., & Wilson, D. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage Publications.

Liu, G-Z., Chiu, W-Y., Lin, C-C., & Barrett, N.E. (2014). English for scientific purposes (EScP): Technology, trends, and future challenges for science education. Journal of Science Education and Technology, 23, 827-839.

Lomax, R.G. (1986). The effect of measurement error in structural equation modeling, The Journal of Experimental Education, 54:3, 157-162.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1(2), 130–149.

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: relations and causal ordering. British Journal of Educational Psychology, 81, 59–77.

Marsh, H.W., et al. (2015). The internal/external frame of reference model of self-concept and achievement relations: age-cohort and cross-cultural differences. American Educational Research Journal, 52(1), 168-202.

Mawang, L.L., Kigen, E.M., & Mutweleli, S.M. (2018). The relationship between musical self-concept and musical creativity among secondary school music students. International Journal of Music Education, 1-13.

Mejía-Rodríguez, A.M., Luyten, H., & Meelissen M.R.M. (2020). Gender differences in mathematics self-concept across the world: an exploration of student and parent data of TIMSS 2015. International Journal of Science and Mathematics Education. doi.org/10.1007/s10763-020-10100-x.

Meskill, C., & Oliveira, A.W. (2019). Meeting the challenges of English learners by pairing science and language educators. Research in Science Education, 49, 1025-1040.

Meyerhöffer, N. & Dreesmann, D.C. (2019). English-bilingual Biology for standard classes development, implementation and evaluation of an English-bilingual teaching unit in standard German high school classes. International Journal of Science Education, 41(10), 1366-1386.

Mierdel, J., & Bogner, F.X. (2019). Is creativity, hands-on modeling and cognitive learning gender-dependent? Thinking Skills and Creativity, 31, 91-102.

Mierdel, J., & Bogner, F.X. (2020). Simply InGEN(E)ious! How creative DNA modeling can enrich classic hands-on experimentation. Journal of Microbiology and Biology Education, 21(2), 21-45.

Miller, A. L. (2014). A self-report measure of cognitive processes associated with creativity. Creativity Research Journal, 26(2), 203–218.

Miller, A. L., & Dumford, A. D. (2016). Creative cognitive processes in higher education. The Journal of Creative Behavior, 50(4), 282–293.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 International Results in Mathematics and Science. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.

Myhill, D., & Warren, P. (2005). Scaffolds or straitjackets? Critical moments in classroom discourse. Educational Review, 57(1), 55-69.

Nation, I. S. P., & Newton, J. (2008). Teaching ESL/EFL listening and speaking. London: Routledge.

Oga-Baldwin, Q.W.L. (2019). Acting, thinking, feeling, making, collaborating: the engagement process in foreign language learning. System, 86, 1-10.

Patall, E.A., Awad, G.H., & Cestone, C.M. (2014). Academic potential beliefs and feelings: Conceptual development and relations with academic outcomes. Self and Identity, 13(1), 58-80.

Pérez-Cañado, M. L. (2012). CLIL Research in Europe: Past, Present, and Future. International Journal of Bilingual Education and Bilingualism, 15(3), 315-341.

Phakiti A. (2018) Confirmatory Factor Analysis and Structural Equation Modeling. In: Phakiti A., De Costa P., Plonsky L., Starfield S. (eds) The Palgrave Handbook of Applied Linguistics Research Methodology. Palgrave Macmillan, London.

Poza, L.E. (2016). The language of ciencia: translanguaging and learning in a bilingual science classroom. International Journal of Bilingual Education and Bilingualism, DOI: 10.1080/13670050.2015.1125849

Prabha, S. (2016). Laboratory experiences for prospective science teachers: A meta-analytic review of issues and concerns. European Scientific Journal, 12, 235–250.

Richards, J., & Cotterall, S. (2016). Exploring creativity in language teaching. In R. H. Jones & J. C. Richards (Eds.), Creativity in language teaching (pp. 97–113). New York, NY: Routledge.

Rinehart, R., Duncan, R., Chinn, C., Atkins, T., & Di Benedetti, J. (2016). Critical design decisions for successful model-based inquiry in science classroom. International Journal of Design for Learning, 7, 17-40.

Robinson, K. (2011). Out of our minds: Learning to be creative. Oxford, UK: Capstone.

Rodenhauser, A. & Preisfeld, A. (2014). Bilingual (German-English) Molecular Biology Courses in an Out-of-School Lab on a University Campus: Cognitive and Affective Evaluation. International Journal of Environmental and Science Education, 10(1), 99-110.

Rohia, A., & Mäntylä, K. (2019). 'It has given me this kind of courage …': the significance of CLIL in forming a positive target language self-concept. International Journal of Bilingual Education and Bilingualism, DOI: 10.1080/13670050.2019.1636761.

Rosenberg, M.B. (2012). Gewaltfreie Kommunikation. Eine Sprache des Lebens [Peaceful communication. The language of life]. Paderborn: Junfermann.

Rost, J. (2004). Lehrbuch Testtheorie–Testkonstruktion [Textbook test theory–test construction] (2nd ed.). Bern: Hans Huber.

Roth, T., Scharfenberg, F-J., Mierdel, J., & F.X. Bogner. (2020). Self-evaluative scientific modeling in an outreach gene technology laboratory. Journal of Science Education and Technology, 29, 725-739. https://doi.org/0.1007/s10956-020-09848-2.

Runco, M.A., Ancar, S., & Cayirdag, N. (2017). A closer look at the creativity gap and why students are less creative at school than outside of school. Thinking Skills and Creativity, 24, 242-249.

Sarmouk, C., et al. (2019). Pre-laboratory online learning resource improves preparedness and performance in pharmaceutical sciences practical classes. Innovations in Education and Teaching International, 57(4), 460-471.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness of-Fit Measures, Methods of Psychological Research Online, 8(2), 23-74.

Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2002). Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO [Scales to capture school self-concept (SESSKO)]. Göttingen: Hogrefe.

Shaw, J.M., Bunch, G.C., & Geaney, E.R. (2010). Analyzing language demands facing English learners on science performance assessments: the SALD framework. Journal of Research in Science Teaching, 47(8), 909-928.

Sisk, D. (1972). Relationship between self-concept and creativity: Theory into practice. Gifted Child Quarterly, 16, 229-234.

Sutton, C. (1996). Beliefs About Science and Beliefs About Language. International Journal of Science Education, 18 (1), 1–18.

Szumski, G., & Karwowski, M. (2019). Exploring the Pygmalion effect: The role of teacher expectations, academic self-concept, and class context in students' math achievement. Contemporary Educational Psychology, 59, doi: https://doi.org/10.1016/j.cedpsych.2019.101787.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington. DC: American Psychological Association.

Tsai, K-C. (2019). Investigating the empirical links between creative and critical thinking. Psychology, Society, & Education, 11(3), 267-280.

Usher, S. (2013). Letters of Note. Correspondence Deserving of a Wider Audience. Edinburgh, Scotland: Canongate.

Vandenberg, R. J. (2006). Statistical and Methodological Myths and Urban Legends. Organizational Research Methods, 9(2), 194-201.

van der Slik, F.W.P., van Hout, R.W.N.M., Schepens, J.J. (2015). The Gender Gap in Second Language Acquisition: Gender Differences in the Acquisition of Dutch among Immigrants from 88 Countries with 49 Mother Tongues. PLOS ONE, 10(11), 1-11.

Walker, J.P. & Sampson, V. (2013). Learning to argue and arguing to learn – driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. Journal of Research in Science Teaching, 50(5), 561-596.

Wang, J., Oliver, J.S., & Staver, J.R. (2008). Self-concept and science achievement: Investigating a reciprocal relation model across the gender classification in a cross-cultural context. Journal of Research in Science Teaching, 45(6), 711-725.

Yore, L.D., & Treagust, D.F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. International Journal of Science Education, 28(2–3), 291–314

.

Electronic Supplemental Materials

The Relevance of School Self-Concept and Creativity for CLIL Outreach Learning

Tamara Roth[1]*, Cathérine Conradty, Franz X. Bogner[1]

[1]Centre of Math & Science Education (Z-MNU), Department of Biology Education, University of Bayreuth, NW1, 95447 Bayreuth, Germany

*Correspondence: tamara.roth@uni-bayreuth.de; ORCID ID: 0000-0002-9062-1489

Table of Contents
(Total pages 5)

Electronic Supplemental Material 1: Language Workbook

DNA-MODELLING

Find the matching words to the definitions and German translations in your word-cloud and write them down!

English	Definition	Deutsch
Example: *clockwise*	**moving around in the same direction as the hands of a clock**	**rechtsgängig; im Uhrzeigersinn**
	not parallel	antiparallel
	the force causing molecules of the same substance to stick together	Zusammenhalt
	a single thin piece of thread, wire, hair, etc.	Stränge
	a chemical element; a gas that is the lightest of all the elements; it combines with oxygen to form water.	Wasserstoff
	strong connection	Bildung
	things that are different but together form a useful or attractive combination of skills, qualities or physical features	einander ergänzend

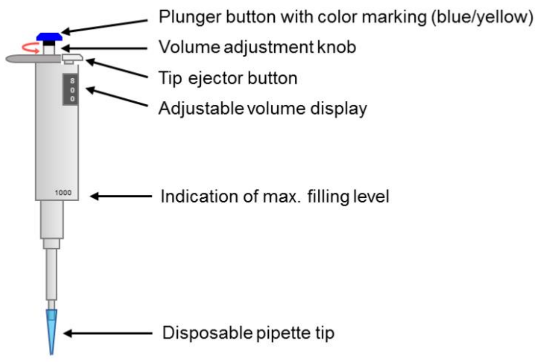
Electronic Supplemental Material 2: Laboratory Manual

WORKING WITH MICROPIPETTES

Using **micropipettes** is one of the most important working techniques in gene technology labs. The handling of such instruments is necessary for several experimentation steps today. Named after its manufacturer, micropipettes are also called ‘Eppendorf pipettes’.

With these special laboratory instruments, it is possible to pipette very low volumes of 2 µl to 1000 µl, depending on the selected micropipette and the previously adjusted volume.

Elements of a micropipette:

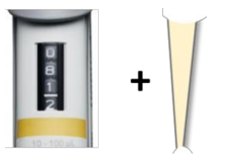
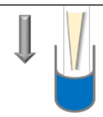
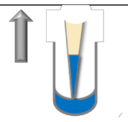
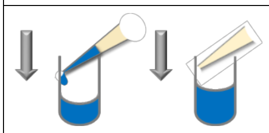


Please note:

Micropipettes are very expensive, sensitive instruments and must be handled carefully!

- **Micropipettes may only be used with a pipette tip attached!**
- **Always keep the tip of the pipettes downwards!**
- **Never use the same pipette tip for different substances!**

How to handle a micropipette:

1		Turn the ***volume adjustment knob*** **(Volumenregler)** to select the desired amount of liquid. The volume is given in µl (said **‘microliter’**) and read from top to bottom. *e.g., 81,2 µl are indicated in the picture on the left* Now put on a suitable ***pipette tip*** **(Pippetenspitze)**.
2		Press the plunger button (Druckknopf) down to the ***first pressure point*** **(erster Druckpunkt)** and hold, then immerse (eintauchen) the tip 0.5 cm in the liquid.
3		Slowly release (loslassen) the plunger button. The medium is sucked in (einsaugen).
4		To eject (auswefen) the medium from the pipette tip, press the plunger button down to the first pressure point again, then down to the ***second pressure point*** **(zweiter Druckpunkt)** so that the tip is completely empty. Always place droplets on the vessel wall (Gefäßwand).

Electronic Supplemental Material 3: SESSKO Scale Item Examples

1.	**Looking at what we have to know in school, I find that I am very … with the tasks**							
	Biology:	uncomfortable	☐	☐	☐	☐	☐	comfortable
	English:	uncomfortable	☐	☐	☐	☐	☐	comfortable
2.	**To me, assignments at school are …**							
	Biology:	more difficult than before	☐	☐	☐	☐	☐	easier than before
	English:	more difficult than before	☐	☐	☐	☐	☐	easier than before
3.	**I am …**							
	Biology:	less intelligent than my peers	☐	☐	☐	☐	☐	more intelligent than my peers
	English:	less intelligent than my peers	☐	☐	☐	☐	☐	more intelligent than my peers
4.	**Looking at what we have to do at school, I perceive myself as …**							
	Biology:	not gifted	☐	☐	☐	☐	☐	very gifted
	English:	not gifted	☐	☐	☐	☐	☐	very gifted
5.	**To me, assignments at school are …**							
	Biology:	more difficult as compared to my peers	☐	☐	☐	☐	☐	easier as compared to my peers
	English:	more difficult as compared to my peers	☐	☐	☐	☐	☐	easier as compared to my peers
6.	**Thinking about what we have to do at school, I perceive myself as …**							
	Biology:	not intelligent	☐	☐	☐	☐	☐	intelligent
	English:	not intelligent	☐	☐	☐	☐	☐	intelligent

Electronic Supplemental Material 4: Creativity Scale Item Examples

	During the last school term …	never	rarely	often	always
1.	… I made a connection between a current problem (task) and a similar situation I already faced.	☐	☐	☐	☐
2.	… While working on something, I tried to fully immerse myself in the experience.	☐	☐	☐	☐
3.	…. I imagined a possible solution to explore its usefulness in my mind.	☐	☐	☐	☐
4.	… when I got stuck on a problem, a solution just cames to me when I set it aside.	☐	☐	☐	☐
5.	… becoming physically involved in my work led me to good solutions.	☐	☐	☐	☐
6.	… While working on something, I tried to generate as many ideas as possible.	☐	☐	☐	☐
7.	… if I got stuck on a problem, I tried to take a different perspective of the situation.	☐	☐	☐	☐
8.	… if I got stuck on a problem, I asked others to help generate potential solutions.	☐	☐	☐	☐
9.	… I could completely lose track of time if I was intensely working.	☐	☐	☐	☐

10.	… If I was intensely working, I was fully aware of "the big picture."	☐	☐	☐	☐
11.	… I have used an old / proven solution to explore a new direction.	☐	☐	☐	☐
12.	… when I was intensely working, I didn't like to stop.	☐	☐	☐	☐
13.	… I got solutions to problems when my mind was relaxed.	☐	☐	☐	☐
14.	… I combined different ideas to develop a new idea.	☐	☐	☐	☐
15.	… while working on something I enjoyed, the work felt automatic and effortless.	☐	☐	☐	☐

6 ANHANG

Fragebogen

Es wird der Fragebogen (1) des Pilotfragebogens mit (A) Kreativität und (B) Persönlichkeit sowie exemplarisch der Vortest des Fragebogens (2) für das CLIL-Modul gezeigt. Dieser enthält alle verwendeten Items und Teilfragebägen zu Wissen (A), Sprachkompetenz (B), Modellverständnis (C), Schulbezogenes Selbstkonzept für Biologie (D), Schulbezogenes Selbstkonzept für Englisch (D), Kreativität (E), Notenskala (F), Persönlichkeit (G). Die Fragen zu A, B, C, D und E wurden zu allen Testzeitpunkten abgefragt. Im Nach- und Behaltenstest wurden die Reihenfolge der Wissensfragen einschließlich der Antwortoptionen sowie die Itemreihenfolge des Wissensfragebogen zufällig vertauscht

Fragebogen 1 (Pilotfragebogen)

Liebe Schülerin, lieber Schüler,

dieser Fragebogen ist Teil einer wissenschaftlichen Untersuchung und **streng vertraulich**.
Er wird **nicht von deiner Lehrkraft eingesehen oder benotet**.

Bitte bearbeite **alle** Fragen alleine, sorgfältig und wahrheitsgemäß.
Bitte fülle den Fragebogen mit einem **dunklen Stift** aus (keine hellen Stifte, Neonfarben oder Bleistifte verwenden).

Wenn du fertig bist, **kontrolliere** bitte, ob du alle Seiten ausgefüllt hast.

Genehmigung des Bayerischen Staatsministeriums für Unterricht und Kultus vom 04.10.2019 mit dem Aktenzeichen: IV.7-BO5106/149/18

Vielen Dank, dass du an dieser Befragung teilnimmst!

Datum								
					2	0		
	TT		MM		JJJJ			

Persönlicher Code:

Durch diesen Code können wir nicht mehr nachvollziehen wer diesen Fragebogen ausgefüllt hat, jedoch die Fragebögen untereinander zuordnen.

Kürze dein Geschlecht mit **M** (männlich) bzw. **W** (weiblich) ab.
Trage den Monat deines Geburtstages ein (z.B. **08** für August, **12** für Dezember).
Trage das Jahr deiner Geburt ein (z.B. **99** für 1999, **00** für 2000).
Trage die zwei ersten Buchstaben des Vornamens deiner Mutter ein (z.B. **CL** für Claudia).
Trage die Hausnummer ein, in der du wohnst (z.B. **003** für Hausnummer 3).

Geschlecht	**Geburtsmonat**		**Geburtsjahr**		**Mutter**		**Hausnummer**		

Beispiel: Maximilian ist männlich, geboren im September 2001, seine Mutter heißt Andrea und er wohnt in Hausnummer 61. ****Sein Code lautet: **M0901AN061*****

Teilfragebogen A - Kreativität: Während des vergangenen Schuljahres –wie oft hast Du Folgendes gemacht?

	nie	selten	manchmal	häufig	immer
1. Wenn ich körperlich in meine Arbeit eingebunden bin erziele ich gute Ergebnisse	☐	☐	☐	☐	☐
2. Wenn ich konzentriert an etwas arbeite, mag ich es nicht unterbrochen zu werden	☐	☐	☐	☐	☐
3. Wenn ich konzentriert arbeite, verliere ich das „große Ganze“ nicht aus den Augen	☐	☐	☐	☐	☐
4. Wenn ich an etwas arbeite, versuche ich so viele Ideen wie nur möglich zu entwickeln	☐	☐	☐	☐	☐
5. Wenn ich bei einer Problemstellung nicht weiterkomme, suche ich in meiner Umgebung nach möglichen Hinweisen	☐	☐	☐	☐	☐
6. Wenn ich bei einer Problemstellung nicht weiterkomme, lege ich diese beiseite die Lösung fällt mir irgendwann von selbst ein	☐	☐	☐	☐	☐
7. Über mehr als nur eine Idee gleichzeitig nachzudenken eröffnet einen anderen Blickwinkel	☐	☐	☐	☐	☐
8. Wenn ich bei einer Problemstellung nicht weiterkomme, versuche ich meine Perspektive auf die Situation zu verändern	☐	☐	☐	☐	☐
9. Arbeiten, die ich gerne verrichte, erscheinen mir intuitiv und mühelos	☐	☐	☐	☐	☐
10. Sich mögliche Lösungen zu einer Problemstellung auszudenken führt zu neuen Erkenntnissen	☐	☐	☐	☐	☐
11. Wenn ich bei einer Problemstellung nicht weiterkomme, frage ich andere nach möglichen Lösungen	☐	☐	☐	☐	☐
12. Die Lösungen zu Problemstellungen fallen mir ein, wenn ich entspannt bin	☐	☐	☐	☐	☐
13. Wenn ich an etwas arbeite, versuche ich vollkommen in meiner Tätigkeit zu versinken	☐	☐	☐	☐	☐

	nie	selten	manchmal	häufig	immer
14. Ich verliere vollkommen das Gefühl für Zeit, wenn ich an etwas konzentriert arbeite	☐	☐	☐	☐	☐
15. Wenn ich bei einer Problemstellung nicht weiterkomme, suche ich nach Details, die ich normalerweise gar nicht bemerken würde	☐	☐	☐	☐	☐

Teilfragebogen B - Persönlichkeit: Inwieweit treffen die folgenden Aussagen auf dich zu?

	Stimmt… gar nicht	eher nicht	eher schon	voll und ganz
1. Ich bin eher zurückhaltend, reserviert.	☐	☐	☐	☐
2. Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	☐	☐	☐	☐
3. Ich bin bequem, neige zur Faulheit.	☐	☐	☐	☐
4. Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	☐	☐	☐	☐
5. Ich habe nur wenig künstlerisches Interesse.	☐	☐	☐	☐
6. Ich gehe aus mir heraus, bin gesellig.	☐	☐	☐	☐
7. Ich neige dazu, andere zu kritisieren.	☐	☐	☐	☐
8. Ich erledige Aufgaben gründlich.	☐	☐	☐	☐
9. Ich werde leicht nervös und unsicher.	☐	☐	☐	☐
10. Ich habe eine aktive Vorstellungskraft, bin phantasievoll.	☐	☐	☐	☐
11. Ich bin rücksichtsvoll zu anderen, einfühlsam.	☐	☐	☐	☐

Fragebogen 2 (CLIL-Modul)

Teilfragebogen A: Beantworte nun die folgenden Fragen zu deinem Wissen. Es gibt immer nur 1 richtige Antwort, deshalb setze bitte nur 1 Kreuz pro Frage. Wenn du die Antwort nicht weißt, kreuze die Frage nicht an!

Teilfragebogen B: Cloze-Test sprachliches Wissen (Robinson 1974) Es gibt immer nur 1 richtige Antwort, deshalb setze bitte nur 1 Kreuz pro Frage. Wenn du die Antwort nicht weißt, kreuze die Frage nicht an!

Teilfragebogen C: Modellverständnis (Treagust, Chittleborough & Mamiala 2002) Bewerte die folgenden Aussagen zu Modellen, indem du im entsprechenden Kästchen 1 Kreuz setzt.

Teilfragebogen D: Das schulbezogene Fähigkeitsselbstkonzept (Schöne et al. 2003) Bewerte die folgenden Aussagen, indem du im entsprechenden Kästchen 1 Kreuz setzt (Auszug aufgrund von Copyright-Bestimmungen)

Teilfragebogen E: Kreativität (Miller & Dumford 2014) Bewerte die folgenden Aussagen, indem du im entsprechenden Kästchen 1 Kreuz setzt.

Teilfragebogen F: Gib deine Schulnote an, indem du im entsprechenden Kästchen 1 Kreuz setzt.

Teilfragebogen G: Persönlichkeit (Rammstedt & John 2007) Bewerte die folgenden Aussagen, indem du im entsprechenden Kästchen 1 Kreuz setzt.

Fragebogen zum Demonstrationslabor Bio-/ Gentechnik der Universität Bayreuth im Kontext „Ein bilinguales Modul zur DNA als Träger der Erbinformation“

Liebe Schülerin, lieber Schüler,

dieser Fragebogen ist Teil einer wissenschaftlichen Untersuchung und **streng vertraulich**.
Er wird **nicht von deiner Lehrkraft eingesehen oder benotet**.

Bitte bearbeite **alle** Fragen alleine, sorgfältig und wahrheitsgemäß.
Bitte fülle den Fragebogen mit einem **dunklen Stift** aus (keine hellen Stifte, Neonfarben oder Bleistifte verwenden).

Wenn du fertig bist, **kontrolliere** bitte, ob du alle Seiten ausgefüllt hast.

Genehmigung des Bayerischen Staatsministeriums für Unterricht und Kultus vom 04.10.2019 mit dem Aktenzeichen: IV.7-BO5106/149/18

Vielen Dank, dass du an dieser Befragung teilnimmst!

Datum [ ][ ] [ ][ ] [2][0][ ][ ]

TT MM JJJJ

Persönlicher Code:

Durch diesen Code können wir nicht mehr nachvollziehen wer diesen Fragebogen ausgefüllt hat, jedoch die Fragebögen untereinander zuordnen.

Kürze dein Geschlecht mit **M** (männlich) bzw. **W** (weiblich) ab.
Trage den Monat deines Geburtstages ein (z.B. **08** für August, **12** für Dezember).
Trage das Jahr deiner Geburt ein (z.B. **99** für 1999, **00** für 2000).
Trage die zwei ersten Buchstaben des Vornamens deiner Mutter ein (z.B. **CL** für Claudia).
Trage die Hausnummer ein, in der du wohnst (z.B. **003** für Hausnummer 3).

Geschlecht	**Geburtsmonat**		**Geburtsjahr**		**Mutter**		**Hausnummer**		

Beispiel: Maximilian ist männlich, geboren im September 2001, seine Mutter heißt Andrea und er wohnt in Hausnummer 61. ****Sein Code lautet: **M0901AN061*****

Teilfragebogen A - Wissen

1	**Ein sogenannter „DNA-Längenstandard“ dient …**
☐	dem Anfärben von DNA-Strängen.
☐	der Messung der Länge eines DNA-Fragments.
☐	der Reparatur von DNA-Abschnitten.
☐	der Verlängerung der DNA-Bereiche.

2	**Bei der Analyse eines DNA-Abschnittes ergibt sich einen Anteil von Guanin mit 30 %. Der Anteil von Adenin ist somit …**
☐	20 %.
☐	70 %.
☐	ebenfalls 30%.
☐	nicht bestimmbar.

3	**Was stimmt <u>nicht</u>? Die Wanderungsgeschwindigkeit eines Moleküls durch das Elektrophoresegel ist abhängig von …**
☐	der Dichte der Probe.
☐	der Dichte des Elektrophoresegels.
☐	der angelegten Spannung.
☐	der Größe der Moleküle.

4	**Die beiden DNA-Stränge sind …**
☐	gegenläufig.
☐	unabhängig voneinander.
☐	identisch.
☐	versetzt voneinander.

5	**Einen DNA-Abschnitt, der die Grundinformation für die Ausbildung eines bestimmten Merkmals trägt, nennt man …**
☐	Genom.
☐	Plasmid.
☐	Chromosom.
☐	Gen.

6	**Die DNA ist Träger der Erbinformation bei …**
☐	allen Organismen außer Bakterien.
☐	allen Organismen.
☐	den Wirbeltieren.
☐	den Menschenaffen.

7	**Die DNA besteht aus folgenden Atomsorten:**
☐	Wasserstoff, Sauerstoff, Schwefel, Kohlenstoff und Stickstoff
☐	Wasserstoff, Sauerstoff, Phosphor, Kohlenstoff und Stickstoff
☐	Wasserstoff, Sauerstoff, Phosphor, Schwefel und Stickstoff
☐	Wasserstoff, Schwefel, Phosphor, Kohlenstoff und Stickstoff

8	**Welche Basenpaarung ist korrekt?**
☐	Guanin paart mit Cytosin.
☐	Adenin paart mit Guanin.
☐	Cytosin paart mit Adenin.
☐	Thymin paart mit Cytosin.

9	**Mit Hilfe der Gelelektrophorese lassen sich Aussagen treffen über …**
☐	die Anzahl der Bindungen eines Moleküls.

10	**Was stimmt nicht?** **Die DNA des Menschen…**
☐	ist ein langes Kettenmolekül.

☐	die Bestandteile eines Moleküls.
☐	die Molekülmasse.
☐	die Atome eines Moleküls.

☐	ist aus Aminosäuren aufgebaut.
☐	ist ein Makromolekül.
☐	ist Träger der Erbinformation.

11	**Um 20 µl einer Flüssigkeit zu einer Probe zu geben, verwendet man …**
☐	eine Mikropipette.
☐	eine Messpipette.
☐	einen Messzylinder.
☐	eine Pasteur-Pipette.

12	**Die Erbinformation der DNA wird verschlüsselt durch die …**
☐	Windungen des DNA-Stranges.
☐	Verschmelzung von Ei- und Spermienzelle bei der Befruchtung.
☐	Bildung verschiedener Chromosomen.
☐	Abfolge der einzelnen Basen.

13	**Die DNA-Basen befinden sich …**
☐	an der Außenseite des DNA-Moleküls an Phosphat gebunden.
☐	im Inneren des DNA-Moleküls an Phosphat gebunden.
☐	im Inneren des DNA-Moleküls an den Zucker gebunden.

14	**Was stimmt <u>nicht</u>?** **Die DNA ist in kaltem Alkohol …**
☐	unlöslich.
☐	löslich.
☐	als fädige Struktur zu erkennen.

☐	an der Außenseite des DNA-Moleküls an den Zucker gebunden.

☐	ein weißer Feststoff.

15	**Die Auftrennung der DNA-Moleküle bei der Elektrophorese beruht auf dem DNA-Bestandteil…**
☐	Cytosin
☐	Phosphat
☐	Zucker
☐	Thymin

16	**Die Abkürzung DNA steht für …**
☐	Oxyribonukleinsäure.
☐	Desoxyribonukleinsäure.
☐	Desoxynukleinsäure.
☐	Didesoxyribonukleinsäure

17	**Mit Hilfe einer Zentrifuge …**
☐	werden feste Stoffe von Flüssigkeiten getrennt.
☐	wird die Probe durchmischt.
☐	können einzelne Moleküle isoliert werden.
☐	werden die Moleküle in Schwingung gebracht.

18	**Welcher der folgenden Bestandteile ist nicht in der DNA enthalten?**
☐	Ribose
☐	Desoxyribose
☐	Guanin
☐	Adenin

19	**Ein DNA-Einzelstrang hat folgende Basenabfolge: AATGGG** *(Großbuchstabe = Anfangsbuchstabe Base?*

20	**Ein positiv geladenes Teilchen wandert im elektrischen Feld …**

	Wie lautet die Basenabfolge des gegenüberliegenden, paarenden DNA-Einzelstrangs?
☐	TTGAAA
☐	GGACCC
☐	TTGCCC
☐	TTACCC

☐	zum negativen Pol.
☐	zum positiven Pol.
☐	überhaupt nicht.
☐	zwischen beiden Polen hin und her.

21	Die molekulare Struktur der DNA lässt sich am besten vergleichen mit …
☐	einem Bindfaden.
☐	einer Pappröhre.
☐	einer eingedrehten Strickleiter.
☐	einer Bahnschiene.

22	Die Gesamtlänge der menschlichen DNA pro Zelle beträgt etwa …
☐	2 cm.
☐	200 m.
☐	2 m.
☐	20 m.

23	1962 erhielten James Watson und Francis Crick den Nobelpreis für Medizin für die Entdeckung …
☐	der Bestandteile der DNA.

24	Das Verhältnis von Zucker zu Phosphat im DNA-Molekül beträgt …
☐	2:1

☐	der Gelelektrophorese.
☐	der DNA im Zellkern.
☐	der Doppelhelixstruktur der DNA.

☐	1:2
☐	3:1
☐	1:1

25	**Der Zusammenhalt der beiden DNA-Stränge entsteht durch Ausbildung von …**
☐	Ionischen Wechselwirkungen.
☐	Wasserstoffbrückenbindungen.
☐	Schwefelbrückenbindungen.
☐	Atombindungen.

26	**Die räumliche Struktur der DNA …**
☐	ist eine linksgängige Doppelhelix.
☐	ist eine abwechselnd rechts- und linksgängige Doppelhelix.
☐	ist eine rechtsgängige Doppelhelix.
☐	besitzt keine Drehrichtung.

27	**Was stimmt <u>nicht</u>?** **Das Sichtbarmachen der DNA-Moleküle bei der Gelelektrophorese wird möglich durch …**
☐	einen Farbstoff, der im UV-Licht leuchtet.
☐	einen Farbstoff, der an DNA-Moleküle bindet.
☐	den blau eingefärbten Auftragspuffer.

28	**Spricht man vom „Rückgrat der DNA“, dann meint man damit …**
☐	die zum Schutz der DNA gebundenen Fettsäuren.
☐	die Paarung der DNA-Basen.
☐	die ringförmige Struktur der DNA.

☐	die Farbstoff-Zugabe in das Gel.

☐	die Kette aus Phosphat abwechselnd mit Desoxyribose als Bestandteil der DNA.

29	In welchem Zellorganell befindet sich die DNA?
☐	im Zellplasma
☐	im Ribosom
☐	in der Vakuole
☐	im Zellkern

30	Wie viele verschiedene DNA-Bausteine gibt es?
☐	4
☐	8
☐	6
☐	2

Teilfragebogen B - Sprachkompetenz: Extracted from Organelles (Discover Biology, 2010, S.227)

Cells contain organelles, which are structures __(1)__ carry out specific tasks within the __(2)__. The name “organelle” means “little organ”. __(3)__ name shows that the relationship between __(4)__, which are often enclosed by membranes, __(5)__ cells is similar to the relationship __(6)__ an organism and its organs. Examples __(7)__ organelles are, inter alia[2], the nucleus, __(8)__, and vacuoles of different sizes. The __(9)__ contains all the information required to __(10)__ the functions of the cell. Mitochondria __(11)__ very important because they are the “__(12)__ stations” of all living cells. Vacuoles __(13)__ different sizes carry out functions of __(14)__, transport and excretion.

(1)	☐	these	☐	which	☐	who	☐	that
(2)	☐	cell	☐	vacuole	☐	plastid	☐	mitochondria
(3)	☐	Those	☐	Her	☐	This	☐	That
(4)	☐	organelles	☐	organs	☐	origamis	☐	organics
(5)	☐	also	☐	or	☐	but	☐	and
(6)	☐	concerned	☐	amid	☐	inter	☐	between
(7)	☐	pro	☐	of	☐	for	☐	from
(8)	☐	blood cells	☐	mitochondria	☐	membrane	☐	liver

[2] Inter alia = unter anderem

(9)	☐	essence	☐	core	☐	centre	☐	nucleus
(10)	☐	control	☐	govern	☐	switch	☐	power
(11)	☐	have been	☐	are	☐	will be	☐	is
(12)	☐	command	☐	strength	☐	electricity	☐	power
(13)	☐	with	☐	for	☐	from	☐	of
(14)	☐	room	☐	accommodation	☐	storage	☐	loading

Teilfragebogen C - Modellverständnis

	Folgender Aussage stimme ich …	**absolut nicht zu**	**nicht zu**	**weder Zustimmung noch Ablehnung**	**zu**	**stark zu**
1.	Modelle werden zur Erklärung wissenschaftlicher Phänomene benutzt.	☐	☐	☐	☐	☐
2.	Modelle werden zur Formulierung von Ideen und Theorien über wissenschaftliche Ereignisse benutzt.	☐	☐	☐	☐	☐
3.	Modelle werden benutzt, um Prognosen über ein wissenschaftliches Ereignis herzustellen und zu testen.	☐	☐	☐	☐	☐
4.	Alles an einem Modell sollte erkennen lassen, was es abbildet.	☐	☐	☐	☐	☐
5.	Modelle werden verwendet, um eine Idee aufzuzeigen.	☐	☐	☐	☐	☐
6.	Modelle helfen dabei, sich wissenschaftliche Geschehnisse gedanklich besser vorzustellen.	☐	☐	☐	☐	☐
7.	Ein Modell kann sich ändern, wenn neue Erkenntnisse vorliegen.	☐	☐	☐	☐	☐
8.	Ein Modell muss nah am Realobjekt sein, sodass es niemand widerlegen kann.	☐	☐	☐	☐	☐

9.	Ein Modell kann sich ändern, wenn neue Änderungen der Daten oder Ansichten auftreten.	☐	☐	☐	☐	☐
10.	Ein Modell kann sich ändern, wenn neue Theorien oder Beweise etwas anderes besagen.	☐	☐	☐	☐	☐
11.	Ein Modell muss nah am Realobjekt sein.	☐	☐	☐	☐	☐
12.	Modelle werden benutzt, um ihre Funktion in wissenschaftlichen Untersuchungen aufzuzeigen.	☐	☐	☐	☐	☐
13.	Ein Modell sollte eine exakte Kopie sein.	☐	☐	☐	☐	☐
14.	Modelle werden benutzt, um etwas physisch oder visuell zu repräsentieren.	☐	☐	☐	☐	☐

Teilfragebogen D - Schulbezogenes Selbstkonzept Biologie und Englisch

1.	**Wenn ich mir angucke, was wir in der Schule können müssen, finde ich, dass ich mit den Aufgaben in der Schule …**							
	Biologie:	nicht gut zurechtkomme	☐	☐	☐	☐	☐	gut zurechtkomme
	Englisch:	nicht gut zurechtkomme	☐	☐	☐	☐	☐	gut zurechtkomme
2.	**Die Aufgaben in der Schule fallen mir …**							
	Biologie:	schwerer als meinen Mitschüler*Innen	☐	☐	☐	☐	☐	leichter als meinen Mitschüler*Innen
	Englisch:	schwerer als meinen Mitschüler*Innen	☐	☐	☐	☐	☐	leichter als meinen Mitschüler*Innen
3.	**Ich bin …**							
	Biologie:	weniger intelligent als meine Mitschüler*Innen	☐	☐	☐	☐	☐	intelligenter als meine Mitschüler*Innen
	Englisch:	weniger intelligent als meine Mitschüler*Innen	☐	☐	☐	☐	☐	intelligenter als meine Mitschüler*Innen
4.	**Wenn ich mir angucke, was wir in der Schule können müssen, halte ich mich für …**							
	Biologie:	nicht begabt	☐	☐	☐	☐	☐	sehr begabt
	Englisch:	nicht begabt	☐	☐	☐	☐	☐	sehr begabt
5.	**Mit den Aufgaben in der Schule komme ich …**							
	Biologie:	schlechter zurecht als meine Mitschüler*Innen	☐	☐	☐	☐	☐	besser zurecht als meine Mitschüler*Innen
	Englisch:	schlechter zurecht als meine Mitschüler*Innen	☐	☐	☐	☐	☐	besser zurecht als meine Mitschüler*Innen

Teilfragebogen E - Kreativität

Im vergangenen Schuljahr …		niemals	manchmal	oft	sehr oft
1.	… habe ich eine alte / bewährte Lösung genutzt, um damit einen neuen Weg einzuschlagen.	☐	☐	☐	☐
2.	… ist mir die Lösung zu einer komplizierten Problemstellung irgendwann von selbst eingefallen, wenn ich die Arbeit beiseitegelegt habe.	☐	☐	☐	☐
3.	… habe ich verschiedene Ideen zusammengefügt, um eine neue Idee zu entwickeln.	☐	☐	☐	☐
4.	… habe ich andere um Hilfe gebeten, um mögliche Lösungen für ein Problem zu entwickeln.	☐	☐	☐	☐
5.	… habe ich eine Verbindung hergestellt zwischen einem aktuellen Problem (Aufgabe) und einer ähnlichen Situation, die ich schon meisterte.	☐	☐	☐	☐
6.	… wollte ich nicht unterbrochen werden, wenn ich konzentriert an etwas arbeitete.	☐	☐	☐	☐
7.	… habe ich mir eine mögliche Lösung vorgestellt, um ihre Brauchbarkeit in Gedanken zu erforschen.	☐	☐	☐	☐
8.	… habe ich beim konzentrierten Arbeiten nie das allgemeine Ziel aus den Augen verloren.	☐	☐	☐	☐

9.	… habe ich bei einer neuen Aufgabe versucht, so viele Ideen wie möglich zu entwickeln.	☐	☐	☐	☐
10.	… habe ich ein Problem oder eine Aufgabe aus einem anderen Blickwinkel betrachtet, um eine Lösung zu finden.	☐	☐	☐	☐
11.	… bin ich komplett in meine Arbeit an einem Problem (Aufgabe) versunken.	☐	☐	☐	☐
12.	… fallen mir Lösungen zu Problemstellungen ein, wenn ich entspannt bin.	☐	☐	☐	☐
13.	… habe ich gefühlt, dass die Arbeit automatisch und mühelos war, während ich eine angenehme Aufgabe erledigte.	☐	☐	☐	☐
14.	… erzielte ich gute Ergebnisse, wenn ich körperlich in meine Arbeit eingebunden war..	☐	☐	☐	☐
15.	… habe ich komplett die Zeit aus den Augen verloren, wenn ich intensiv an etwas gearbeitet habe.	☐	☐	☐	☐

Teilfragebogen F - Schulnoten

	Notenskala	1	2	3	4	5	6
1.	Schulnote Biologie	□	□	□	□	□	□
2.	Schulnote Englisch	□	□	□	□	□	□

Teilfragebogen G - Persönlichkeit

Inwieweit treffen die folgenden Aussagen auf dich zu?

	Stimmt… gar nicht	eher nicht	eher schon	voll und ganz
1. Ich bin eher zurückhaltend, reserviert.	☐	☐	☐	☐
2. Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	☐	☐	☐	☐
3. Ich bin bequem, neige zur Faulheit.	☐	☐	☐	☐
4. Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	☐	☐	☐	☐
5. Ich habe nur wenig künstlerisches Interesse.	☐	☐	☐	☐
6. Ich gehe aus mir heraus, bin gesellig.	☐	☐	☐	☐
7. Ich neige dazu, andere zu kritisieren.	☐	☐	☐	☐
8. Ich erledige Aufgaben gründlich.	☐	☐	☐	☐
9. Ich werde leicht nervös und unsicher.	☐	☐	☐	☐
10. Ich habe eine aktive Vorstellungskraft, bin phantasievoll.	☐	☐	☐	☐
11. Ich bin rücksichtsvoll zu anderen, einfühlsam.	☐	☐	☐	☐

(Eidesstattliche) Versicherungen und Erklärungen

(§ 8 Satz 2 Nr. 3 PromO Fakultät)
Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).

(§ 8 Satz 2 Nr. 3 PromO Fakultät)
Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.

(§ 8 Satz 2 Nr. 4 PromO Fakultät)
Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. –vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.

(§ 8 Satz 2 Nr. 7 PromO Fakultät)
Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung der Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.

(§ 8 Satz 2 Nr. 8 PromO Fakultät)
Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

………………………………………..	……………………………………
Ort, Datum	Unterschrift