

# **Untersuchung zur Wirksamkeit hybrider und virtueller Lernumgebungen zum naturwissenschaftlichen Arbeiten in der Biotechnologie**

Julia Margareta Stich

Vollständiger Abdruck der von der TUM School of Social Sciences and Technology der Technischen Universität München zur Erlangung einer Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitz: Prof. Dr. phil. Maria Bannert

Prüfende der Dissertation:

1. Prof. Dr. rer. nat. Claudia Nerdel
2. Prof. Dr. Steffen Schaal

Die Dissertation wurde am 20.11.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Social Sciences and Technology am 21.02.2024 angenommen.



---

## Zusammenfassung

Infolge der COVID-19 Pandemie sind Begriffe wie "Homeschooling" und "Flipped Classroom" für Schüler\*innen und Lehrkräfte zum Alltag geworden. Der Einsatz digitaler Medien in unterschiedlicher Qualität kennzeichnete die Zeit der geschlossenen Schulen. Studien zeigen, dass der Lernfortschritt oder -verlust der Schüler\*innen während dieser Zeit den Sommerferien entspricht (Hammerstein et al., 2021). Nur in Studien (Spitzer & Musslick, 2021; van der Velde et al., 2021), in denen eine Art Online-Lernsoftware zur Bewertung der Schülerleistungen während dieser Zeit eingesetzt wurde, konnte ein positiver Lernzuwachs gemessen werden. Neben dem normalen Fachunterricht entfielen auch die in den MINT-Fächern wichtigen Profilstunden, die naturwissenschaftliche Arbeitsweisen fördern. Für die Schulen war es daher notwendig, alternative Lehrformate zu entwickeln, um den entstandenen Lernausfall für die Schüler\*innen zu kompensieren. Virtuelle Lernumgebungen können dies teilweise leisten. So können virtuelle Experimente sogar einen Mehrwert gegenüber realen Experimenten bieten, indem sie den Schüler\*innen ermöglichen, unbeobachtbare Phänomene zu erforschen und beobachtbare und unbeobachtbare Phänomene miteinander zu verbinden (de Jong et al., 2013). Überdies kann ein ständiges Feedback in einer digitalen Lernumgebung das Verständnis der Schüler\*innen für Konzepte und Prozesse fördern (Thisgaard & Makransky, 2017).

In dieser Studie soll daher der Frage nachgegangen werden, inwiefern unterschiedlich gestaltete digitale Lernumgebungen (hybrid oder virtuell) das naturwissenschaftliche Lernen bzw. den Kompetenzerwerb in praxisorientierten Fächern effektiv unterstützen können. Um die Wirksamkeit von vier ökologisch validen Varianten in einem einfaktoriellen Versuchsdesign zu untersuchen, wurden die in Präsenz (P) und digital (D) präsentierten Lernumgebungen systematisch nach Schulsituation variiert. Daraus ergaben sich die Szenarien PP "Präsenzunterricht", DP "Flipped Classroom", PD "Wechselunterricht" und DD "Homeschooling". Diese Varianten werden am Beispiel der molekularbiologischen Methoden PCR und Agarose-Gelelektrophorese durch den Einsatz eines interaktiven Buches im Vergleich zum klassischen Klassenunterricht und eines virtuellen Labors für Biotechnologie im Vergleich zum entsprechenden analogen Schülerversuch umgesetzt. Die Umsetzung erfolgte im Rahmen des "Kriminalfalls" im Themenschwerpunkt Genetik und Gentechnik.

Die virtuellen Lernumgebungen Interactive Book und Digitales Labor wurden in H5P erstellt. Bei ihrer Erstellung wurden die verwendeten Materialien in Bezug auf Umfang, Ausstattung, Illustrationen, Durchführungsdauer etc. parallel zum Klassenunterricht und zum Schülerexperiment aufgebaut. Der Klassenunterricht wurde von der Biologielehrkraft des jeweiligen Kurses durchgeführt, die die vorgegebenen Inhalte umsetzte. Die Teilnehmer\*innen

---

der Studie im Sommer 2021 und 2022 waren Schüler\*innen von Fachoberschulen und Gymnasien der 11. Jahrgangsstufe (Gymnasium N = 117; FOS N = 37, ♀ = 45 %, ♂ = 55 %). Um für alle Schüler\*innen gleiche Voraussetzungen zu gewährleisten und alle Schüler\*innen unabhängig von ihrem Interesse zu erreichen, wurde die Studie an einem Schulvormittag durchgeführt. Um die Veränderung des Fachwissens (FW) und des praktischen Wissenszuwachses (NWA) im Zusammenhang mit der PCR und der Agarose-Gelelektrophorese zu überprüfen, wurde eine Prä-, Post- und Follow-up-Testung mithilfe eines Fragebogens (MC, 2 aus 5) durchgeführt.

Die FW- und NWA-Skalen wurden mit einer Rasch-Analyse überprüft und zeigten gute Item-Reliabilitäten (>0.9) und Item-Separationen (>4). Die Fragebögen wurden geankert, um eine gleiche Skala zu erhalten und daraus die Personenfähigkeiten der Lernenden zu berechnen. Der kognitive Fähigkeitstest zum ersten Messzeitpunkt und die Schulform wurden als Kovariaten verwendet, um Gruppenunterschiede zu kontrollieren. Zur Überprüfung der Gruppenunterschiede zu den jeweiligen Messzeitpunkten wurden die Verfahren der einfaktoriellen ANCOVA (einfaktorielle Kovarianzanalyse) und zum Vergleich der Messzeitpunkte der einzelnen Gruppen eine ANOVA mit Messwiederholung und Kovariaten verwendet.

Aus diesen Daten geht hervor, dass die Betrachtungen zu den Messzeitpunkten in der Regel keine Unterschiede zwischen den Gruppen, insbesondere hinsichtlich der Nachhaltigkeit des Lernens, ergaben. Die Entwicklung der Gruppen über die Zeit zeigte jedoch signifikante Unterschiede. Das hybride Design des „Flipped Classroom“, die Kombination der Vermittlung des Fachwissens über eine virtuelle Lernumgebung mit dem in Präsenz durchgeführten Schülerversuch zur Vermittlung naturwissenschaftlicher Arbeitsweisen, hatte in dieser Studie den nachhaltigsten Effekt auf die Entwicklung der Personenfähigkeiten. Dies lässt den Schluss zu, dass in vorliegender Studie der Einsatz von hybriden oder virtuellen Formaten dem klassischen Präsenzunterricht, bezogen auf seine Nachhaltigkeit ähneln. Die in dieser Studie erzielten Ergebnisse sind für den Einsatz virtueller Lernumgebungen als mögliches Standardmedium für den Biologieunterricht von großer Bedeutung, da die Studie mit Schüler\*innen mit unterschiedlicher Motivation und Vorkenntnissen und unter realen Bedingungen durchgeführt wurde.

## Abstract

As a result of the COVID-19 pandemic, terms such as "homeschooling" and "flipped classroom" have become part of everyday life for pupils and teachers. The use of digital media of varying quality characterised the period of school closures. Studies show that students' learning progress or loss thereof during this time corresponds to the summer holidays (Hammerstein et al., 2021). Only in studies (Spitzer & Musslick, 2021; van der Velde et al., 2021) in which a type of online learning software was used to assess student performance during this time was it possible to measure a positive learning curve. In addition to the normal subject lessons, the profile lessons, which are important in the STEM subjects and promote scientific working methods, were also cancelled. It was therefore necessary for schools to develop alternative teaching formats to compensate the loss of learning experienced by students. Virtual learning environments can do this to some extent. Virtual experiments can even offer added value compared to real experiments by enabling students to explore unobservable phenomena and combine observable and unobservable phenomena (de Jong et al., 2013). Furthermore, constant feedback in a digital learning environment can promote students' understanding of concepts and processes (Thisgaard & Makransky, 2017).

This study therefore aims to investigate the extent to which differently designed digital learning environments (hybrid or virtual) can effectively support science learning and the acquisition of competences in practice-oriented subjects. In order to investigate the effectiveness of four ecologically valid variants in a single-factor experimental design, the learning environments presented in presence (P) and digitally (D) were systematically varied according to the school situation. This resulted in the scenarios PP "face-to-face teaching", DP "flipped classroom", PD "alternating teaching" and DD "homeschooling". These variants are implemented using the example of the molecular biology methods PCR and agarose gel electrophoresis through the use of an interactive book in comparison to traditional classroom teaching and a virtual laboratory for biotechnology in comparison to the corresponding analogue student experiment. The implementation took place as part of the "criminal case" in the subject area of genetics and genetic engineering.

The virtual learning environments Interactive Book and Digital Lab were created in H5P. During their creation, the materials used were developed in parallel to the classroom lessons and the student experiment in terms of scope, equipment, illustrations, implementation time, etc. The class lessons were conducted by the biology teacher of the respective course, who implemented the specified content. The participants in the study in the summer of 2021 and 2022 were students from 11th grade technical colleges and grammar schools (grammar school N = 117; FOS N = 37, ♀ = 45 %, ♂ = 55 %). In order to ensure the same conditions for all pupils and to reach all pupils regardless of their interest, the study was conducted on a school

morning. A pre-, post- and follow-up test was carried out using a questionnaire (MC, 2 out of 5) to assess the change in specialised knowledge (FW) and practical knowledge (NWA) in connection with PCR and agarose gel electrophoresis.

The FW and NWA scales were checked with a Rasch analysis and showed good item reliabilities ( $>0.9$ ) and item separations ( $>4$ ). The questionnaires were anchored to obtain the same scale and to calculate the learners' personal abilities. The cognitive ability test at the first measurement point and the type of school were used as covariates to control for group differences. Single-factor ANCOVA (single-factor analysis of covariance) was used to analyse the group differences at the respective measurement points and ANOVA with repeated measures and covariates was used to compare the measurement points of the individual groups.

These data show that the observations at the measurement points generally revealed no differences between the groups, particularly with regard to the sustainability of learning. However, the development of the groups over time showed significant differences. The hybrid design of the "flipped classroom", the combination of teaching specialised knowledge via a virtual learning environment with the face-to-face student experiment to teach scientific working methods, had the most lasting effect on the development of personal skills in this study. This leads to the conclusion that in this study, the use of hybrid or virtual formats is similar to traditional face-to-face teaching in terms of its sustainability. The results obtained in this study are of great importance for the use of virtual learning environments as a possible standard medium for biology lessons, as the study was conducted with students with different motivation and prior knowledge and under real conditions.

---

# Inhaltsverzeichnis

<b>1 Impulse aus der COVID-19 Pandemie für das digitale Lehren und Lernen</b>	<b>1</b>
<b>2 Theoretischer Hintergrund</b>	<b>4</b>
2.1 Lernumgebungen im Unterricht	4
2.1.1 Modelle der Unterrichtsqualität	4
2.1.2 Methodische und mediale Gestaltungsmöglichkeiten von Unterricht	8
2.1.2.1 Methodische Gestaltungsformen	8
2.1.2.2 Mediale Gestaltungsformen	14
2.1.3 Lerntheorien und ihre Auswirkungen auf hybride und virtuelle Unterrichtsgestaltung	19
2.2 Fachkompetenzen im Biologieunterricht	22
2.2.1 Fachwissen	23
2.2.2 Naturwissenschaftliche Arbeitsweisen	24
<b>3 Fachwissenschaftlicher Hintergrund</b>	<b>28</b>
3.1 PCR (Polymerasekettenreaktion)	28
3.2 Gelelektrophorese	30
3.3 Zystische Fibrose	31
3.4 Fachliche Analyse des Lerngegenstandes	34
<b>4 Ziele und übergeordnete Fragestellung</b>	<b>44</b>
<b>5 Pilotierung</b>	<b>49</b>
5.1 Material und Methoden	49
5.1.1 Stichprobe	49
5.1.2 Erhebungsinstrumente und Skalendokumentation	50
5.1.2.1 Erhebungsinstrumente	50
5.1.2.2 Skalendokumentation	52
5.2 Ergebnisse	59
5.3 Diskussion	69

---

<b>6 Material und Methoden der Hauptstudie</b>	<b>70</b>
6.1 Stichprobe	70
6.2 Studiendesign	70
6.3 Durchführung der Studie	72
6.4 Konzeptionierung der Lernumgebungen	74
6.5 Erhebungsinstrumente	79
6.6 Lernumgebungen	80
6.6.1 Analoge Lernumgebungen	80
6.6.1.1 Klassenunterricht	80
6.6.1.2 Labor	81
6.6.2 Digitale Lernumgebungen	87
6.6.2.1 Interactive Book	88
6.6.2.2 Digitales Labor	92
6.7 Statistische Analysemethoden	97
6.7.1 Das eindimensionale Rasch-Modell	97
6.7.2 Klassische Testtheorie	104
<b>7 Ergebnisse der Hauptstudie</b>	<b>108</b>
7.1 Analyse der Itemschwierigkeiten und Personenfähigkeiten	108
7.1.1 Skala zum Fachwissen Genetik	108
7.1.2 Zwischendiskussion Fachwissensskala	123
7.1.3. Skala zur naturwissenschaftlichen Arbeitsweise	127
7.1.4 Zwischendiskussion NWA-Skala	135
7.2 Fachwissen	139
7.2.1 Vergleich der Gruppen	142
7.2.2 Vergleich der Messzeitpunkte	145
7.3 Naturwissenschaftliche Arbeitsweise	150
7.3.1 Vergleich der Gruppen	153
7.3.2 Vergleich der Messzeitpunkte	155

---

<b>8 Diskussion</b>	<b>160</b>
8.1 Methodische Diskussion	160
8.2 Inhaltliche Diskussion	162
<b>9 Ausblick und Relevanz</b>	<b>171</b>
<b>10 Verzeichnisse</b>	<b>173</b>
10.1 Abbildungsverzeichnis	173
10.2 Tabellenverzeichnis	178
10.3 Literaturverzeichnis	180
<b>11 Anhang</b>	<b>208</b>
11.1 Erhebungsinstrumente	208
11.1.1 Fragebogen Fachwissen	208
11.1.2 Fragebogen naturwissenschaftliche Arbeitsweise	211
11.2 Ergebnisse: Skalendokumentation, Winsteps Datentabellen	218
11.2.1 Prä-Test Fachwissen	218
11.2.2 Post-Test Fachwissen	222
11.2.3 Follow-up-Test Fachwissen	226
11.2.4 Post-Test naturwissenschaftliche Arbeitsweise (NWA)	230
11.2.5 Follow-up-Test naturwissenschaftliche Arbeitsweise (NWA)	234
11.3 Ergebnisse: Fachwissensfragebögen Datentabellen	238
11.3.1 Modellprüfung	238
11.3.2 Vergleich der Gruppen	239
11.3.3 Vergleich der Messzeitpunkte	242
11.4 Ergebnisse: Naturwissenschaftliche Arbeitsweise Datentabellen	246
11.4.1 Modellprüfung	246
11.4.2 Vergleich der Gruppen	247
11.4.3 Vergleich der Messzeitpunkte	249
11.5 Aktivität im wissenschaftlichen Umfeld	253
<b>Danksagung</b>	<b>254</b>



# 1 Impulse aus der COVID-19 Pandemie für das digitale Lehren und Lernen

Im Frühjahr 2020 beschlossen viele Regierungen, aufgrund der COVID-19 Pandemie, Schulen für mehrere Wochen zu schließen. Obwohl Schulschließungen als eine der wirksamsten Maßnahmen zur Eindämmung der Virusausbreitung gelten (Haug et al., 2020), warnten bereits zu diesem Zeitpunkt Pädagog\*innen und Forscher\*innen vor den Konsequenzen, die sich aus den durch die Schulschließungen entstehenden unterschiedlichen Lernbedingungen für die Schüler\*innen ergeben würden (Haeck & Lefebvre, 2020; Wößmann, 2021). Die empirischen Belege für die Auswirkungen von COVID-19 bedingten Schulschließungen auf die schulischen Leistungen, als auch die im alltäglichen Schulbetrieb beobachtbaren Situationen, sind ernüchternd (Hammerstein et al., 2021; Institut für die Qualitätsentwicklung im Bildungswesen (IQB) in Berlin, 2022; McKinsey & Company, 2022). Die Ergebnisse der Untersuchung von Hammerstein et al. (2021) – einer Metaanalyse von 11 Studien, die während der Pandemie im Zeitraum der ersten Schulschließungen durchgeführt wurden – deuten darauf hin, dass ein negativer Effekt von Schulschließungen auf die Schülerleistungen besteht; insbesondere bei jüngeren Schüler\*innen aus Familien mit niedrigem sozioökonomischem Status. Hammerstein et al. (2021) schließen ihre Metaanalyse mit dem Satz:

„This indicates that most remote learning measures implemented during the first school closures in spring 2020 were not effective for student learning; there was no difference between them and the absence of systematic teaching during summer vacation.”  
Hammerstein et al. (2021, S. 6).

Für Schulen und Lehrkräfte ergaben und ergeben sich aus der Pandemie weiterhin besondere Herausforderungen bzgl. der methodischen und medialen Gestaltung von schulischen Lernumgebungen. Dies wird u.a. durch Studien untermauert, in denen Lehrkräfte als Hauptprobleme für das schwache individuelle Lernen der Schüler\*innen zu Hause die mangelnde digitale Kompetenz der Lehrkräfte sowie die verbesserungsbedürftige Digitalisierung der Schulen angeben (Forsa Politik- und Sozialforschung GmbH, 2020). Der Einsatz digitaler Medien in unterschiedlicher Qualität kennzeichnete somit die Zeit der geschlossenen Schulen. In den Ergebnissen der Metaanalyse von Hammerstein et al. (2021) wird ebenso sichtbar, dass nur Studien, in denen eine Art von Online-Lernsoftware zur Leistungsbewertung eingesetzt wurde, einen positiveren Lernzuwachs zeigten (Spitzer & Musslick, 2021; van der Velde et al., 2021). Diese leistungsbewertenden und systematisch

strukturierten Online-Lernmaterialien und -Lernsoftwares standen jedoch nicht für alle Fächer und Schulen in dieser Zeit zur Verfügung (Kennedy et al., 2022).

Neben dem normalen Unterricht in Präsenz konnten die in den MINT-Fächern wichtigen Profilstunden zur Stärkung der grundlegenden Bildungsziele des naturwissenschaftlichen Unterrichts nicht mehr durchgeführt werden. Die Förderung des Verständnisses für naturwissenschaftliche Phänomene, Methoden und Denkweisen sowie die Entwicklung einer naturwissenschaftlichen Grundbildung, die die Schüler\*innen zu verantwortungsbewusstem und reflektiertem Handeln in einer technisch-naturwissenschaftlich geprägten Welt befähigt entfielen in ihrer praktischen Anwendung weitgehend. Dies wäre aber wichtig, um die Begeisterung der Schüler\*innen für naturwissenschaftliche Themen zu wecken und sie zu selbstständigem und kritischem Denken zu motivieren, damit sie in der Lage sind, auf der Grundlage naturwissenschaftlicher Erkenntnisse Entscheidungen im persönlichen und gesellschaftlichen Kontext zu treffen. (Staatsinstitut für Schulqualität und Bildungsforschung [ISB], 2004b, 2023a)

Somit waren neue Vermittlungsformate an den Schulen für die Schüler\*innen dringend erforderlich, um diesen Ausfall aufzufangen. Bereits vor der COVID-19 Pandemie haben sich Bildungseinrichtungen und Unternehmen mit der Erstellung von digitalen Animationen und digitalen Lernumgebungen in diesem Bereich beschäftigt; durch die Pandemie und dem damit verbundenen Druck der Digitalisierung von Unterricht wurde dieser Prozess aber beschleunigt (Meier et al., 2023). Leider sind diese Beschleunigungen eher im Forschungsbereich zu finden. Selbst heute sind die kostenlos verfügbaren digitalen Lernumgebungen, die über einfache Simulationen (z.B. PhET), WebQuests oder kleine interaktive, multimediale Elemente (z.B. LearningApps.org) hinausgehen, für Lehrkräfte recht begrenzt. Virtuelle Labore, die Visualisierungs- und Erforschungsmöglichkeiten bieten, weisen aber eine sehr hohe Lernwirksamkeit auf (Kramer et al., 2019), sind jedoch oft sehr kostenintensiv (wie z.B. Labster) oder bisher hauptsächlich im Forschungs- oder universitären Kontext zugänglich. Daher tendieren Lehrkräfte, auch post-pandemisch, dazu, eigenständig kleinere digitale Lernumgebungen zu erstellen. Es gibt ein breites Fortbildungsangebot für diese Zwecke, das von der Erstellung einfacher Quizze, wie Kahoot!, bis hin zur Programmierung kompletter digitaler Lernumgebungen, etwa mit H5P, reicht.

Im Verlauf der Pandemie wurde es nach der Phase des Homeschoolings wieder möglich, kleinere Schülergruppen in die Schule zurückzubringen, während die andere Hälfte der Klasse daheim weiterhin selbstständig mit den vorgegebenen, oft digitalen Materialien arbeitete. Der Wechselunterricht war mehrere Wochen die Lehr- und Lernform der Wahl. Wechselunterricht oder das Prinzip des Flipped Classrooms gehen auf das Prinzip des Blended Learnings zurück. Blended Learning, eine Fusion aus traditionellem Präsenzunterricht und Online-Lernansätzen (z.B. Learning Management Systeme), welches in den 1990er-Jahren in der

Hochschullehre entstanden ist, hat sich durch seine Renaissance in der Pandemie rasant weiterentwickelt (Zhonggen, 2015) und erreicht einen neuen Stand der Popularität. Learning Management Systeme (LMS), softwarebasierte Plattformen, die zum Verwalten, Durchführen und Bewerten von Lernprozessen und -inhalten in Bildungseinrichtungen und Unternehmen eingesetzt werden (z.B. mebis), erfahren zunehmend Förderung und Anerkennung und obwohl sie bereits in geringerem Maße im täglichen Schulbetrieb Einzug gefunden haben, bieten sie wichtige Möglichkeiten für den zukünftigen Unterricht.

Die Bedeutung und Herausforderung, die mit den digitalen Lernumgebungen einhergeht, hatte die Wissenschaft bereits vor der COVID-19 Pandemie erkannt. Seit Längerem beschäftigt sie sich damit, die Lernwirksamkeit analoger oder digitaler Lernumgebungen zu überprüfen und zu verbessern. In Studien werden oft rein digitale mit rein analogen Lernumgebungen verglichen (Fleissner-Martin et al., 2023; Lester & King, 2009) und kommen zu unterschiedlichsten Ergebnissen. Somit kann festgehalten werden, dass allein die Unterscheidung in „digitale“ oder „analoge“ Lernumgebung zu eng gefasst ist. Wichtig ist, dass die Darbietungsform der Lernumgebung, die Situation und die zu bearbeitenden Inhalte entsprechend gewählt werden (Hammerstein et al., 2021). Die Anzahl der wissenschaftlichen Untersuchungen, die sich mit dieser speziellen Thematik auseinandersetzen, ist jedoch gering. Darüber hinaus gibt es wenig Erkenntnisse darüber, wie sich eine unterschiedliche methodische Gestaltung (hybrid oder virtuell) von Lernumgebungen auf den Kompetenzerwerb, insbesondere in Bezug auf das naturwissenschaftliche Arbeiten, auswirkt.

Die vorliegende Arbeit zielt daher auf die Klärung der Frage, inwieweit unterschiedlich stark digital gestaltete Lernumgebungen das Lernen von naturwissenschaftlichen Arbeitsweisen und den Erwerb von Fachwissen in praxisorientierten Fachbereichen effektiv unterstützen können. Spezifisches Ziel der hier berichteten Untersuchungen war es, mit Blick auf die Fragestellung zu prüfen, inwiefern sich die Variation in der Darbietungsform der Inhalte (in Präsenz oder digital), im Vergleich zum regulären Unterricht in Präsenz (Präsenzunterricht), auf den Lernerfolg und die Behaltensleistung des Fachwissens und der naturwissenschaftlichen Arbeitsweisen der Schüler\*innen gewinnbringend auswirkt.

## 2 Theoretischer Hintergrund

### 2.1 Lernumgebungen im Unterricht

In den Fachwissenschaften gibt es viele Modelle, welche die Qualität von Unterricht diskutieren. Unabhängig von der Herangehensweise haben alle diese Modelle gemeinsam, dass sie einen maximalen Ertrag der Kompetenzentwicklung der Schüler\*innen erzeugen sollen. Die Effektivität des Unterrichts wird daran gemessen, inwiefern er die Fähigkeiten und Fertigkeiten der Lernenden verbessert – kognitive, wie auch psychomotorische und affektive Erträge sind hierbei von Bedeutung (s. *2.1.1 Modelle der Unterrichtsqualität*).

Um diese Kompetenzentwicklung zu ermöglichen, werden verschiedene unterrichtliche Maßnahmen ergriffen. Diese Maßnahmen werden allerdings z.B. durch äußere strukturelle Aspekte, wie einer (partiellen) Schulschließung während der COVID-19 Pandemie, beeinflusst. Die daraus entstehenden methodischen und medialen Gestaltungsmöglichkeiten von Unterricht in Präsenz oder digitalem Unterricht müssen mit Blick auf deren Lernwirksamkeit bzw. Kompetenzförderung gezielt gewählt sein und umgesetzt werden; insbesondere für die Vermittlung komplexer biologischer Sachverhalte wie im Bereich der Biotechnologie. Durch die COVID-19 Pandemie ergaben sich je nach aktueller Pandemielage zwingende methodische Überlegungen für die Unterrichtsorganisation: Präsenzunterricht, Wechselunterricht, Flipped Classroom und Homeschooling (s. *2.1.2.1 Methodische Gestaltungsformen*). Aber auch die medialen Gestaltungsmöglichkeiten dieser Unterrichtsformen und der dabei genutzten Medien mussten differenziert betrachtet werden, um die Kompetenzförderung im Unterricht zu ermöglichen (s. *2.1.2.2 Mediale Gestaltungsformen*). Die Grundlage all dieser methodischen und didaktischen Überlegungen stellt das Lernen auf kognitionspsychologischer Ebene dar. Die evidenzbasierten Theorien des multimedialen Lernens wie R. E. Mayers kognitive Theorie des multimedialen Lernens und Swellers Theorie der kognitiven Belastung (s. *2.1.3 Lerntheorien und ihre Auswirkungen auf hybride und virtuelle Unterrichtsgestaltung*) spielen hierbei eine entscheidende Rolle.

#### 2.1.1 Modelle der Unterrichtsqualität

Ein in der deutschsprachigen Bildungsforschung weitverbreitetes Modell zur Erklärung von Schulerfolg geht auf die Arbeiten von Fend (1981) und Helmke (2003) zurück. Eine vereinfachte Darstellung stammt von Lipowsky (2009) und ist in Abbildung 1 dargestellt. Es fasst für den Schulerfolg bedeutsame, schulische und außerschulische Faktoren auf einem höheren Abstraktionsniveau zu komplexen Variablenclustern zusammen. Dieser Ansatz schafft einen Modellrahmen, der durch spezifische Elemente und theoretisch begründete

Hypothesen weiter spezifiziert werden muss. Das Modell betrachtet den Bildungserfolg als Ergebnis verschiedener Faktoren, die miteinander interagieren. Es umfasst nicht nur den Lern- und Leistungszuwachs, sondern auch die emotionale, motivationale und persönlichkeitsbezogene Entwicklung der Schüler\*innen (s. Abbildung 1) (Lipowsky, 2009).

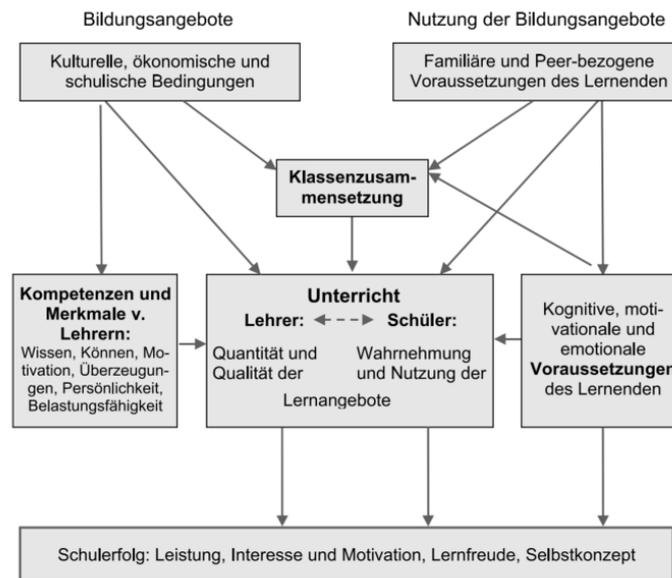


Abbildung 1: Darstellung des vereinfachten Angebots-Nutzungs-Modells  
Übernommen aus Lipowsky (2009, S. 81) © 2009 Springer Nature

Das vereinfachte Modell von Lipowsky (2009) enthält noch die wesentlichen Inhalte des ursprünglichen Modells von Helmke, das von diesem im Laufe der Jahre auch immer wieder angepasst wurde. Die Wirksamkeit eines Unterrichtsangebots und damit die Lernaktivität der Schüler\*innen hängt von Vermittlungsprozessen ab. Bei Lehrpersonen wird die fachwissenschaftliche und fachdidaktische Expertise sowie die Expertise in Klassenführung und Diagnostik als entscheidend angesehen (= unterrichtsrelevante Expertise) (Helmke, 2022). Der Unterricht (das Angebot) steht in Abhängigkeit von der Lehrperson. Dabei ist die Prozessqualität des Unterrichts sowohl fachspezifisch als auch fächerübergreifend von Bedeutung. Auch die Unterrichtsquantität und die damit direkt verbundene tatsächlich nutzbare Unterrichtszeit sowie die aktive Lernzeit der Lernenden sind hier relevant (Helmke, 2022). Hattie (2009) konnte zeigen, dass die Unterrichtszeit und der Leistungszuwachs einen positiven linearen Zusammenhang aufweisen. Aber auch die Qualität der Lehr- und Lernmaterialien spielt eine Rolle (Helmke, 2022). Der Zusammenhang zwischen dem familiären Umfeld der Schüler\*innen und ihrem Lernpotenzial wurde u. a. von Hattie (2009, S. 61–71) untersucht und weist hohe Zusammenhänge auf. Vor allem der Bildungshintergrund, der Anregungsgehalt der häuslichen Lernumgebung, die elterliche Unterstützung, ihre Vorstellungen und Forderungen sind wesentliche Punkte (Hattie, 2009).

Somit bilden die „kognitiven, motivationalen und volitionalen Lernvoraussetzungen“ bei den Schüler\*innen die Grundlage, die über ihren Lernerfolg im Unterricht entscheidet (Helmke, 2022).

Ein wesentlicher Bestandteil des Unterrichts ist der Kontext, der eine Unterrichtsstunde maßgeblich bestimmt. Hierzu gehören die zu unterrichtende Altersstufe, die Schulform, in der unterrichtet wird und auch das Fach, welches als fachwissenschaftliche Grundlage für die vermittelten Inhalte dient (Helmke, 2022). Des Weiteren ist auch der Klassenkontext von großer Bedeutung. Hierzu zählt man unter anderem das Fähigkeits- und Vorkenntnisniveau der Klasse, die Schichtzusammensetzung, die sprachlichen Fähigkeiten, die Klassengröße und das Klassenklima (Helmke, 2022). Diese Kriterien und Ergebnisse der allgemeinen Lehr-Lernforschung werden seit Jahren diskutiert; es wurde jedoch noch kein einfaches und einheitliches Konzept für die Komplexität von Unterricht gefunden.

In der aktuellen didaktischen Forschung treten Überlegungen auf, welche Kriterien guten Biologieunterricht ausmachen und wie diese gruppiert werden können (Neuhaus, 2021). Maßgebend sind die geläufigen drei Basisdimensionen nach Klieme et al. (2001), modifiziert nach Kunter et al. (2011, S. 98): „Klassenführung“, „Potenzial zur kognitiven Aktivierung“ und „konstruktive Unterstützung“. Neben dem Ansatz der Erweiterungen dieses Konzepts, über das Professionswissen von Lehrkräften, wie von Neuhaus (2021) beschrieben und umgesetzt, gibt es den nachfolgend beschriebenen Ansatz, die bestehenden drei Basisdimensionen zu erweitern. Dies wurde von Praetorius et al. (2020) unter Berücksichtigung der Aspekte der allgemeinen Lehr-Lern-Forschung in einem Syntheserahmen mit fachspezifischen Dimensionen und im Sinne der Lehr-Lernforschung ergänzt und gebündelt (Neuhaus, 2021). Das Angebot-Nutzungs-Modell von (Helmke, 2007, 2022) wurde hierbei in die Überlegungen einbezogen. Bei den definierten Dimensionen handelt es sich um „Auswahl und Thematisierung von Inhalten und Fachmethoden, kognitive Aktivierung, Unterstützung des Übens, formatives Assessment, Unterstützung der aktiven Beteiligung, sozio-emotionale Unterstützung und Klassenführung und Differenzierung und Adaptivität“ (Praetorius & Gräsel, 2021, S. 182). Auf Basis des erweiterten Syntheserahmens wurde von Charalambous und Praetorius (2020) das in Abbildung 2 dargestellte MAIN-TEACH-Modell (*multi-layered and integrated in conceptualizing the quality of teaching*) entwickelt, „das eine [...] stärkere Verknüpfung zwischen Unterrichtsdimensionen, Lernprozessen und Outcomes ermöglichen soll“ (Praetorius & Gräsel, 2021, S. 182).

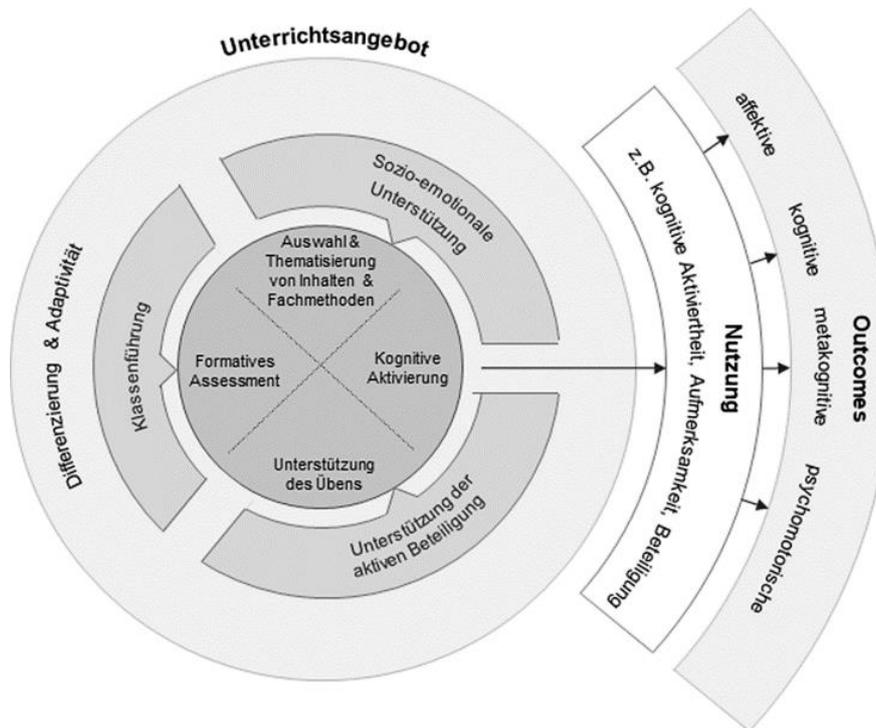


Abbildung 2: Das MAIN-TEACH-Modell (nach Charalambous und Praetorius (2020))  
 Übernommen aus Praetorius und Gräsel (2021), S.182, CC BY 4.0

Dieses Modell des Unterrichtsangebots umfasst drei Schichten, mit jeweils unterschiedlichen Funktionen. Die äußerste Schicht im Modell „Differenzierung und Adaptivität“ stellt die Grundlage dar. Sie verknüpft diese und die anderen Dimensionen. In der mittleren Schicht befinden sich die Dimensionen des Unterrichts. Diese wird gegliedert in die drei Dimensionen „Klassenführung, sozio-emotionale Unterstützung, Unterstützung der aktiven Beteiligung“. Sie stellen eine mögliche unterstützende Funktion auf die innerste Schicht dar. Diese innerste Schicht beinhaltet die Dimensionen, welche die direkten Lernprozesse der Lernenden unterstützen soll. Die vier Dimensionen der innersten Ebene „Auswahl und Thematisierung von Inhalten und Fachmethoden, kognitive Aktivierung, Unterstützung des Übens und formatives Assessment“ (s. Abbildung 2) sind eng durch ihre Bedeutung für den Lernprozess verbunden (Charalambous & Praetorius, 2020).

Das Modell folgt der Idee, dass Lernen im Unterricht durch verschiedene Faktoren beeinflusst wird. Im Kern des Modells steht der direkte Kontakt zwischen dem, was im Unterricht angeboten wird, und dem, was die Schüler\*innen daraus lernen. Die Faktoren der äußeren Schichten beeinflussen das Lernen zudem auf indirektem Wege über die innerste Schicht. Durch das Modell soll verstanden werden, wie das Lernen verbessert werden kann, um bestimmte Outcomes bei den Schüler\*innen zu erreichen. Diese Outcomes sind unterteilt in „kognitive, meta-kognitive, psychomotorische und affektive Outcomes“ (s. Abbildung 2). Die verschiedenen Dimensionen des Modells helfen dabei, den Unterricht aus verschiedenen Blickwinkeln zu betrachten und zu verstehen, welche Aspekte des Unterrichts in einem

bestimmten Lernabschnitt besonders wichtig sind und welche weniger wichtig sein könnten. So kann das Modell dabei helfen, bessere Unterrichtsmethoden zu entwickeln und den Lernprozess für die Schüler\*innen zu verbessern (Praetorius & Gräsel, 2021).

## **2.1.2 Methodische und mediale Gestaltungsmöglichkeiten von Unterricht**

Um das eben beschriebene Unterrichtsangebot und die damit erhofften positiven Outcomes bei den Schüler\*innen zu ermöglichen, wird das Unterrichtsangebot unterschiedlich gestaltet. Die methodischen und medialen Gestaltungsmöglichkeiten eröffnen eine hohe Flexibilität in der Unterrichtsgestaltung, erzeugen aber auch damit verbundene Grenzen in der Umsetzbarkeit. Dabei spielen externe Faktoren eine nicht zu unterschätzende Rolle.

### **2.1.2.1 Methodische Gestaltungsformen**

#### ***Unterrichtsformen mit computergestützten Lernumgebungen***

Um die Entwicklung von Kompetenzen zu fördern, werden im Unterricht verschiedene Maßnahmen eingesetzt. Allerdings können externe strukturelle Bedingungen, wie z.B. die (teilweise) Schließung von Schulen während der COVID-19 Pandemie, diese beeinträchtigen. Die Pandemie machte es notwendig, über verschiedene Unterrichtsmethoden nachzudenken, welche über den normalen Präsenzunterricht an Schulen hinausgingen. Alternierender Unterricht (Wechselunterricht und Flipped Classroom) oder das Lernen zu Hause (Homeschooling) waren von der jeweiligen Situation der Pandemie abhängige Möglichkeiten.

Angesichts dieser neuen Unterrichtsmethoden für den Schulunterricht, die sich aus den Herausforderungen der Pandemie ergeben haben, gewinnt die Digitalisierung des Lernens an Bedeutung. Dabei steht insbesondere eine Organisationsform im Mittelpunkt der Diskussion: das Homeschooling. Der Begriff **Homeschooling** wird nachfolgend verwendet, wie er sich originär aus der COVID-19 Pandemie ergeben hat. H. Meyer (2020) definiert dies wie folgt: „Homeschooling ist ein durch die Schule organisierter Fernunterricht, in dem das gemeinsame Arbeiten in der Klasse/im Lernverband zeitlich befristet aufgehoben und durch individualisierte Hausarbeit ersetzt wird. Sie wird in unterschiedlichem Umfang von den Eltern bzw. Erziehungsberechtigten beaufsichtigt und von der Schule durch die Arbeit mit Bildungsservern und den Einsatz digitaler Medien unterstützt.“<sup>1</sup> Betrachtet man die methodische Gestaltung von Homeschooling, wie es in der COVID-19 Pandemie praktiziert wurde, so ist die Nutzung von Online-Plattformen und -Tools essenziell. Hierbei handelt es sich um

---

<sup>1</sup> Im angloamerikanischen Raum weicht die Definition des Homeschoolings von dieser ab (Murphy, 2014).

Lernmanagementsysteme (LMS), Videokonferenz-Tools und andere interaktive Lernplattformen, die das synchrone und asynchrone Lernen zu ermöglichen (Hodges et al., 2020). Von den Schüler\*innen, ihren Eltern und auch den Lehrkräften muss die Anpassungsfähigkeit und Flexibilität im Umgang mit diesen LMS geleistet werden (Daniel, 2020). Synchrones Lernen, Live-Online-Unterricht über Plattformen wie Zoom, Microsoft Teams oder Visavid, bei denen Lehrkräfte und Schüler\*innen in Echtzeit miteinander interagieren können, stellen einen sehr strukturierten und klassisch schulischen Teil des Homeschoolings dar. Das asynchrone Lernen, die Bereitstellung von aufgezeichneten Lerneinheiten, Lernmaterialien, Videos oder Leseaufgaben, die die Schüler\*innen in ihrem eigenen Tempo und zur eigens gewählten Zeit durchgehen können, führt zu einem der wesentlichsten Punkte in der methodischen Gestaltung des Homeschoolings. Die Selbstregulierung des Arbeitens durch die Schüler\*innen fordert diese in einem neuen Umfang (Daniel, 2020; Zimmerman, 2002). Um die Schüler\*innen im Homeschooling zu unterstützen, sind einige didaktische Maßnahmen möglich (H. Meyer, 2020). In Bezug auf das selbstregulierende Lernen sollen mit den Schüler\*innen Strategien erlernt werden, ihre eigenen Lernwege zu erkennen und zu verfolgen (Zimmerman, 2002). Schon das Bereitstellen einer klaren Tages- und Wochenstruktur kann den Schüler\*innen helfen, den Lernalltag besser zu organisieren (Hodges et al., 2020). Auch ist der Einsatz von interaktiven Tools und Methoden sinnvoll, wenn sie gewissenhaft gewählt werden und zu den Bedürfnissen der Schüler\*innen passen, um Schüler\*innen aktiv zu beteiligen und das Lernen ansprechender zu gestalten (Deterding et al., 2011; H. Meyer, 2020). In diesem Zusammenhang ist es einfach möglich, regelmäßiges und konstruktives Feedback in automatischer oder manueller Form an die Schüler\*innen zu geben (H. Meyer, 2020; Narciss, 2018) und Lernmaterialien und Lernaktivitäten an die individuellen Bedürfnisse und Fähigkeiten der Schüler\*innen anzupassen (Letzel & Pozas, 2022). Nicht zuletzt sollte auch noch in die Planung des Unterrichts die Förderung der sozialen Interaktion integriert werden, indem Möglichkeiten für Schüler\*innen geschaffen werden, sich mit ihren Mitschüler\*innen auszutauschen, um soziale Isolation zu vermindern (H. Meyer, 2020; L. Meyer & Stein, 2022).

Eine weitere Organisationsform, die im Verlauf der COVID-19 Pandemie in aller Munde war, ist der **Wechselunterricht**. Beim Wechselunterricht wechseln sich Präsenz- mit Fernunterrichtsphasen ab, wenn z.B. nur eine begrenzte Anzahl von Schüler\*innen gleichzeitig in der Schule sein darf. In solchen Fällen kann eine Gruppe von Schüler\*innen an bestimmten Tagen in der Schule sein, während die andere Gruppe von zu Hause aus lernt. An anderen Tagen wird dann gewechselt. Der Wechsel kann täglich, wöchentlich oder nach einem anderen Zeitplan erfolgen. (Rogge & Seifert, 2023) Betrachtet man die methodischen Gestaltungskriterien für den Wechselunterricht, so ist hier vor allem eine klare Struktur und Organisation wichtig, damit die Schüler\*innen wissen, wann sie im Präsenzunterricht und

wann im Fernunterricht sind und dies auf klaren Wegen kommuniziert wird (z.B. E-Mail, Lernmanagement-Systeme, ...) (Linder, 2017).

In Bezug auf die didaktischen Maßnahmen des Wechselunterrichts ist es wichtig, dass die Unterrichtsinhalte so didaktisch angepasst werden, dass sie dem jeweiligen Format (Präsenz- oder Fernunterricht) gerecht werden. Nicht jede Aktivität, die im Klassenzimmer funktioniert, eignet sich für den Fernunterricht und umgekehrt. Hierbei gelten für den Fernunterricht die didaktischen Überlegungen, welche auch im Homeschooling (s.o.) genannt wurden, als auch die didaktischen Maßnahmen des Präsenzunterrichts. Dabei sind die **10 Merkmale guten Unterrichts** von H. Meyer (2004) konkrete Anhaltspunkte für die Gestaltung eines effektiven und erfolgreichen Unterrichts. Diese Merkmale sollen nicht als starrer Kanon verstanden werden, sondern vielmehr als Orientierungshilfe für die Unterrichtspraxis dienen (Helmke, 2007). Nachfolgend wird nur kurz auf die Merkmale eingegangen. Eine genauere Beschreibung ist bei H. Meyer (2004) oder Helmke (2022) nachzulesen. Das erste Merkmal ist die **Klare Strukturierung des Unterrichts**. Damit ist gemeint, dass der Ablauf des Unterrichts für die Schüler\*innen nachvollziehbar sein sollte. Ein roter Faden, der sich durch die Stunde zieht, erleichtert das Lernen und Verstehen. Des Weiteren betont ein **Hoher Anteil echter Lernzeit**, die Notwendigkeit, dass Schüler\*innen aktiv und ohne Unterbrechung arbeiten können. Vermeidbare Störungen oder lange Organisationsphasen sollten reduziert werden, um eine maximale Lernzeit zu gewährleisten. Mit dem dritten Punkt, **Lernförderliches Klima**, hebt Meyer hervor, wie wichtig eine positive und angstfreie Atmosphäre im Klassenzimmer ist. Ein solches Klima fördert nicht nur die Motivation, sondern auch die aktive Beteiligung und das Selbstvertrauen der Schüler\*innen. Stetiges Feedback spielt hierbei eine Rolle, was durch Hattie (2009) ebenso bestätigt wurde. Ein erfolgreicher Unterricht zeichnet sich durch **Inhaltliche Klarheit** durch verständliche Aufgaben, logischen Themenverlauf und klare Ergebnissicherung aus. Dies manifestiert sich in informativen Einstiegen, konsequentem Fokus auf das Thema, sorgfältiger Nutzung von Unterrichtsmedien, Einbeziehung von Schülervorkenntnissen, konstruktivem Umgang mit Fehlern und regelmäßigen Zusammenfassungen durch die Schüler\*innen. **Sinnstiftende Kommunikation** beinhaltet sowohl die Art und Weise, wie die Lehrkraft den Stoff vermittelt, als auch die Kommunikation zwischen den Schüler\*innen und der Lehrkraft (u.a. Feedback) sowie unter den Schüler\*innen selbst. Das sechste Merkmal, **Methodenvielfalt**, betont die Bedeutung verschiedener Lehr- und Lernmethoden. Die Variation von Methoden hält den Unterricht abwechslungsreich und spricht unterschiedliche Lernstile an. Als siebtes Merkmal formuliert Meyer **Individuelles Fördern**. Das bedeutet, dass auf die individuellen Bedürfnisse und Voraussetzungen der Schüler\*innen eingegangen wird. Eine Differenzierung des Unterrichts kann hierbei hilfreich sein. **Intelligentes Üben** als achttes Merkmal hebt hervor, dass nicht nur die Quantität, sondern vor allem die Qualität des Übens entscheidend ist. Übungen sollten

gezielt ausgewählt und sinnvoll in den Unterricht integriert werden. Das neunte Merkmal, **Transparente Leistungserwartungen**, welches auch von Hattie (2009) bestätigt wird, bedeutet, dass Schüler\*innen genau wissen sollten, welche Leistungen von ihnen erwartet werden und wie sie diese erbringen können. Zuletzt betont Meyer die Vorbereitung auf die Leistungsfeststellung als zehntes Merkmal. Das bedeutet, dass Prüfungen oder Tests nicht überraschend kommen sollten, sondern dass Schüler darauf vorbereitet werden und wissen, wie sie sich am besten darauf vorbereiten können. Zuletzt wird die **Vorbereitete Umgebung** genannt. Eine effektive Lernumgebung wird durch Klassen- und Fachräume geschaffen, die durch ihre gut durchdachte Ordnung, funktionale Einrichtung und bereitgestelltes Lernwerkzeug charakterisiert sind. Es ist essenziell, dass sowohl Lehrer\*innen als auch Schüler\*innen die Möglichkeit haben, sich den Raum zu eigen zu machen und eine wirkungsvolle Raumgestaltung umzusetzen, um erfolgreich zu arbeiten und zu lernen.

Eine weitere Organisationsform und zugleich ein didaktisches Konzept ist der **Flipped Classroom**. Hierbei wird der traditionelle Unterricht "umgedreht" (J. Bishop & Verleger, 2013). Das bedeutet, dass die Schüler\*innen sich die Lehrinhalte zuerst zu Hause (meist über Videos oder andere Online-Ressourcen) aneignen. Die eigentliche "Hausaufgabe" (also die Anwendung des Gelernten, Diskussionen, Projekte etc.) findet dann im Klassenraum statt, wo die Lehrkraft unterstützend und beratend zur Seite steht. Dieses Modell legt den Fokus darauf, die Präsenzzeit für aktives Lernen und individuelle Unterstützung zu nutzen, während die Informationsvermittlung außerhalb des Klassenraums stattfindet. (Bergmann & Sams, 2012; Lutz et al., 2023) Ein Ziel des Flipped Classroom besteht somit darin, den Präsenzunterricht interaktiver und anwendungsorientierter zu gestalten und Raum für Diskussionen und Fragen zu geben (Al-Samarraie et al., 2020). Didaktische Maßnahmen, die zum Gelingen des Flipped Classrooms nötig sind, sind ähnlich denen des Wechselunterrichts. Hervorzuheben ist die Bedeutung, die dem vorbereitenden digitalen Material zukommt. Lehrkräfte stellen Videos, Lesematerial oder interaktive Module bereit, mit denen sich die Schüler\*innen zu Hause auf den kommenden Unterricht vorbereiten können. Diese Materialien sind oft selbst produziert oder sorgfältig ausgewählt, um den Kerninhalt zu vermitteln. (Bergmann & Sams, 2012) Die aktiven Lernzeiten in der Klasse werden für vertiefende Diskussionen, praktische Übungen, Gruppenarbeit und individuelle Unterstützung genutzt; Schüler\*innen können direkt Fragen stellen und Feedback zu ihrem Verständnis und ihren Arbeitsweisen erhalten (Bergmann & Sams, 2012). In den vorbereitenden Einheiten zu Hause können die Schüler\*innen die vorbereitenden Materialien in ihrem eigenen Tempo bearbeiten und bei Bedarf wiederholen. Die Lehrkräfte haben die Möglichkeit, auf die individuellen Bedürfnisse der Schüler\*innen während des Unterrichts in Präsenz einzugehen. (Bergmann & Sams, 2012; Lutz et al., 2023) Zudem wird kollaboratives Lernen während der Klassenunterrichtszeit gefördert, in dem Gruppenprojekte, Peer-to-Peer-Lernen, Think-Pair-Share usw. als Unterrichtsmethoden

eingesetzt werden (Bishop & Verleger, 2013). Die Schüler\*innen lernen voneinander durch Diskussionen und gemeinsame Problemlösung (Lutz et al., 2023). Zudem können Lehrkräfte kontinuierlicheres Feedback geben, da sie mehr Zeit haben, sich individuell mit den Schüler\*innen zu beschäftigen. Der Erfolg des Flipped Classrooms hängt von der sorgfältigen Planung und Vorbereitung durch die Lehrkraft, der Verfügbarkeit von Ressourcen und der Bereitschaft der Schüler\*innen ab, sich auf diese aktive Lernweise einzulassen. Die Integration von Flipped Classroom-Konzepten kann zu einer tieferen Verarbeitung von Lerninhalten und zu einer stärkeren Beteiligung der Schüler\*innen führen (Bergmann & Sams, 2012; Bishop & Verleger, 2013).

Das Konzept des Flipped Classrooms geht auf das didaktische Konzept des **Blended Learning** zurück, welches eine Lehr- und Lernmethode ist, bei welcher der traditionelle Präsenzunterricht mit Online-Lernen kombiniert wird. Blended Learning hat seinen Ursprung in der Hochschullehre. Es ist eine Mischung aus face-to-face-Interaktionen und digitalen Lernaktivitäten. Die zeitliche Verteilung kann je nach Bedarf variieren, und Online-Elemente können vor, während oder nach dem Präsenzunterricht eingesetzt werden. Ein Ziel von Blended Learning besteht darin, die Vorteile von Präsenzunterricht und Online-Lernen zu nutzen, um die Effektivität des Lernens zu steigern, die Flexibilität zu erhöhen und auf verschiedene Lernstile einzugehen. (Bokolo et al., 2022)

### ***Wirksamkeit hybrider und virtueller Lernumgebungen***

Die vorgestellten Unterrichtsformen integrieren teilweise digitale Ressourcen und Methoden in unterschiedlichem Ausmaß. Nachfolgend wird darauf eingegangen, wie diese digitalen und analogen Lernumgebungen mit den beschriebenen Unterrichtskonzepten und der Lernwirksamkeit in Verbindung stehen und diese beeinflussen.

Die Wirksamkeit digitaler Lernumgebungen im Vergleich zum traditionellen Präsenzunterricht wurde in den vergangenen Jahren in zahlreichen Studien und Forschungsarbeiten überprüft, die auf verschiedene Weisen die positiven Auswirkungen des digitalen Lernens auf die Bildungsergebnisse von Lernenden betonen. In den Studien von Gunbas (2015) (Mathematik), Chang et al. (2008) (Physik), Buckley et al. (2004) (Biologie) oder Frailich et al. (2009) (Chemie) zeigte sich, dass die Verwendung digitaler Werkzeuge das Lernen und Lehren in den MINT Fächern verbessern kann. Die folgenden Gründe wurden in den verschiedenen Studien identifiziert. Digitale Lernplattformen können das Lernen individualisieren, indem sie den Fortschritt jedes Lernenden verfolgen und personalisierte Lernpfade erstellen. Dies ermöglicht es, auf die spezifischen Bedürfnisse und Fähigkeiten jedes Schülers und jeder Schülerin einzugehen. Somit wird den Lernenden ermöglicht, die Lernmaterialien eigenständig und in ihrem individuellen Tempo wiederholt zu bearbeiten

(Lodge & Horvath, 2017). Dies wird auch durch die vielfältige Auswahl interaktiver Elemente unterstützt, darunter Videos, Simulationen, Quizfragen und Spiele, die das Interesse und die Motivation der Lernenden aktivieren können. Interaktive Inhalte fördern das aktive Lernen (Mayrose, 2012). Digitale Plattformen ermöglichen es Lernenden, online miteinander zu kommunizieren und zusammenzuarbeiten, unabhängig von ihrer geografischen Lage (Kollaboration). Dies fördert den Austausch von Ideen und die Zusammenarbeit, was zu einem tieferen Verständnis führen kann (Benbunan-Fich & Hiltz, 1999). Sofortiges Feedback, z.B. durch automatische Antworten auf Quizfragen ist in diesen möglich, um Fehler in schriftlichen Aufgaben zu markieren oder diese zu bewerten. Stetiges Feedback in einer digitalen Lernumgebung fördert das Verständnis der Schüler\*innen für Konzepte und Prozesse (Thisgaard & Makransky, 2017). Dies ermöglicht den Lernenden, schnell aus ihren Fehlern zu lernen (M. Brown, 2013; X. Chen & DeBoer, 2015). Der Einsatz spielerischer Elemente in digitale Lernumgebungen, wie Belohnungen, Abzeichen und Ranglisten (Gamification), kann die Motivation der Schüler\*innen steigern und sie ermutigen, sich intensiver mit dem Lernmaterial auseinanderzusetzen (Ab Rahman et al., 2019; Rahman et al., 2018). Digitale Lernumgebungen ermöglichen es den Lernenden, jederzeit und von jedem Ort aus auf Inhalte zuzugreifen. Dies fördert die Flexibilität und ermöglicht es den Lernenden, ihren eigenen Zeitplan zu erstellen (Gilhespy, 2011). Ein längerfristiger Vorteil ist, dass Daten über den Lernfortschritt und das Verhalten der Lernenden gesammelt werden können und so eine Lernanalyse ermöglicht wird. Lehrer\*innen und Bildungseinrichtungen können diese Daten nutzen, um Einblicke in den Fortschritt der Schüler\*innen zu gewinnen und ihre Lehrmethoden anzupassen (Seiler et al., 2019). Hillmayr et al. (2020) zeigten in ihrer Metaanalyse im MINT-Bereich, dass sich der Einsatz digitaler Werkzeuge positiv auf die Lernergebnisse der Schüler\*innen und ihre Einstellung zum Unterrichtsfach auswirkt.

Andere Studien hingegen zeigten, dass konventionelle Klassenzimmer effektiver sind als digitale Lernumgebungen. Das physische Berühren von Büchern, Schreibutensilien und anderen analogen Materialien kann dazu beitragen, Informationen besser zu verinnerlichen. Das haptische Erlebnis kann das Gedächtnis und das Verständnis fördern (Tanhua-Piironen et al., 2010). Analoge Lernumgebungen sind in der Regel weniger anfällig für Ablenkungen durch Benachrichtigungen, soziale Medien oder andere digitale Störungen. Dies kann dazu beitragen, die Konzentration und das Lernen zu verbessern (Kostić & Randelović, 2022; Pérez-Juárez et al., 2023). Auch fördern sie oft die direkte verbale Kommunikation zwischen Lehrer\*innen und Schüler\*innen oder zwischen Lernenden untereinander. Dies kann zu tieferen Diskussionen, schnellerer Hilfe und einem besseren Verständnis führen (Mottarella et al., 2005; A. Y. Wang & Newlin, 2000). Klassenzimmer oder Gruppenarbeitsräume fördern zudem ein Gefühl der Gemeinschaft und des Zusammenhalts unter den Lernenden. Dies kann die Motivation und das soziale Lernen unterstützen (Attié et al., 2022; Baepler et al., 2014). In

einigen Fällen kann langsames, vertieftes Lernen in analoger Form effektiver sein als das schnelle Durchklicken von digitalen Inhalten. Analoge Materialien erfordern oft mehr Zeit und Aufmerksamkeit (Mueller & Oppenheimer, 2014).

Wieder andere Studien (Fleissner-Martin et al., 2023; Kemp & Grieve, 2014; Waschull, 2001) zeigen, dass die Unterschiede bezogen auf den Lernzuwachs der Lernenden zwischen den beiden Lehrstrategien unbedeutend sind. Sowohl die digitalen als auch die analogen Lernmodule verbesserten den Wissensstand in den jeweiligen Studien.

Die Kombination aus analogen und digitalen Lernumgebungen zeigt in Studien, dass Studierende, die das entdeckende Lernen mit einer Kombination aus Präsenz und digitalem Unterricht anwenden, bessere Ergebnisse als im konventionellen Entdeckungsunterricht erzielten (Brata et al., 2021). Auch die Beiträge u. a. von Hafeez und Akhter (2021), Castro-Rodríguez et al. (2021) und Ireland et al. (2009) stellen die Effektivität von Blended Learning bezogen auf den akademischen Erfolg und die kritischen und kreativen Fähigkeiten heraus.

Somit ergibt sich über viele Studien hinweg kein einheitliches Bild bzgl. der „besseren“ Lernwirksamkeit analoger oder digitaler Lernumgebungen. Jede Form und Kombination weist gewisse Vor- und Nachteile auf.

Darüber hinaus darf nicht außer Acht gelassen werden, dass neben der Wahl der Präsentationsform auch die mediale Aufbereitung der Inhalte einen maßgeblichen Einfluss auf die Lerneffizienz und den Lernerfolg hat.

### **2.1.2.2 Mediale Gestaltungsformen**

#### ***Mediale Gestaltung von Lernumgebungen***

R. E. Mayer (2017) hat zwölf forschungsbasierte Grundsätze für die Gestaltung computergestützter multimedialer Lehrmaterialien zur Förderung des akademischen Lernens entwickelt, die darauf abzielen, das Lernen in digitalen Umgebungen zu optimieren. Diese Prinzipien sind wichtige Leitlinien für die Gestaltung von multimedialen Inhalten und digitalen Lernmaterialien (Clark & Mayer, 2016). Die ursprünglichen **zwölf Prinzipien zum multimedialen Lernen** mit den jeweiligen zentralen Aussagen werden nachfolgend kurz beschrieben. Diese Prinzipien wurden in verschiedenen Studien von Mayer und seinen Kolleg\*innen, aber auch teilweise von anderen Wissenschaftler\*innen bestätigt. (R. E. Mayer, 2017; R. E. Mayer & Fiorella, 2022a)

Das Multimedia-Prinzip besagt, dass Menschen besser mit Worten und Bildern in computergestützten Lehrmaterialien lernen, als mit Worten allein. Die Kombination von Text mit relevanten Bildern, Animationen oder Diagrammen kann das Verständnis und die Merkfähigkeit verbessern.

Zu den Prinzipien, die darauf abzielen, irrelevante Verarbeitungsprozesse – sogenannte extrinsische kognitive Belastungen, die nicht mit dem Lernziel in Zusammenhang stehen – zu reduzieren, gehören Kohärenz, Signalisierung, Redundanz, räumliche Kontiguität und zeitliche Kontiguität. Fremdmaterial wie dekorative Grafiken oder nicht zusammenhängende Inhalte sollten auf ein Minimum reduziert werden, da sie von der Hauptbotschaft ablenken und das Lernen behindern können (**Kohärenzprinzip**) (Fiorella & Mayer, 2022c). Wichtige Informationen sollten hervorgehoben oder signalisiert werden, um die Aufmerksamkeit der Lernenden auf wichtige Konzepte oder Ideen zu lenken (**Signalisierungsprinzip**) (Fiorella & Mayer, 2022c; van Gog, 2022). Dieselben Informationen gleichzeitig in visueller und auditiver Form zu präsentieren, soll vermieden werden, da dies die kognitive Verarbeitung der Lernenden überfordern kann (**Redundanzprinzip**) (Fiorella & Mayer, 2022c; Kalyuga & Sweller, 2022). Zusammenhängende Grafiken und Texte sollten auf dem Bildschirm oder der Seite nahe beieinander präsentiert werden, um den kognitiven Aufwand für die Integration der Informationen zu verringern (**Prinzip der räumlichen Kontiguität**). Informationen, die gleichzeitig in visueller und auditiver Form präsentiert werden, sollten synchronisiert werden, um den Lernenden zu helfen, die Informationen effektiver zu verknüpfen (**Prinzip der zeitlichen Kontiguität**) (Fiorella & Mayer, 2022c). Ergänzt wird dies durch das **Prinzip des bearbeiteten Beispiels**. Menschen lernen besser, wenn sie beim Erlernen kognitiver Fähigkeiten ausgearbeitete Beispiele sehen (Renkl, 2022).

Zu den Prinzipien, die die wesentliche Verarbeitung – d.h. die mentale Repräsentation des wesentlichen Materials – steuern, gehören Segmentierung, Vorwissen und Modalität. Komplexes Material sollte in kleinere, überschaubare Segmente unterteilt werden, um eine kognitive Überlastung zu vermeiden und schrittweises Lernen zu unterstützen (**Prinzip der Segmentierung**) (R. E. Mayer & Fiorella, 2022b). Wenn die Lernenden einführende Informationen zum Thema erhalten, bevor sie sich mit komplexeren Inhalten beschäftigen, kann dies die Lernergebnisse verbessern (**Prinzip des Vorwissens**) (R. E. Mayer & Fiorella, 2022b). Bei komplexem Material (z. B. Animation oder Video) ist ein multimediales Format anstelle von reinem Text zu wählen, z.B. Ton, um das Verständnis zu erleichtern (**Modalitätsprinzip**) (Paas & Sweller, 2022).

Prinzipien zur Förderung generativer Aktivität beim Multimedia-Lernen (d.h. der kognitiven Verarbeitung, die darauf abzielt, dem Material einen Sinn zu geben), die auf sozialen und affektiven Merkmalen beruhen, gehören Personalisierung, Stimme, Bild und das Verkörperungsprinzip (R. E. Mayer, 2017). Der Einsatz eines unterhaltsamen Umgangstons statt einer formellen oder akademischen Sprache (**Personalisierungsprinzip**), sowie eines freundlichen, unterhaltsamen Tons mit menschenähnlicher Stimme (**Prinzip der Stimme**) kann Engagement und das Verständnis der Lernenden fördern. Es wird hingegen nicht effektiver gelernt, wenn das starre Gesicht des virtuellen Instruktor in einer multimedialen

Präsentation zu sehen ist (**Bildprinzip**). Menschen lernen besser, wenn der virtuelle Instruktor auf dem Bildschirm menschenähnliche Verhaltensweisen zeigt, wie z. B. Gestik, Blickkontakt oder Mimik (**Verkörperungsprinzip**). (Fiorella & Mayer, 2022b) Diese Prinzipien wurden um das Immersionsprinzip, das Kollaborationsprinzip, das Animationsprinzip und das emotionale Gestaltungsprinzip ergänzt (R. E. Mayer & Fiorella, 2022a, Part VI). Menschen lernen nicht unbedingt besser mit Medien mit höherer Immersion (z. B. immersive virtuelle Realität) als mit Medien mit geringerer Immersion (z. B. Video auf dem Bildschirm) (**Immersionsprinzip**) (Makransky, 2022), aber sie lernen besser in der Gruppe als allein (**Kollaborationsprinzip**) (Jeroen Janssen et al., 2022). Auch fördern dynamische Grafiken das Lernen mehr, als statische Grafiken (**Animationsprinzip**) (Lowe et al., 2022). Auch sind multimediale Einheiten lernwirksamer, die Elemente mit Gesichtsausdruck und warmen Farben enthalten (**Emotionales Gestaltungsprinzip**) (Plass & Hovey, 2022).

Zusätzlich zu den oben beschriebenen Prinzipien wurden einige weitere Prinzipien zusammengetragen, die zu generativer Aktivität beim multimedialen Lernen führen (Fiorella & Mayer, 2022a). Dabei handelt es sich um Gestaltungsprinzipien, die darauf abzielen, den Lernenden zu generativen Lernaktivitäten anzuregen, d.h. zu Verhaltensweisen, die der Lernende während des Lernens mit der Absicht ausführt, das Lernen zu verbessern (R. E. Mayer & Fiorella, 2022a, Part VII). Es zeigt sich, dass Menschen besser lernen, wenn sie während des Lernens aufgefordert werden, generative Lernstrategien anzuwenden (z.B. Zusammenfassen, Abbilden, Zeichnen, Vorstellen, Selbsterklären, Selbsttesten, Erklären für andere und Umsetzen) (**Prinzip der generativen Aktivität**) (Fiorella & Mayer, 2022a), als auch wenn sie ermutigt werden, während des Lernens Konzeptkarten, Wissenskarten oder grafische Organisatoren zu erstellen (**Mapping-Prinzip**) (Adesope et al., 2022). Zudem wird das Lernen unterstützt, wenn die Lernenden beim Lesen eines erklärenden Textes dazu aufgefordert werden, Zeichnungen anzufertigen (**Zeichnungsprinzip**) (Leutner & Schmeck, 2022) und sich Zeichnungen vorzustellen, während sie einen erklärenden Text lesen (**Vorstellungsprinzip**) (Leopold, 2022). Weitere Prinzipien, die das Lernen verbessern, sind das **Selbsterklärungsprinzip**, was besagt, dass Menschen besser lernen, wenn sie ermutigt werden, während des Lernens Selbsterklärungen zu erstellen (Chi, 2022) und das **Prinzip der angeleiteten Entdeckung**, welches beinhaltet, dass es lernwirksam ist, wenn eine Anleitung in entdeckungsbasierte Multimedia-Umgebungen integriert ist (de Jong, 2022). Menschen lernen besser aus multimedialen Lektionen, wenn sie erklärendes Feedback zu ihrer Leistung erhalten (**Feedback-Prinzip**) (Johnson & Marraffino, 2022) und wenn sie aufgefordert werden, Kontrolle über ihre Lernprozesse auszuüben (**Selbstmanagement-Prinzip**) (S. Zhang et al., 2022). Jedoch gilt dies nicht, wenn sie mehr Kontrolle über die Auswahl und Organisation des Materials haben (**Prinzip der Lernerkontrolle**) (Scheiter, 2022).

### ***Lernen durch praktisches Arbeiten in Präsenz und virtuellen Laboren***

Die im Biologieunterricht angewendeten wissenschaftlichen Methoden dienen nicht nur dem besseren Verständnis der Inhalte, sondern können auch als eigenständiges Unterrichtsthema oder primäres Unterrichtsziel behandelt werden (J. Mayer, 2002): zum Beispiel durch den Einsatz von Schülerversuchen im Unterricht. Die Schüler\*innen lernen durch die Beschäftigung mit verschiedenen Methoden und Techniken, wie beispielsweise genaues Beobachten, das Prüfen von Hypothesen und sorgfältiges Arbeiten funktioniert. Diese Fähigkeiten befähigen sie, selbstständig neue Probleme zu lösen (J. Mayer, 2002). Diese fachspezifischen Arbeitsmethoden, zusammen mit praktischen Laborerfahrungen, tragen wesentlich zur Vertiefung und Erweiterung des Verständnisses der Schüler\*innen für das wissenschaftliche Fachwissen bei. Es bietet den Lernenden die Möglichkeit, über die theoretischen Konzepte hinaus ein tieferes Verständnis für die Biologie zu entwickeln (*America's Lab Report*, 2006; le Plat et al., 2018). Auch fördert die praktische kritische Auseinandersetzung mit den genutzten Methoden, ein aufgeklärteres Verhältnis zu wissenschaftlichen Erkenntnissen (J. Mayer, 2002; Otsetov, 2020). Das wissenschaftliche Arbeiten, im Schülerversuch oder Labor, ermöglicht zudem einen selbstständigen Wissenserwerb, der sich positiv auf die Lernmotivation und Interessenentwicklung der Schüler\*innen auswirkt (*America's Lab Report*, 2006; J. Mayer, 2002; Meier & Mayer, 2014). Zudem unterstützen reale Experimente das Verständnis der Erkenntnistheorie und der Prozesse wissenschaftlichen Arbeitens (de Jong et al., 2013). Schließlich fördert die Arbeit in Kleingruppen während des wissenschaftlichen Arbeitens soziale Kompetenzen wie Teamfähigkeit und Kommunikation (J. Mayer, 2002). Auch die Interaktion zwischen Lehrer\*in und Schüler\*innen, um den Lernenden die Ziele und die Bedeutung der Laborarbeit zu vermitteln, ist hierbei von Bedeutung (Högström et al., 2010). Trotz dieser bedeutenden Aspekte für das Lernen durch praktisches Arbeiten ist es nicht immer möglich, dies umzusetzen. Häufig wird hierbei durch die Lehrkräfte mangelnde Zeit und fehlende Ausstattung an Schulen angegeben. Auch die Erreichbarkeit von Schülerlaboren und außerschulischen Lernorten ist nicht überall in einem angemessenen Zeitraum möglich. (Breuer et al., 2020)

Durch die Corona-Pandemie gefördert entstehen immer mehr digitale Angebote durch außerschulische Lernorte (Meier et al., 2023). Das Lernen im digitalen Labor ist somit ein bedeutender Ansatz in der modernen Bildung. Digitale Labore ermöglichen es, praktische Erfahrungen und experimentelles Lernen in einem virtuellen Umfeld durchzuführen. Sie bieten Lernmöglichkeiten, die über geografische Grenzen hinweg zugänglich sind. Lernende können auf virtuelle Labore zugreifen, ohne physisch anwesend zu sein. Dies ermöglicht eine erweiterte Flexibilität und die Anpassung an individuelle Zeitpläne und Bedürfnisse. Zudem werden die Bildungschancen für Menschen in abgelegenen Regionen verbessert. Des

Weiteren reduzieren digitale Labore das Risiko, das mit physischen Experimenten verbunden ist. Schüler\*innen können gefährliche Experimente ohne physische Gefahren durchführen (The European Business Review, 2021). Digitale Labore ermöglichen die individuelle Anpassung von Experimenten. Lernende können Experimente wiederholen und modifizieren, um ein besseres Verständnis zu erlangen. Dies fördert ein tieferes Verständnis und ermöglicht das Lernen im eigenen Tempo (Govender, 2023). Schülerzentrierte Ansätze, bei denen die Schüler\*innen selbstständig explorative naturwissenschaftliche Untersuchungen durchführen, sind effektiv, um naturwissenschaftliches Denken und Wissenserwerb zu fördern (de Jong et al., 2023). Überdies kann stetiges Feedback in einer digitalen Lernumgebung das Verständnis der Schüler\*innen für Konzepte und Prozesse fördern (Thisgaard & Makransky, 2017) und für den Erwerb praktischer Fähigkeiten förderlich sein, wenn es auf den Aktionen und Entscheidungen der Schüler\*innen basiert und somit Missverständnisse klären und den Lernprozess unterstützt (van Joolingen & de Jong, 1997). Digitale Labore ermöglichen eine einfache Integration von multimedialen Inhalten und erleichtern die Datenerfassung und Datenanalyse von Experimenten und Versuchen. Studierende können Videos, Simulationen und Echtzeitdaten verwenden, um komplexere Konzepte zu verstehen (Berkaoui et al., 2023). Virtuelle Labore, die eine visualisierende und explorative Funktion haben, sind bei Lernenden sehr beliebt (Kramer et al., 2019). So können virtuelle Experimente einen Mehrwert gegenüber realen Experimenten bieten, indem sie den Schüler\*innen ermöglichen, unbeobachtbare Phänomene zu erforschen und beobachtbare und unbeobachtbare Phänomene miteinander zu verbinden (de Jong et al., 2013) Zudem können gut gestaltete computerbasierte Simulationen den Wissenserwerb in den Naturwissenschaften unterstützen. Sie bieten den Schüler\*innen die Möglichkeit, mit virtuellen Experimenten zu interagieren und so ihre Verständnis von wissenschaftlichen Konzepten zu vertiefen (de Jong, 2006; de Jong et al., 2013). Eine der schwierigsten Herausforderungen digitaler Labore ist der Mangel an praktischer Erfahrung. Trotz der fortschrittlichen Simulationen und virtuellen Experimente können sie physische Labore nicht vollständig ersetzen (The European Business Review, 2021). Reale Laborerfahrungen sind somit wichtig, um die Fähigkeiten und das Verständnis der Schüler\*innen zu vertiefen.

Das Lernen im digitalen Labor bietet eine Vielzahl von Vorteilen, wie Zugänglichkeit, Sicherheit und Individualisierung. Dennoch sollten die Herausforderungen, wie der Mangel an praktischer Erfahrung und technische Hürden, nicht vernachlässigt werden.

### 2.1.3 Lerntheorien und ihre Auswirkungen auf hybride und virtuelle Unterrichtsgestaltung

Um effektive Lernprozesse sowohl im digitalen als auch im Präsenzlernen zu verstehen und zu gewährleisten, ist es wichtig, sich mit theoretischen Modellen und Rahmenbedingungen vertraut zu machen, die die zugrunde liegenden Prinzipien und Methoden veranschaulichen. Diese Modelle dienen als Leitfaden für das Verständnis komplexer Zusammenhänge und erleichtern das systematische Erlernen neuer Fähigkeiten und Kenntnisse.

Das **Mehrspeichermodell des Gedächtnisses**, entwickelt von Atkinson und Shiffrin (1968), beschreibt das menschliche Gedächtnis als einen Prozess, der in drei Hauptkomponenten unterteilt ist. In der ersten Stufe des Gedächtnisses, dem sensorischen Gedächtnis, werden Informationen aus der Umgebung vorübergehend gespeichert. Diese Informationen bleiben nur für sehr kurze Zeit erhalten, unter einer Sekunde, und werden in den meisten Fällen nicht bewusst wahrgenommen. Nur relevante Informationen werden an die zweite Stufe, das Kurzzeitgedächtnis, weitergegeben. Dieses wird auch als Arbeitsgedächtnis bezeichnet. Hier werden die relevanten Informationen für eine begrenzte Zeitspanne von wenigen Sekunden aktiv gehalten und können bewusst verarbeitet werden. Das Kurzzeitgedächtnis hat eine begrenzte Kapazität und erfordert Wiederholung oder bewusste Aufmerksamkeit, um Informationen längerfristig zu speichern. Der dritte Speicher ist das Langzeitgedächtnis. Informationen, die im Kurzzeitgedächtnis aufrechterhalten und durch Wiederholung oder tiefe Verarbeitung verstärkt werden, können ins Langzeitgedächtnis überführt und dort über einen sehr langen Zeitraum gespeichert werden. (Strobach & Wendt, 2019) Verglichen mit dem Langzeitgedächtnis hat das Arbeitsgedächtnis nur eine begrenzte Kapazität zur Speicherung und Verarbeitung von Informationen. Aus diesem Grund kann die kognitive Belastung eines Lernenden mit der kognitiven Aktivität des Arbeitsgedächtnisses verglichen werden (Baddeley, 1992). Das Mehrspeichermodell des Gedächtnisses wurde in einigen seiner Elemente revidiert, dient aber weiterhin als didaktische Struktur, um Gedächtniskomponenten zu strukturieren (Strobach & Wendt, 2019).

Die **Cognitive Load Theory** von Chandler und Sweller (1991; Paas & Sweller, 2022) ist eine Theorie aus dem Bereich der kognitiven Psychologie, die sich mit der Art und Weise befasst, wie Informationen in das Arbeitsgedächtnis eines Menschen gelangen und verarbeitet werden. Die Theorie wurde entwickelt, um das Lernen und die damit verbundene kognitive Belastung zu erklären. Die Grundidee der Cognitive Load Theory ist, dass das Arbeitsgedächtnis eine begrenzte Kapazität hat, um Informationen gleichzeitig zu verarbeiten. Wenn zu viele Informationen oder komplexe Informationen auf einmal präsentiert werden, kann dies zu einer Überlastung des Arbeitsgedächtnisses führen, was das Lernen und die Problemlösung erschweren. Sweller et al. (2019, S. 263) unterscheidet zwischen drei Arten von kognitiver Belastung „intrinsic“ (i), „extraneous“ (ii) und „germane“ (iii). Die intrinsische kognitive

Belastung (i) ist definiert durch die inhärente Schwierigkeit und Komplexität des Lernmaterials. Diese wird sowohl durch die Komplexität der Informationen als auch durch das Wissen der Person, die diese Informationen verarbeitet, bestimmt. Die extrinsische kognitive Belastung (ii) ergibt sich aus der Art und Weise, mit welcher die Informationen präsentiert werden. Klarheit in Gestaltung und Darbietung des Lernmaterials und der Lerninstruktion verringern die extrinsische kognitive Belastung. Bei der lernbezogenen Belastung (iii) handelt es sich um eine Umverteilungsfunktion und keine eigene Belastung. Dabei werden Arbeitsgedächtnisressourcen von irrelevanten Aktivitäten auf direkt lernrelevante Aktivitäten umverteilt, indem sie für Informationen genutzt werden, die für die Lernaufgabe relevant sind. Diese Belastung ist positiv, da sie mit dem eigentlichen Lernen und der Verarbeitung von Informationen in Verbindung steht. Es handelt sich um die geistige Anstrengung, die für das Verstehen und Einprägen von Konzepten notwendig ist (Sweller et al., 2019). Die Cognitive Load Theory schlägt vor, das effektive Lernen durch die Reduzierung der extrinsischen Belastung zu verbessern. Dies ist z. B. durch die Verwendung von Informationen in der grafischen Darstellung anstatt daneben (*the split attention effect*) und der Vermeidung von Redundanz in Diagramm und Text (*the redundancy effect*) möglich. Zudem sollte die Optimierung der intrinsischen Belastung z. B. durch Nutzung von ausgearbeiteten (*the worked example effect*) und teilweise ausgearbeiteten Arbeitsbeispielen (*the completion effect*), sowie durch die Isolierung interagierender Elemente (*the isolated interacting elements effect*) erreicht werden (Chong, 2005; Paas & Sweller, 2022; Sweller, 2016; Sweller et al., 2019; Zumbach, 2010). Schnotz und Bannert (2003) weisen durch ihr „Modell des Text- und Bildverstehens“ auch darauf hin, dass die Präsentation von Grafiken nicht immer vorteilhaft für den Wissenserwerb ist. Während für Aufgaben angemessene Grafiken das Lernen unterstützen können, können für Aufgaben nicht angemessene Grafiken den Aufbau des mentalen Modells beeinträchtigen. Somit kann durch klare und gut strukturierte Präsentationen und die schrittweise Einführung von Informationen versucht werden, das Arbeitsgedächtnis von Lernenden nicht zu überlasten.

R. E. Mayer (2022) erweitert mit seiner **Cognitive Theory of Multimedia Learning** die Ideen der Cognitive Load Theory und erklärt die Reduzierung der extrinsischen kognitiven Belastung in der Gestaltung von multimedialen Lernmaterialien. Multimediale Lerninhalte sollten so konzipiert sein, dass sie während des Lernprozesses eine adäquate kognitive Verarbeitung fördern, ohne die kognitiven Kapazitäten des Lernenden zu überlasten (R. E. Mayer, 2014). Die kognitive Theorie des multimedialen Lernens beruht auf drei grundlegenden kognitionswissenschaftlichen Prinzipien: 1) das menschliche Informationsverarbeitungssystem verfügt über zwei Kanäle für die visuelle/bildliche und die auditive/verbale Verarbeitung – bekannt als das Prinzip der Zweikanaligkeit; 2) jeder dieser Kanäle hat eine begrenzte Verarbeitungskapazität, was als das Prinzip der begrenzten Kapazität bezeichnet wird; 3) aktives Lernen beinhaltet die Durchführung einer koordinierten Abfolge kognitiver Prozesse

während des Lernens, womit das Prinzip der aktiven Verarbeitung beschrieben wird. (R. E. Mayer, 2022; R. E. Mayer & Fiorella, 2022a)

Neben den Theorien zur Vermeidung kognitiver Überlastungen beim Lernen gibt es weitere Modelle, welche versuchen, die Lernwirksamkeit digitaler Anwendungen im Unterricht zu beschreiben. Hierzu zählt das **ICAP-Modell** von Chi und Wylie (2014), welches aufzeigt, wie die Aktivitäten von Lernenden je nach genutzter Intensitätsstufe unterschiedlich starke kognitive Beteiligung und damit verbundene Lerneffektivität fördern (Chi et al., 2018). Es dient dazu, Lernumgebungen so zu gestalten, dass Schüler\*innen über die passive Aufnahme von Informationen hinausgehen und durch aktive Beteiligung und kollaborative Interaktion ein tieferes Verständnis der Lerninhalte entwickeln. Hierbei unterscheidet man vier Stufen der kognitiven Beteiligung während der Schüleraktivitäten: passiv, aktiv, konstruktiv, interaktiv. Dieser Rahmen entstand aus einer umfassenden Analyse von Lernaktivitäten und geht davon aus, dass die kognitiven Prozesse und die daraus resultierenden Ergebnisse eng mit der Art des Engagements der Schüler\*innen verbunden sind. (Chi, 2009; Chi & Wylie, 2014) Im passiven Modus sind die Schüler\*innen lediglich Empfänger von Informationen, ein Szenario, das häufig in traditionellen Vorlesungen zu beobachten ist. Der aktive Modus zeichnet sich dadurch aus, dass die Schüler\*innen die gegebenen Informationen bearbeiten, aber nicht unbedingt neue Inhalte erstellen, z. B. durch das Hervorheben von Text oder Notizen. Eine Verlagerung hin zu einer tieferen kognitiven Verarbeitung ist im konstruktiven Modus zu beobachten, in dem die Lernenden auf der Grundlage ihres Verständnisses neue Interpretationen oder Verbindungen herstellen, die im Quellenmaterial nicht ausdrücklich vorgesehen sind, z. B. durch eigenständiges Anfertigen einer Ergebniszusammenfassung. Der Höhepunkt, der interaktive Modus, zeigt die Zusammenarbeit der Lernenden, in der Regel in Form von Diskussionen in der Gruppe, die eine synergetische Konstruktion und Verfeinerung des Wissens ermöglichen. (Chi, 2009; Chi & Menekse, 2015; Chi & Wylie, 2014) Empirische Belege bestätigen die Behauptung, dass beim Übergang von passiver zu interaktiver Bearbeitung eines Lerngegenstandes durch die Schüler\*innen die Tiefe des Lernens vergrößert wird, was Pädagog\*innen einen Einblick in die Entwicklung effektiver pädagogischer Strategien ermöglicht (Sánchez et al., 2022; Wiggins et al., 2017).

## 2.2 Fachkompetenzen im Biologieunterricht

Scientific Literacy, also die wissenschaftliche Grundbildung, stellt einen unerlässlichen Baustein in der heutigen Bildungslandschaft dar. Sie befähigt Individuen, naturwissenschaftliche Konzepte und Prozesse zu verstehen, um informierte Entscheidungen zu treffen und in einer technologisch fortgeschrittenen Gesellschaft kompetent agieren zu können (Gräber & Nentwig, 2002). Die Scientific Literacy unterteilt sich nach Bybee (2002) in vier verschiedene Stufen. Aufsteigend handelt es sich hierbei um die nominale, die funktionale, die konzeptionelle sowie prozedurale und die multidimensionale Scientific Literacy.

Nerdel (2017) fasst die Inhalte der vier Stufen nach Bybee (2002) wie folgt zusammen. Die nominale Scientific Literacy markiert eine Basisebene im Verständnis naturwissenschaftlicher Konzepte und Prozesse. Auf dieser Ebene sind Individuen zwar in der Lage, bestimmte Begriffe und Fragestellungen als naturwissenschaftlich zu erkennen, jedoch zeigen sie eine falsche oder lückenhafte Kenntnis und Verständnis dieser. Dies manifestiert sich in irrtümlichen Auffassungen von naturwissenschaftlichen Konzepten und Prozessen, mangelhaften oder unpassenden Erklärungen naturwissenschaftlicher Phänomene sowie einer simplen Interpretation naturwissenschaftlicher Zusammenhänge. Die nominale Scientific Literacy zeigt eine wesentliche Begrenztheit in der Tiefe und Breite des Verständnisses der Naturwissenschaften, was häufig zu Missverständnissen und falschen Deutungen führt. Die funktionale Scientific Literacy bildet die Grundlage für das Verständnis der Naturwissenschaften. Sie umfasst die korrekte Nutzung von naturwissenschaftlichem Vokabular und die genaue Definition naturwissenschaftlicher Begriffe. Auch die Kenntnis technischer Begriffe ist Teil dieses Bereichs. Durch die funktionale Scientific Literacy werden die Basiskenntnisse bereitgestellt, um naturwissenschaftliche Informationen zu deuten und zu kommunizieren. Die konzeptionelle und prozedurale Scientific Literacy erweitert das Verständnis für naturwissenschaftliche Konzepte und Prozesse. In diesem Bereich geht es um das Verstehen dieser Konzepte und das Erlangen naturwissenschaftlicher Fertigkeiten. Zudem ist es wichtig, die Verbindungen zwischen den verschiedenen Bereichen einer naturwissenschaftlichen Disziplin und ihrer konzeptionellen Gesamtstruktur zu begreifen. Das Verständnis der grundlegenden Prinzipien und Prozesse einer Naturwissenschaft ist auch ein zentraler Aspekt der konzeptionellen und prozeduralen Scientific Literacy. Durch die Verschmelzung von konzeptionellem und prozeduralem Wissen sind Individuen in der Lage, komplexe naturwissenschaftliche Zusammenhänge zu erfassen, zu analysieren und in verschiedenen Kontexten anzuwenden. Die multidimensionale Scientific Literacy repräsentiert hingegen eine weit entwickelte Ebene des Verständnisses und der Wertschätzung der Naturwissenschaften. Individuen mit multidimensionaler Scientific Literacy verstehen die Einzigartigkeit der Naturwissenschaften und können klar zwischen den Naturwissenschaften

und anderen wissenschaftlichen Disziplinen unterscheiden. Sie sind auch mit der Geschichte und der Natur der Naturwissenschaften vertraut, was ein tieferes Verständnis der Entwicklung und der Grundlagen dieser Disziplinen ermöglicht. Ferner verstehen sie die Naturwissenschaften in ihrem gesellschaftlichen Kontext und erkennen die Auswirkungen und Interaktionen zwischen wissenschaftlichen Entdeckungen und gesellschaftlichen Entwicklungen. Die multidimensionale Scientific Literacy ermöglicht eine fundierte Auseinandersetzung mit naturwissenschaftlichen Themen und fördert die Fähigkeit, komplexe Zusammenhänge zu analysieren und zu interpretieren, sowohl in der wissenschaftlichen Gemeinschaft als auch in der breiteren Gesellschaft (Bybee, 2002; Nerdel, 2017).

Ähnlich zu den Angaben der Scientific Literacy fordert auch die zweite Ebene des bayerischen Lehrplans (ISB, 2004a, 2023a), welche Informationen zu den Fachprofilen enthält, die Vermittlung und Förderung der empirischen Erkenntnisgewinnung, des Abstraktionsvermögens und des logischen Denkens sowie die Bedeutung des Selbstverständnisses des Menschen und der Natur. Dadurch wird das Selbstverständnis des Faches als auch sein Beitrag zur gymnasialen Bildung und Persönlichkeitsentwicklung erläutert (ISB, 2004a). Hierbei wird auf der Profilebene, genauso wie in den oberen Stufen des Modells der Scientific Literacy ein hohes Maß an Fach-, Methoden-, Sozial- & Selbstkompetenz in den Naturwissenschaften gefordert (Nerdel, 2017).

Im Biologieunterricht spielen neben den oben aufgeführten Kompetenzen die Sach- und Erkenntnisgewinnungskompetenz eine entscheidende Rolle, um Schüler\*innen ein tiefgehendes Verständnis für die biologischen Prozesse und Phänomene zu vermitteln. Die Bildungsstandards für die allgemeine Hochschulreife im Fach Biologie in Bayern, die im Jahr 2020 von der Kultusministerkonferenz veröffentlicht wurden legen den Rahmen fest, nach dem der Biologieunterricht gestaltet wird. Diese Standards bilden die Grundlage für den Erwerb von Fachkompetenzen und setzen hohe Anforderungen an die Schüler\*innen.

### **2.2.1 Fachwissen**

Bereichsspezifisches Fachwissen ist von entscheidender Bedeutung, wenn es um die Erreichung der Bildungsstandards für die allgemeine Hochschulreife im naturwissenschaftlichen Bereich geht. Die Relevanz des bereichsspezifischen Fachwissens kann in mehrfacher Hinsicht erklärt werden (KMK, 2020). Ohne ein fundiertes Verständnis und anwendbares Wissen der jeweiligen Fachgebiete, sei es in Physik, Chemie, Biologie oder Mathematik, ist es unmöglich, die geforderten Kompetenzen zu erlangen. Dieses Wissen bildet die Grundlage, damit die Schüler\*innen komplexe Probleme analysieren und lösen können. Des Weiteren ermöglicht bereichsspezifisches Fachwissen eine kritische Reflexion und Anwendung von wissenschaftlichen Konzepten und Methoden. Dies ist von großer Bedeutung,

da die Bildungsstandards nicht nur das Wissen, sondern auch die Fähigkeit zur kritischen Beurteilung und Anwendung desselben betonen. Ohne ein solides Fundament in den Naturwissenschaften können Schüler\*innen diese Anforderungen nicht erfüllen. Schließlich ist das spezifische Fachwissen in den Naturwissenschaften von großer Relevanz für die Vorbereitung auf ein Hochschulstudium. Die meisten Studiengänge im naturwissenschaftlichen Bereich setzen ein gewisses Vorwissen voraus. Insgesamt ist bereichsspezifisches Fachwissen also von essenzieller Bedeutung, um eine solide Grundlage für weiterführende Bildungs- und Berufswege zu schaffen. Bereits McNamara (1991) sowie Calderhead und Miller (1986) erörtern die Bedeutung des Fachwissens der Schüler\*innen und dessen Anwendung im Unterricht und betonen hierbei die Notwendigkeit, dass sich Lehrerausbilder\*innen mit Fragen wie dem Umfang des Fachwissens und dessen Zusammenschluss mit dem Verständnis des Kindes befassen. Auch Wilson (1989) unterstreicht die zentrale Bedeutung des Fachwissens für den Unterricht und die Wechselwirkung zwischen inhaltlichem Wissen und Pädagogik. K. Bishop und Denley (1997) betont die grundlegende Rolle des Fachwissens im naturwissenschaftlichen Unterricht und schlägt Strategien vor, um dessen Stellenwert in der Lehrerausbildung zu erhöhen. Insgesamt wird in diesen Beiträgen der Wert von Fachwissen für Schüler\*innen hervorgehoben und betont, dass Pädagog\*innen Fachwissen in ihrer Unterrichtspraxis berücksichtigen und verbessern sollen. Für das Lernen der Schüler\*innen ist es von Bedeutung, dass das Fachwissen kumulativ im Verlauf des Unterrichts erarbeitet wird und somit kein träges Wissen darstellt, welches unter sich verändernden Bedingungen keine Anwendung mehr findet (Harms, 2018). Zusammenfassend lässt sich die Bedeutung des Fachwissens im Biologieunterricht damit begründen, dass das Fachwissen die inhaltliche Dimension eines Lerngegenstandes darstellt, die Erkenntnisgewinnung, die Kommunikation und die Bewertung hingegen Handlungsdimensionen sind (Nerdel, 2017).

### **2.2.2 Naturwissenschaftliche Arbeitsweisen**

Naturwissenschaftliche Phänomene und Prozesse beschreiben und erklären zu können, ist ein Ziel, welches Schüler\*innen in der Schule erlernen sollen. Hierzu ist es nötig, die naturwissenschaftlichen Arbeitsweisen für den Erkenntnisgewinn zu verstehen und sinnvoll zu nutzen, d.h. für die Problemlösung der Fragestellung heranzuziehen. Somit sind die naturwissenschaftlichen Arbeitsweisen grundlegend für das Verständnis der Naturwissenschaften, da sie das Verständnis wichtiger fachlicher Aspekte sicherstellen (Nerdel, 2017). Des Weiteren sollen durch sie fachspezifische methodische und prozessbezogene Fähigkeiten und Fertigkeiten vermittelt werden (Millar, 2006).

Die Kompetenz der Schüler\*innen in der naturwissenschaftlichen Erkenntnisgewinnung zeigt sich im Verständnis für wissenschaftliche Denk- und Arbeitsmethoden. Dies wird deutlich daran, wie sicher sie diese Methoden darlegen, erläutern und miteinander in Beziehung setzen können. Dies befähigt sie, wissenschaftliche Erkenntnisprozesse entweder nachzuvollziehen oder eigenständig zu entwickeln sowie deren Potenzial und Begrenzungen kritisch zu hinterfragen (ISB, 2004a).

Das wissenschaftliche Arbeiten in der Biologie folgt dem hypothetisch-deduktiven Modell, das auf die systematische Erforschung biologischer Phänomene ausgerichtet ist. Ausgehend von einem beobachteten Phänomen beginnt der Prozess mit der Formulierung präziser Fragestellungen. Darauf aufbauend werden Hypothesen entwickelt, die als vorläufige Antworten auf die gestellten Fragen dienen. Diese Hypothesen sind spezifisch und testbar formuliert, sodass sie durch gezielte Untersuchungen überprüft werden können. Der nächste Schritt umfasst die Planung und Durchführung von Untersuchungen. Diese sollen so konzipiert sein, dass sie die Möglichkeit bieten, die Hypothesen zu stützen oder zu widerlegen. Nach der Datensammlung folgt die Auswertung, bei der die Ergebnisse analysiert und interpretiert werden. Diese Phase ist entscheidend, um zu ermitteln, inwieweit die erhobenen Daten die Hypothesen unterstützen. Abschließend erfolgt eine methodische Reflexion, die kritisch betrachtet, ob und wie die Untersuchungen zur Widerlegung oder Unterstützung der Hypothesen beigetragen haben und inwiefern die ursprünglichen Fragestellungen beantwortet werden konnten. Dieser Prozess fördert nicht nur ein tiefes Verständnis für biologische Konzepte, sondern schult auch die Fähigkeit, wissenschaftliche Methoden kritisch zu reflektieren und anzuwenden. (ISB, 2023a; Wellnitz & Mayer, 2016)

Der naturwissenschaftliche Erkenntnisweg, wie er zuvor skizziert wurde (Wellnitz & Mayer, 2016), wird nicht in jedem Schülerversuch in allen Schritten durchlaufen. Die Komplexität und die Anforderungen an die Planung der Durchführung von Experimenten überschreitet häufig das schulische Vorwissen der Schüler\*innen zu einem Sachverhalt. Gerade in der Genetik und Gentechnik ist der Umfang, den die Schüler\*innen aus den fachlichen Inhalten des Lehrplans lernen, viel zu gering, um damit selbstständig einen Versuch planen zu können. Die Planung einer praktischen Untersuchung von DNA mittels PCR und anschließender Agarose-Gelelektrophorese mit dem damit verbundenen Primer-Design und weiteren Aspekten ist hierbei nicht leistbar. In solchen Fällen ist eine gezielte Unterstützung unerlässlich, damit die Lernenden erfolgreich die Brücke von der Theorie zur Praxis schlagen können. Hierbei helfen didaktische Hilfestellungen und angeleitetes Lernen, um die Schüler\*innen durch den anspruchsvollen Prozess der Versuchsplanung zu führen. Diese Unterstützung ermöglicht es den Schüler\*innen, sich kompetent mit der Erhebung und Analyse von Daten zu befassen und aus den Ergebnissen fundierte Schlussfolgerungen zu ziehen. In dieser Arbeit liegt somit der Schwerpunkt auf den Schritten der Datenerhebung und der Interpretation der Ergebnisse.

Diese Phasen sind essenziell, um aus der Theorie greifbare Schlüsse ziehen zu können. (Kirschner et al., 2006) Deshalb ist es entscheidend, den Schüler\*innen Methoden an die Hand zu geben, die es ihnen ermöglichen, über das theoretische Verständnis hinaus praktische Erfahrungen zu sammeln. Dies ebnet den Weg für ein tieferes Verständnis und die Entwicklung von Beobachtungs- und Vergleichsfähigkeiten.

Die Erkenntnismethoden des Beobachtens und Vergleichens ist besonders in der Genetik und Gentechnik für Schüler\*innen von Bedeutung, da hier Veränderungen oft nicht unmittelbar sichtbar sind. Folgende Ansätze können helfen, die Beobachtungskompetenz und die Erkenntnismethode des Vergleichens in diesen Bereichen lernwirksam zu fördern. Schüler\*innen können durch die Anwendung von Mikroskopie z.B. Zellteilung oder die Struktur von DNA und Chromosomen beobachten (Killermann et al., 2020). Methoden wie die Polymerase-Kettenreaktion (PCR) und Agarose-Gelelektrophorese, die in der Schule durchgeführt oder demonstriert werden können, ermöglichen die Beobachtung von genetischen Markern und DNA-Fragmenten. Hierbei werden die Schüler\*innen aufgefordert, auf subtile Unterschiede in den Bandenmustern im Gel zu achten, was eine genaue und sorgfältige Beobachtung erfordert. Sie beobachten den gesamten Vorgang der Agarose-Gelelektrophorese, von der Vorbereitung der Proben über das Einbringen in die Taschen des Gels bis hin zur eigentlichen Trennung durch elektrischen Strom. Dies fördert ein tieferes Verständnis für experimentelle Abläufe. Durch die Interpretation der Ergebnisse, also das Auswerten der Banden im Gel nach der Färbung, entwickeln die Schüler\*innen analytische Fähigkeiten, die für das Verständnis von genetischen Informationen notwendig sind. Die Lernenden können vorhersagen, wie sich Moleküle unterschiedlicher Größe im Gel verhalten, und ihre Hypothesen mit den tatsächlichen Ergebnissen vergleichen. Sie üben sich darin, die Qualität der Gelelektrophorese kritisch zu beurteilen, beispielsweise die Gleichmäßigkeit der Banden und die Auflösung der Trennung. Zudem werden die Schüler\*innen dazu angehalten, mögliche Fehlerquellen, die die Ergebnisse beeinflussen könnten, zu identifizieren und zu diskutieren, was zur Problemlösungskompetenz beiträgt. (Schöppner et al., 2022) Weitere praktische Erfahrungen mit gentechnischen Verfahren, wie Restriktionsenzymverdau oder DNA-Transformation, ermöglichen es Schüler\*innen, die Ergebnisse gentechnischer Eingriffe direkt zu vergleichen (Carson et al., 2020). Durch das Vergleichen von Genotypen mit den resultierenden Phänotypen können Schüler\*innen lernen, wie genetische Information die Merkmale eines Organismus beeinflusst (Schwanewedel, 2006). Auch computergestützte Simulationen, Software und virtuelle Labore, die genetische Prozesse simulieren, ermöglichen Beobachtungen, die im realen Labor nicht möglich oder ethisch bedenklich wären (Killermann et al., 2020). Die Nutzung von Bioinformatik-Tools zur Sequenzanalyse ermöglicht es Schüler\*innen, genetische Sequenzen von verschiedenen Organismen zu vergleichen (Martins et al., 2020). Durch die Kombination dieser verschiedenen Ansätze können Lehrkräfte

den Schüler\*innen helfen, die komplexen Prozesse in der Genetik und Gentechnik nicht nur zu beobachten, sondern auch zu verstehen, zu interpretieren, zu formulieren und zu reflektieren (ISB, 2023a). Vor allem die immer stärker werdenden digitalen Möglichkeiten zur Verwendung von erklärenden Modellen und Simulationen unterstützen eine weitere Erkenntnismethode, das Modellieren (Wellnitz et al., 2017). Moderne Unterrichtstechnologie kann genutzt werden, um virtuelle Modelle zu erstellen oder zu simulieren. Programme wie *Molecular Workbench* oder *PhET Interactive Simulations* bieten interaktive Möglichkeiten, komplexe biologische Systeme zu modellieren. Simulationen oder virtuelle Experimentierräume stellen eine Ersatzmöglichkeit für reale Experimente dar und können zur Erforschung von Kausalzusammenhängen genutzt werden (de Jong et al., 2013).

Die fach- und sachgerechte Durchführung der Erkenntnismethoden verlangt danach, dass Arbeitstechniken beherrscht werden. Die sachgemäße Handhabung wissenschaftlicher Geräte wie Mikroskope, Präparierbesteck und Mikroliterpipetten stellt einen wesentlichen Aspekt wissenschaftlicher Methodik dar. Diese Arbeitstechniken erfordern primär manuelle Geschicklichkeit und sind essenziell für das Durchführen naturwissenschaftlicher Untersuchungen wie Beobachtungen, Vergleichen und Experimentieren. Diese wiederum bilden das Fundament für ein hypothetisch-deduktives Herangehen, das zur Lösung wissenschaftlicher Fragestellungen notwendig ist. (ISB, 2023a; Wellnitz & Mayer, 2013)

Die naturwissenschaftlichen Arbeitsweisen erhalten auch in den Bildungsstandards im Fach Biologie für die allgemeine Hochschulreife (KMK, 2020) einen hohen Stellenwert. In den Bereichen der Sach-, Kommunikations- und Bewertungskompetenz sind sie grundlegend enthalten. Aber vor allem im Bereich der Erkenntnisgewinnungskompetenz wird ihre große Bedeutung ersichtlich. Es wird gefordert, dass die Schüler\*innen „Fragestellungen und Hypothesen auf Basis von Beobachtungen und Theorien entwickeln“, „Fachspezifische Modelle und Verfahren charakterisieren, auswählen und zur Untersuchung von Sachverhalten nutzen“, „Erkenntnisprozesse und Ergebnisse interpretieren und reflektieren“ und „Merkmale wissenschaftlicher Aussagen und Methoden charakterisieren und reflektieren“ (KMK, 2020, S. 7). In den Bildungsstandards im Fach Biologie für den mittleren Schulabschluss (KMK, 2005) des Bereichs Erkenntnisgewinnung wird neben der Anwendung experimenteller Schritte auch deutlich der Bildungsanspruch auf das Verständnis und die Handhabung weiterer charakteristischer Methoden wie Beobachtung und Vergleich hervorgehoben (KMK, 2005; Wellnitz & Mayer, 2013).

## 3 Fachwissenschaftlicher Hintergrund

### 3.1 PCR (Polymerasekettenreaktion)

Die Polymerasekettenreaktion (PCR, *polymerase chain reaction*), die 1983 von Kary Mullis entdeckt wurde, ist eine Technologie von großer Bedeutung. Sie stellt ein grundlegendes Werkzeug in vielen wissenschaftlichen Forschungsfeldern dar (Domingues, 2017), da sie es ermöglicht, DNA-Fragmente milliardenfach zu vervielfältigen. Dadurch ist es möglich, diese zu charakterisieren und zu manipulieren. Gängige Einsatzgebiete reichen von der Diagnose genetischer Erkrankungen und Krankheitserregern bis hin zur Identifizierung des Besitzers eines an einem Tatort gefundenen Haars (Berg et al., 2018).

Die Reaktionsabläufe der natürlichen Replikation dienen hierbei als theoretische Grundlage der Vervielfältigungsprozesse von Nucleinsäuren mittels PCR. In der PCR synthetisiert eine DNA-Polymerase einen neuen DNA-Strang vom 3'-Ende aus, an der einzelsträngigen Template-DNA, wodurch ein Doppelstrang entsteht. Die DNA-Polymerase bindet hierzu zunächst an Primer, welche synthetische DNA-Oligonukleotide sind, die an die einzelsträngig vorliegenden Template-DNA hybridisieren. Durch Nutzung eines gegenläufig orientierten Oligonukleotid-Primerpaares wird die DNA-Sequenz im Bereich zwischen den beiden Primern vervielfältigt (K. Mullis et al., 1986; H. Schmidt & Rothhämel, 2012).

Der Reaktionsansatz einer PCR enthält neben der Template-DNA und den beiden Oligonukleotid-Primern noch alle vier Desoxynukleotide (dNTPs = Desoxynukleotid-triphosphate), eine hitzestabile DNA-Polymerase in einem Puffer, der  $Mg^{2+}$ -Ionen als Kofaktor enthält (Kück, 2005). Die thermostabile DNA-Polymerase (z. B. *Taq*-Polymerase) bleibt auch bei hohen Temperaturen stabil, welche für die Trennung der doppelsträngig vorliegenden DNA in einzelsträngige Template-DNA notwendig ist (T. A. Brown, 2011; H. Schmidt & Rothhämel, 2012). Neben der *Taq*-Polymerase sind weitere DNA-Polymerasen aus thermophilen Bakterien isoliert, ggf. modifiziert und rekombinant exprimiert worden, die Vorteile gegenüber der *Taq*-Polymerase haben. Einige weisen eine relativ starke 3' → 5' Exonucleaseaktivität auf. Diese Exonucleaseaktivität bezeichnet man auch als „Korrekturlese-Aktivität“, was zur Nutzung dieser Polymerasen für eine akkurate und fehlerfreie Amplifikation führt (Müller & Prange, 2016).

Die PCR basiert auf der zyklischen Wiederholung von drei Temperaturschritten (Melting, Annealing und Elongation), was dazu führt, dass die Template-DNA exponentiell vervielfältigt wird, solange ausreichend dNTPs, Primer etc. vorliegen. Ein PCR-Zyklus beginnt mit dem Schritt *Melting*, was eine thermische Denaturierung des DNA-Doppelstranges bewirkt. Bei 90–95 °C wird die DNA aufgeschmolzen, indem die Wasserstoffbrücken gespalten werden, was zur Bildung von einzelsträngiger Template-DNA führt. Nachfolgend kommt es zum *Annealing*, der

Primer-Hybridisierung an die einzelsträngige DNA durch Basenpaarung. Hierbei binden die synthetischen Oligonukleotide bei Temperaturen von  $\geq 50\text{ }^{\circ}\text{C}$  an die komplementären Sequenzen der Template-DNA (H. Schmidt & Rothhämel, 2012). Für ein angemessenes Ergebnis ist es wichtig, dass die Schmelztemperaturen der Primer sehr ähnlich sind, um eine korrekte Bindung beider bei einer ähnlichen Hybridisierungstemperatur zu ermöglichen. Die Spezifität der Primer ist zudem einer der wichtigsten Faktoren für eine gute PCR, da die Primer nur an der Zielsequenz hybridisieren sollten (Domingues, 2017). Als dritter Schritt erfolgt die *Elongation*, die Amplifikation des zwischen den Primern liegenden Sequenzabschnittes bei  $72\text{ }^{\circ}\text{C}$  durch das bei dieser Temperatur aktive Enzym *Taq*-Polymerase in  $5' - 3'$  Richtung. Durch die hitzestabile DNA-Polymerase können die PCR-Zyklen kontinuierlich durchgeführt werden, ohne erneute Enzymzugabe bei jedem Zyklus. (H. Schmidt & Rothhämel, 2012)

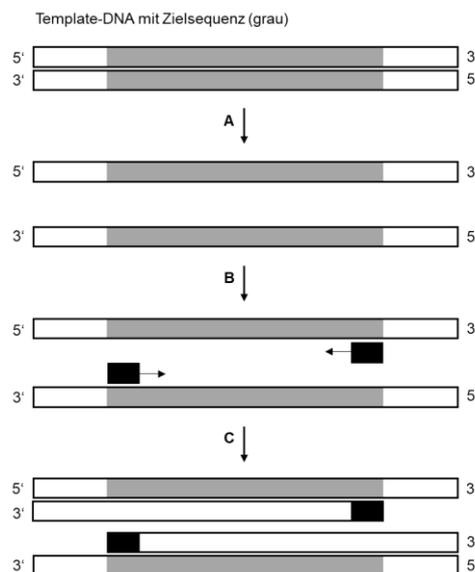


Abbildung 3: Schematischer Ablauf eines Zyklus der Polymerasekettenreaktion (PCR)

A) Melting: Die Template-DNA wird durch thermische Denaturierung in einzelsträngige DNA überführt. B) Annealing: Hybridisierung der Oligonukleotid-Primer am  $3'$ -OH. C) Elongation: Vom Primer ausgehend synthetisiert die DNA-Polymerase die komplementären Stränge in  $5' - 3'$  Richtung. (verändert nach H. Schmidt und Rothhämel, 2012, S.136)

Die beschriebenen drei Schritte stellen einen PCR-Zyklus dar und werden ca. 25- bis 40-mal wiederholt. Die in den Zyklen zuvor entstandenen DNA-Stränge können in den darauffolgenden Zyklen ebenfalls als Template-DNA dienen. (Berg et al., 2018; Kück, 2005; Müller & Prange, 2016) Die Nutzung einer höheren Zyklenzahl ( $>30$ ) erhöht die Wahrscheinlichkeit, dass fehlerhafte Nukleotide eingebaut oder unspezifische Fragmente amplifiziert werden (Müller & Prange, 2016). Vor dem ersten Schritt wird meist ein länger initialer Denaturierungsprozess durchgeführt und der letzte Zyklus durch einen längeren finalen Elongationsschritt erweitert (Committee AMCTB, 2013; Domingues, 2017).

Durchgeführt wird die PCR in PCR-Reaktionsgefäßen, welche in einen Thermocycler gestellt werden. Dieser ist ein temperierbaren Reaktionsraum, welcher automatisch gesteuert ist und die Temperatur zwischen den Stufen reguliert. Die spezifischen Zeiten pro Reaktionsschritt und die entsprechenden Temperaturen werde individuell programmiert (H. Schmidt & Rothhämel, 2012).

Erweiterungen des PCR-Verfahrens gibt es inzwischen einige, auf die hier nicht näher eingegangen werden sollen. Nachzulesen ist dies in einschlägige Literatur: bspw. bei H. Schmidt und Rothhämel (2012), Müller und Prange (2016) und Berg et al. (2018).

## 3.2 Gelelektrophorese

Beim Verfahren der Gelelektrophorese werden elektrisch geladene Stoffe, wie DNA anhand ihrer Ladung und Größe voneinander getrennt (McDonnell et al., 1977; C. Schmidt & Dietrich, 2022). Die Anwendung trägergestützter Elektrophoresen ist weitverbreitet (Flachbettverfahren). Als Trennstrecke dient hierbei eine horizontal oder vertikal angeordnete Platte aus Glas oder Folie, worauf sich ein Trenngel aus Polyacrylamid- oder Agarose-Gel, möglichst homogen und mit ebenmäßiger Schichtdicke aufgetragen befindet (Gey, 2021; McDonnell et al., 1977). Die Einwirkung von elektrischem Strom führt zu einer Wanderung der DNA (z. B. PCR-Produkte) durch das gelartige Medium. Als Medium wird i. d. R. Agarose verwendet. Agarose ist ein weißes Pulver und wird aus den Algengattungen *Gelidium* und *Gracilaria* isoliert. Dieses natürliche Polymer besteht in seiner chemischen Struktur abwechselnd aus  $\beta$ -D-Galaktose- und 3,6-Anhydro-L-Galaktose-Einheiten der Agarobiose (Merck KGaA, 2023a). Durch die Zugabe von Puffer lässt sich das Agarose-Pulver mittels Wärme lösen und es entsteht eine zähe, transparente und farblose Flüssigkeit. Diese geliert nach einigen Minuten bei Raumtemperatur. Hierbei verbinden sich die Agarose-Polymere nicht-kovalent und es entsteht ein Geflecht aus Bündeln. Diese Struktur ergibt eine hohe Stabilität bei einem relativ großen Porendurchmesser, welcher wiederum über die Siebeigenschaften des Gels bestimmt. Im oberen Bereich des Gels befinden sich Vertiefungen, die dem Auftrag der Proben dienen (Ableitner, 2018; P. Y. Lee et al., 2012; Merck KGaA, 2023a; Westermeier & Görg, 2022). Zur elektrophoretischen Trennung von DNA-Molekülen werden meist Tris-Acetat-EDTA-Puffer (TAE) oder Tris-Borat-EDTA-Puffer (TBE) genutzt (Gey, 2021; Green & Sambrook, 2019), da die DNA durch die schwach basische Eigenschaft des Puffermediums vollständig deprotoniert vorliegt ist die Wanderung zur Anode hin begünstigt. EDTA-Puffer dient als Komplexbildner, um mögliche Co-Faktoren wie  $Mg^{2+}$ -Ionen zu chelatisieren und so mögliche Reaktionen durch Nukleasen zu inhibieren (Alberts et al., 2017).

Die DNA-Fragmente, welche durch das Phosphatgerüst negativ geladen sind, wandern im elektrischen Feld zur positiv geladenen Anode. Da die DNA ein einheitliches Masse/Ladungsverhältnis aufweist, werden die DNA-Moleküle in einem Agarose-Gel nach ihrer Größe getrennt, bei dem die zurückgelegte Strecke umgekehrt proportional zum Logarithmus ihres Molekulargewichts ist. Die kleinen Fragmente wandern somit schneller durch das Agarose-Gel und befinden sich in der Nähe zum unteren Rand (Anode) des Gels, wohingegen die größeren Fragmente langsamer wandern und in der Nähe zum oberen Rand des Gels (Kathode) bleiben. Dadurch entsteht ein Muster aus den DNA-Banden, dessen Fragmentgrößen im Vergleich zu einem Größenstandard abgeschätzt werden können (Ableitner, 2018; Berg et al., 2018; Löffler & Petrides, 2022; C. Schmidt & Dietrich, 2022).

Die Agarosekonzentration des Gels orientiert sich an den Längen der aufzutrennenden DNA-Proben. Je größer der prozentuale Anteil an Agarose im Gel ist, desto kleinere DNA-Fragmente können getrennt werden. Üblicherweise werden für die Auftrennung von DNA-Fragmenten im Bereich von 210 bp bis 2100 bp Agarosekonzentrationen von 0,5 % bis 2 % genutzt (Lexikon der Biologie - Spektrum, 2018; Wagner & Beckmann, 2022).

Neben der Agarose-Konzentration im Gel und der DNA-Fragmentgrößen sind weitere Faktoren für die Wandergeschwindigkeit der DNA im Agarose-Gel verantwortlich. Die DNA-Konformation, die angelegte Spannung, das Vorhandensein von Interkallationsfarbstoffen, die Art der Agarose und der gewählte Elektrophoresepuffer spielen eine Rolle (P. Y. Lee et al., 2012).

Die einzelnen DNA-Banden können im Gel nach der Auftrennung angefärbt und unter UV-Licht sichtbar gemacht werden (Löffler & Petrides, 2022). Neben der Möglichkeit der Färbung des Gels nach der Gelelektrophorese kann man auch den Farbstoff dem noch flüssigen Agarose-Gel zufügen. Die Interkallationsfarbstoffmoleküle lagern sich zwischen die Basen der DNA-Moleküle. Anstelle von hochtoxischem Ethidiumbromid wird das Gel heute meist mit Nilblau, GelGreen® oder anderen Interkallationsfarbstoffen oder fluoreszierenden Nukleinsäure-Farbstoff angefärbt (Adkins & Burmeister, 1996; Merck KGaA, 2023b).

### 3.3 Zystische Fibrose

Die zystische Fibrose (*Cystic Fibrosis* – CF), welche auch unter dem Namen Mukoviszidose bekannt ist, ist die häufigste angeborene, autosomal rezessiv vererbte, Stoffwechselerkrankung in Mitteleuropa. Ursache hierbei ist eine Mutation des *CFTR*-Gens, welches auf dem menschlichen Chromosom 7 sitzt (Elborn, 2016; Riordan et al., 1989). Das *CFTR*-Protein (*Cystic fibrosis transmembrane conductance regulator*) ist ein Membrantransportprotein und zeichnet sich dadurch aus, dass es als Ionenkanal fungiert und

seine membranüberspannenden Domänen eine Pore bilden, durch die Chloridionen geleitet werden können (Stratford & Bear, 2006; Z. Zhang et al., 2018).

Je nach Art der Mutation kommt es zu einer Fehlfunktion dieses Kanals und einer damit verbundenen verminderten Chloridionen-Leitfähigkeit an der apikalen Membran der Epithelzellen (Riordan et al., 1989). Auch gibt es Fälle, in denen der Ionenkanal überhaupt nicht mehr vorhanden ist (z.B. Mutation  $\Delta F508$ ). Mehr als 2000 Mutationen, welche zystische Fibrose verursachen, wurden bisher identifiziert (Elborn, 2016; Ensink et al., 2020). Zwei Drittel aller zystischen Fibrose-Erkrankungen lassen sich auf eine einzelne Punktmutation im *CFTR*-Gen zurückführen (Gadsby et al., 2006). Die bekannteste Mutation ist hier  $\Delta F508$ , welche in der Nukleotidbindedomäne 1 (NBD1) lokalisiert ist und durch Verlust (Deletion) der Aminosäure Phenylalanin an Position 508 der Aminosäuresequenz (s. Abbildung 5) zu einer fehlerhaften Faltung des Proteins führt (Du et al., 2005; Gadsby et al., 2006; H. A. Lewis et al., 2004; Riordan et al., 1989). In Abbildung 4 ist schematische Darstellung des intakten *CFTR*-Gens und Proteins dargestellt und zeigt die Position der häufigsten Mutation am *CFTR*-Gen, die  $\Delta F508$ -Deletion im Protein.

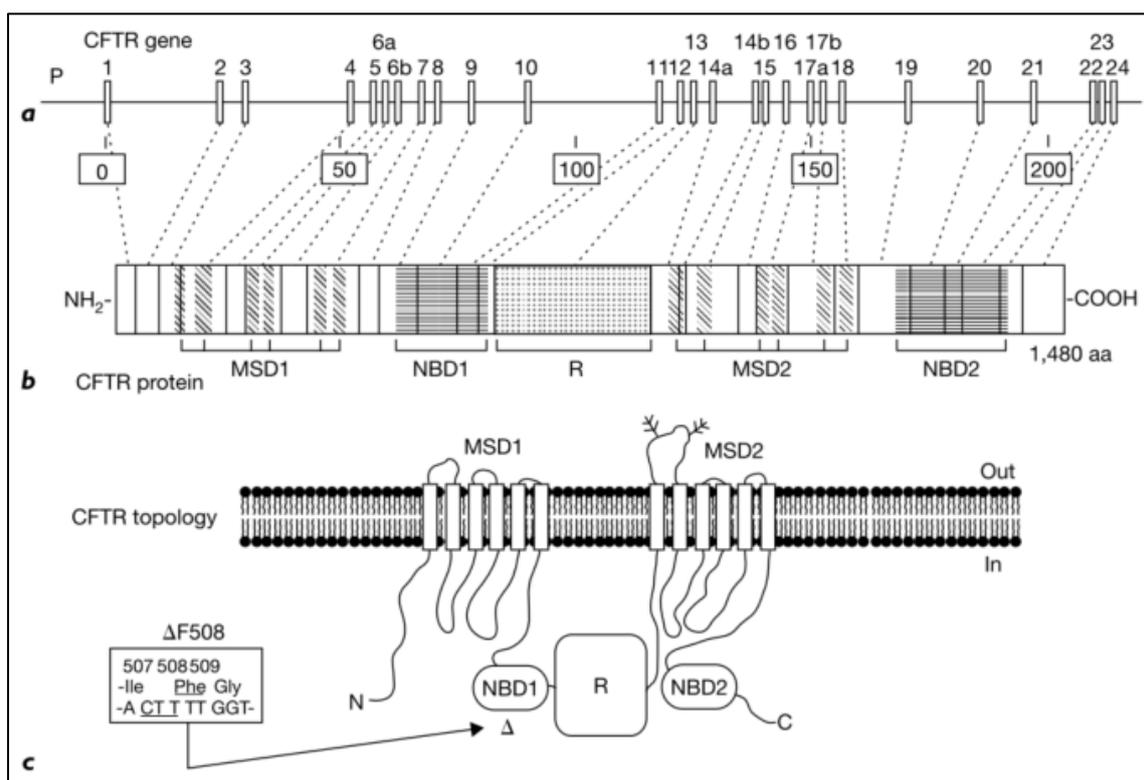


Abbildung 4: Schematische Darstellung des *CFTR*-Gens und Proteins

a) Struktur des *CFTR*-Gens, bestehend aus Promotorregion (P) und 27 Exons. B) *CFTR*-Protein mit vorhergesagten Domänen (hervorgehoben). c) Topologie des *CFTR*-Proteins im Verhältnis zur Cytoplasmamembran und Position der häufigsten Mutation,  $\Delta F508$ . Kasten: Deletion von 3 Nukleotiden, CTT (unterstrichen), und anschließender Verlust von Phenylalanin 508 (unterstrichen).

Übernommen aus Nissim-Rafinia et al. (2006, S. 3) © S. Karger AG, Basel [angepasst aus Zielenski (2000, S. 120) © S. Karger AG, Basel]

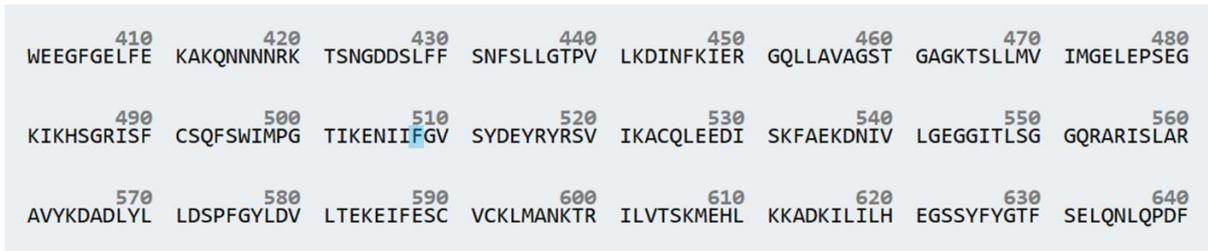


Abbildung 5: Ausschnitt aus der AS-Sequenz des *CFTR*-Gens (Die AS 1–400 und 461–1480 sind nicht dargestellt). Blau hervorgehoben ist die AS 508 Phenylalanin, welche bei der  $\Delta F508$  Mutation fehlt. Übernommen und verändert nach (UniProt, 2007)

Diese fehlerhafte Faltung des Proteins sorgt anschließend für den mit dem endoplasmatischen Retikulum verbundenen Abbau des  $\Delta F508$ -CFTR-Proteins durch den Ubiquitin-Proteasom-Weg (Ward et al., 1995).

Die Mutationen des *CFTR*-Gens werden in Mutationsklassen (I – VI) eingeteilt und sind in Abbildung 6 dargestellt (Boyle & Boeck, 2013). Die gerade beschriebene Mutation  $\Delta F508$  wird der Mutationsklasse II zugeordnet (Elborn, 2016).

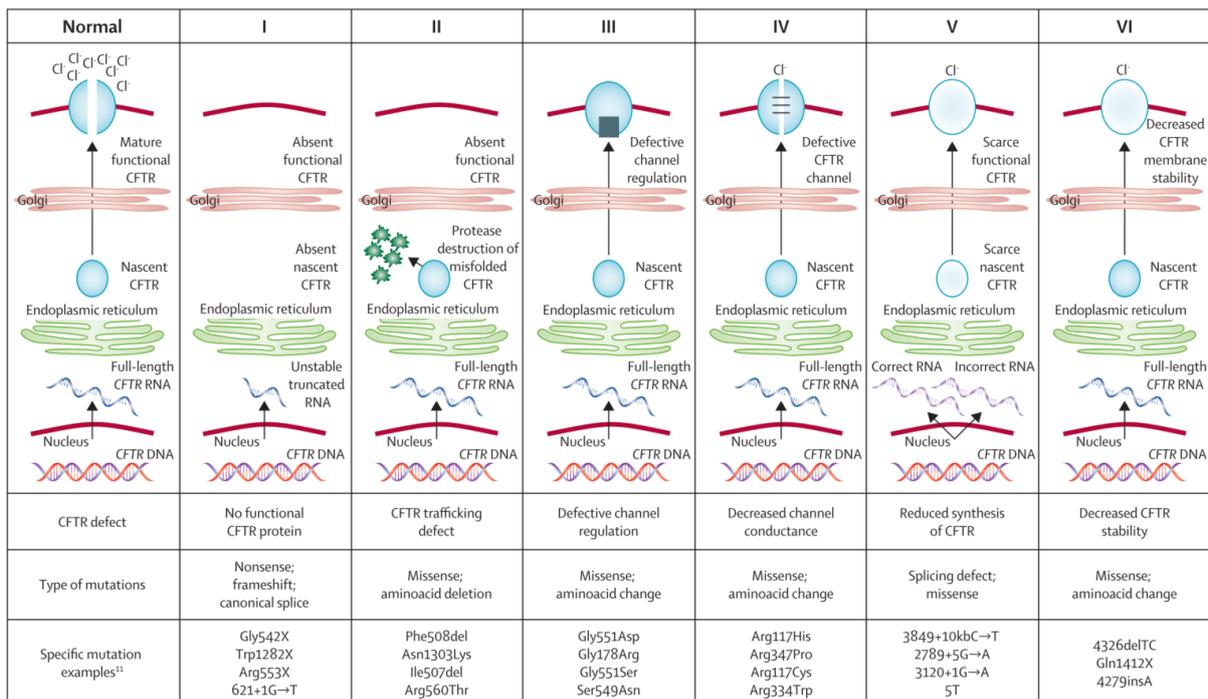


Abbildung 6: Normale CFTR-Protein-Bildung und die Mutationsklassen I bis VI Mutationen der Klasse II (z. B.  $\Delta F508$  – Deletion der Aminosäure Phenylalanin an Position 508) führen zur Bildung eines fehlgefalteten Proteins im endoplasmatischen Retikulum und zum anschließenden Abbau durch das Proteasom. Übernommen nach Boyle und Boeck (2013) S. 159 © 2013 Elsevier Ltd.

Das mutierte CFTR-Protein führt zu einer Dysfunktion des Wasser- und Salztransports in der Plasmamembran und den Epithelzellen. Folglich kommt es zu einer vermehrten Natriumionenabsorption und der Abtransport von Chloridionen ist eingeschränkt oder kommt

zum Erliegen. Somit ändert sich die Zusammensetzung einiger Sekrete in Atemwegen, Verdauungstrakt und Fortpflanzungsorganen zu viskösem Schleim. Dies führt in den betroffenen Organen zu verschiedenen Funktionsstörungen wie exokriner Pankreasinsuffizienz, Verdauungsstörungen, die zur Mangelernährung führen können, Leberzirrhosen und Gallensteinen sowie Unfruchtbarkeit bei Männern (L. Krüger, 2018). Besonders charakteristisch sind chronische bakterielle Infektionen der oberen und unteren Atemwege (Elborn, 2016). Die meisten Todesfälle sind auf chronische bakterielle Infektionen oder Lungenentzündungen zurückzuführen, die durch Mukusbildung entstehen (Gadsby et al., 2006; L. Krüger, 2018). Bis heute ist es nicht möglich, die Krankheit zu heilen. Patienten mit zystischer Fibrose haben eine Lebenserwartung von circa 40 Jahren, mit steigender Tendenz (L. Krüger, 2018; Z. Zhang et al., 2018).

### **3.4 Fachliche Analyse des Lerngegenstandes**

#### ***Einbettung des Studientages in den Lehrplan***

Der Studientag gliedert sich im G8 Lehrplan der Qualifikationsphase 11 (Q11) am Gymnasium in den Bereich B11.2 Genetik und Gentechnik im Unterpunkt Gentechnik ein. Gefordert sind hier bedeutende Methoden der Gentechnik, u. a. die Methode der PCR und die Anwendung der Gentechnik durch den genetischen Fingerabdruck (ISB, 2004b). Ähnliches gilt für den Lehrplan Plus der Fachoberschule im Bereich B11 Lernbereich 6: Grundlagen der Cytogenetik, in Kombination mit dem Lernbereich 1: Naturwissenschaftliche Denk- und Arbeitsweisen und Lernbereich 8: Biologisches Praktikum (ISB, 2023b).

Vorausgegangen ist dem Studientag die Lerneinheit der Molekulargenetik und Cytogenetik, welche die Verständnisgrundlagen liefert. Aufbau der DNA, „die DNA als Speicher der genetischen Information“, „Ursachen und Folgen von Genmutationen“ usw. (ISB, 2004b) und „Mechanismus der semikonservativen Replikation“ etc. (ISB, 2023b). An den Studientag anschließend können weitere Anwendungsbeispiele der Gentechnik, welche im Lehrplan des ISB (2004b) gefordert sind, besprochen werden oder auch die erhaltenen Erkenntnisse bzgl. Chancen und Risiken und den damit verbundenen gesellschaftlichen Perspektiven und Problemen genauer betrachtet und beurteilt werden (ISB, 2023b). Die Beschreibung der Lehrplanbezüge bezieht sich im Folgenden auf das Gymnasium in Bayern; andere schulische Bildungsgänge weisen ähnliche Strukturen auf.

Auch in weiteren Unterrichtsfächern spielt der beschriebene Lerninhalt eine Rolle, auf die im Folgenden exemplarisch eingegangen wird. Bereits im Lehrplan der 10. Jahrgangsstufe für katholische Religionslehre unter „K 10.1 Gewissen konkret: Verantwortung für das Leben übernehmen“ (ISB, 2004e) wird die Thematik im Unterpunkt „Lebensrecht und

Menschenwürde auf dem Prüfstand: Chancen und Risiken moderner Medizin und Humangenetik, Diagnostik, Fortpflanzungsmedizin, Organersatz“ angesprochen. In der Jahrgangsstufe 11 bietet neben dem Biologieunterricht auch das Fach Englisch „E 11/12.3 Interkulturelles Lernen und Landeskunde“ (ISB, 2004d) mit dem Sachbereich „Umwelt, Natur, Wissenschaft und Technik: moderne Entwicklungen in Wissenschaft und Technik und damit verbundene ethische Fragen, Umweltprobleme und mögliche Lösungen [...]“ die Möglichkeit genetische Innovationen und Verfahren einzubinden. Schließlich kann das erworbene Wissen in der 12. Jahrgangsstufe im Fach katholische Religionslehre unter dem Punkt „K 12.3 Dimensionen der Zukunft – Gestaltungsauftrag für die Gegenwart (ca. 12 Stunden)“ (ISB, 2004f) für die Diskussion zum Thema „Die Geschichte der Zukunft hat schon begonnen: Lebensvisionen junger Erwachsener unter den Vorzeichen der Zukunft, z.B. Lebenswissenschaften und Medizin (Fortpflanzungsmedizin und Gentechnik), [...]“ genutzt werden.

### ***Bedeutung des Themas für die Schüler\*innen***

Im Lehrplan der 11. Jahrgangsstufe Biologie unter B 11.2 Genetik und Gentechnik (ISB, 2004b) findet sich folgende Aussage zur Bedeutung molekularbiologischer Methoden: „Auf diesen Grundlagen der Molekulargenetik aufbauend, haben die Schüler\*innen die Möglichkeit, elementare Methoden der Gentechnik zu verstehen und Anwendungsbeispiele oder Zukunftsaspekte unter Einbeziehung aktueller Entwicklungen und Wertvorstellungen zu reflektieren.“

Die Genetik und vor allem die Gentechnik stellen aktuelle Forschungsgebiete der Biologie dar. Sie bieten viel Gesprächsstoff, da Gentechnik zum einen medial präsent ist und zum anderen jeden einzelnen in seinem Alltag begegnet, z. B. im Kontext der Produktion von Nahrungsmitteln, Medikamenten oder Erkrankungen (Luger et al., 2017; Renneberg et al., 2023). Globale Herausforderungen, wie die Nahrungsmittelknappheit oder die COVID-19 Pandemie, können durch die Nutzung der Gentechnik angegangen bzw. untersucht werden. Hierzu ist ein grundlegendes Verständnis der entsprechenden Prozesse und deren Auswirkungen von Bedeutung. Zudem ermöglicht ein Verständnis der Arbeitsweisen in der Genetik den Schüler\*innen, die Grundlagen wissenschaftlicher und medialer Diskussionen besser zu verstehen und deren Aussagen kritisch zu beurteilen (Luger et al., 2017).

### **Schülervorstellung**

Im Rahmen des Modells der Didaktischen Rekonstruktion nach Kattmann et al. (1997) wird versucht, die fachlichen Klärungen des Unterrichtsgegenstandes und die pädagogischen Aspekte bzw. das Lernpotenzial der Schüler\*innen gleichermaßen in der Vorbereitung zu berücksichtigen und daraus lernförderliche Unterrichtseinheiten zu entwickeln (Kattmann, 2007, 2022). Gerade die Schülervorstellungen spielen für den Erwerb von Wissen und Fertigkeiten eine entscheidende Rolle, da sie von den Lernenden nicht leichtfertig gegen neue Vorstellungen ausgetauscht werden („conceptual change“). Sie stellen für den Lernenden bereits ernsthafte Überlegungen und Bemühungen dar, natürliche Phänomene zu verstehen (Kattmann, 2015, 2022). Studien wie von J. Lewis und Wood-Robinson (2000) zeigen, dass es Lernenden schwerfällt, Prozesse und Begrifflichkeiten im Bereich der Genetik fachgerecht zu verwenden. Dieses fehlende Wissen und die damit verbundenen Unsicherheiten münden in Fehlvorstellungen der Konzepte. Nach Kattmann (2022) ist es wichtig, die Ursachen des Denkfehlers zu verstehen und daraus wertvolle Erkenntnisse für das pädagogische Handeln zu erhalten. Schülervorstellungen sind für die Schüler\*innen von Bedeutung, weil sie für die Lernenden eine Hilfe beim Verstehen von Prozessen im Alltag waren (Gropengießer & Kattmann, 2018). Durch die Nutzung von Anthropomorphismen konnte nachgewiesen werden, dass das Lernen biologischer Sachverhalte erleichtert wurde und der Lernerfolg größer und anhaltender ist (Kattmann, 2005). Durch die Reflexion von Alltagsvorstellungen im Unterricht zeigt sich ein größerer Lernzugewinn als ohne deren Nutzung (Hamann & Asshoff, 2023; Kattmann, 2015). Alltagsvorstellungen von Schüler\*innen können, je nach Ausprägung, auch genutzt werden, um als Anknüpfungspunkt, als Kontrast, als Brücke oder durch einen Perspektivwechsel zu einem „conceptual change“ zu führen. Hierbei ist der Begriff „conceptual reconstruction“ der treffendere (Kattmann, 2015). Somit ist es im Sinne der didaktischen Rekonstruktion sinnvoll, die Ansichten der Schüler\*innen mit fachlich geklärten Konzepten in Beziehung zu setzen, um lernförderlichen Unterricht zu ermöglichen.

In der Studie von Schwanewedel (2006) zu „[Schüler-]vorstellungen zu Gesundheit und Krankheit im Kontext von Genetik und genetischer Diagnostik“ zeigt sich, dass Lernende Gesundheit mehrdimensional auf körperlicher, psychischer und sozialer Ebene wahrnehmen. Entscheidend ist für sie, wie gut eine Person mit ihrer Krankheit umgehen kann und wie stark sie in ihrer Leistungs- und Handlungsfähigkeit eingeschränkt ist. In Schülervorstellungen werden Erbkrankheiten v.a. durch Missbrauch von Medikamenten oder Drogen verursacht, welche in der nachfolgenden Generationen Ausprägung findet. Die Krankheit ist für die Lernenden die Abweichung von einer Norm des Genotyps (Schwanewedel, 2006; Schwanewedel et al., 2008). Es ist möglich, die mehrperspektivischen Vorstellungen der Lernenden in der Unterrichtspraxis gewinnbringend einzusetzen, um die unterschiedlichen Aspekte von Gesundheit und Krankheit zu betrachten (Schwanewedel, 2006).

Eine wesentliche Schülervorstellung im Bereich der Genetik ist die Gleichsetzung von Genotyp und Phänotyp. Wenn Gene mit Merkmalen gleichgesetzt werden, gibt es keine klare Unterscheidung zwischen Genotyp und Phänotyp. Somit besteht keine Notwendigkeit, einen Mechanismus zu berücksichtigen, durch den ein Gen im Phänotyp ausgedrückt werden könnte. Die Betrachtung der Ebene der Proteine entfällt. Schüler\*innen weisen dem Gen die Eigenschaften des Merkmals zu, dass dieses krank sei oder die Farbe Rot besitzt (J. Lewis & Kattmann, 2004; Schwanewedel et al., 2008). Diese Ebenenstruktur ist für die Lerneinheit in Abbildung 7 dargestellt, welche für die entsprechenden Rückschlüsse auf die unterschiedlichen Ebenen entsprechendes Wissen voraussetzt (Schöppner et al., 2022).

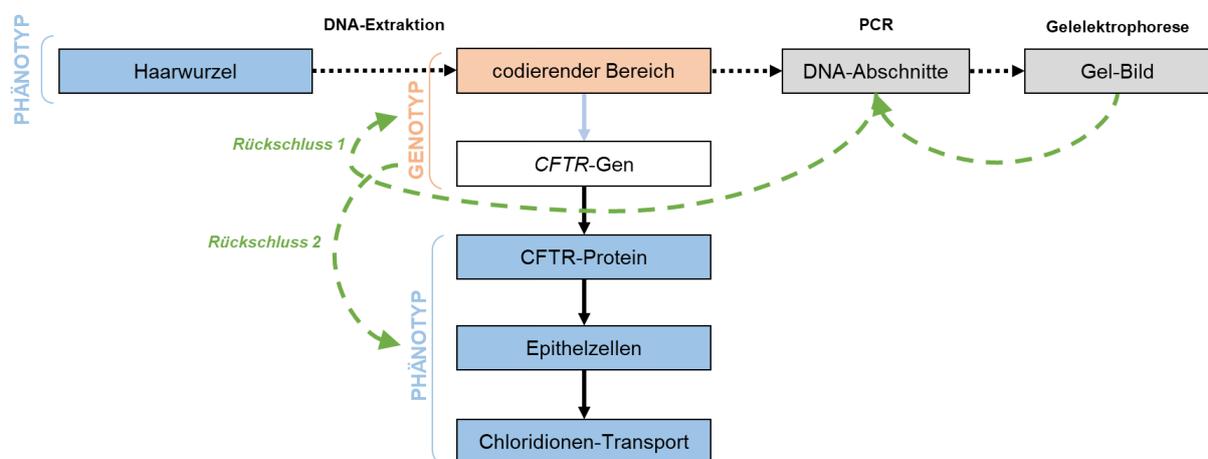


Abbildung 7: Zusammenhänge von Genotyp und Phänotyp und deren Verknüpfung über die drei verschiedenen Analyseschritte DNA-Extraktion, PCR und Agarose-Gelelektrophorese. (Verändert nach Schöppner et al., 2022, S. 81)

Die Gleichsetzung von Gen und Merkmal ist im Sinne des anthropomorphen Prinzips der Kontinuität. Diese Alltagsvorstellung entspricht der Anschauung der Vererbung von Genen als Weitergabe von Gütern (Kattmann, 2022). Zudem fehlt den Schüler\*innen dadurch die Vorstellung des Gens als chemischer Stoff, mit entsprechender regulierter Genaktivität. Sie sind Teile des Systems Zelle und zeigen entsprechende Wechselwirkungen (Kattmann, 2022; Schwanewedel et al., 2008). Dem Gen wird die Eigenschaft „defekt“ oder „funktionierend“ durch die Lernenden zugesprochen, obwohl dies der Phänomen-Ebene zugehört (Kattmann, 2022). Auch sind Gene Strukturen, welche nur schwer veränderbar sind und unverändert von Generation zu Generation vererbt werden (Kattmann, 2022).

### ***Grundwissen und Vorwissen – Kumulativer Lernprozess***

Aus den vorausgegangenen Jahren sollten die Schüler\*innen ein gewisses Vorwissen in die 11. Jahrgangsstufe mitgebracht haben. Inhalte wie den Bau und die Bedeutung von Proteinen, den Aufbau und die Bedeutung der DNA und Grundlagen im Bereich der Gentechnik sind ihnen bereits aus vorausgegangenen Jahrgangsstufen bekannt (ISB, 2004c, 2023c). Dieses Vorwissen wird in den Unterrichtseinheiten und -stunden vor der in dieser Studie untersuchten Thematik, durch die Lehrplaninhalte der 11. Jahrgangsstufe nochmals vertieft und ausgebaut (ISB, 2004b).

### ***Bildungsstandards und Basiskonzepte***

Betrachtet man die Inhalte der Lernumgebungen hinsichtlich ihrer Eignung zur Vermittlung der **Kompetenzen der Bildungsstandards** (KMK, 2005, 2020), so wird die Förderung der **Sachkompetenz** der Schüler\*innen bezüglich „naturwissenschaftlicher Konzepte, Theorien und Verfahren“ durchgeführt. Auch die Fähigkeit, diese Konzepte zu „beschreiben und zu erklären“, wird ausgebildet und getestet (KMK, 2020, S. 10). Neben der Sachkompetenz liegt ein zentraler Fokus der Lernumgebungen auf der **Erkenntnisgewinnungskompetenz** der Lernenden. Die „Kenntnis von naturwissenschaftlichen [...] Arbeitsweisen“ und „die Fähigkeit diese zu beschreiben und zu erklären“ wird den Lernenden eingerichtet. Eine Verknüpfung der Inhalte und somit ein Nachvollziehen der Erkenntnisprozesse sind möglich, um somit „Möglichkeiten und Grenzen [dieser Arbeitsweisen] zu reflektieren“ (KMK, 2020, S. 10). Die **Kommunikationskompetenz** der Schüler\*innen wird durch die Erweiterung der Fachsprache und der damit verbundenen Argumentationsstruktur geschult (KMK, 2020). In spezifischem Maße wird die **Bewertungskompetenz** der Lernenden erweitert und ermöglicht ihnen, die Aussage und Deutung von spezifischen Daten zu beurteilen und daraus gewonnene Entscheidungsprozesse zu reflektieren (KMK, 2020). Erweiternd wird auf die Bildungsstandards, welche in den Lernumgebungen umgesetzt werden, unter 2.2 *Fachkompetenzen im Biologieunterricht* eingegangen werden.

Die Beschlüsse der Kultusministerkonferenz zu den Bildungsstandards im Fach Biologie für die allgemeine Hochschulreife (KMK, 2020), geben an, dass die Schüler\*innen die geforderten Kompetenzen an konkreten Inhalten erwerben sollen. Hier werden im Inhaltsbereich „Vielfalt des Lebens“ (KMK, 2020) die in dieser Arbeit genutzten Lerneinheit vermittelten Inhalte angegeben. Zum Erreichen des Grundniveaus sind dabei die fachlichen Inhalte des Themenbereichs „Molekulargenetische Grundlagen des Lebens“ (KMK, 2020, S. 22) u. a. der Bau der DNA, die semikonservative Replikation, Genmutation, Zusammenhänge zwischen genetischem Material, Genprodukten und Merkmal, als auch die Genetik menschlicher Erkrankungen zu verwenden. Zur Erlangung des erhöhten Anforderungsniveaus, welches dem

Oberstufenniveau entspricht, werden die fachlichen Verfahren der PCR und Gelelektrophorese als Fachinhalt ergänzt (KMK, 2020, S. 22).

Die Beschreibung von Sachverhalten in den Naturwissenschaften basiert auf fachlichen Gemeinsamkeiten, die durch Basiskonzepte strukturiert werden können. Durch die Basiskonzepte des jeweiligen Faches wird die Vernetzung fachlicher Inhalte und deren Betrachtung aus unterschiedlichen Perspektiven ermöglicht. Die **Basiskonzepte** der Bildungsstandards für den mittleren Schulabschluss (KMK, 2005) und des bayerischen Lehrplans für das Fach Biologie (ISB, 2004g) entsprechen sich. Die Basiskonzepte für die allgemeine Hochschulreife (KMK, 2020) sind prinzipiell ähnlich und unterscheiden sich nur in ihrer Zusammenstellung. Diese Arbeit orientiert sich an den Basiskonzepten der Bildungsstandards für die Oberstufe (KMK, 2020). Das Basiskonzept „**Struktur und Funktion**“, welches das Zusammenspiel von Strukturen auf verschiedenen Systemebenen mit den daraus folgenden Funktionen beschreibt, ist in dieser Arbeit eines der relevantesten Basiskonzepte. Es findet an einigen Stellen Anwendung und wird, wie auch die anderen Basiskonzepte, nachfolgend an Beispielen beschrieben. Die Mutation des CFTR-Gens hat eine veränderte Proteinstruktur und somit eine Veränderung der Struktur des Ionenkanals zur Folge. Dies wiederum hat Auswirkungen auf die Funktionalität des Ionenkanals, woraus die Symptome der Krankheit (Mukoviszidose) entstehen (s. 3.3 *Zystische Fibrose*). Auch das Prinzip der komplementären Basenpaarung beim Bau der DNA, welches in der Unterrichtseinheit wiederholt wird, ist diesem Basiskonzept zugeordnet. Das Verfahren der Gelelektrophorese bzw. der Aufbau des Gels in diesem, ist ebenso Basiskonzept zugehörig. Die Agarose-Konzentration bei der Herstellung des Gels bestimmt im Gel über die Größe der Poren und somit über die Siebstärke des Gels (s. 3.2 *Gelelektrophorese*). Das Basiskonzept „**Individuelle und evolutive Entwicklung**“ beschäftigt sich u. a. mit der genetischen Variation, Rekombination und Entwicklung. Es beinhaltet das Prinzip der Replikation von DNA, bei der der vorhanden DNA-Stang als Vorlage zur Neusynthese weiterer DNA-Stränge verwendet wird. Entsprechendes gilt für das Verfahren der PCR, in welchem in vitro DNA vervielfältigt wird (s. 3.1 *PCR (Polymerasekettenreaktion)*). Auch die Mutation des CFTR-Gens stellt eine spontane, ungerichtete Veränderung dar. Sie bewirkt eine Neuausprägung des Merkmals und damit verbundene neue Eigenschaften. „**Information und Kommunikation**“ als Basiskonzept, findet beim genetischen Code Anwendung. Die Abfolge der Basen auf der DNA bestimmt die Aminosäuresequenz und die damit verbundene räumliche Struktur im Protein. Proteine, wie das CFTR-Protein, steuern wiederum Stoffwechselfvorgänge in der Zelle.

### **Fachliche Klärung**

Die **Polymerase-Kettenreaktion (PCR)** ist ein vielseitiges molekularbiologisches Verfahren. Die Fähigkeit der PCR, spezifische DNA-Sequenzen *in vitro* exponentiell zu amplifizieren, ist ein Eckpfeiler für ihrer Effizienz und hohen Ausbeute. Die Verwendung spezifischer Primer-Sequenzen ermöglicht die gezielte Amplifikation von DNA-Regionen (Saiki et al., 1985). Ein genauere fachlicher Einblick in die Funktionsweise ist unter **3.1 PCR (Polymerasekettenreaktion)** beschrieben. Es ist zu beachten, dass die PCR nur in Kombination mit der Gelelektrophorese oder einem anderen Bildgebungsverfahren sinnvoll eingesetzt werden kann. Der Nutzen der PCR in der medizinischen Diagnostik zum Nachweis von Infektionskrankheiten und genetischen Störungen ist allgemein bekannt. So wurde etwa das diagnostische Potenzial der PCR zum Nachweis pathogener Mikroorganismen bereits von Narita (1993) aufgezeigt. In der Forensik wird die PCR zur Analyse von DNA-Spuren und zur Identifizierung von Personen eingesetzt (Anderson, 1992; Berg et al., 2018). Die Rolle der PCR bei der Analyse von genetischen Variationen, Mutationen und Polymorphismen ist von entscheidender Bedeutung für Krankheitsstudien und die Evolutionsforschung (Berg et al., 2018; J. Wang et al., 2009). Die quantitative Realtime-PCR (qPCR) ist ein wichtiges Werkzeug zur Quantifizierung der Genexpression und zur Untersuchung der Genregulation (Berg et al., 2018; T. A. Brown, 2011). Das Verfahren der PCR hat aber auch Grenzen. Es kann zu einer ungleichmäßigen Amplifikation kommen, die bestimmte DNA-Regionen begünstigt und die Ergebnisse verfälscht (Degrelle et al., 2008). Die Bedeutung der Primerspezifität wurde bereits von Saiki et al. (1988) hervorgehoben, wo die Autoren die Bedeutung präziser Primer zur Vermeidung unspezifischer Amplifikation diskutieren. Die Herausforderung bei der Amplifikation sehr langer DNA-Fragmente sowie die Bereitstellung von Polymerasen, die eine geringe Replikationsfehlerrate aufweisen, werden wissenschaftlich kontinuierlich verbessert (Jansohn & Rothhämel, 2012). Je nach Anwendungsgebiet ist es zudem wichtig, die Eignung verschiedener DNA-Polymerasen für bestimmte PCR-Anwendungen vorab zu überprüfen (Berg et al., 2018; Saiki et al., 1988). Zusammenfassend lässt sich sagen, dass die PCR zwar leistungsstarke Möglichkeiten bietet, ihre Grenzen jedoch sorgfältig abgewogen und optimiert werden müssen.

Die soeben beschriebenen wissenschaftlichen Erkenntnisse der PCR, sollen noch einmal in einen genaueren Kontext gestellt und ihre Bedeutung hervorgehoben werden. In Szenarien, die die Vervielfältigung kleinster Mengen von DNA erfordern, steht die PCR im Mittelpunkt, wie die archäologischen Bemühungen und die forensischen Untersuchungen zur Analyse von DNA zeigen. Als Grundlage der Gentechnik ist die PCR ein unersetzliches Werkzeug für die Isolierung, Modifizierung und nahtlose Integration von DNA-Sequenzen. In der medizinischen Diagnostik hat sich die PCR als Schlüssel zur Identifizierung von Krankheitserregern wie Viren und Bakterien etabliert. Ihre Fähigkeit, selbst Spuren von DNA zu amplifizieren, unterstreicht

ihre diagnostische Potenz (Berg et al., 2018). Die Bedeutung der PCR reicht bis in den Bereich der Evolutionsforschung und ermöglichen es den Forschern, genetische Informationen zu extrahieren und zu untersuchen, um evolutionäre Beziehungen zu analysieren (Atawodi et al., 2010). Das Aufkommen weiterer PCR-Techniken, wie der quantitativen PCR (qPCR) hat Studien zur Genexpression revolutioniert, indem sie die Quantifizierung von RNA-Mengen und Einblicke in die Genaktivität ermöglicht (Berg et al., 2018).

Die Polymerase-Kettenreaktion geht über ihren technischen Bereich hinaus und umfasst komplexe Ebenen des wissenschaftlichen, epistemologischen und ethischen Diskurses. Verschiedene Positionen unterstreichen diese Mehrdimensionalität, wobei jede Perspektive die Bewertung und Interpretation der PCR prägt. Eine empirische Position vertritt die Auffassung, dass die PCR strengen Validierungs- und Kontrollverfahren unterliegt und ihre Anwendung auf empirischer Forschung und experimenteller Evidenz beruht. Diese Ansicht, die von Saiki et al. in ihrer Arbeit „Enzymatic Amplification of  $\beta$ -globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia“ (1985) vertreten wird, betont die Notwendigkeit von Kontrollen, statistischen Analysen und Standardisierungen, um die Zuverlässigkeit und Reproduzierbarkeit von PCR-Ergebnissen zu erhöhen. Eine technologieoptimistische Haltung proklamiert die PCR als revolutionären Durchbruch, der die biologische Forschung, Diagnostik und Medizin revolutionierte. Bereits 1987 erkannten K. B. Mullis und Faloona das Potenzial der PCR. Sie unterstreichen die Möglichkeit, genetische Informationen schnell und präzise zu analysieren, was zu tiefgreifenden Erkenntnissen über Krankheiten, Evolution und Genetik geführt hat. Die kritische epistemologische Perspektive zielt darauf ab, mögliche Vorbehalte oder Einschränkungen der PCR zu untersuchen. Man muss sich immer der Unsicherheit bewusst sein, die durch Kontamination, experimentelle Fehler oder eingeschränkte Reproduzierbarkeit entstehen (Bezold et al., 2000; Stillfried & Boor, 2021). Überlegungen zum Umgang mit Proben von Menschen, Tieren oder Umweltproben sind relevant. Fragen der informierten Einwilligung bei Menschen, des Umgangs mit genetischem Material und der möglichen Nutzung von Forschungsergebnissen sind zu berücksichtigen. Es ist daher wichtig, die Ergebnisse der PCR sorgfältig zu analysieren, um Fehlinterpretationen zu vermeiden. (Murphy, 2013) Ethische und gesellschaftliche Fragen sind eng mit der erkenntnistheoretischen Natur der PCR verbunden, wie z. B. diejenigen, die sich mit dem Schutz der Privatsphäre, der Bioethik und den ethischen Auswirkungen der Gentechnik befassen (Herget et al., 2020; McQueen, 1998). Diese Punkte unterstreichen, dass wissenschaftliche Entdeckungen untrennbar mit ihren breiten gesellschaftlichen und ethischen Implikationen verbunden sind.

Viele Lebensbereiche sind heute direkt oder indirekt von den Ergebnissen der PCR betroffen, mit noch unbekanntem Folgen. Die Rolle der PCR in der genetischen Diagnose ist durch den Nachweis von Variationen, Mutationen und Krankheitserregern aufgedeckt, verbessert

die Früherkennung und Prävention von Krankheiten. Aber auch der Einfluss genetischer Informationen auf persönliche und familiäre Entscheidungen ist zu berücksichtigen (Schnieder, 2015). Der forensische Einsatz der PCR zur Täteridentifizierung bietet Möglichkeiten zur Verbrechensaufklärung, wirft aber auch zu Fragen des Datenschutzes und möglicher Voreingenommenheit auf. Auch die Anhäufung genetischer Daten zu Forschungszwecken führt zu umfangreichen Gendatenbanken, die Diskussionen über Datenschutz, Eigentumsrechte und eine mögliche kommerzielle oder voreingenommene Nutzung hervorrufen (Krimsky & Simoncelli, 2012; McQueen, 1998; Murphy, 2013). Die Zugänglichkeit von PCR-Technologien und der diagnostischen Tests könnte soziale Ungleichheiten aufrechterhalten, indem sie bestimmten Gruppen den Zugang verwehrt und damit potenzielle Ungleichheiten in der Gesundheitsversorgung verschärft (Horneber & Möller, 2021). Lebensweltliche Vorstellungen beziehen sich auf die Alltagskonzepte und intuitive Überlegungen, die Menschen aufgrund ihrer Erfahrungen entwickeln. Wissenschaftler\*innen verwenden Metaphern, die auf Alltagserfahrungen basieren, um die PCR zu erklären. Beispielsweise kann die DNA-Amplifikation mit dem Kopieren von Texten oder dem Duplizieren von Fotos verglichen werden (Bachhiesl, 2011; Jessen, 2019). Zur Erklärung der PCR können auch Vergleiche mit natürlichen Prozessen herangezogen werden. Beispielsweise kann die Vervielfältigung der DNA mit der natürlichen Reproduktion von Zellen oder Organismen verglichen werden (Bayrhuber & Kull, 2005; Markl, 2010)

Die **Agarose-Gelelektrophorese** ist eine in der Molekularbiologie weitverbreitete Technik zur Auftrennung von DNA- oder RNA-Fragmenten nach ihrer Größe. Dies wird dadurch erreicht, dass kleinere Fragmente in einem Agarosegel schnell wandern als größere. Ein genauerer fachlicher Einblick in die Funktionsweise ist unter 3.2 *Gelelektrophorese* beschrieben. Zu den Vorteilen dieser Technik gehört die Visualisierung der Fragmente durch Färbung mit Farbstoffen wie Ethidiumbromid, Nilblau oder GelGreen® was eine qualitative Analyse und eine Quantifizierung anhand der Intensität des Farbstoffsignals ermöglicht (Adkins & Burmeister, 1996; Berg et al., 2018). Allerdings hat diese Methode auch ihre Grenzen. Die Auflösung ist begrenzt, die quantitative Genauigkeit ist eingeschränkt und eine genaue qualitative Analyse der Fragmentverhältnisse kann schwierig sein (Berg et al., 2018; T. A. Brown, 2011). Die Wandergeschwindigkeit der Fragmente kann durch Faktoren wie die Gelkonzentration und die Spannung beeinflusst werden, was zu Schwankungen führt (Berg et al., 2018; Gey, 2015; Green & Sambrook, 2019; P. Y. Lee et al., 2012). Kleine Fragmente können schwer zu trennen und sichtbar zu machen sein, während große Fragmente möglicherweise ineffizient wandern können (Wagner & Beckmann, 2022). Trotz ihrer Vorteile ist die Agarose-Gelelektrophorese zeitaufwendig, insbesondere wenn hochauflösende Trennungen erforderlich sind (Armstrong & Schulz, 2015). Die moderne Molekularbiologie bietet fortschrittlichere Alternativen, um einige dieser Einschränkungen zu überwinden.

Die Agarose-Gelelektrophorese ist in der Molekularbiologie und Biochemie von großer Bedeutung. Sie ist der Eckpfeiler der Technik zur Trennung von DNA-Fragmenten nach Größe, die für die DNA-Sequenzierung und den Nachweis von Mutationen unerlässlich ist (T. A. Brown, 2021). Ebenso wichtig ist sie für die RNA-Analyse, die bei Studien zur Genexpression hilfreich ist (C.-S. B. Lee et al., 2003). In der Diagnostik hilft die Agarose-Gelelektrophorese außerdem bei der Identifizierung genetischer Störungen durch Mutationsuntersuchungen (T. A. Brown, 2021). Ihre vielseitige Anwendung hat wesentlich zum Verständnis von DNA und RNA beigetragen. Daher ist sie auch heute noch ein unverzichtbares Instrument in der biologischen Forschung und Diagnostik.

Die Agarose-Gelelektrophorese ist nicht so sehr mit ethischen und gesellschaftlichen Bedenken verbunden wie einige andere biotechnologische Methoden. Es liegt in der Verantwortung der Forscher, die Daten sorgfältig zu verwenden und zu interpretieren, um mögliche Fehlinterpretationen zu vermeiden (Shamoo et al., 2007). In Bildungseinrichtungen sollten ethische Aspekte wie die Verwendung von Proben, Datenschutz und angemessene Kommunikation beachtet werden, wenn die Agarose-Gelelektrophorese als Lehrmittel eingesetzt wird (National Academies Press (US), 2003). Die Anwendungsbereiche der Agarose-Gelelektrophorese entsprechen weitestgehend denen der PCR (siehe oben), da die Agarose-Gelelektrophorese als Trennverfahren und als bildgebendes Verfahren im Anschluss an die PCR genutzt werden kann.

Wissenschaftliche Quellen zur Agarose-Gelelektrophorese zeigen das Vorhandensein lebensweltlichen Vorstellungen, die das Verständnis erleichtern. Analogien zum Trennen von Partikeln in einer Suspension oder zum Sortieren von Gegenständen nach Größe stellen den Bezug zur Alltagswelt her. Beschreibungen von »Bändern« können aufgegriffen werden, um zu veranschaulichen, wie sich Moleküle unterschiedlicher Größe in einem Gel verteilen und trennen. Die Alltagsvorstellungen von Polarität und Ladung können verwendet werden, um zu erklären, warum geladene Moleküle zur Anode oder Kathode wandern und wie dies zur Trennung beiträgt. (Kattmann, 2022) Es ist wichtig anzumerken, dass diese Analogien aus dem Alltag oft als didaktische Hilfsmittel verwendet werden, um Laien oder Studierenden den Zugang zu dieser Methode zu erleichtern.

## 4 Ziele und übergeordnete Fragestellung

Virtuelle, digital präsentierte Lerneinheiten und Lernumgebungen wurden in den letzten Jahren mit unterschiedlichsten Ansätzen auf ihre Lernwirksamkeit hin untersucht. Dabei stand häufig der Vergleich mit einer entsprechenden analogen Lernumgebung im Mittelpunkt der Forschungsfrage (Buckley et al., 2004; Chang et al., 2008; Fleissner-Martin et al., 2023; Frailich et al., 2009; Gunbas, 2015; Kemp & Grieve, 2014; Pérez-Juárez et al., 2023; Tanhua-Piironen et al., 2010). Die Studien zeigen, wie unter *2.1.2.1 Methodische Gestaltungsformen* beschrieben, unterschiedliche Ergebnisse. Auch die Lernwirksamkeit von virtuellen Lernumgebungen für das naturwissenschaftliche Arbeiten, wie z.B. virtuelle Labore, wurde verstärkt untersucht, um deren Möglichkeiten gegenüber Schülerversuchen aus dem Präsenzunterricht zu verdeutlichen (de Jong, 2006; de Jong et al., 2023; de Jong et al., 2013; Govender, 2023; Thisgaard & Makransky, 2017; van Joolingen & de Jong, 1997)

Die COVID-19 Pandemie hat Unterrichtsmethoden wie Homeschooling und Wechselunterricht als digitale Notlösungen für den nicht mehr durchführbaren Präsenzunterricht hervorgebracht. Wie sich diese Unterrichtsformen auf das Lernverhalten von Schüler\*innen in einem naturwissenschaftlichen Kontext auswirken, wurde bisher wenig untersucht. Lediglich das Prinzip des Flipped Classrooms, als Konzept einer hybriden Lernumgebungen, ist in einigen Studien untersucht und auf seine Lernwirksamkeit hin bestätigt worden (Brata et al., 2021; Castro-Rodríguez et al., 2021; Hafeez & Akhter, 2021; Ireland et al., 2009).

Mit Hilfe dieser Arbeit soll untersucht werden, inwiefern sich die Variation der Präsentationsform von Inhalten in Lernumgebungen (hybrid oder rein digital) im Vergleich zum regulären Unterricht in Präsenz (Präsenzunterricht) auf das Fachwissen und die naturwissenschaftlichen Arbeitsweisen in der Biotechnologie auswirkt. Der Schwerpunkt liegt hierbei auf der Vergleichbarkeit der in Präsenz und/oder digital dargebotenen Inhalte und wie sich diese je nach Kombination (Präsenzunterricht, Flipped Classroom, Wechselunterricht und Homeschooling) gewinnbringend auf den Lernerfolg und die Behaltensleistung der Schüler\*innen auswirkt.

### ***Forschungsfragen und Hypothesen***

Die erste Fragestellung beschäftigt sich damit, welche Auswirkungen die Gestaltung der Lernumgebung auf die Förderung und die Behaltensleistung des Fachwissens bei Schüler\*innen im Bereich der Biotechnologie hat. Virtuelle Lernumgebungen zeichnen sich durch ihre Fähigkeit aus, das Lernen effektiv und ansprechend zu gestalten. Eine wesentliche Stärke liegt in der Anpassungsfähigkeit dieser Systeme, die es ermöglichen, Lernmaterialien

und -aktivitäten auf die individuellen Bedürfnisse und Fähigkeiten der Lernenden zuzuschneiden (Letzel & Pozas, 2022). Dies wird durch die gewissenhafte Auswahl von interaktiven Tools und Methoden unterstützt, die das Engagement der Schüler\*innen erhöhen und ihnen helfen, aktiv am Lernprozess teilzunehmen (Deterding et al., 2011; H. Meyer, 2020). Individualisierung des Lernens wird durch digitale Lernplattformen weiter verstärkt, die in der Lage sind, den Fortschritt jedes Lernenden zu verfolgen und personalisierte Lernpfade zu kreieren. Solche Plattformen erlauben es den Schüler\*innen, das Material in ihrem eigenen Tempo zu bearbeiten, wodurch das Lernen zu einer selbstbestimmten Aktivität wird (Lodge & Horvath, 2017). Um kognitive Überlastungen zu vermeiden, werden komplexe Materialien in kleinere, überschaubare Segmente unterteilt, was das schrittweise Lernen unterstützt und das Verständnis erleichtert – ein Ansatz, der als Segmentierungsprinzip bekannt ist (R. E. Mayer & Fiorella, 2022b). Das Prinzip des Vorwissens besagt, dass das Verständnis für komplexe Inhalte verbessert wird, wenn Lernende zunächst einführende Informationen erhalten (R. E. Mayer & Fiorella, 2022b). Des Weiteren lenkt das Signalisierungsprinzip die Aufmerksamkeit auf wichtige Konzepte oder Ideen, indem diese hervorgehoben oder signalisiert werden, was die Wissensaufnahme optimiert (Fiorella & Mayer, 2022c; van Gog, 2022). Das Prinzip der Lernerkontrolle limitiert die Auswahl und Organisation des Materials, sodass Lernende nicht überfordert werden (Scheiter, 2022). Zudem erleichtert das Prinzip der angeleiteten Entdeckung das Lernen in entdeckungsbasierten Multimedia-Umgebungen, indem es Anleitung und Struktur bietet (de Jong, 2022). Durch die Flexibilität, jederzeit und überall auf Lerninhalte zugreifen zu können, können Lernende ihren eigenen Lernrhythmus entwickeln, was die Selbstständigkeit und Eigenverantwortung fördert (Gilhespy, 2011). Das Selbstmanagement-Prinzip bekräftigt, dass Lernende effektiver lernen, wenn sie Kontrolle über ihre Lernprozesse haben (S. Zhang et al., 2022). Die Effektivität digitaler Lernumgebungen wird durch das Feedback-Prinzip weiter verstärkt, da Lernende regelmäßig und konstruktiv Rückmeldungen erhalten, die automatisiert sind und somit unmittelbares Feedback auf Leistungen bieten (Johnson & Marraffino, 2022; H. Meyer, 2020; Narciss, 2018). Dieses sofortige Feedback hilft den Schüler\*innen, rasch aus Fehlern zu lernen und ihr Verständnis für Konzepte und Prozesse zu vertiefen (M. Brown, 2013; X. Chen & DeBoer, 2015; Thisgaard & Makransky, 2017). Interaktive Elemente wie Videos, Simulationen, Quizfragen und Spiele, die in digitalen Lernumgebungen integriert sind, tragen maßgeblich zur Motivationssteigerung bei. Sie regen die Schüler\*innen dazu an, aktives Lernen zu betreiben und generative Lernstrategien wie das Zusammenfassen oder Zeichnen zu verwenden (Fiorella & Mayer, 2022a; Mayrose, 2012). Die Vorteile dynamischer gegenüber statischer Grafiken, bekannt als Animationsprinzip, tragen ebenfalls dazu bei, das Lernen zu intensivieren (Lowe et al., 2022). Abschließend erhöht der Einsatz von Gamification-Elementen wie Belohnungen, Abzeichen und Ranglisten die Motivation der Schüler\*innen, sich

eingehender mit dem Lernmaterial zu beschäftigen (Ab Rahman et al., 2019; Rahman et al., 2018).

Aus diesen und weiteren Erkenntnissen sollte eine digital dargebotene virtuelle Lernumgebung einen größeren Lernerfolg im Fachwissen zeigen und zu einer besseren Behaltensleistung führen. Somit leitet sich die Forschungsfrage F1 mit den Hypothesen H<sub>1.1</sub> und H<sub>1.2</sub> ab:

**F 1 Welche Auswirkung hat die Gestaltung der Lernumgebung auf die Förderung des Fachwissens bei Schüler\*innen im Bereich der Biotechnologie?**

H<sub>1.1</sub>: Eine virtuelle Lernumgebung zum Fachwissen im Bereich der Biotechnologie zeigt bei Schüler\*innen den größeren Lernerfolg.

H<sub>1.2</sub>: Eine virtuelle Lernumgebung zum Fachwissen im Bereich der Biotechnologie zeigt bei Schüler\*innen eine bessere Behaltensleistung.

Die zweite Fragestellung beschäftigt sich mit den naturwissenschaftlichen Arbeitsweisen im Bereich der Biotechnologie und inwiefern die Gestaltung der Lernumgebung Einfluss darauf hat, einen nachhaltigeren Lerngewinn zu erzeugen. In Präsenz durchgeführte Schülerversuche und Laborarbeiten sind von unschätzbarem Wert für den Lernprozess. Die aktive Auseinandersetzung mit verschiedenen Methoden und Techniken, wie sie J. Mayer (2002) beschreibt, ermöglicht es Schüler\*innen, essenzielle wissenschaftliche Fähigkeiten wie genaues Beobachten, das Überprüfen von Hypothesen und sorgfältiges Arbeiten zu erlernen. Durch diese Praxis erwerben sie die Kompetenz, unabhängig neue Probleme zu lösen. Darüber hinaus stärkt die praktische, kritische Auseinandersetzung mit den verwendeten Methoden das Verständnis und das aufgeklärte Verhältnis zu wissenschaftlichen Erkenntnissen. (J. Mayer, 2002; Otsetov, 2020) Dies fördert nicht nur den selbstständigen Wissenserwerb, sondern wirkt sich auch positiv auf die Lernmotivation und die Entwicklung von Interessen aus (Meier & Mayer, 2014). De Jong et al. (2013) beschreiben die Bedeutung von realen Experimenten für das Verständnis der Erkenntnistheorie und der Prozesse des wissenschaftlichen Arbeitens. Wenn Schüler\*innen selbst experimentieren, erleben sie Wissenschaft nicht als abstraktes Konzept, sondern als praktischen Prozess, der greifbare Ergebnisse liefert. Die sozialen Aspekte des Lernens werden ebenfalls verbessert. Die Arbeit in Kleingruppen stärkt soziale Kompetenzen wie Teamfähigkeit und Kommunikation, die in der modernen Arbeitswelt unerlässlich sind (J. Mayer, 2002). Die persönliche Interaktion zwischen Lehrkräften und Schüler\*innen ist ein weiterer entscheidender Faktor, der von Högström et al. (2010) unterstrichen wird. Diese Interaktionen erleichtern tiefere Diskussionen, ermöglichen

schnelle Hilfe und tragen zu einem besseren Verständnis bei (Mottarella et al., 2005; A. Y. Wang & Newlin, 2000). Ebenso wichtig ist die individuelle Betreuung durch Lehrkräfte, die in Präsenzveranstaltungen möglich ist. Lehrkräfte können im direkten Unterricht besser auf die Bedürfnisse und Probleme einzelner Schüler\*innen beim Durchführen des Versuchs eingehen. (Bergmann & Sams, 2012; Lutz et al., 2023) Die für Laborarbeiten in Präsenz durchaus längere erforderliche Zeit und Aufmerksamkeit kann zu einem langsamen, tieferen Lernprozess führen, der in einer analogen Umgebung oft effektiver ist als schnelles, oberflächliches Lernen auf einem iPad o. Ä. (Mueller & Oppenheimer, 2014). Schließlich verweisen Eckert (2020) und Trendel und Lübeck (2018) auf die Vorteile des Konzeptes des Erfahrungslernens. Durch eigenes Erleben und anschließende Reflexion erzielen Schüler\*innen ein tieferes und nachhaltigeres Verständnis des Lernmaterials. Die Verwendung besonderer Materialien im Labor kann zudem eine zusätzliche Motivation darstellen und so die Lernbereitschaft sowie die kognitive Aktivität anregen (Eckert, 2020).

Insgesamt ist die Lernwirksamkeit von in Präsenz durchgeführten Laborarbeiten und Schülerversuchen das Ergebnis einer Kombination aus praktischer Erfahrung, persönlicher Interaktion, individueller Unterstützung und der Möglichkeit zur tiefgreifenden Reflexion und sollte somit den nachhaltigsten Lernerfolg zeigen. Somit leitet sich die Forschungsfrage F2 mit der zugehörigen Hypothesen H<sub>2</sub> ab

## **F 2 Inwiefern hat die Gestaltung der Lernumgebung Auswirkung auf die Nachhaltigkeit des Lernens der naturwissenschaftlichen Arbeitsweise im Bereich der Biotechnologie?**

H<sub>2</sub>: In Präsenz dargebotene Inhalte im Bereich des praktischen Arbeitens in der Biotechnologie zeigen bei Schüler\*innen einen nachhaltigeren Lernerfolg.

Die Kombination aus virtuellem und praktischem Lernen, wie durch Forschungsfrage F1 und F2 nahegelegt, maximiert die Lerneffektivität bei Schüler\*innen. Indem sie zuerst digital eigenständig fachwissenschaftliche Aspekte lernen und dann das Gelernte in Präsenzversuchen anwenden, erlangen sie ein tiefgreifendes Verständnis des Stoffs. Das hybride Konzept, bei dem Lernenden zunächst über digitale Medien lernen und anschließend praktisch einen Versuch durchführen, steigert die Lernwirksamkeit. Diese Kombination fördert sowohl die Selbstständigkeit als auch die praktische Anwendung des Wissens. Studien bestätigen, dass ein solch integrativer Ansatz zu einer nachhaltigeren Bildung beiträgt (Lutz et al., 2023; Ulbrich, 2021).

**F 3 Welchen Einfluss hat die Kombination der Darbietungsform der Lernumgebungen zum Fachwissen und naturwissenschaftlichen Arbeiten auf die Nachhaltigkeit des Lernens bei Schüler\*innen?**

H<sub>3</sub>: Die selbstständige Erarbeitung des Fachwissens in einer virtuellen Lernumgebung in Kombination mit in Präsenz dargebotenen Inhalte im Bereich des praktischen Arbeitens in der Biotechnologie zeigen bei den Schüler\*innen einen nachhaltigen Lernerfolg

## 5 Pilotierung

Um die Fragebögen zu testen und die praktische Nutzung der digitalen Lernumgebungen (s. Abbildung 8) unter realen Bedingungen, sowohl in Bezug auf die technische Umsetzung, als auch die Benutzerfreundlichkeit zu erproben, wurden im Frühjahr 2021 die Pilotierungen dieser durchgeführt.

Die Erprobung der rein digitalen Lernumgebungen entspricht dem in der Hauptstudie genutzten Begriff des Homeschoolings.

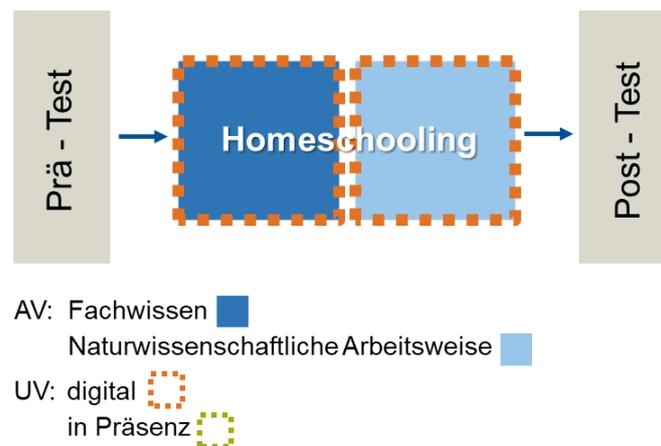


Abbildung 8: Design der Pilotierung der Fragebögen und der digitalen Lernumgebungen. An den Prä-Test schließt sich die Durchführung der digital dargebotenen Lernumgebungen an. Abgeschlossen wird die Pilotierung mit einem Post-Test.

### 5.1 Material und Methoden

#### 5.1.1 Stichprobe

Die Pilotierungsstichprobe umfasst 59 Teilnehmer\*innen der 9. Jahrgangsstufe des Gymnasiums (♀ = 41 %, ♂ = 59 %, Alter: 14,75 Jahre,  $SD = 0,54$  Jahre). Die Schüler\*innen zeigen, gemessen an ihren letzten Zeugnisnoten (1 „sehr gut“ bis 6 „ungenügend“) befriedigende bis gute Leistungen in den Naturwissenschaften, Mathematik und Sprachen (Biologie:  $M = 2,12$ ,  $SD = 0,94$ , Chemie:  $M = 2,32$ ,  $SD = 0,89$ , Mathe:  $M = 2,83$ ,  $SD = 1,12$ , Deutsch:  $M = 2,46$   $SD = 0,81$ , Englisch:  $M = 2,34$ ,  $SD = 0,91$ ). Zum Zeitpunkt des Post-Tests waren nicht mehr alle Schüler\*innen verfügbar und somit verringert sich die Zahl der Datensätze, die in Prä- und Post-Test vorliegen, auf 38. Im Lehrplan des bayerischen Gymnasiums (G8) sind in der 9. Jahrgangsstufe die Themenbereiche „Grundlagen der Genetik“ und „Angewandte Biologie: Grundlagen der Gentechnik“ (ISB, 2004c) eingebunden,

weshalb diese Jahrgangsstufe bezogen auf grundlegendes Wissen in diesem Themenbereich als „vergleichbar“ zur 11. Jahrgangsstufe eingeschätzt wurde.

## **5.1.2 Erhebungsinstrumente und Skalendokumentation**

### **5.1.2.1 Erhebungsinstrumente**

Der entwickelte Fragebogen entspricht den von Moosbrugger und Kelava (2011) formulierten „Qualitätsanforderungen an einen psychologischen Test“. Das Gütekriterium der Objektivität ist erfüllt, da die Fragebögen unter kontrollierten Bedingungen durchgeführt wurden (iPad, eine technisch geschulte Aufsichtsperson, die identische Darbietungsreihenfolge der Fragen und Antworten, ein Antwortzwang bei den Aufgaben, keine Einschränkungen für körperlich benachteiligte Personen, ...). Die Auswertung kann auch durch unabhängige Personen erfolgen, da das Testformat keinen Interpretationsspielraum zulässt (Moosbrugger & Kelava, 2011). Das Kriterium der Reliabilität wird durch die interne Konsistenz überprüft und teilweise erfüllt (Moosbrugger & Kelava, 2011), aber steht aufgrund des vorliegenden Wissenstests und der geplanten Auswertung durch die IRT nicht im Vordergrund der Untersuchung. Die inhaltliche Validität kann durch die Repräsentativität mittels Experten (Fachlehrer\*innen und Biochemiker\*innen) bestätigt werden (Cronach & Meehl, 1955; Moosbrugger & Kelava, 2011). Die gewonnenen Daten sind intervallskaliert und werden nach der Erhebung mit dem Raschmodell auf ihre Güte hin überprüft (Moosbrugger & Kelava, 2011). Die praktische Relevanz des Tests überwiegt den finanziellen und zeitlichen Aufwand. Auch die Aspekte der Zumutbarkeit (Bearbeitungszeit ca. 25min), der Unverfälschbarkeit und der Fairness wurden überprüft und sind gegeben (Moosbrugger & Kelava, 2011).

Bei dem entwickelten Test handelt es sich um einen Leistungstest (Rost, 2004), mit dem die unidimensionalen Merkmale (molekulargenetisches Fachwissen und naturwissenschaftlicher Erkenntnisgewinn in der Molekulargenetik) in einem Multiple-Choice-Design (MC) erfasst werden (Jonkisz et al., 2011). Diese Leistungstests können, aufgrund mangelnder Testmotivation, insbesondere in Feldstudien „nach unten“ verzerrt werden (Rost, 2004). Durch das Antwortformat „2 von 5 Antworten richtig“ wird versucht, dem Erraten von richtigen Antworten entgegenzuwirken (Kubinger et al., 2010). Dieses Antwortformat ist nach Kubinger et al. (2010) einem offenen Antwortformat gleichzusetzen. Die Ratewahrscheinlichkeit liegt hier bei 10 % (Rost, 2004). Die Codierung erfolgt, indem nur vollständig richtige Antworten als richtig gewertet werden und entspricht somit den üblichen Regeln von Leistungstests (Rost, 2004). Daraus ergibt sich eine Codierung von 1 = richtig und 0 = falsch. Die MC-Items wurden in Anlehnung an Haladyna et al. (2002) erstellt. Es wurde darauf geachtet, dass die Items kurz und eindeutig formuliert sind und keine Verneinungen verwendet werden. Die Plausibilität der

Distraktoren ist gegeben und es wurde versucht, die Antwortmöglichkeiten unabhängig voneinander zu formulieren. Es wurde eine klare Fachsprache auf dem Niveau von Lernenden der 11. Jahrgangsstufe verwendet (Haladyna et al., 2002).

### **Skala zur Erfassung des Fachwissens**

Ein von Peter (2019) im Rahmen einer Masterarbeit entwickelter Fragebogen, zu den Themen DNA-Struktur, Proteinbiosynthese und biotechnologische Methoden, der bei Studierenden des Lehramts sehr gute Reliabilitäten aufwies, diente als Grundlage für die Weiterentwicklung zu einem Fachwissenstest Genetik für die 11. Jahrgangsstufe. Durch teilweise leichte Modifikationen dieser Items entstanden die Items 02, 03, 04, 05, 07, 08 und 11. Der Fragebogen wurde um weitere molekularbiologische Items (Item 01, 06, 09 und 10, 12 und 13) ergänzt. Inhaltlich decken die Items das Wissen ab, das von Lernenden der 11. Jahrgangsstufe im Fach Biologie als Grundwissen vorausgesetzt wird bzw. in dieser Jahrgangsstufe neu erworben wird. Grundlage hierzu ist der Lehrplan Biologie (ISB, 2004b). Zur Orientierung des Aufgabenniveaus und der Fachsprache wurden neben gängigen Lehrwerken wie Stryer Biochemie (Berg et al., 2018) auch Schulbücher herangezogen. Hier wurde unter anderem Linder Biologie (Bayrhuber & Kull, 2005), Markl Biologie (Markl, 2010) und Nautilus (Bochter et al., 2009) verwendet. Als Beispiel ist in Abbildung 9 das Item 04 dargestellt.

<p>Bitte wählen Sie bei nachfolgender Frage 2 aus 5 Antworten aus.</p> <p><u>Die DNA-Polymerase ...</u></p> <p>... synthetisiert den Komplementärstrang.</p> <p>... ist Bindestelle für Primer.</p> <p>... katalysiert die Polymerisation von Nukleotiden.</p> <p>... spaltet die Doppelhelix in zwei komplementäre Stränge auf.</p> <p>... ist ein Primer.</p>
---

Abbildung 9: Item 04 im Bereich Fachwissen

### **Skala zur Erfassung der naturwissenschaftlichen Arbeitsweise bzw. Erkenntnisgewinns**

Der Fragebogen zur praktischen Arbeitsweise bei dem naturwissenschaftlichen Erkenntnisgewinn umfasst 12 Items. Diese gliedern sich mit jeweils vier Items in die Anwendung der PCR (Item 21–24), der Agarose-Gelelektrophorese (Item 25–28) und des Pipettierens mit einer Mikroliterpipette (Item 29–32). Das Item 22 ist beispielhaft in Abbildung

10 dargestellt. Der Aufbau der Aufgaben orientiert sich an bewährten methodischen Prinzipien, ähnlich denen, die man auch in der Schule bei Tests anwenden würde. Der Anwendungsbezug der Items zielt darauf ab, die Vorgänge und Abläufe des jeweiligen Verfahrens und den damit verbundenen Erkenntnisgewinn zu überprüfen. Alle Items des Fragebogens sind nach diesem Prinzip erstellt. Die Skala umfasst zum Post- und Follow-up-Testzeitpunkt jeweils alle 12 Items.

Auf einem Agarosegel wurden ein DNA-Längenstandard und drei DNA-Proben aufgetragen. Nach 25 Minuten angelegter Spannung an das Gel in der Gelkammer erhalten Sie nachstehendes Ergebnis.

Markieren Sie die beiden Aussagen die als Ursachen für ein solches Ergebnis vorliegen können.

Der DNA-Längenstandard (DNA-Leiter) passt nicht zur DNA-Fragmentlänge.  
 Der Fluoreszenzfarbstoff (Stain) wurde bei der Gelherstellung vergessen.  
 Die Pole des Stromkreises sind falsch angeschlossen.  
 Der Auftragspuffer wurde bei den Proben vergessen.  
 Es wurden zu große Mengen (Volumina) der Proben in die Taschen aufgetragen.

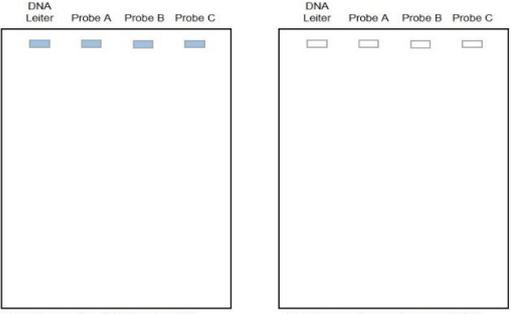


Abb.: Agarose-Gel mit Proben und DNA-Leiter **vor** der Agarose-Gelelektrophorese bei Tageslicht betrachtet.

Abb.: Agarose-Gel mit Proben und DNA-Leiter **nach** der Agarose-Gelelektrophorese bei Blau-Licht betrachtet.

Abbildung 10: Item 22 im Bereich naturwissenschaftlicher Erkenntnisgewinn Agarose-Gelelektrophorese

### 5.1.2.2 Skalendokumentation

#### a) Skala zur Erfassung des Fachwissens im Prä-Test

Quelle: verändert nach Peter (2019) und eigene Skala

Anzahl der Items: 12

Antwortformat: 2 richtige aus 5 Antworten (Kubinger et al., 2010).

Bewertung: zwei richtige Antworten = 1; eine oder keine richtige Antwort = 0

Ein von Peter (2019) im Rahmen einer Masterarbeit entwickelter Test, zu den Themenbereichen Aufbau der DNA, Proteinbiosynthese und biotechnologische Methoden, welcher bei Lehramtsstudierenden sehr gute Reliabilitäten aufwies, wurde als Grundlage für die Weiterentwicklung zu einem Fachwissenstest Genetik für die 11. Jahrgangsstufe an Gymnasium und Fachoberschule verwendet. Die Items lassen sich in drei Unterbereiche der des Themenbereichs Genetik gliedern: 1) die Molekulargenetik, die mit den Items 1, 2, 4, 5, 6, 7 und 8 vertreten ist; 2) der Bereich der Cytogenetik mit den Items 3, 9 und 10 und 3) den Bereich der Biotechnologie mit den Items 11 und 13. Die statistischen Kennwerte der Skala aus der Pilotierung befinden sich in Tabelle 1. Die Cronbachs Alpha Reliabilität der Pilotierung ist unzureichend, mit  $\alpha = 0,14$ . Die Trennschärfen der Items sind nicht immer über 0,25 (Döring & Bortz, 2015), was eine schlechte Korrelation der Items mit der Skala darstellt.

Tabelle 1: Skala zum Fachwissen in der Biotechnologie (Prä-Test-Pilotierung)

*Deskriptive Statistiken Items Fachwissensskala Prä-Test-Pilotierung - Allgemeine Genetik*

Item	Mittelwert	Std.-Abweichung	$r_{it}$	$\alpha$
pr_1	0,66	0,48	0,19	0,02
pr_2	0,34	0,48	0,29	0,14
pr_3	0,12	0,33	0,18	0,10
pr_4	0,10	0,30	0,15	0,25
pr_5	0,15	0,36	0,37	0,08
pr_6	0,37	0,49	0,17	0,20
pr_7	0,31	0,46	0,32	-0,04 <sup>a</sup>
pr_8	0,03	0,18	0,27	0,18
pr_9	0,37	0,49	0,17	0,18
pr_10	0,19	0,39	0,19	0,04
pr_11	0,05	0,22	0,26	0,14
pr_13	0,05	0,22	0,16	0,16
Skala Fachwissen Prä-Test			N = 59	M = 0,23
pr_fw		Cronbachs	$\alpha = 0,14$	SD = 0,15

a. Der Wert ist negativ aufgrund einer negativen mittleren Kovarianz zwischen den Items.

Item	Itemformulierung
pr_1	Die DNA ...
pr_2	Bestandteile der DNA sind ...
pr_3	Ein Gen ...
pr_4	Die DNA-Polymerase
pr_5	Die Assoziierung zweier DNA-Stränge erfolgt durch
pr_6	An den Enden eines DNA-Stranges befinden sich ...
pr_7	Markieren Sie die richtigen Basenpaarungen in der DNA ...
pr_8	Bei der Replikation ...
pr_9	Ein SNP ...
pr_10	Ein rezessives Allel ...
pr_11	Die PCR ...
pr_13	Die Agarose-Gelelektrophorese ...

### **b) Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test**

Quelle: eigene Skala

Anzahl der Items: 6

Antwortformat: 2 richtige aus 5 Antworten (Kubinger et al., 2010).

Bewertung: zwei richtige Antworten = 1; eine oder keine richtige Antwort = 0

Grundlegende naturwissenschaftlichen Arbeitsweise in der Biologie und Chemie werden genutzt, um das allgemeine Verständnis und Wissen in diesem Bereich zu testen. Die Items 15, 16 und 18 prüfen Inhalte des Anfangsunterrichts Chemie. Das Item 20 ist dem Bereich der Elektrochemie zugeordnet. Das Item 17 stellt die grundlegende Kompetenz des Mikroskopierens in der Biologie dar und das Item 19 verbindet chemisches Wissen und die praktische Anwendung in der Biologie. Die statistischen Kennwerte der Skala aus der Pilotierung befinden sich in Tabelle 2. Die Cronbachs Alpha Reliabilität der Pilotierung ist unzureichend, mit  $\alpha = 0,36$ . Die Trennschärfen der Items liegen unter 0,25 (Döring & Bortz, 2015).

Tabelle 2: Skala zur allgemeinen naturwissenschaftlichen Arbeitsweise (Prä-Test-Pilotierung)

*Deskriptive Statistiken Items NWA-Skala Prä-Test-Pilotierung - NWA-Allgemein*

Item	Mittelwert	Std.-Abweichung	$r_{it}$	$\alpha$
pr_15	0,05	0,22	0,24	0,268
pr_16	0,32	0,47	0,15	0,298
pr_17	0,17	0,38	0,05	0,318
pr_18	0,25	0,44	0,23	0,100
pr_19	0,12	0,33	0,07	0,423
pr_20	0,10	0,30	0,07	0,423
Skala allg. NWA Prä- Test			N = 59	M = 0,17
pr_nwa_1		Cronbachs	$\alpha = 0,36$	SD = 0,13

Item	Itemformulierung
pr_15	Universalindikator kommt in flüssiger oder fester Form vor. Ein gängiges Verfahren ist die Nutzung von pH-Papier (s. Abbildung), um den pH-Wert von Stoffen zu bestimmen.
pr_16	Ein wichtiges Gerät für chemische Versuche ist der Gasbrenner. Geben Sie an, welche beiden Aussagen zur Nutzung des Gasbrenners richtig sind
pr_17	Beim Mikroskopieren sind einige Aspekte zu beachten.
pr_18	Nachfolgende Abbildungen zeigen unterschiedliche Aufbauten zu Destillationsapparaturen. Wählen Sie die Aussagen über die Aufbauten aus, die zu einem sinnvollen Ergebnis einer Destillation führen.
pr_19	Chromatographie ist ein wichtiges Trennverfahren, was in vielen Bereichen der Analyse seinen Einsatz findet. Versuche mit Farbstoffen lassen sich einfach durchführen. Markieren Sie die beiden Antworten, die sich auch nebenstehender Abbildung bestätigen lassen.
pr_20	Die Elektrolyse einer Zink-Iodid-Lösung ( $ZnI_2(aq)$ ) sollte Ergebnis A liefern. Bei Ihrem Versuch erhalten Sie allerdings nach einiger Zeit das Ergebnis B (Kabel u. ä. sind nicht dargestellt). ...

### c) Skala zur Erfassung des Fachwissens im Post-Test

Quelle: verändert nach Peter (2019) und eigene Skala

Anzahl der Items: 9

Antwortformat: 2 richtige aus 5 Antworten (Kubinger et al., 2010).

Bewertung: zwei richtige Antworten = 1; eine oder keine richtige Antwort = 0

Die Items entsprechen der Skala zur Erfassung des Fachwissens im Prä-Test-Pilotierung. Die statistischen Kennwerte der Skala aus der Pilotierung befinden sich in Tabelle 3. Die Cronbachs Alpha Reliabilität der Pilotierung ist unzureichend, mit  $\alpha = -0,108$ . Die Trennschärfen der Items sind nicht immer über 0,25 (Döring & Bortz, 2015).

Tabelle 3: Skala zum Fachwissen in der Biotechnologie (Post-Test-Pilotierung)

<i>Deskriptive Statistiken Items Fachwissensskala Post-Test-Pilotierung - Allgemeine Genetik</i>				
Item	Mittelwert	Std.-Abweichung	$r_{it}$	$\alpha$
po_1	0,47	0,51	0,35	-0,008 <sup>a</sup>
po_2	0,45	0,50	0,40	-0,410 <sup>a</sup>
po_4	0,18	0,39	0,12	0,02
po_7	0,39	0,50	0,22	0,03
po_8	0,00	0,00		
po_11	0,08	0,27	0,12	-0,137 <sup>a</sup>
po_12	0,18	0,39	0,14	-0,093 <sup>a</sup>
po_13	0,13	0,34	0,34	-0,050 <sup>a</sup>
po_14	0,08	0,27	0,11	-0,203 <sup>a</sup>
Skala Fachwissen Post-Test			N = 38	M = 0,25
po_fw		Cronbachs	$\alpha = -0,108$	SD = 0,17

a. Der Wert ist negativ aufgrund einer negativen mittleren Kovarianz zwischen den Items.

Item	Itemformulierung
po_1	Die DNA ...
po_2	Bestandteile der DNA sind ...
po_4	Die DNA-Polymerase
po_7	Markieren Sie die richtigen Basenpaarungen in der DNA ...
po_8	Bei der Replikation ...
po_11	Die PCR ...
po_12	Bei der PCR ...
po_13	Die Agarose-Gelelektrophorese ...
po_14	Bei der Agarose-Gelelektrophorese

**d) Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Post-Test**

Quelle:	eigene Skala
Anzahl der Items:	6
Antwortformat:	2 richtige aus 5 Antworten (Kubinger et al., 2010).
Bewertung:	zwei richtige Antworten = 1; eine oder keine richtige Antwort = 0

Die Items entsprechen der Skala zur Erfassung der allgemeinen naturwissenschaftlichen Arbeitsweise in der Prä-Test-Pilotierung. Die statistischen Kennwerte der Skala abefinden sich in Tabelle 4. Die Cronbachs Alpha Reliabilität der Pilotierung ist unzureichend, mit  $\alpha = -0,193$ . Die Werte der Trennschärfen der Items liegen unter 0,25 (Döring & Bortz, 2015).

Tabelle 4: Skala zur allgemeinen naturwissenschaftlichen Arbeitsweise (Post-Test-Pilotierung)

<i>Deskriptive Statistiken Items NWA-Skala Post-Test-Pilotierung - NWA-Allgemein</i>				
Item	Mittelwert	Std.-Abweichung	$r_{it}$	$\alpha$
po_15	0,08	0,27	0,09	-,166a
po_16	0,29	0,46	0,12	-,306a
po_17	0,29	0,46	0,22	0,096
po_18	0,21	0,41	0,13	-,325a
po_19	0,13	0,34	0,12	-,165a
po_20	0,26	0,45	0,11	-,147a
Skala allg. NWA Post- Test			N = 38	M = 0,21
po_nwa_1		Cronbachs	$\alpha = -0,193$	SD = 0,16

a. Der Wert ist negativ aufgrund einer negativen mittleren Kovarianz zwischen den Items.

Item	Itemformulierung
po_15	Universalindikator kommt in flüssiger oder fester Form vor. Ein gängiges Verfahren ist die Nutzung von pH-Papier (s. Abbildung) um den pH-Wert von Stoffen zu bestimmen.
po_16	Ein wichtiges Gerät für chemische Versuche ist der Gasbrenner. Geben Sie an, welche beiden Aussagen zur Nutzung des Gasbrenners richtig sind
po_17	Beim Mikroskopieren sind einige Aspekte zu beachten.
po_18	Nachfolgende Abbildungen zeigen unterschiedliche Aufbauten zu Destillationsapparaturen. Wählen Sie die Aussagen über die Aufbauten aus, die zu einem sinnvollen Ergebnis einer Destillation führen.
po_19	Chromatographie ist ein wichtiges Trennverfahren was in vielen Bereichen der Analyse seinen Einsatz findet. Versuche mit Farbstoffen lassen sich einfach durchführen. ...
po_20	Die Elektrolyse einer Zink-Iodid-Lösung( $ZnI_2(aq)$ ) sollte Ergebnis A liefern. Bei Ihrem Versuch erhalten Sie allerdings nach einiger Zeit das Ergebnis B ...

### e) Skala zur Erfassung des naturwissenschaftlichen Erkenntnisgewinns in der Biotechnologie im Post-Test

Quelle:	eigene Skala
Anzahl der Items:	12
Antwortformat:	2 richtige aus 5 Antworten (Kubinger et al., 2010).
Bewertung:	zwei richtige Antworten = 1; eine oder keine richtige Antwort = 0

Die Skala untergliedert sich in drei wesentliche Arbeitsweisen. Item 21–24 prüft Inhalte zum Verständnis der Agarose-Gelelektrophorese. Item 25–28 befasst sich mit den Prinzipien der PCR und die Items 29–32 überprüft das Verständnis im Umgang mit einer Mikroliterpipette. Die statistischen Kennwerte der Skala aus der Pilotierung befinden sich in Tabelle 5. Die Cronbachs Alpha Reliabilität der Pilotierung ist in Ordnung, mit  $\alpha = 0,616$ . Die Trennschärfen der Items liegen über 0,25 (Döring & Bortz, 2015).

Tabelle 5: Skala zur naturwissenschaftlichen Arbeitsweise biotech. (Post-Test-Pilotierung)

*Deskriptive Statistiken Items NWA-Skala Post-Test-Pilotierung - NWA biotech.*

Item	Mittelwert	Std.-Abweichung	$r_{it}$	$\alpha$
po_21	0,08	0,27	0,36	0,62
po_22	0,05	0,23	0,29	0,65
po_23	0,26	0,45	0,46	0,59
po_24	0,11	0,31	0,41	0,63
po_25	0,18	0,39	0,33	0,57
po_26	0,26	0,45	0,49	0,53
po_27	0,00	0,00		
po_28	0,11	0,31	0,34	0,61
po_29	0,50	0,51	0,40	0,57
po_30	0,16	0,37	0,48	0,56
po_31	0,18	0,39	0,45	0,58
po_32	0,34	0,48	0,43	0,59
Skala biotech. NWA Post- Test			N = 38	M = 0,20
po_nwa_2		Cronbachs	$\alpha = 0,616$	SD = 0,15

Item	Itemformulierung
po_21	Auf einem Agarosegel wurden ein DNA-Längenstandard und zwei DNA-Proben aufgetragen. Nach 25 Minuten angelegter Spannung an das Gel in der Gelkammer erhalten Sie nachstehendes Ergebnis. Markieren Sie die beiden Aussagen, die als Ursachen für ein solches Ergebnis vorliegen können

- po\_22 Auf einem Agarosegel wurden ein DNA-Längenstandard und drei DNA-Proben aufgetragen. Nach 25 Minuten angelegter Spannung an das Gel in der Gelkammer erhalten Sie nachstehendes Ergebnis. Markieren Sie die beiden Aussagen, die als Ursachen für ein solches Ergebnis vorliegen können.
- po\_23 In der Abbildung sehen Sie das Schema eines Gels nach der Gelelektrophorese. Nachfolgend sind fünf Aussagen über das Gel getätigt. Markieren Sie die beiden richtigen Aussagen die sich aus dem Gel-Bild ableiten lassen.
- po\_24 Nachfolgend ist das Ergebnis nach dem Restriktionsverdau aller Proben (A –C) und der PCR auf einem Agarosegel dargestellt. Entscheiden Sie, welche beiden Aussagen sich aus der Auswertung des Gels treffen lassen
- po\_25 Nachfolgend finden Sie vier PCR Protokolle. Wählen Sie aus, welche Protokolle zu einem aussagekräftigen Ergebnis führen würden.
- po\_26 26 In der nachfolgend dargestellten Abbildung sind fünf Temperatur-Zeit-Diagramme dargestellt. Entscheiden Sie, welche der nachfolgenden Diagramme Teile eines typischen PCR-Zeit-Temperaturdiagramms sein können.
- po\_27 Im nachfolgenden Temperatur-Zeit-Diagramm sieht man die verschiedenen Schritte der PCR sehr deutlich. Markieren Sie die beiden richtigen Aussagen zu diesem Diagramm.
- po\_28 Im nachfolgenden Diagramm ist die Anzahl der DNA-Kopien in Abhängigkeit von der Zyklusanzahl beschrieben. Wählen Sie die beiden Antworten, deren Aussagen hierzu zutreffend sind.
- po\_29 Für einen Versuch müssen verschiedene flüssige Substanzen pipettiert werden. Hierzu stehen Ihnen vier gängige Pipetten Varianten zur Verfügung. Wählen Sie die beiden Varianten, die für das angegebene Volumen die sichersten und sinnvollsten sind.
- po\_30 Folgender Ansatz soll pipettiert werden. Entscheiden Sie, welche beiden Reihenfolgen für einen Pipettiervorgang am sinnvollsten sind, bei der praktischen Umsetzung.
- po\_31 Nachfolgend sind 5 Abbildungen angeführt, welche verschiedene Pipettierszenarien darstellen. Entscheiden Sie, welche der Szenarien ohne Fehler sind.
- po\_32 DNA-Proben müssen vor der Gelelektrophorese auf das Agarose-Gel (in der Gelkammer mit Pufferlösung überschichtet) aufgetragen werden. Damit keine Fehler passieren, muss man einige Sachen beachten. Wählen Sie die beiden richtigen Antworten aus, deren Aussagen richtiges Pipettierverhalten zeigen.
-

## 5.2 Ergebnisse

Die Qualität der Fachwissens- und NWA-Fragebögen des Prä- und Post-Tests und deren zugehörigen Itemschwierigkeiten und daraus folgende Personenfähigkeiten der Testteilnehmer\*innen wurden durch Raschanalysen bestimmt. Hierzu werden, wie unter 6.7.1 *Das eindimensionale Rasch-Modell* beschrieben, die entsprechenden Kennwerte überprüft und anschließend diskutiert.

### a) Skala zur Erfassung des Fachwissens im Prä-Test

Der Infit und Outfit der Items liegt zwischen 0,5 und 1,5, nur der Outfit von Item pr\_4 liegt bei einem Wert von 2,38 und bei Item pr\_8 bei 1,57. Bei der Betrachtung der unerwarteten Personenantworten wurden bei Item pr\_4 zwei Antworten mit einem Z-Residual von 6 und bei Item PR\_08 eine Antwort mit einem Z-Residual von 7 aus dem Datensatz entfernt. Der INFIT MNSQ liegt im Durchschnitt bei  $M = 1,00$  ( $SD = 0,6$ ) und der OUTFIT MNSQ bei  $M = 1,05$  ( $SD = 0,24$ ) (s.

Tabelle 6).

Tabelle 6: Übersichtstabelle der Skala zur Erfassung des Fachwissens im Prä-Test; sortiert nach Item-Reihenfolge

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	39	59	-2.74	.32	.92	-.55	.93	-.28	.55	.49	78.2	73.9	PR_1
2	20	59	-1.06	.30	1.04	.38	1.06	.41	.36	.40	67.3	68.7	PR_2
3	7	59	.47	.42	.94	-.12	1.26	.68	.27	.26	87.3	87.3	PR_3
4	4	57	1.13	.53	1.07	.30	1.33	.68	.10	.20	92.7	92.8	PR_4
5	9	59	.15	.38	.94	-.19	.88	-.22	.34	.29	83.6	83.7	PR_5
6	22	59	-1.24	.30	1.10	.93	1.06	.45	.34	.41	63.6	67.8	PR_6
7	18	59	-.87	.31	.84	-1.30	.85	-.79	.51	.38	76.4	70.5	PR_7
8	1	58	2.61	1.02	1.04	.36	1.19	.58	.04	.10	98.1	98.2	PR_8
9	22	59	-1.24	.30	1.10	.90	1.22	1.42	.31	.41	67.3	67.8	PR_9
10	11	59	-.12	.36	.90	-.45	.75	-.77	.41	.31	83.6	80.2	PR_10
11	3	59	1.45	.61	1.03	.23	.90	.12	.15	.17	94.5	94.6	PR_11
12	3	59	1.45	.61	1.05	.27	1.24	.54	.11	.17	94.5	94.6	PR_13
MEAN	13.3	58.7	.00	.45	1.00	.06	1.05	.24			82.3	81.7	
P. SD	10.8	.6	1.44	.20	.08	.60	.18	.62			11.3	11.2	

Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft und ist bei allen Items unter 0,7 und über - 0,7. Die Hauptkomponentenanalyse der Residuen (PCAR) ergibt einen Eigenwert von 1,66 im 1. Kontrast und bestätigt die Eindimensionalität. Der Datensatz enthält vier Personen mit dem Minimalscore von 0 und keine Personen mit dem Maximalscore von 12. Der Personen INFIT MNSQ liegt im Durchschnitt bei  $M = 1,00$  ( $SD = 0,06$ ) und der

OUTFIT MNSQ bei  $M = 1,05$  ( $SD = 0,24$ ). Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten, sind die Personenfähigkeiten der Itemkategorie 0 kleiner als die der Itemkategorie 1.

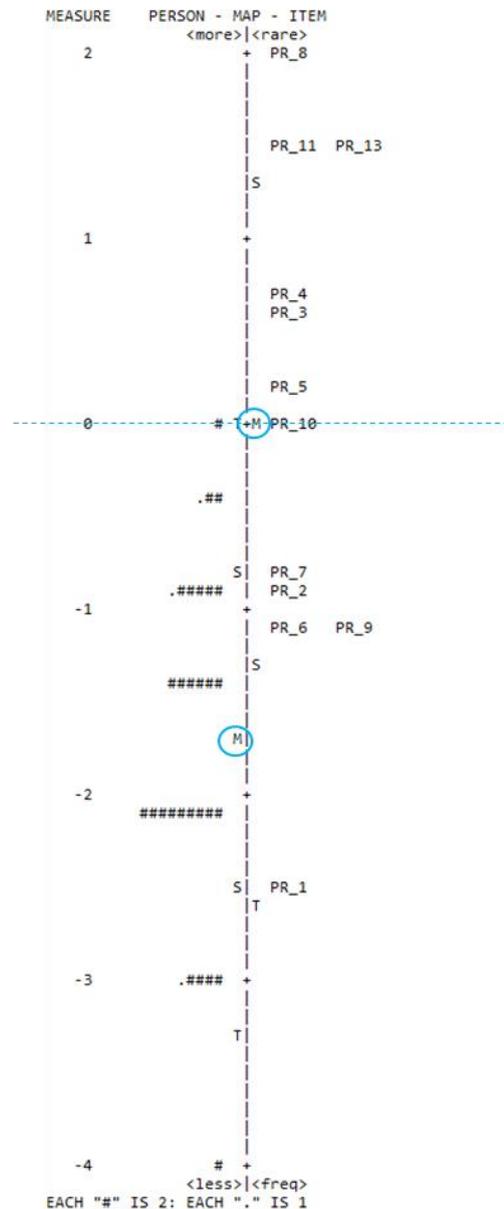


Abbildung 11: Wright Map der Skala zur Erfassung des Fachwissens im Prä-Test (59 Personen, 12 Items, 2 Kategorien: 0/1)

Der Mittelwert der Itemschwierigkeiten „M“ ist auf den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit („M“ linke Seite) ist mit -1,71 Logits geringer. Die Item-Separation von 2,67 ( $> 4,00$ ) und die Item-Reliabilität mit 0,88 ( $> 0,90$ ) der Skala erfüllen die Zielwerte nicht. Die Personen-Separation von 0,42 ( $> 1,50$ ) und die Personen-Reliabilität von 0,15 ( $> 0,50$ ) liegen unterhalb der gewünschten Kennwerte.

### b) Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test

Der Infit und Outfit der Items liegt zwischen 0,5 und 1,5. Der INFIT MNSQ liegt im Durchschnitt bei  $M = 0,99$  ( $SD = -0,09$ ) und der OUTFIT MNSQ bei  $M = 1,03$  ( $SD = -0,10$ ) (s. Tabelle 7). Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft, ist bei allen Items unter 0,7 und über -0,7. Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,76 im 1. Kontrast und bestätigt die Eindimensionalität. Der Datensatz enthält 25 Personen mit dem Minimal score von 0 und keine Personen mit dem Maximalscore von 6. Der Personen INFIT MNSQ liegt im Durchschnitt bei  $M = 1,00$  ( $SD = -0,01$ ) und der OUTFIT MNSQ bei  $M = 0,96$  ( $SD = -0,15$ ).

Tabelle 7: Übersichtstabelle der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test; sortiert nach Item-Reihenfolge

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	3	59	1.45	.62	.76	-.40	.38	-1.03	.44	.28	91.2	91.1	PR_15
2	19	59	-1.41	.37	1.08	.70	1.06	.43	.61	.65	58.8	66.2	PR_16
3	10	59	-.15	.40	1.00	.05	.91	-.32	.50	.49	64.7	72.9	PR_17
4	15	59	-.87	.37	.78	-1.87	.75	-1.85	.70	.58	76.5	64.7	PR_18
5	7	59	.38	.45	1.16	.69	1.49	1.36	.29	.42	76.5	79.9	PR_19
6	6	59	.59	.47	1.20	.77	1.17	.53	.29	.39	79.4	82.6	PR_20
MEAN	10.0	59.0	.00	.45	1.00	-.01	.96	-.15			74.5	76.2	
P.SD	5.5	.0	.94	.09	.17	.93	.35	1.06			10.4	9.3	

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten, sind die Personenfähigkeiten der Itemkategorie 0 kleiner als die der Itemkategorie 1.

Der Mittelwert der Itemschwierigkeiten beträgt den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit ist mit -1,12 Logits geringer. Die Item-Separation von 1,74 ( $> 4,00$ ) und die Item-Reliabilität mit 0,75 ( $> 0,90$ ) der Skala erfüllen die Zielwerte nicht. Die Personen-Separation und die Personen-Reliabilität lag bei 0,00.



Abbildung 12: Wright Map der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test (59 Personen, 6 Items, 2 Kategorien: 0/1)

### c) Skala zur Erfassung des Fachwissens im Post-Test

Der Infit und Outfit der Items liegt zwischen 0,5 und 1,5. Der INFIT MNSQ liegt im Durchschnitt bei  $M = 1,00$  ( $SD = 0,07$ ) und der OUTFIT MNSQ bei  $M = 0,96$  ( $SD = 0,01$ ) (s. Tabelle 8).

Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft, ist bei allen Items unter 0,7 und über - 0,7.

Tabelle 8: Übersichtstabelle der Skala zur Erfassung des Fachwissens im Post-Test; sortiert nach Item-Reihenfolge

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	18	38	-1.39	.36	1.09	.76	1.07	.54	.36	.43	57.1	63.7	PO_1
2	17	38	-1.26	.36	.81	-1.71	.76	-1.85	.61	.43	65.7	64.0	PO_2
3	7	38	.22	.44	1.09	.45	1.16	.54	.21	.31	80.0	79.9	PO_4
4	15	38	-1.00	.36	1.10	.83	1.15	1.00	.31	.41	65.7	66.1	PO_7
5	0	38	3.69	1.82	MAXIMUM MEASURE				.00	.00	100.0	100.0	PO_8
6	3	38	1.27	.62	.97	.09	.77	-.16	.26	.21	91.4	91.4	PO_11
7	7	38	.22	.44	.99	.03	1.04	.22	.30	.31	80.0	79.9	PO_12
8	5	38	.66	.50	1.06	.28	.89	-.09	.24	.26	85.7	85.6	PO_13
9	3	38	1.27	.62	.92	-.02	.56	-.56	.33	.21	91.4	91.4	PO_14
MEAN	8.3	38.0	.41	.61	1.00	.09	.92	-.04			77.1	77.7	
P. SD	6.3	.0	1.51	.44	.10	.74	.20	.82			12.0	11.0	

Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 2,10 im 1. Kontrast (oder die erste PCA-Komponente in der Korrelationsmatrix der Residuen) und widerspricht der Annahme der Eindimensionalität. Der etwas zu hohe Anteil an nicht über die Daten erklärbarer Varianz ist erwartungswidrig. Die Analyse der entsprechenden Daten ergab, dass die beiden errechneten Kontraste rein statistisch eine Multidimensionalität in den Daten zulassen. Auch wenn auf Grundlage der PCAR rechnerisch eine zweite Dimension vorhanden sein könnte, zeigt die Inhaltsanalyse der entsprechenden Items und die Betrachtung der sekundären Dimension, die eine Stärke unter zwei Items hat, dass hier keine zweite Dimension vorliegt (Boone & Staver, 2020; Linacre, 2023a).

Der Datensatz enthält drei Personen mit dem Minimalscore von 0 und keine Person mit dem Maximalscore von 9. Der Personen INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD 0,10) und der OUTFIT MNSQ bei 0,92 (ZSTD 0,10). Die Item-Separation von 1,87 ( $> 4,00$ ) und die Item-Reliabilität mit 0,78 ( $> 0,90$ ) der Skala erfüllen die Zielwerte nicht. Die Personen-Separation und die Personen-Reliabilität liegen bei 0,00.

Der Mittelwert der Itemschwierigkeiten „M“ ist auf den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit („M“ linke Seite) ist mit -1,32 Logits geringer.

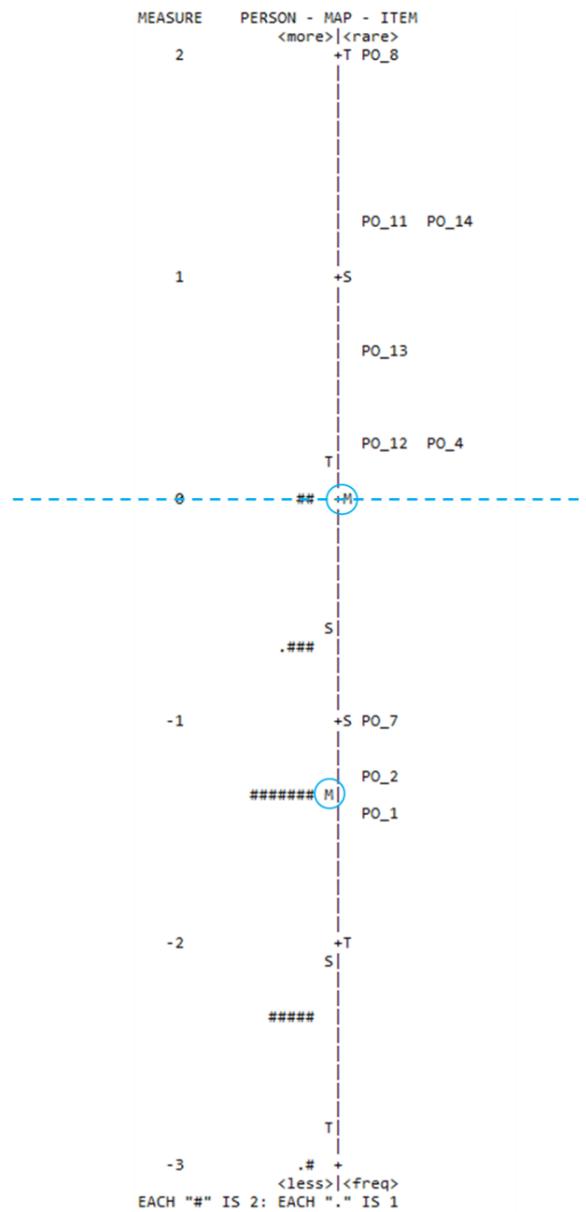


Abbildung 13: Wright Map der Skala zur Erfassung des Fachwissens im Post-Test (38 Personen, 9 Items, 2 Kategorien: 0/1)

**d) Skala zur Erfassung der allgemeinen naturwissenschaftlichen Arbeitsweisen im Post-Test**

Der Infit und Outfit der Items liegt zwischen 0,5 und 1,5. Der INFIT MNSQ liegt im Durchschnitt bei  $M = 1,00$  ( $SD = 0,08$ ) und der OUTFIT MNSQ bei  $M = 1,01$  ( $SD = 0,10$ ) (s. Tabelle 9).

Tabelle 9: Übersichtstabelle der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Post-Test; sortiert nach Item-Reihenfolge

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	3	38	1.17	.62	1.00	.16	1.06	.29	.23	.24	90.0	90.0	PO_15
2	11	38	-.60	.40	.92	-.54	.91	-.53	.51	.43	70.0	67.1	PO_16
3	11	38	-.60	.40	1.17	1.13	1.17	1.05	.29	.43	56.7	67.1	PO_17
4	8	38	-.09	.43	.90	-.43	.82	-.68	.48	.38	73.3	73.8	PO_18
5	5	38	.55	.51	.99	.06	1.07	.30	.30	.31	83.3	83.3	PO_19
6	10	38	-.44	.41	1.00	.07	1.02	.18	.41	.42	73.3	69.4	PO_20
MEAN	8.0	38.0	.00	.46	1.00	.08	1.01	.10			74.4	75.1	
P. SD	3.1	.0	.66	.08	.09	.54	.12	.58			10.5	8.7	

Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft, ist bei allen Items unter 0,7 und über - 0,7. Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,77 im 1. Kontrast und bestätigt die Eindimensionalität. Der Datensatz enthält acht Personen mit dem Minimal score von 0 und keine Personen mit dem Maximalscore von 12. Der Personen INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD 0,14) und der OUTFIT MNSQ bei 1,01 (ZSTD 0,15). Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten, sind die Personenfähigkeiten der Itemkategorie 0 kleiner als die der Itemkategorie 1.

Der Mittelwert der Itemschwierigkeiten beträgt 0 Logit. Die mittlere Personenfähigkeit ist mit -1,58 Logits geringer. Die Item-Separation von 0,98 ( $> 4,00$ ) und die Item-Reliabilität mit 0,49 ( $> 0,90$ ) der Skala erfüllen die Zielwerte nicht. Die Personen-Separation von 0,42 ( $> 1,50$ ) und die Personen-Reliabilität von 0,15 ( $> 0,50$ ) liegen unterhalb der gewünschten Kennwerte. Die Personen-Separation und die Personen-Reliabilität liegen bei 0,00.

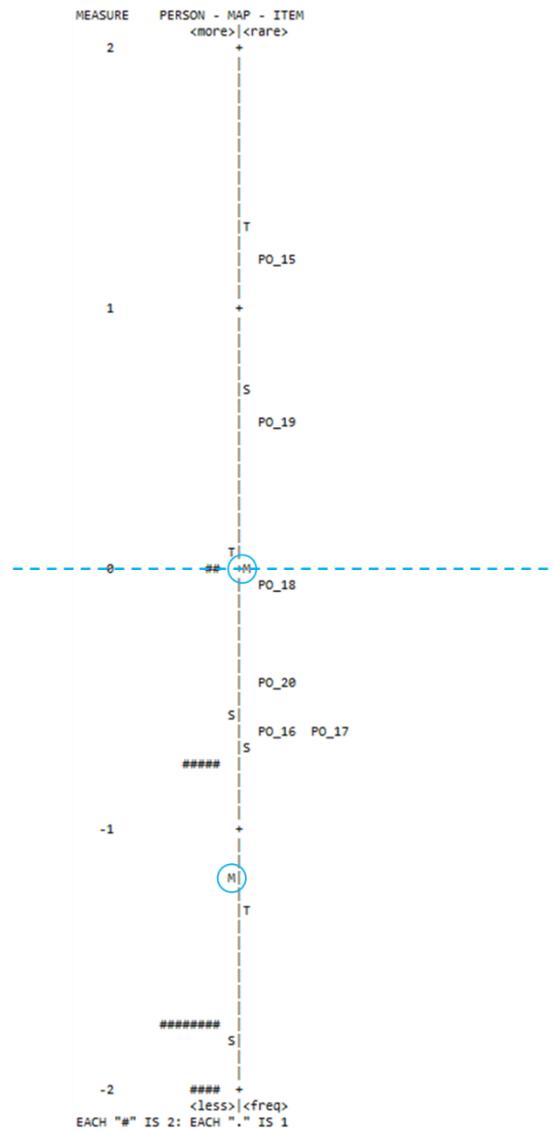


Abbildung 14: Wright Map der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Arbeitsweise im Post-Test (38 Personen, 6 Items, 2 Kategorien: 0/1)

### e) Skala zur Erfassung der naturwissenschaftlichen Arbeitsweisen in der Biotechnologie im Post-Test

Der Infit und Outfit der Items liegt zwischen 0,5 und 1,5, nur der Outfit von Item po\_22 liegt bei einem Wert von 2,12 und bei Item po\_24 bei 1,85. Bei der Betrachtung der unerwarteten Personenantworten wurden bei Item po\_22 zwei Antworten mit einem Z-Residual von 6 und bei Item po\_24 zwei Antworten mit einem Z-Residual größer 5 aus dem Datensatz entfernt. Der INFIT MNSQ liegt im Durchschnitt bei  $M = 0,99$  ( $SD = 0,2$ ) und der OUTFIT MNSQ bei  $M = 0,95$  ( $SD = 0,13$ ) (s. Tabelle 10).

Tabelle 10: Übersichtstabelle der Skala zur Erfassung des naturwissenschaftlichen Erkenntnisgewinns in der Biotechnologie im Post-Test; sortiert nach Item-Reihenfolge

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	3	38	1.30	.65	1.22	.62	.93	.21	.25	.33	89.3	89.2	PO_21
2	0	36	3.83	1.83	MAXIMUM MEASURE				.00	.00	100.0	100.0	PO_22
3	10	38	-.56	.45	1.13	.71	1.18	.66	.49	.56	71.4	72.8	PO_23
4	2	36	1.79	.77	.86	-.07	.49	-.18	.35	.27	92.9	92.8	PO_24
5	7	38	.09	.49	.90	-.34	1.15	.48	.50	.48	85.7	78.1	PO_25
6	10	38	-.56	.45	.81	-.98	.69	-1.06	.65	.56	78.6	72.8	PO_26
7	0	38	3.83	1.83	MAXIMUM MEASURE				.00	.00	100.0	100.0	PO_27
8	4	38	.92	.58	1.06	.27	.83	-.01	.36	.38	89.3	85.9	PO_28
9	19	38	-2.27	.45	1.07	.41	.98	.07	.66	.68	71.4	74.1	PO_29
10	6	38	.33	.51	.84	-.56	.59	-.78	.54	.45	82.1	80.4	PO_30
11	7	38	.09	.49	.94	-.17	1.51	1.14	.46	.48	85.7	78.1	PO_31
12	13	38	-1.13	.43	1.05	.32	1.17	.75	.58	.61	78.6	71.3	PO_32
MEAN	6.8	37.7	.64	.74	.99	.02	.95	.13			82.5	79.6	
P.SD	5.4	.7	1.77	.50	.13	.51	.30	.65			7.0	7.1	

Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 2,56 im 1. Kontrast (oder die erste PCA-Komponente in der Korrelationsmatrix der Residuen) und widerspricht der Annahme der Eindimensionalität. Der etwas zu hohe Anteil an nicht über die Daten erklärbarer Varianz ist erwartungswidrig. Die Analyse der entsprechenden Daten ergab, dass die beiden errechneten Kontraste rein statistisch eine Multidimensionalität in den Daten zulassen. Auch wenn auf Grundlage der PCAR rechnerisch eine zweite Dimension vorhanden sein könnte, zeigt die Inhaltsanalyse der entsprechenden Items und die Betrachtung der geringen Faktorladungen, dass hier keine zweite Dimension vorliegt (Boone & Staver, 2020; Linacre, 2023a).

Der Personen INFIT MNSQ liegt im Durchschnitt bei  $M = 0,98$  ( $SD = 0,02$ ) und der OUTFIT MNSQ bei  $M = 0,95$  ( $SD = 0,11$ ) für die 28 Personen ohne Extremwerte. Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten, sind die Personenfähigkeiten der Itemkategorie 0 kleiner als die der Itemkategorie 1.

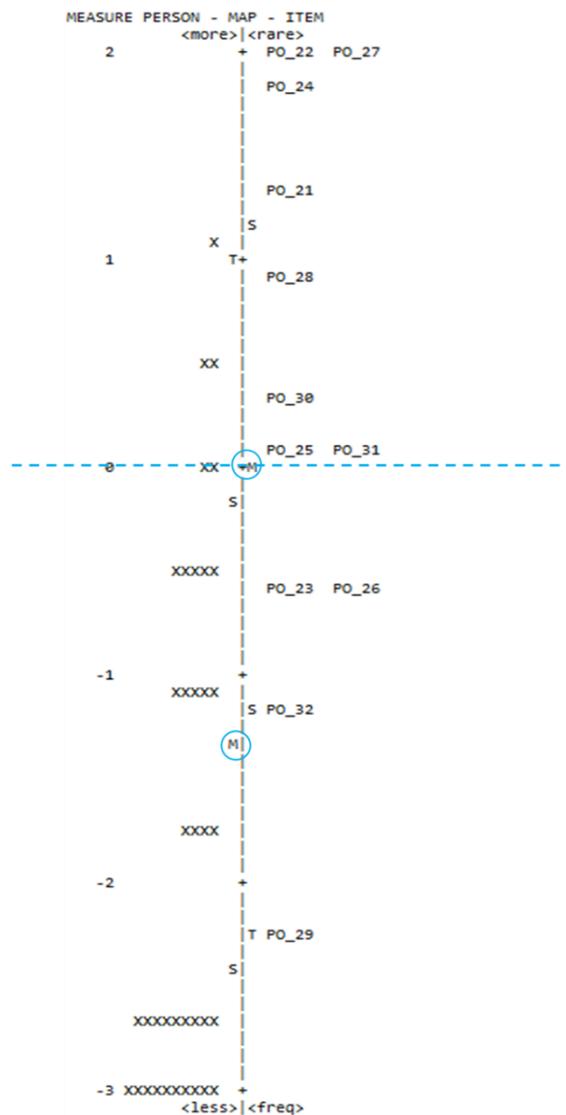


Abbildung 15: Wright Map der Skala zur Erfassung des naturwissenschaftlichen Erkenntnisgewinns in der Biotechnologie im Post-Test (38 Personen, 6 Items, 2 Kategorien: 0/1)

Der Mittelwert der Itemschwierigkeiten liegt bei 0 Logit und die mittlere Personenfähigkeit bei -1,58 Logits. Der Datensatz enthält zehn Personen mit dem Minimalscore von 0 und keine Person mit dem Maximalscore von 12.

Die Item-Separation von 1,68 ( $> 4,00$ ) und die Item-Reliabilität mit 0,74 ( $> 0,90$ ) der Skala erfüllen die Zielwerte nicht. Die Personen-Separation von 0,72 ( $> 1,50$ ) und die Personen-Reliabilität von 0,34 ( $> 0,50$ ) liegen unterhalb der gewünschten Kennwerte.

## 5.3 Diskussion

Es zeigte sich, dass die gute Beantwortung der Items und die hohe Skalen-Korrelation der Masterarbeit (Peter, 2019) nicht im gewünschten Umfang reproduziert werden konnte. Die Ergebnisse der Rasch-Analyse zeigen über alle Skalen hinweg, dass die Verteilung der Items über die Wright Map akzeptabel bis gut ist. Allerdings alle Skalen eine zu hohe Itemschwierigkeit haben. Dies mag hauptsächlich an der gewählten Stichprobe liegen, da Schüler\*innen der 9. Jahrgangsstufe über weniger Wissen im Bereich der molekularen Genetik verfügen, als die für die Hauptstudie geplante Testgruppe einer 11. Jahrgangsstufe. Aufgrund von COVID-19 Beschränkungen war es nicht möglich, Schüler\*innen höheren Alters für eine Pilotierung zu gewinnen, da die geworbenen Klassen der 11. Jahrgangsstufe für die Haupterhebung verwendet werden mussten. **Die Skalen werden für die Hauptstudie nicht verändert, da sie auf das zu erwartende Wissen und die Fähigkeiten von Schüler\*innen der 11. Jahrgangsstufe ausgelegt sind.** Die Items 15–20, welche das allgemeine naturwissenschaftliche Verständnis abfragen, wurden für die Hauptstudie verworfen, da sie bei den Lernenden der 9. Jahrgangsstufe, welche diese grundlegenden Thematiken aktuell vertieft behandeln, geringe Leistungen zeigten und somit keine bessere Beantwortung der Items in den höheren Jahrgangsstufen zu erwarten ist. Zudem zeigen die Items keinen direkten Bezug zur Genetik oder Biotechnologie.

Die Testung der digitalen Lernumgebungen zeigte, dass die Schüler\*innen der 9. Jahrgangsstufe sehr gut mit diesen umgehen konnten und kaum technische Unterstützung durch die Betreuung notwendig war. Die Rückmeldungen ergaben ein durchwegs positives Bild zur Nutzerfreundlichkeit und Umsetzung der Inhalte. Die Verständlichkeit in der Darstellung der Inhalte wurde von allen Schüler\*innen positiv bewertet. Es stellte sich heraus, dass die erforderliche Internetbandbreite nicht überall aufrechterhalten werden konnte, so dass für die Hauptstudie entsprechende technische Ergänzungen (WLAN-Router) angeschafft werden mussten.

## 6 Material und Methoden der Hauptstudie

### 6.1 Stichprobe

Die im Sommer 2021 und 2022 durchgeführte Datenerhebung fand am Werner-Heisenberg-Gymnasium Garching, dem Städtischen Luisengymnasium München und an der staatlichen Fachoberschule Ingolstadt statt. An der Studie nahmen 154 Schüler\*innen der 11. Jahrgangsstufe teil (Gymnasium  $N = 117$ ; FOS  $N = 37$ , ♀ = 45 %, ♂ = 55 %, Alter: 17,20 Jahre,  $SD = 0,77$  Jahre). Die Schüler\*innen zeigen, gemessen an ihren letzten Oberstufenzeugnisnoten (15–13 Punkte „sehr gut“, 12–10 Punkte „gut“, 9–7 Punkte „befriedigend“, ..., 0 Punkte „ungenügend“) befriedigende bis gute Leistungen in den Naturwissenschaften, Mathematik und Sprachen (Biologie:  $M = 9,56$ ,  $SD = 3,02$ , Chemie:  $M = 9,70$ ,  $SD = 3,06$ , Mathematik:  $M = 8,33$ ,  $SD = 3,08$ , Deutsch:  $M = 9,24$ ,  $SD = 2,73$ , Englisch:  $M = 10,62$ ,  $SD = 2,67$ ). Bei den vier Testgruppen handelt es sich um Klumpenstichproben (Blanz, 2021) wobei die einzelnen Biologie-Kurse als Klumpen betrachtet werden. Die Schüler\*innen sind innerhalb einer Gruppe weitestgehend normalverteilt und die Gruppen lassen sich bzgl. Geschlechtsverhältnis und Alter vergleichen.

### 6.2 Studiendesign

Mit einem quasi-experimentellen einfaktoriellen Feldexperiment wird die Entwicklung der beiden abhängigen Variablen Fachwissen (AV1) und die naturwissenschaftliche Arbeitsweise (AV2) im Bereich der Biotechnologie, genauer der PCR und der Agarose-Gelelektrophorese, untersucht. Als unabhängige Variablen werden als Gestaltungsmerkmale der Lernumgebung die Darbietungsformen (digital (UV1) oder in Präsenz (UV2)) betrachtet. Um die Wirksamkeit der ökologisch validen Varianten zu untersuchen, werden die in Präsenz (P) und digital (D) dargebotenen Lernumgebungen je nach Schulsituation systematisch variiert. In jedem durchgeführten Szenario wird zunächst die Theorie vermittelt und anschließend die erlernten Inhalte in der Durchführung eines Schülerversuchs vertieft. Beide Teile können hierbei in Präsenz (P) oder in einem digitalen Format (D) dargeboten und entsprechend kombiniert werden. Daraus ergeben sich die vier schulischen Szenarien PP »Präsenzunterricht«, PD »Wechselunterricht«, DP »Flipped Classroom« und DD »Homeschooling«. Diese Varianten werden durch den Einsatz eines interaktiven Buches im Vergleich zum klassischen Unterricht und eines virtuellen Labors für Biotechnologie im Vergleich zum entsprechenden analogen Schülerexperiment umgesetzt. Die Umsetzung erfolgte im Rahmen der Unterrichtseinheit

„Genetischer Fingerabdruck“ im Themenschwerpunkt „Genetik und Gentechnik“. Mithilfe eines Fragebogens wurde in einem Prä-Post-Design untersucht, wie sich die Sachkompetenz und die Erkenntnisgewinnungskompetenz in Bezug auf PCR und Agarose-Gelelektrophorese durch die Präsentation der verschiedenen Lernumgebungen entwickelt haben. Ein Follow-up-Fragebogen untersucht abschließend die Behaltensleistung der verschiedenen Versuchsgruppen.

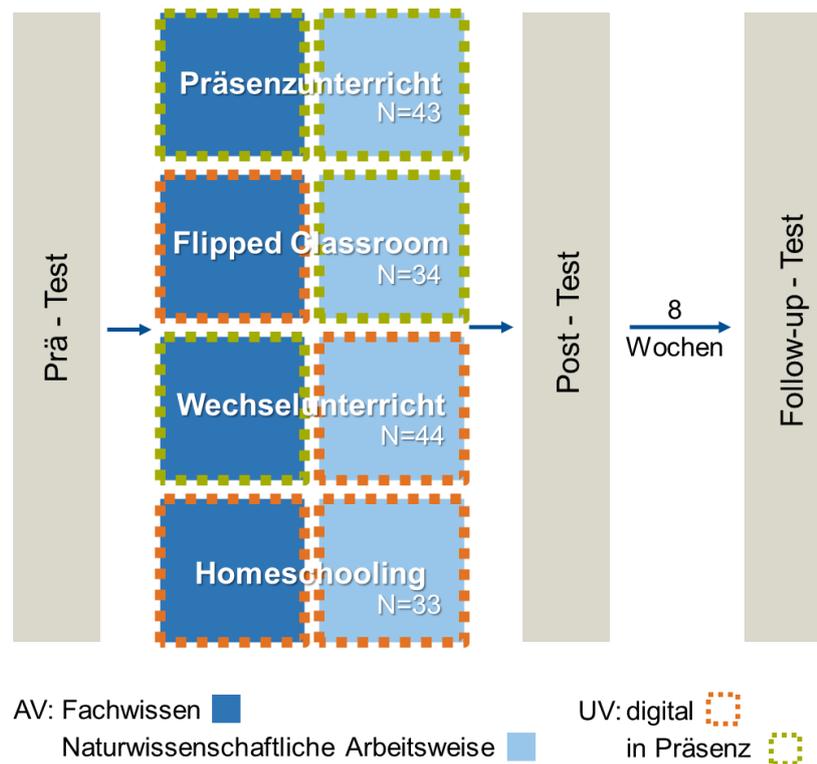


Abbildung 16: Design der Studie. An den Prä-Test schließt sich die Durchführung der in Präsenz und/oder digitalen dargebotenen Lernumgebungen an. Abgeschlossen wird der Studientag mit einem Post-Test, bevor acht Wochen später ein Follow-up-Test genutzt wird. Abhängige Variable Wissen (Fachwissen – dunkelblau; Naturwissenschaftliche Arbeitsweise – hellblau); unabhängige Variable Darbietungsform (digital – orange; in Präsenz – hellgrün)

Die digitalen Lernumgebungen, das Interactive Book und das digitale Labor, wurden in H5P (H5P.org, 2023b) erstellt. Bei ihrer Erstellung wurden die verwendeten Materialien in Bezug auf Umfang, Ausstattung, Illustrationen, Durchführungsdauer etc. parallel zum Klassenunterricht und zum Schülerversuch (Labor) aufgebaut. Die Lernenden wurden entsprechend ihrer Klassenzusammensetzung einer Experimentalgruppe zugeordnet, da eine randomisierte Zuweisung der Teilnehmer, aus organisatorischer Umsetzbarkeit und pädagogischen Vorstellungen nicht sinnvoll waren, woraus sich Klumpenstichproben ergaben (Döring & Bortz, 2015). Der Klassenunterricht zur Vermittlung des fachwissenschaftlichen

Hintergrundes wurde von der jeweiligen Biologielehrkraft durchgeführt, die die vorgegebenen Inhalte mit bereit gestellten Materialien umsetzte, ebenso wie die digitalen Lernumgebungen. Der Schülerversuch (Labor) wurde von einer wissenschaftlichen Mitarbeiterin der TUM betreut, welche den Schüler\*innen aus der Schule bekannt war.

Zum Prä- und Post-Erhebungszeitpunkt teilen sich die Schüler\*innen wie nachfolgend in Tabelle 11 dargestellt auf die vier Testgruppen auf. Zum Zeitpunkt des Follow-up-Tests waren nicht mehr alle Schüler\*innen verfügbar (geänderte Kursabwahl, Schule verlassen, krank, ...), wie in Tabelle 11 zu sehen; somit verringert sich die Zahl der Teilnehmer\*innen.

Tabelle 11: Zusammensetzung der Testgruppen nach Schulart und Testzeitpunkt.

*Personenzusammensetzung der Gruppen zu den verschiedenen Messzeitpunkten*

Gruppen- einteilung	Prä und Post			Follow-up		
	N(Gesamt)	N(Gymnasium)	N(FOS)	N(Gesamt)	N(Gymnasium)	N(FOS)
PP	43	43	-	32	32	-
DP	34	34	-	27	27	-
PD	44	26	18	29	15	14
DD	33	14	19	30	11	19

PP = Präsenzunterricht; DP = Flipped Classroom; PD = Wechselunterricht; DD = Homeschooling

### 6.3 Durchführung der Studie

Um gleiche fachliche Voraussetzungen für alle Schüler\*innen zu gewährleisten und alle Schüler\*innen unabhängig von ihrem Interesse zu erreichen, wurde die Studie an einem Schulvormittag durchgeführt.<sup>2</sup> Die Schüler\*innen wurden jeweils als gesamter Biologie-Kurs einer Versuchsbedingung zugewiesen. Um die Veränderung des Fachwissens (FW) und den praktischen Wissenszuwachs (NWA) im Zusammenhang mit grundlegendem Wissen im Bereich der Genetik, der PCR und der Gelelektrophorese zu überprüfen, wurde diese vor und nach der Behandlung der Lernumgebungen sowie an einem Follow-up-Termin mit den unter 6.5 *Erhebungsinstrumente* berichteten Erhebungsinstrumenten getestet. Im Prä-Test werden zudem die personenbezogenen Daten (Alter, Geschlecht, Schulart, Schulnoten, ...) und nonverbales anschauungsgebundenes Denken abgefragt.

<sup>2</sup> Die Studie wurde am 14.06.2021 durch das Bayerische Staatsministerium für Unterricht und Kultus unter dem Zeichen IV.7-BO5106/210/14 genehmigt.

Am Studientag wird die Klasse begrüßt und auf organisatorische Besonderheiten des Tages aufmerksam gemacht. Anschließend, wie in Abbildung 17 dargestellt, erfolgt die erste Datenerhebung in Form des Prä-Tests auf dem iPad mit einer Bearbeitungszeit von 40 Minuten. An diesen schließt sich die erste Lerneinheit zur Vermittlung der theoretischen Grundlagen von PCR und Agarose-Gelelektrophorese (Fachwissen) an. Diese wird je nach Versuchsbedingung entweder in Präsenz oder digital von der Biologielehrkraft der Klasse mit den Schüler\*innen behandelt (s. 6.2 *Studiendesign*). Nach dieser ersten Lerneinheit folgt eine Pause von 15 Minuten gemäß Stundenplan der Schule. Anschließend folgt die zweite Lerneinheit zur Vermittlung der naturwissenschaftlichen Arbeitsweise im Bereich der Biotechnologie in der Praxis. Je nach Gruppe erfolgt auch dies in Präsenz oder digital. Die Präsenzform des Labors ist auf 120 Minuten angesetzt, da Wartezeiten auftreten. In diesen Wartezeiten können die Schüler\*innen selbst gewählt Pausen einlegen. Das digital dargebotene Labor erfordert keine Wartezeiten und ist somit in 90 Minuten zu bearbeiten. Auch hier können die Schüler\*innen selber frei darüber entscheiden, wann sie Pausen einlegen wollen. Nach Abschluss der zweiten Lerneinheit bearbeiten die Schüler\*innen den Posttest.



Abbildung 17: Ablauf der Datenerhebung am Studientag

Während der Studiendurchführung befinden sich die Schüler\*innen an der ihnen vertrauten Schule und werden von der ihnen bekannten Lehrkraft in die Lernumgebungen eingewiesen. Lediglich das in Präsenz durchgeführte Labor wird von einer Lehrperson angeleitet, die nicht aktiv in der Klasse unterrichtet, aber den Schüler\*innen aus der Schule bekannt ist.

Nach 8 Wochen (Sommerferien) wurden die Schüler\*innen der Biologie-Kurse, wie in Abbildung 18 zu sehen, erneut in einem Follow-up-Test bzgl. ihrer Behaltensleistung des Fachwissens und dem praktischen Wissen im Zusammenhang mit PCR und Gelelektrophorese befragt.

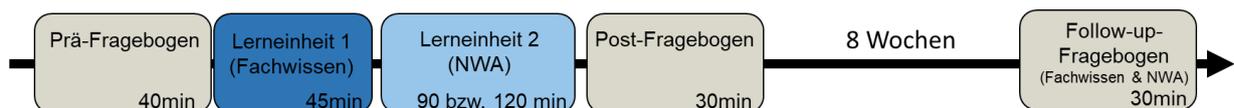


Abbildung 18: Ablauf der Gesamtstudie

Die Bearbeitungszeit der Fragebögen betrug für den ersten Testzeitpunkt 27,76 Minuten,  $SD = 5,79$  Minuten. Für den zweiten Testzeitpunkt 13,45 Minuten,  $SD = 6,26$  Minuten. Und für den Dritten 13,94 Minuten,  $SD = 8,75$  Minuten. Auch die für die digitalen Lerneinheiten

bereitgestellten Bearbeitungszeiten von 45 und 90 Minuten waren angemessen. Entsprechendes gilt für die in Präsenz dargebotenen Lerneinheiten.

## **6.4 Konzeptionierung der Lernumgebungen**

### ***Vergleichbarkeit der Lernumgebungen***

Bei der Gestaltung der Lernumgebungen wird großer Wert auf die Vergleichbarkeit der Lerninhalte und deren Vermittlung gelegt. Bezogen auf die virtuelle Lernumgebung des digitalen Labors wurden u.a. folgende Aspekte bei der Gestaltung berücksichtigt. Der bestehende Schülerversuch im Präsenzunterricht wurde in seiner Struktur analysiert und optimiert. Die sich aus der Analyse ergebenden Eckpfeiler und Besonderheiten des Schülerversuchs wurden in der digitalen Lernumgebung entsprechend verarbeitet. So ist ein wesentlicher Faktor die jederzeit mögliche Kontrolle des Versuchsablaufs oder die Möglichkeit, Aspekte zu wiederholen, um auftretende Verständnisfragen zu klären. Auch die kontextuelle Einbettung des Experiments und die Bedeutung der einzelnen Techniken für das praktische Arbeiten ziehen sich wie ein roter Faden durch beide Lernumgebungen. Entsprechendes gilt für die Lernumgebungen zur Vermittlung des Fachwissens in analoger und digitaler Form. Der Klassenunterricht in Präsenz und das digitale Interactive Book sind inhaltlich aufeinander abgestimmt. Hierbei war es wichtig, den Schüler\*innen in beiden Lernumgebungen die wesentlichen Grundlagen zur Wiederholung bereitzustellen und die Verknüpfung dieser zu den neu zu erlernenden Inhalten zu ermöglichen. Somit sind die digitalen und in Präsenz dargebotenen Lernumgebungen sowohl im Bereich des Fachwissens als auch bzgl. Vermittlung der naturwissenschaftlichen Arbeitsweise inhaltsgleich. Auch die Bearbeitungszeit in den entsprechenden Lernumgebungen kann als identisch angesehen werden.

### ***Didaktische Strukturierung***

Die didaktische Strukturierung wird hier für die Lerneinheit des „Kriminalfalls“ betrachtet. Die fachlichen Inhalte im fachwissenschaftlichen und didaktischen Bereich sind unter *3 Fachwissenschaftlicher Hintergrund* einzusehen. Die Lerneinheit umfasst eine Unterrichtsstunde zur Vermittlung des Fachwissens und drei Unterrichtsstunden zum Erlernen der zugehörigen naturwissenschaftlichen Arbeitsweisen.

Der inhaltliche Aufbau der Lernumgebungen richtet sich unter anderem nach den Aspekten der didaktischen Analyse nach Kattmann (2007) und den grundlegenden Prinzipien des biologiedidaktischen Denkens und Handelns, die unter *3.4 Fachliche Analyse des Lerngegenstandes* beschrieben werden. Hierbei wurden die allgemein vorherrschenden

Schülervorstellungen in diesem Lernbereich in den Lernumgebungen berücksichtigt. Eine wesentliche Schülervorstellung im Bereich der Genetik ist die Gleichsetzung von Genotyp und Phänotyp. In der Lernumgebung wird dies thematisiert, in dem konkret unterschieden wird zwischen dem Phänotyp, also dem entstehenden Protein und der damit verbundenen Krankheit und dem Genotyp des codierenden Bereichs durch den vorliegenden DNA-Abschnitt, welcher durch das Ergebnis im Agarose-Gelelektrophorese-Bild zu erkennen ist. Auch Alltagsvorstellungen der Schüler\*innen wurde aufgenommen und unter anderem genutzt, um Phänomene wie die Siebwirkung eines Agarosegels zu erklären, aber auch auf Fehlvorstellungen einzugehen. Die Lernziel-, Inhalts- und Methodenentscheidungen wurden aufeinander abgestimmt, um einen schlüssigen Unterrichtsverlauf sowohl in Präsenz als auch im digitalen Raum zu ermöglichen.

### ***Mediale Gestaltung der Lernumgebungen***

Die digitalen Lernumgebungen wurden nach den in *2.1.2 Methodische und mediale Gestaltungsmöglichkeiten von Unterrichtsbeschriebenen* Grundsätze für die Gestaltung computergestützter multimedialer Lehrmaterialien zur Förderung des Lernens (R. E. Mayer, 2017; R. E. Mayer & Fiorella, 2022a) erstellt, welche für eine minimale extrinsische kognitive Belastung der Lernenden sorgen sollen. So wurde z.B. das Segmentierungsprinzip im Interactive Book (digitale Lernumgebung zur Vermittlung von Fachwissen) genutzt, um einerseits die verschiedenen fachlichen Lerninhalte räumlich und zeitlich voneinander abzugrenzen, es aber auch dem Lernenden ermöglichen, Bereiche in eigens benötigter Tiefe zu wiederholen bzw. zu erarbeiten und somit die extrinsische Belastung zu reduzieren. Das Prinzip der räumlichen Kontiguität wird unter anderem in den Videos zur praktischen Durchführung im digitalen Labor (digitale Lernumgebung zur Vermittlung der naturwissenschaftlichen Arbeitsweisen) angewendet, um die Schüler\*innen in den mentalen Prozessen zu unterstützen. Auch wird durch eine einfache Nutzung der Oberflächenstruktur der digitalen Lernumgebungen den Erkenntnissen der Cognitive Load Theory nach Chandler und Sweller (1991) Rechnung getragen. So wurden z. B. Grafiken und Videos speziell für die vorliegenden Lernumgebungen erstellt, um eine einheitliche und zielgerichtete Darstellung zu gewährleisten. Bei den digitalen Lernumgebungen handelt es sich nach Roth und Wiesner (2014) um Lernpfade. Lernpfade sind in verschiedene Abschnitte oder Module unterteilt, die jeweils spezifische Themen oder Konzepte behandeln. Das Interactive Book (digitale Lernumgebung zur Vermittlung von theoretischem Fachwissen) ist z.B. in unterschiedliche Seiten, entsprechend einem Buch, aufgeteilt. Jede Seite thematisiert einen Lerninhalt, wie die Wiederholung des Aufbaus der DNA oder die Funktionsweise der Agarose-Gelelektrophorese. Zwischen den eher Text- und Bild-bestimmten Lerninhalten sind Aktivitäten und Aufgaben wie

ein Memory, Quizfragen und Lückentexte eingebettet, welche die Lernenden aktiv in die Lernumgebung einbinden. Hierbei erhalten die Lernenden direktes Feedback nach der Bearbeitung zu ihren Leistungen. Auch im digitalen Labor werden diese Möglichkeiten genutzt. In dem Branching Scenario werden von den Schüler\*innen stetig Entscheidungen zum weiteren Vorgehen eingefordert. Dies führt zu einer aktiven Beteiligung und einem direkten Feedback zu jeder Entscheidung bezogen auf die weitere Aktivität. Auch in dieser Lernumgebung sind die oben beschriebenen Aktionen wie Quizfragen, Zuordnungsaufgaben, Lückentexte und andere vorhanden. Die Präsentation und Durchführung der digitalen Lernumgebungen erfolgt an einem digitalen Endgerät. In diesem Fall auf einem iPad.

Die beiden in Präsenz dargebotenen Lernumgebungen Unterricht zur Vermittlung von Fachwissen und Labor zur Vermittlung der naturwissenschaftlichen Arbeitsweisen sind in ihrer Struktur identisch zu den digitalen Lernumgebungen aufgebaut, erfüllen aber nicht die Ansprüche eines Lernpfads. Der Anteil digitaler Inhalte beschränkt sich hierbei auf die Präsentation von Inhalten (z.B. Power Point-Präsentation) über das Medium Beamer oder Smart Board. Des Weiteren werden analoge Arbeitsblätter genutzt, um mit den Schüler\*innen Inhalte zu erarbeiten und zu sichern. Auch hier werden die unter *2.1.2 Methodische und mediale Gestaltungsmöglichkeiten von Unterricht* beschriebenen Aspekte berücksichtigt. Das entscheidende Medium in den in Präsenz dargebotenen Lernumgebungen stellt die Lehrkraft dar. Wesentliche Aufgabe ist das fundierte Fachwissen, welches die Lehrkraft besitzt, mit der Fähigkeit, dieses Wissen effektiv an die Lernenden zu vermitteln, zu verbinden (Harms & Kattmann, 2018). Dabei werden zudem pädagogische Prinzipien und Techniken angewendet, um die Aufmerksamkeit der Lernenden zu fesseln, sie zu motivieren und eine interaktive Lernumgebung zu schaffen, um damit den Lernprozess von Schüler\*innen zu unterstützen und zu fördern. Darüber hinaus muss die Lehrkraft sich an unterschiedliche Lernstile, Bedürfnisse und Situationen anpassen und damit verbunden die Fähigkeit besitzen, konstruktives Feedback zu geben, um die Lernentwicklung der Schüler\*innen zu unterstützen (Gropengießer, 2018). Zudem muss die Lehrkraft in der Lage sein, moderne Technologien und Medien zu nutzen, um den Unterricht zu bereichern und den Lernprozess zu unterstützen. Die Aufgabendarstellung der Lehrkraft ist hier relativ allgemein gehalten, da jede Klasse der Studie sicher andere spezifische Forderungen an ihre Lehrkraft gestellt hat.

Alle Lernumgebungen sind so gestaltet, dass sie möglichst barrierefrei sind und auch von Schüler\*innen mit Einschränkungen genutzt werden können.

### ***Lernpfade als Beispiel digitaler Lernumgebungen***

Digitale Lernumgebungen sind ein breites und vielseitiges Feld in der Bildungstechnologie, das in den vergangenen Jahren zunehmend an Bedeutung gewonnen hat. Sie umfassen alle digitalen Ressourcen, Tools und Plattformen, die dazu dienen, den Lernprozess zu unterstützen und zu verbessern. Diese Umgebungen reichen von Lernmanagementsystemen und E-Learning-Plattformen hin zu interaktiven Online-Kursen und Simulationen. Digitale Lernumgebungen bieten den Vorteil der Flexibilität, Zugänglichkeit und Individualisierung des Lernens, was sie zu einem wichtigen Bestandteil moderner Bildungssysteme macht.

Im Kontext digitaler Lernumgebungen nehmen Lernpfade eine herausragende Rolle ein. Ein Lernpfad ist nach Roth und Wiesner (2014, S. 1003–1004) „... eine internetbasierte Lernumgebung, die mit einer Sequenz von aufeinander abgestimmten Arbeitsaufträgen strukturierte Pfade durch interaktive Materialien (z. B. Applets) anbietet, auf denen Lernende handlungsorientiert, selbsttätig und eigenverantwortlich auf ein Ziel hin arbeiten“. Diese Pfade sind nicht linear, sondern können an die individuellen Bedürfnisse und das Lerntempo der Lernenden angepasst werden. Sie bieten eine organisierte Möglichkeit, Bildungsinhalte zu präsentieren und Lernziele zu erreichen. Jeder Lernpfad beginnt mit einer klaren Einführung und Zielsetzung. Hier erfahren die Lernenden, was sie in diesem Pfad erreichen und welche Kenntnisse oder Fähigkeiten sie am Ende erwerben sollen. Lernpfade sind in verschiedene Abschnitte oder Module unterteilt, die jeweils spezifische Themen oder Konzepte behandeln. Diese Inhalte können aus Texten, Videos, interaktiven Übungen und anderen Materialien bestehen. Zwischen den Lerninhalten sind Aktivitäten und Aufgaben eingebettet, die die Lernenden aktiv einbinden. Dies können Quizze, Diskussionsforen, Gruppenprojekte oder praktische Übungen sein. Während des Pfads erhalten die Lernenden Feedback zu ihren Leistungen. Dies kann automatisch durch die Plattform oder durch den Lehrenden erfolgen. Die Bewertung erfolgt oft kontinuierlich, um den Lernfortschritt zu verfolgen. Ein entscheidender Aspekt von Lernpfaden ist ihre Flexibilität. Sie passen sich den individuellen Bedürfnissen und dem Lernfortschritt der Lernenden an, wodurch der Heterogenität einer Lerngruppe entgegengewirkt werden kann. Wenn ein Schüler etwa Schwierigkeiten mit einem bestimmten Konzept hat, kann der Pfad zusätzliche Ressourcen oder Übungen anbieten. (Roth, 2015; Roth & Wiesner, 2014; R. Schmidt et al., 2015)

### ***Technische Umsetzung der digitalen Lernumgebungen***

Die digitalen Lernumgebungen der Studie wurden mit der Open-Source-Software H5P („HTML5-Package“) der Firma Joubel erstellt (H5P.org, 2023b). Die Programmierumgebung H5P ermöglicht die Erstellung eigener interaktiver Lehr-Lern-Inhalte und die Darstellung auf Online-Plattformen wie mebis (Online-Lernplattform des bayerischen Kultusministeriums für bayerische Schulen) auf eine einfache und intuitive Art und Weise (mebis-Redaktion, 2020). Zur Erstellung der H5P-Dateien wurde die Open-Source-Software Lumi der Firma Lumi Education GbR (2022) genutzt. Die Software ermöglicht offline und kostenlos, ohne einen H5P-Account, die Erstellung der H5P-Elemente, welche anschließend als H5P-Dateien exportiert werden können (Lumi Education GbR, 2022). Die Abbildungen wurden mit der Software BioRender Academic der Firma BioRender (2023) selbst angefertigt.

Es wurden die beiden H5P Optionen Interactive Book zur Vermittlung des Fachwissens und das Branching Scenario zur Darstellung der naturwissenschaftlichen Arbeitsweisen im Labor in Zusammenhang mit der PCR, der Agarose-Gelelektrophorese und der Arbeitsmethode des Pipettierens gewählt. Diese beiden H5P-Programmierungsumgebungen ermöglichen es den Schüler\*innen eine Vielzahl von interaktiven Inhalten und Wahlmöglichkeiten zu präsentieren (H5P.org, 2023a). Sie ermöglichen die Mischung aus Informationsinhalten und Aufgaben (H5P.org, 2023c). Beim Interactive Book kann der Ersteller zwischen verschiedensten Bausteinen wählen (H5P.org, 2023c). Beim Branching Scenario ist die Wahl auf Kurs-Präsentation mit den entsprechenden Aktivitäten, Text, Bild, Bild-Hotspots, interaktives Video und Video beschränkt (H5P.org, 2023a).

Die Strukturierung des digitalen Labors in Form eines Branching Scenarios zielt darauf ab, naturwissenschaftliche Arbeitsweisen wie den Umgang mit einer Pipette, sauberes und sinnvolles Arbeiten im Labor, Umgang mit Gerätschaften aus dem Labor (Zentrifuge, Thermocycler, Gelkammern, ...), zeitliche Planung und Strukturierung eines Versuchs einzuüben. Die Schüler\*innen haben in dieser Lernumgebung nicht immer die Möglichkeit, Arbeitsschritte zu wiederholen. Der Rückblick auf die benötigten Versuchsprotokolle mit den Arbeitsanweisungen steht ihnen aber jederzeit zur Verfügung. Bei Fehlern in ihrer Wahl werden sie auf diese aufmerksam gemacht und erhalten Hinweise, wie sie weiter vorgehen sollten.

Ein stabiler Internetzugang stellt die Voraussetzung für die Durchführung der digitalen Lernumgebungen dar. Besonders an Schulen sollte dies vor der Nutzung im Klassenverband getestet werden..

## 6.5 Erhebungsinstrumente

Die unter 5.1.2 *Erhebungsinstrumente und Skalendokumentation* beschriebenen Fragebögen werden in der Hauptstudie unverändert übernommen. Hinzu kommen die Items aus dem Kognitiven Fähigkeitstest KFT (s. unten) nach Heller und Perleth (2000). Die KFT-Items werden unverändert verwendet.

Die Anzahl und die Auswahl der verwendeten Items der Skalen des Fachwissens und der naturwissenschaftlichen Arbeitsweise variiert je nach Erhebungszeitpunkt. Tabelle 12 gibt einen Überblick über die Zusammensetzung der Items in den Fragebögen über die Messzeitpunkte.

Tabelle 12: Übersicht über die genutzten Items je Erhebungszeitpunkt

### Fragebogenübersicht mit Itemzuordnung

	Prä-Test		Post-Test		Follow-up-Test	
	FW	KFT	FW	NWA	FW	NWA
Itemauswahl	1 - 11, 13	N2A (25 Items)	1, 2, 4, 7, 8, 11 - 14	21 - 32	1 - 14	21 - 32

FW: Fachwissen; NWA: Naturwissenschaftliche Arbeitsweise

### Kovariaten: KFT (Kognitiver Fähigkeiten Test) und Schulart

Ein wesentlicher Bestandteil im Prä-Fragebogen ist der Test zum „anschauungsgebundenen (figuralen) Denken (N)“ von Heller und Perleth (2000) aus dem „Kognitiven Fähigkeitstest für 4. bis 12. Klassen (KFT 4–12 + R)“. Es handelt es sich um einen Test, der das logisch-räumliche Denken erfasst, indem die dargestellten Formen zueinander in Beziehung gesetzt werden. Die Items werden wie im Testheft für den Subtest N2A für die 11. Jahrgangsstufe genutzt. Eine Beispielaufgabe ist in Abbildung 19 dargestellt. Die Auswertung erfolgt, wie im Testheft beschrieben. Das Ergebnis des KFT geht als Kovariate in die Berechnungen ein.

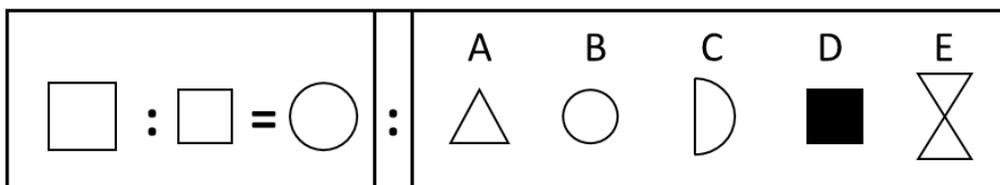


Abbildung 19: Beispielaufgabe aus dem N2A-Test. Logische Erweiterung der vorgegebenen Abfolge. (eigene Darstellung nach Heller und Perleth (2000); Beispielaufgabe zu Test N2A).

Im Prä-Test wird bei der Erhebung der personenbezogenen Daten neben Alter, Geschlecht und Schulnoten auch die Schulart erfasst. Im Verlauf der Auswertung zeigte sich, dass die

Leistungen der Schüler\*innen u.a. durch die Schulart beeinflusst werden. Diese wurde daher als Kovariate mit in die Berechnungen einbezogen.

## 6.6 Lernumgebungen

### 6.6.1 Analoge Lernumgebungen

#### 6.6.1.1 Klassenunterricht

##### **Vorbereitung**

Für den Klassenunterricht wurde den Lehrkräften Material zur Durchführung der Einheit zur Verfügung gestellt. Hierbei handelt es sich um eine PowerPoint-Präsentation und ein darauf abgestimmtes Arbeitsblatt. Das Klassenzimmer muss für die Durchführung über eine digitale Projektionsfläche (Beamer oder Bildschirm) verfügen und das Arbeitsblatt muss den Schüler\*innen als Kopie oder digital zur Verfügung gestellt werden. Die Lehrkraft benötigt ein digitales Endgerät.

##### **Durchführung**

Anhand der PowerPoint-Präsentation erarbeitet die Lehrkraft mit den Schüler\*innen die Thematik der Stunde, die Vermittlung der theoretischen Grundlagen zur PCR und Agarose-Gelelektrophorese. Die Lernziele werden besprochen und visuell dargeboten (s. *3.4 Fachliche Analyse des Lerngegenstandes*). Daraufhin wiederholt die Lehrkraft mit den Schüler\*innen gemeinsam wichtige Grundlagen zum Aufbau von DNA und der natürlichen Replikation dieser. Von diesem Wissen über die Replikation von DNA ausgehend, wird in Partnerarbeit erarbeitet, welche Materialien für einen Laborversuch mit entsprechendem Ziel benötigt werden. Eine gemeinsame Sicherung hierzu erfolgt durch ein Lehrer-Schüler-Gespräch (Plenum) auf dem Arbeitsblatt. Die Lehrkraft beauftragt anschließend die Lernenden mit der Aufgabe, aus den zur Verfügung stehenden Angaben (Text und Abbildung) auf dem Arbeitsblatt die Abbildung eines PCR-Diagramms mit Abbildungen auf Teilchenebene in Einzelarbeit zu beschriften und anschließend ihrem Sitznachbarn in Partnerarbeit den Ablauf der PCR auf Grundlage ihrer vorausgegangenen Arbeit zu beschreiben. Hierbei auftretende Unstimmigkeiten in der Lösung der Schüler\*innen sollen in Partnerarbeit diskutiert und verbessert werden, bevor die Inhalte im Lehrer-Schüler-Gespräch gesichert werden. Eine gemeinsame Besprechung der Bedeutung dieser Thematik beendet den ersten Teil der Stunde.

Das zweite Lernziel der Stunde umfasst die Agarose-Gelelektrophorese. Hierbei werden den Schüler\*innen zunächst zwei DNA-Fragmente unterschiedlicher Länge vorgestellt und nochmals auf den Stundenbeginn und die geplanten Überlegungen zurückgegriffen. Durch genauere Betrachtung des molekularen Aufbaus der DNA und der damit verbundenen

Erkenntnis, dass diese negativ geladen ist, wird im Lehrervortrag der Aufbau der Agarose-Gelelektrophoresekammer und des Agarosegels erklärt um die DNA über ihre Ladung zu trennen. Im Lehrer-Schüler-Gespräch wird die Funktionsweise der Elektrophoresekammer besprochen. Anschließend ergänzen die Schüler\*innen den Lückentext zur Funktionsweise der Gelelektrophorese auf dem Arbeitsblatt in Einzelarbeit, bevor die Richtigkeit gemeinsam gesichert wird. Abschließend wird die Färbung des Gels und die damit entstehenden DNA-Banden auf diesem besprochen. Die Schüler\*innen sollen in Partnerarbeit schriftlich formulieren, welche Beschreibung zu dem Agarose-Gelbild gehört und welche Deutung sich daraus ergibt. Nach der Besprechung dieser Ergebnisse im Plenum endet auch der zweite Teil der Stunde.



Abbildung 20: Klassenunterricht im Klassenzimmer zum Fachwissen in Präsenz.

### 6.6.1.2 Labor

#### **Vorbereitung**

Die Durchführung des Schülerversuchs erfolgte mit den Materialien der Firma miniPCR®. Geräte und Reagenzien wurden über diese Firma bezogen und der „Kriminalfall“ an den Materialien des Sets miniPCR® Crime Lab: Missy Baker Missing aufgebaut (miniPCR, 2022; Schöppner & Nerdel, 2015). Die genutzten Glasgeräte, Pipetten und weitere Materialien wurden durch die entsprechende Schule oder die TU München (Professur für Fachdidaktik Life Sciences) bereitgestellt.

Der praktische Versuch wird im Klassenzimmer durchgeführt, welches für diesen Zweck einige Vorbereitungen benötigt. Jeder Tisch, der anschließend in Partnerarbeit genutzt wird, wird mit demselben Equipment ausgestattet. Hierbei handelt es sich um zwei Pipetten (100  $\mu\text{L}$ ; Eppendorf Research®), eine Box mit Pipettenspitzen (10–200  $\mu\text{L}$ ), eine Styroporbox mit Eis und ein Rack für Reaktionsgefäße (1 mL und PCR-Cups), für den Kriminalfall benötigte Flüssigkeiten (vier DNA-Proben, einen PCR-Primer-Mix und einen Mastermix) sowie zwei 1

mL Reaktionsgefäße mit farblosem und gefärbtem Wasser. Des Weiteren benötigt wird ein Müllgefäß, zwei Folienstifte und ein Versuchsverlaufsschema. Zwei Gruppen teilen sich zudem eine Tischzentrifuge und Zentrifugenzubehör (Rotorblätter je nach eingesetztem Volumen) und eine Box mit Reaktionsgefäßen (1 mL und PCR-Cups). Handschuhe verschiedener Größen werden an einer zentralen Stelle im Raum zur Verfügung gestellt.

An einer Stelle des Raums wird die Station der PCR-Cycler (Thermocycler miniPCR®) und der dazugehörigen iPads aufgebaut. Jedem Thermocycler wird ein iPad zugewiesen.

In einem weiteren Klassenzimmer werden die Tische zum Gießen der Agarose-Gele vorbereitet. Hierbei teilen sich zwei Gruppen einen Tisch, auf welchem eine Waage ( $d = 0,01$ ), ein Weithals-Erlenmeyerkolben, ein Messzylinder (100 mL), ein Spatel, die Reagenzien für das Agarosegel (TBE-Puffer, Agarose, gel stain dye), zwei Pipetten (10  $\mu$ L und 100  $\mu$ L) mit entsprechenden Spitzen und eine Elektrophoresekammer (blueGel™ (miniPCR)) vorliegen.



Abbildung 21: Räumlichkeiten zur praktischen Durchführung des Schülerversuchs.  
Links: Raum zur Durchführung der PCR und der damit verbundenen vorbereitenden Schritte. Rechts: Raum zur Durchführung der Agarose-Gelelektrophorese.

Des Weiteren müssen beide Räume über eine Projektionsfläche bzw. -medium (Beamer oder Bildschirm) und die Lehrperson über ein digitales Endgerät verfügen.

### **Durchführung**

Bevor die Schüler\*innen mit dem eigentlichen Kriminalfall und seinen Besonderheiten in Kontakt kommen, erfolgt durch die Lehrperson eine Sicherheitsbelehrung. Hierbei werden die wesentlichen Besonderheiten des Übungssaales der Schule, die Grundlagen der Sicherheit und die persönlichen Schutzmaßnahmen besprochen. Um in das praktische biotechnologische Arbeiten einzuführen, wird anschließend der Umgang mit einer Mikroliterpipette (Eppendorf Research®) und den in der Biotechnologie genutzten Reaktionsgefäßen (Eppendorf PCR Tubes®) eingeübt. Während sämtlicher Arbeiten mit Reagenzien werden von den

Schüler\*innen Handschuhe getragen. Dies soll dazu dienen, ihnen die Prinzipien des sauberen und sicheren Arbeitens näherzubringen.

Anschließend wird den Schüler\*innen der Sachverhalt des „Kriminalfalls“ vorgestellt. Der „Kriminalfall“ handelt davon, dass die Schulsekretärin Clara Müller (im Original Missy Baker) vermisst wird. Die Schüler\*innen werden durch einen Zeitungsartikel über den Sachverhalt aufgeklärt und von der Polizei aufgefordert zu untersuchen, ob Clara Müller im Werkstattkombi oder im Lieferwagen entführt wurde. Dies ist in Abbildung 22 dargestellt. Dabei wird ihnen im Zeitungsartikel der Umstand aufgezeigt, dass Clara Müller an einer genetisch bedingten Erkrankung leidet (miniPCR, 2022).

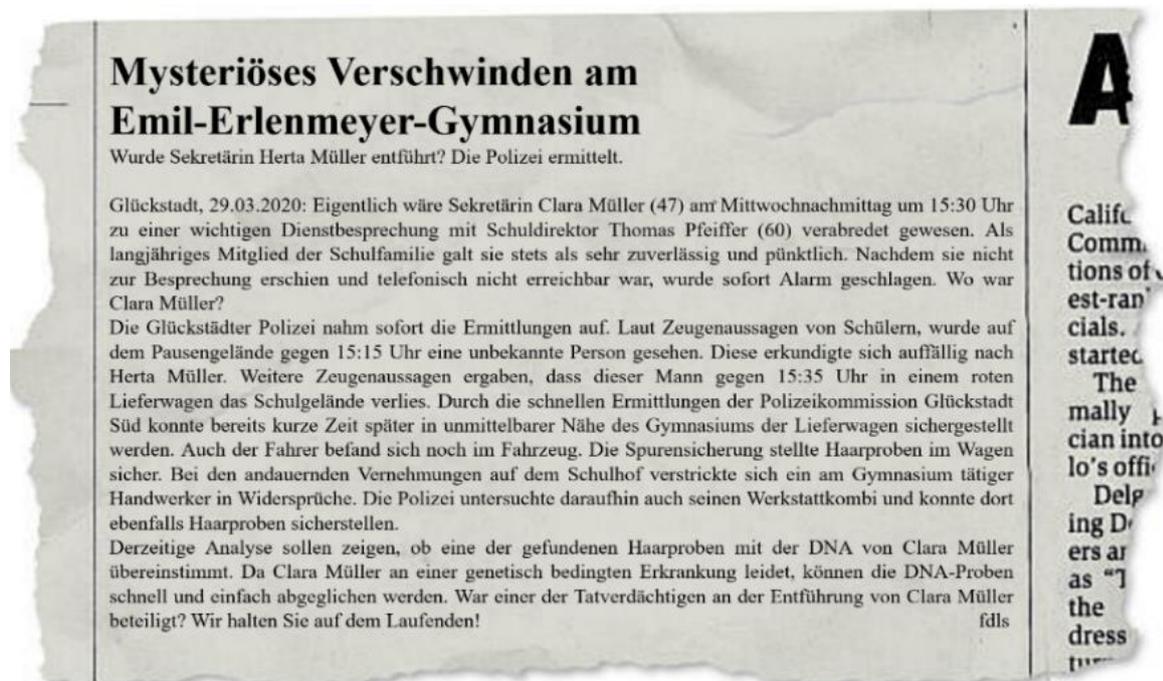


Abbildung 22: Zeitungsartikel zum Verschwinden von Clara Müller  
(Verändert nach Schöppner und Nerdel (2015), S.32)

Die Praktikumsleitung bespricht anschließend mit den Schüler\*innen die genetische Erkrankung der Mukoviszidose (= zystischen Fibrose (*Cystic fibrosis*)). Hierbei handelt es sich um eine autosomal-rezessiv vererbte Erkrankung, welche Organe wie die Lunge, Atemwege, Schweißdrüsen und Bauchspeicheldrüse betrifft (Elborn, 2016; Riordan et al., 1989) (s. 3.3 *Zystische Fibrose*). Durch Mutation kommt es zu einer Fehlfunktion des CFTR-Proteins und einer damit verbundenen verminderten Chloridionen-Leitfähigkeit an der apikalen Membran der Epithelzellen durch den Kanal (Riordan et al., 1989). Dadurch ist der Wasser- und Ionen-transport in den Epithelzellen gestört und somit bilden sich visköse Sekrete. Die häufigste Mutation des CFTR-Gens ist eine Deletionsmutation, die auf Proteinebene zum Verlust der Aminosäure Phenylalanin führt (Elborn, 2016). Da das Fehlen einer einzelnen Aminosäure mit einer in der Schule zur Verfügung stehenden Agarose-Gelelektrophorese

nicht analysiert werden kann, wird der Deletionsabschnitt auf 400 bp (Basenpaaren) erweitert, sodass sich für das intakte CFTR-Gen eine Länge von 800 bp und für das mutierte Gen mit der Deletion eine Länge von 400 bp ergibt (miniPCR, 2022).

Nachdem die theoretischen Grundlagen des CFTR-Gens und seiner Mutation besprochen wurden, wird im Lehrer-Schüler-Gespräch (Plenum) das vorausgegangene Wissen zur PCR wieder aufgegriffen und auf den aktuellen Fall angewandt. Anschließend beginnen die Schüler\*innen in Gruppen von zwei Personen mit der Herstellung der PCR-Proben. Hierzu stehen ihnen die DNA-Proben der Tatverdächtigen aus dem Lieferwagen (Probe A) und dem Werkstattkombi (Probe B) zur Verfügung. Zudem werden zur besseren Überprüfbarkeit der Ergebnisse am Agarosegel die Proben G (Vergleichsprobe eines intakten CFTR-Gens) und D (Vergleichsprobe mit mutiertem CFTR-Gen) mit untersucht. Nach einer gemeinsamen Vorbesprechung bzgl. des Ziels der nachfolgenden Tätigkeiten und mithilfe eines laminierten Versuchsverlaufsschemas pipettieren die Schüler\*innen die PCR-Ansätze, wie in Tabelle 13 zu sehen, in entsprechend beschriftete Reaktionsgefäße (Eppendorf PCR Tubes®).

Tabelle 13: Pipettierschema für die vier Reaktionsansätze des Kriminalfalls

<i>Pipettierschema PCR-Ansatz</i>				
	<b>A</b>	<b>B</b>	<b>G</b>	<b>D</b>
DNA-Probe	DNA A	DNA B	Kontrolle "gesund"	Kontrolle "krank"
	5 µL	5 µL	5 µL	5 µL
Primer Mix <sup>a</sup>	10 µL	10 µL	10 µL	10 µL
Master Mix <sup>b</sup>	15 µL	15 µL	15 µL	15 µL
Gesamtes Volumen	30 µL	30 µL	30 µL	30 µL

a. Crime Lab Primer Mix, Load Ready™ (miniPCR™, USA): Primervorwärts, Primerrückwärts

b. EZ PCR Master Mix (2x), Load Ready™ (miniPCR™, USA): PCR-Puffer, Mg<sup>2+</sup> Ionen, Taq-Polymerase, dNTPs, DNA-Auftragspuffer (2x)

Anschließend werden die Reagenzien in den einzelnen Ansätzen durch Auf- und Abpipettieren vorsichtig gemischt. Nach einer kurzen Einführung der Lernenden in die Nutzung einer Tischzentrifuge werden die Proben für ca. 3 Sekunden zentrifugiert (Abbildung 23).



Abbildung 23: Schüler\*innen beim Pipettieren der PCR-Proben in der Lernumgebung Labor

Die zentrifugierten PCR-Proben werden von den Gruppen jeweils in einen Thermocycler (miniPCR®) gestellt und anschließend das passende PCR-Programm über die Nutzeroberfläche des Thermocyclers am iPad eingegeben und gestartet. Dieses ist nachfolgend in Tabelle 14 dargestellt (miniPCR, 2022).

Tabelle 14: PCR-Protokoll zum Kriminalfall

<i>PCR-Protokoll</i>		
	<b>Temperatur [°C]</b>	<b>Zeit [sec]</b>
Initiale Denaturierung	94	30
Denaturierung	94	5
Annealing	57	5
Extension	72	8
Finale Extension	72	30
Anzahl an Zyklen	30	

Der Thermocycler ist mit dem iPad verbunden, auf welchem nun in Echtzeit die PCR-Reaktion schematisch auf Teilchenebene verfolgt werden kann (s. Abbildung 24). Nach etwa 40 Minuten ist die PCR abgeschlossen. Dieser Wert ergibt sich aus den angesetzten 30 Zyklen mit je einer Dauer von 18 Sekunden und unter Einberechnung der initialen Denaturierung, den Änderungen der Temperatur zwischen den Schritten, sowie der finalen Extension, wie in Tabelle 14 beschrieben. Die Reaktionsgefäße (Eppendorf PCR Tubes®) werden aus dem Thermocycler entfernt und kurz auf Eis gelagert.



Abbildung 24: Durchführung des Schülerversuchs in Präsenz.

Links: PCR-Cycler (Thermocycler miniPCR®) mit iPad. Rechts: Erklärung zur praktischen Durchführung der Agarose-Gelelektrophorese durch die Lehrperson.

In dieser Zeit wird die Funktionsweise der Agarose-Gelelektrophorese gemeinschaftliche wiederholt, die praktische Herstellung besprochen und das Laden der Gele mit den DNA-Proben gezeigt (s. Abbildung 24). Zur Herstellung des 1,6%-igen Agarose-Gels werden 30,00

mL 1x TBE-Puffer mit 0,48 g Agarose versetzt (s. Tabelle 15). Durch kurzes Aufkochen in der Mikrowelle und mehrmaligem Schwenken entsteht eine homogene, klare Lösung. Diese wird gekühlt, bis sie Handtemperatur hat und wird anschließend mit 3  $\mu\text{L}$  *gel staining dye* (miniPCR™) versetzt. Die Agarose-Lösung wird in die Gel-Gießapparatur (blueGel™) überführt und der Probenkamm an die entsprechende Position gesetzt. Der Probenkamm ist ein Platzhalter für die später benötigten Geltaschen. Die vollständige Polymerisierung des Agarosegels dauert etwa 15 Minuten.

Tabelle 15: Zusammensetzung 1,6%-iges Agarosegel

<i>Agarose-Gel 1,6%</i>	
Agarose für DNA-Elektrophorese	0,48 g
1x TBE-Puffer	30 mL
DNA-Stain GelGreen® (nucleic acid gel stain)	3 $\mu\text{L}$

Das Agarose-Gel wird anschließend im vorhandenen Schlitten in die Elektrophoresekammer überführt. Bevor die Proben aufgetragen werden, wird das Agarosegel jeweils vorsichtig mit 30 mL 1x TBE-Puffer überschichtet. Jede Gruppe trägt ihre vier PCR-Proben (A, B, G und D) mit jeweils 15  $\mu\text{L}$  und 10  $\mu\text{L}$  eines DNA-Größenstandards auf das Gel auf. Der Deckel der Elektrophoresekammer (blueGel™) wird aufgelegt und dadurch der Stromkreislauf geschlossen. Mittels des *run*-Knopfs wird die Elektrophorese gestartet. Die Auftrennung der DNA-Proben während der Gelelektrophorese kann durch den *Glühlampe*-Knopf auf der Vorderseite des Geräts (blueGel™-Transilluminator) betrachtet werden (s. Abbildung 25). Nach ca. 20 Minuten, wenn sich die Banden des DNA-Größenstandards im Bereich 400–800 bp gut sichtbar voneinander getrennt haben, ist die Agarose-Gelelektrophorese beendet und die Apparatur wird ausgeschaltet.

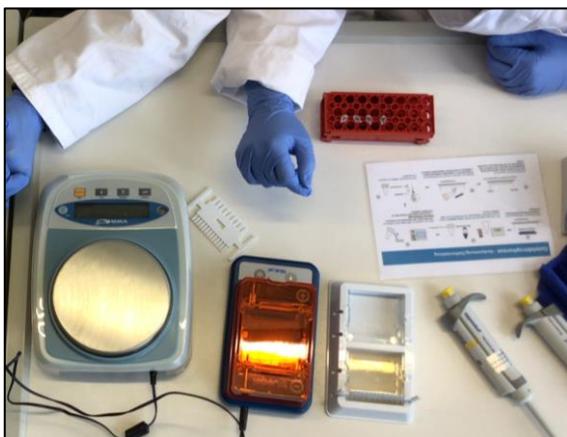


Abbildung 25: Durchführung der Agarose-Gelelektrophorese durch Schüler\*innen.

Links: Ausgehärtetes Agarosegel in der Gießapparatur ist bereit zur weiteren Verwendung. Rechts: Betrachtung des Ergebnisses während der Agarose-Gelelektrophorese während der Auftrennung der Proben.

Abschließend kann das Ergebnis betrachtet und mit dem Smartphone für genauere Betrachtungen und Analysen fotografiert werden. Durch den Vergleich der Banden der in der PCR vervielfältigten DNA-Proben mit den definierten Banden des DNA-Größenstandards kann die Größe der DNA-Fragmente in den aufgetrennten Proben geschätzt werden. Durch den Vergleich der Laufweiten der DNA-Proben A und B mit den Kontrollproben G und D ist nun eine Interpretation des Ergebnisses möglich (s. Abbildung 26). Dies ermöglicht einen begründeten Verdacht auf den Tathergang. Dieser wird von jeder Gruppe einzeln mit der Lehrperson besprochen und diskutiert.

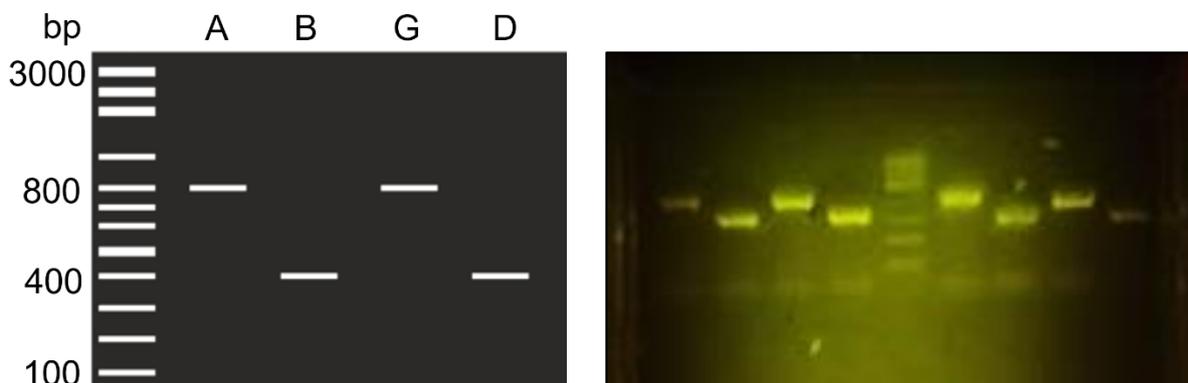


Abbildung 26: Ergebnis der Gelelektrophorese.

Links: Theoretische Darstellung (Längenstandard und die vier Proben).

Rechts: Originalbild (Zwei Gruppen, dazwischen der Längenstandard).

### 6.6.2 Digitale Lernumgebungen

In den digitalen Lernumgebungen wurde auf Audioelemente verzichtet, um die Umsetzung in der Schule sicherzustellen. An der Studie waren keine Schüler\*innen mit visuellen Beeinträchtigungen beteiligt, sodass keine Benachteiligungen entstanden sind.

Um die Lerninhalte (s. 6.4 *Konzeptionierung der Lernumgebungen*) den Schüler\*innen zugänglich zu machen, ist es nötig, auf der mebis Lernplattform, die Online-Lernplattform des bayerischen Kultusministeriums für bayerische Schulen, einen Kurs einzurichten. Die H5P Elemente können hier als „Interaktive Inhalte“ hochgeladen werden und stehen anschließend in dem erstellten mebis-Kurs als digitale Lernumgebung zur Verfügung. Der Kurs ist nur für Lehrkräfte sichtbar, die eine Berechtigung zur Nutzung des Kurses erhalten haben. Die Lehrkräfte können nur Schüler\*innen ihrer eigenen Schulen zu Kursen hinzufügen, sodass dies in der Studie von den entsprechenden Lehrkräften der Schulen übernommen wurde. Nachdem die Lehrkraft ihre Schüler\*innen dem Kurs hinzugefügt hat, ist dieser auf dem mebis-Schreibtisch der mebis Lernplattform des Schülers bzw. der Schülerin sichtbar. Alle bayerischen Schüler\*innen haben hierfür einen Zugang und können auf die Lernplattform ohne

weitere Anmeldungen und Bekanntgabe personenbezogener Daten zugreifen (mebis-Redaktion, 2023). Nach Anmeldung auf mebis wird den Schüler\*innen der Kurs „Ein Ausflug in die Forensik“ auf ihrem digitalen mebis-Schreibtisch angezeigt und kann genutzt werden.

### 6.6.2.1 Interactive Book

#### Vorbereitung

Die Lehrkraft führt die digitale Unterrichtseinheit zur Vermittlung des Fachwissens mit ihren Schüler\*innen durch. Bei der ihnen zur Verfügung stehenden Zeit handelt es sich um eine Unterrichtsstunde mit einem Zeitumfang von 45 Minuten. Hierbei übernimmt die Lehrkraft allein eine begleitende Tätigkeit. Sie zeigt den Schüler\*innen bei Bedarf, wo und wie sie die digitale Lerneinheit in mebis finden und weist auf wichtige organisatorische Sachen wie den zeitlichen Rahmen hin. Auch die Aufgabe des technischen Supports, sollten Probleme bei der Internetverbindung oder Ähnliches auftreten, übernimmt die Lehrkraft. Ein Eingreifen zur Klärung von fachlichen Problemen ist im Studiendesign nicht vorgesehen.

Um auf die digitale Lernumgebung zugreifen zu können, melden sich die Schüler\*innen über iPads auf der Website <https://www.mebis.bayern.de/> mit ihren persönlichen Zugangsdaten an. Anschließend können sie auf den Kurs „Ein Ausflug in die Forensik“ auf ihrem digitalen mebis-Schreibtisch zugreifen. Der Kurs untergliedert sich in die zwei Kursteile „Theorie“ und „Praxis im Labor“, wie in Abbildung 27 zu sehen. In diesem Fall wird die „Theorie“ von den Schüler\*innen geöffnet und kann anschließend bearbeitet werden.

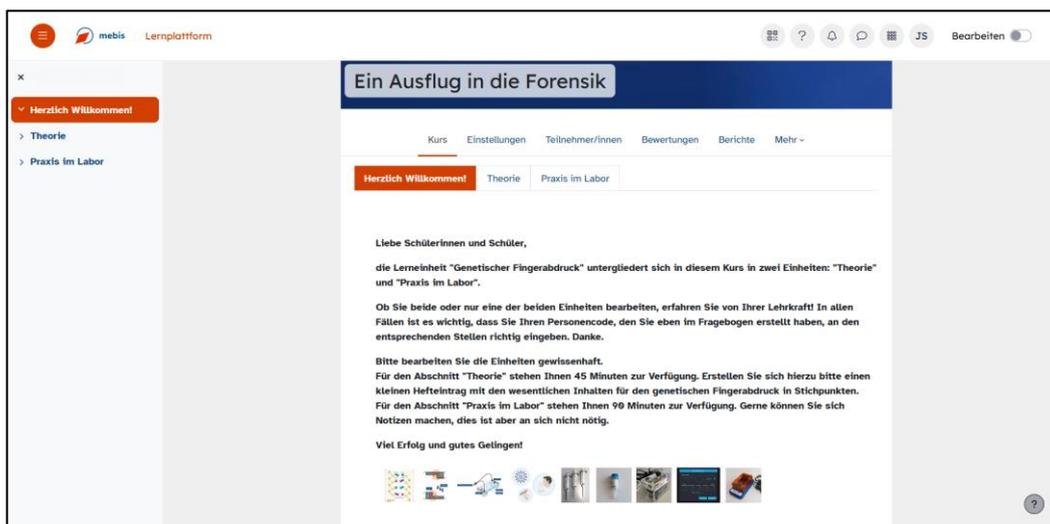


Abbildung 27: Startseite des mebis-Kurses „Ein Ausflug in die Forensik“

## Durchführung

Die Strukturierung der „Theorie“ in Form eines interaktiven Buches zielt darauf ab, Inhalte wie den Aufbau der DNA und die Funktionsweise einer Polymerase zu wiederholen, um anschließend neue Inhalte (PCR und Gelelektrophorese) darauf aufzubauen. Die Schüler\*innen haben in dieser Lernumgebung immer die Möglichkeit, Aufgaben erneut durchzuführen und zwischen den einzelnen Inhalten uneingeschränkt wechseln zu können. Im nachfolgenden Text in Klammern angegeben, sind die Bezeichnungen der genutzten Inhaltstypen (Content Types), die in der Lernumgebung erscheinen.

Zunächst wiederholen die Schüler\*innen eigenständig durch Texte und interaktive Elemente (Image Hotspots) die Grundlagen zum Aufbau der DNA, bevor sie ihr Wissen in einem Memory (Memory Game) spielerisch testen könnten. Hierbei werden nach dem Finden der passenden Paare erneut kurz die wesentlichen Aspekte der abgebildeten Darstellung aufgeführt (s. Abbildung 28).

Die Abbildung zeigt zwei Screenshot-Ausschnitte aus einer interaktiven Lernumgebung zur DNA.

Links: Ein Screenshot des 'Interactive Books' mit dem Titel 'DNA'. Die Seite enthält Text, Diagramme und eine Liste von Themen. Ein Diagramm zeigt die Struktur der DNA mit Beschriftungen: Chromosom, Zucker-Phosphat-Rückgrat, Base, Basen (Adenin, Thymin, Guanin, Cytosin). Ein Textblock erklärt die Zusammensetzung der DNA: 'Die DNA ist eine Nukleinsäure, die aus einzelnen Bausteinen, den Nucleotiden, aufgebaut ist [3]. Ein Nucleotid wiederum besteht aus einem Einfachzucker Desoxyribose, einer Phosphatgruppe und einer Base [4]. Viele aneinander gereihete Nucleotide ergeben zusammen einen Nucleotidstrang. Die einzelnen Nucleotide werden immer vom C3-Atom der Desoxyribose mit der Phosphatgruppe des nächsten Nucleotids verbunden. Die DNA ist aus zwei Nucleotidsträngen aufgebaut, die zusammen eine Doppelhelix formen [4]. Dabei liegen die Phosphat- und Zuckermoleküle außen, die Basen liegen innen und zeigen zueinander, analog zu den Sprossen einer Strickleiter [1]. Aufgrund spezieller Bindungseigenschaften der Basen sind die beiden DNA-Stränge antiparallel angeordnet [4].'

Rechts: Ein Screenshot eines Memory-Spiels mit dem Titel 'Nachfolgend finden Sie ein Memory. Suchen Sie zu jeder Abbildung die passende Bezeichnung!'. Das Spiel besteht aus einer 3x3-Matrix von Karten. Eine Karte zeigt die chemische Struktur von Guanin und Cytosin, die andere zeigt die chemische Struktur von Adenin und Thymin. Die Bezeichnungen sind: 'Guanin und Cytosin', 'DNA Rückgrat', 'Negativ geladene DNA'. Ein Textblock erklärt: 'Guanin und Cytosin sind komplementäre Basen. Sie bilden drei Wasserstoffbrücken miteinander aus.' Die Spielstatistik zeigt: 'Benötigte Zeit: 0:27' und 'Züge: 14'.

Abbildung 28: Beispiel des "Interactive Books" zur Theorie: Die DNA

Links: Ausschnitt zur Wiederholung des Aufbaus der DNA; Rechts: Ausschnitt aus dem Memory zum Aufbau der DNA

Bei der Erarbeitung der Inhalte zur Funktionsweise einer Polymerase wird das Vorwissen der Schüler\*innen aus vorangegangenen Unterrichtsstunden aktiviert (Image Slider) und erweitert. Die bekannten Inhalte werden um die Besonderheit der *Taq*-Polymerase ergänzt, welche bei der anschließenden PCR Anwendung findet. Durch einen kurzen „Richtig oder Falsch“ Test (True/False Question) können die Schüler\*innen ihr Wissen überprüfen und werden bei einem schlechten Ergebnis dazu aufgefordert, die Fachinhalte zu wiederholen (s. Abbildung 29).

Mutter- und Tochterzelle vorliegt. Ein für die DNA-Replikation wichtiges Enzym ist die Polymerase. Sie verknüpfen Nucleotide in einer durch den komplementären Nucleinsäure-Strang vorgegebenen Reihenfolge, der als Vorlage genutzt wird, zu einer Nucleinsäurekette [7].

Nachfolgend finden Sie einige Informationen zur Wirkung von DNA-Polymerasen auf molekularer Ebene in der Zelle.

Zu Beginn der Replikation muss die DNA-Doppelhelix durch das Enzym Topoisomerase entspiralisiert werden, bevor die Helicase die beiden über Wasserstoffbrücken verknüpften DNA-Stränge trennt. Somit liegt ein Stück der DNA offen zugänglich vor (Replikationsgabel) welches zwei gegensätzlich verlaufende DNA-Stränge

**Testen Sie Ihr Wissen!**  
Nachfolgend finden Sie ein Quiz. Bewerten Sie die Aussagen mit richtig oder falsch und drücken sie auf "Überprüfen".

Die DNA-Polymerase katalysiert die Spaltung der DNA vor der Replikation.  
 Wahr  Falsch

Du hast 0 von 1 Punkten erreicht.

Die DNA-Polymerase katalysiert die Spaltung der DNA vor der Replikation.  
 Wahr  Falsch

Du hast 1 von 1 Punkten erreicht.

Abbildung 29: Beispiel des "Interactive Books" zur Theorie: Polymerase  
Links: Ausschnitt zur Wiederholung: Bedeutung der DNA-Polymerase bei der Replikation;  
Rechts: Ausschnitt aus dem „Richtig oder Falsch“ Test

Anschließend beginnt die neue Lerneinheit der Stunde, welche die Grundlagen und die Funktionsweise der PCR erklärt (Course Presentation). Hier wird neben dem Bezug zum natürlichen Ablauf der Replikation auch die Dreiteilung (Annealing, Elongation und Termination) des Biotechnologieversuchs hervorgehoben. Durch abschließende Fragen zur PCR (Summary) erhalten die Schüler\*innen erneut Feedback zu ihrem Wissensstand und eine finale Zusammenfassung der wichtigsten Informationen (s. Abbildung 30).

PCR DNA-Fragmente vermehren. In der Forensik können Vaterschaftsnachweise oder DNA-Vergleiche zur Aufklärung von Kriminalfällen mittels einer PCR und anderen anschließenden Verfahren erfolgen (vgl. 6. Anwendungen) [1].

Wichtig ist es, dass die DNA im Labor unter physiologischen Bedingungen vervielfältigt wird. Analog zur Replikation wird der DNA-Doppelstrang zunächst geöffnet. Anschließend werden Primer als Start für die Vervielfältigung angelegt, um der DNA-Polymerase die Neusynthese eines komplementären DNA-Strangs zur DNA-Vorlage zu ermöglichen. [1]

Nachfolgend sind die einzelnen Schritte der Polymerase-Kettenreaktion dargestellt und beschrieben:

**3. Neusynthese / Elongation**  
Nun folgt die Synthese des neuen Strangs bei 72 °C. In diesem Schritt synthetisiert die Taq-Polymerase, vom Primer ausgehend in 5'-3'-Richtung, den komplementären DNA-Strang. [1],[8]

**Testen Sie ihr Wissen!**

Die PCR gründet sich auf drei wesentliche Schritte in ihrer Durchführung. ✓

Bei der PCR werden die natürlich ablaufenden Schritte der Replikation in leicht abgewandelter Form genutzt. ✓

Bei der Denaturierung werden die Wasserstoffbrücken zwischen den Basen gespalten. ✓

Beim Annealing lagern sich die Primer, entsprechend ihrer Basenabfolge, an die einzelsträngige DNA an. ✓

Nachfolgend sind jeweils [drei] Aussagen formuliert. Wählen Sie die Richtige aus! ✓ Fortschritt: 4/5  
✗ Falsche Antworten: 3

Mit der Elongation beginnt ein neuer Zyklus der PCR. ✗

Mit der Elongation endet ein Zyklus der PCR.

Abbildung 30: Beispiel des "Interactive Books" zur Theorie: PCR  
Links: Ausschnitt zum Inhalt: Bedeutung und Ablauf der PCR; Rechts: Ausschnitt aus der Zusammenfassung.

Anschließend wird den Schüler\*innen auch das zweite Verfahren, die Agarose-Gelelektrophorese, erläutert. Der Aufbau der Apparatur, die Besonderheit des Agarose-Gels und seine Wirkung werden hierbei u.a. angesprochen. Die Auswertung des Gels wird durch eine Zeitleiste (Timeline) dargestellt, welche interaktiv dazu einlädt, die örtlichen Veränderungen der Banden auf dem Gel im Verlauf der Agarose-Gelelektrophorese

nachzuvollziehen. Zur Überprüfung des Wissens wird in diesem Abschnitt ein Drag and Drop-Format in Textform gewählt (Drag the Words) (s. Abbildung 31).

**Timeline zur Agarose-Gelelektrophorese**

In nachfolgender Timeline können Sie den Fortschritt der DNA-Auftrennung im Agarosegel durch die Gelelektrophorese beobachten.

**20 Gel nach 20 Minuten**

... Nun lassen sich beim Längenstandard (DNA-Leiter) bereits einzelne Banden erkennen. Von einer klaren Trennung kann man allerdings noch nicht reden. ...

**Testen Sie ihr Wissen!**

Ziehen Sie die Wörter in die richtigen Felder und drücken Sie anschließend auf "Überprüfen".

Mit Hilfe der Gelelektrophorese können DNA-Gemische nach ihrer Größe aufgetrennt werden. Dabei macht man sich die negative  der DNA bei einem schwach basischen pH-Wert zu Nutze. Wird eine  an die Elektrophoresekammer angelegt, so wandern die DNA-Fragmente in unterschiedlicher Geschwindigkeit von der Kathode zur Anode. Die Wanderungsgeschwindigkeit wird von der  der DNA-Fragmente und der  des Agarose-Gels bestimmt. Zur Sichtbarmachung der farblosen DNA wird dem Agarosegel meist ein  zugesetzt. Dieser kann nach der Gelelektrophorese in UV-Licht sichtbar gemacht werden. Um eine Referenz für die Größe des gesuchten DNA-Fragments zu haben, lässt man eine Kontrollsubstanz (DNA-Längenstandard) mitlaufen. Darüber kann die Größe des DNA-Fragments annähernd bestimmt werden [1].

Überprüfen

Mit Hilfe der Gelelektrophorese können DNA-Gemische nach ihrer Größe aufgetrennt werden. Dabei macht man sich die negative  der DNA bei einem schwach basischen pH-Wert zu Nutze. Wird eine  an die Elektrophoresekammer angelegt, so wandern die DNA-Fragmente in unterschiedlicher Geschwindigkeit von der Kathode zur Anode. Die Wanderungsgeschwindigkeit wird von der  der DNA-Fragmente und der  des Agarose-Gels bestimmt. Zur Sichtbarmachung der farblosen DNA wird dem Agarosegel meist ein  zugesetzt. Dieser kann nach der Gelelektrophorese in UV-Licht sichtbar gemacht werden. Um eine Referenz für die Größe des gesuchten DNA-Fragments zu haben, lässt man eine Kontrollsubstanz (DNA-Längenstandard) mitlaufen. Darüber kann die Größe des DNA-Fragments annähernd bestimmt werden [1].

VII solltest du dir die Inhalte nochmal anschauen

3/5 Lösung anzeigen Wiederholen

Abbildung 31: Beispiel des "Interactive Book" zur Theorie: Agarose-Gelelektrophorese  
Links: Ausschnitt zum Inhalt: Timeline zur Darstellung des Gelverlaufs während der Durchführung; Rechts: Kurzzusammenfassung des Lerninhalts durch eine Drag and Drop-Aufgabe.

Als letzte Lerneinheit erhalten die Schüler\*innen einen kurzen Überblick darüber, wo die PCR und die Gelelektrophorese Anwendung finden (s. Abbildung 32). Hierbei soll die Bedeutung des Verfahrens und der Lebensweltbezug der Schüler\*innen – zumindest zu den Ergebnissen, die daraus resultieren – noch einmal hervorgehoben werden.

**Anwendungen**

Der genetische Fingerabdruck

- Personencode
- DNA
- Polymerase
- PCR
- Agarose-Gelelekt...
- Anwendungen
- Periodische Virenre...

Zusammenfassung und Einsenden

**Anwendungen**

**Nachweis von SARS-CoV-2**

Der Nachweis des Coronavirus SARS-CoV-2 kann mit Hilfe eines PCR-Tests erfolgen. Für den Test wird beispielsweise eine Probe aus den Schleimhäuten der Atemwege entnommen, welche menschliche Zellen enthält. Ist ein Mensch mit dem Coronavirus infiziert, wird mittels der PCR nicht die menschliche DNA, sondern die Erbinformation des Virus vervielfältigt [16]. Durch die Vervielfältigung können auch geringe Mengen an Virusmaterial nachgewiesen werden.

Abbildung 32: Beispiel des "Interactive Books": Anwendungsbereiche der PCR und Gelelektrophorese  
Links: Überblickansicht der angeführten Anwendungsbereiche mit Image Hotspots;  
Rechts: Information des Image-Hotspots der SARS-CoV-2-Abbildung

Am Ende der digitalen Lernumgebung „Theorie“ erhalten die Schüler\*innen eine Punkteübersicht über die von ihnen bearbeiteten Aufgaben und die dabei erreichten Punkte.

### 6.6.2.2 Digitales Labor

#### Vorbereitung

Die Vorbereitungen für das digitale Labor entsprechen denen des Interactive Books, wie unter 6.6.2.1 *Interactive Book* beschrieben. Der mebis-Kurs untergliedert sich in die zwei Kursteile „Theorie“ und „Praxis im Labor“, wie in Abbildung 27 zu sehen. In diesem Fall wird der „Praxis im Labor“ Teil von den Schüler\*innen geöffnet und kann anschließend bearbeitet werden. Bei der den Schüler\*innen zur Verfügung stehenden Zeit handelt es sich um zwei Unterrichtsstunden mit einem gesamten Zeitumfang von 90 Minuten. Die Schüler\*innen haben in dieser Lernumgebung immer die Möglichkeit, Aufgaben erneut durchzuführen, können aber nicht uneingeschränkt zwischen den einzelnen Inhalten wechseln. Im nachfolgenden Text in Klammern angegeben, sind die Bezeichnungen der genutzten Inhaltstypen (Content Types), die in der Lernumgebung erscheinen. Aufgrund des immensen Umfangs an interaktiven Aktivitäten werden diese nachfolgend beispielhaft genannt.

#### Durchführung

Zunächst betreten die Schüler\*innen das Labor und können wählen, ob sie sich erneut mit den theoretischen Grundlagen vertraut machen wollen oder mit der Laborarbeit bzw. den Aufträgen der Polizei beginnen möchten (s. Abbildung 32).



Abbildung 33: Beispiel des „Digitalen Labors“: Links: Willkommensseite; Rechts: Entscheidungsfragen

Pro Entscheidung, welche die Schüler\*innen treffen müssen, stehen immer mehrere Antwort- bzw. Wahlmöglichkeiten zur Verfügung (s. Abbildung 32). Bei der Wahl einer nicht unbedingt sinnvollen bzw. falschen Option erhalten die Schüler\*innen Rückmeldung hierzu und werden auf eine sinnvollere Alternative hingewiesen bzw. dazu aufgefordert, ihre Wahl zu überdenken (s. Abbildung 34).

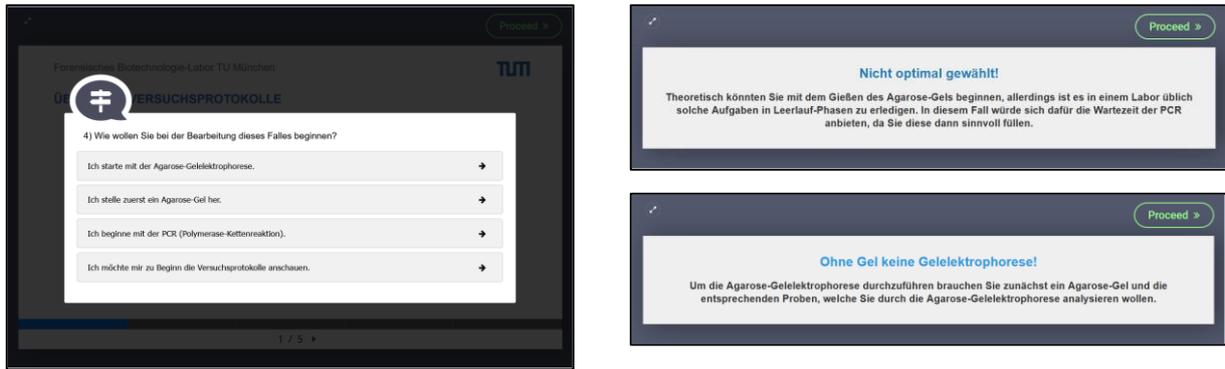


Abbildung 34: Beispiel des „Digitalen Labors“: Entscheidungssituation zu Beginn der Fallbearbeitung  
Links: Entscheidungsfragen; Rechts: Feedback zu falschen bzw. ungünstigen Antworten

Zunächst wird die PCR vorbereitet (Course Presentation). Die Schüler\*innen machen sich mit den Materialien, die im Labor zur Verfügung stehen vertraut und suchen sich anschließend die benötigten Materialien zusammen (Drag and Drop, Videos ...) (s. Abbildung 35).

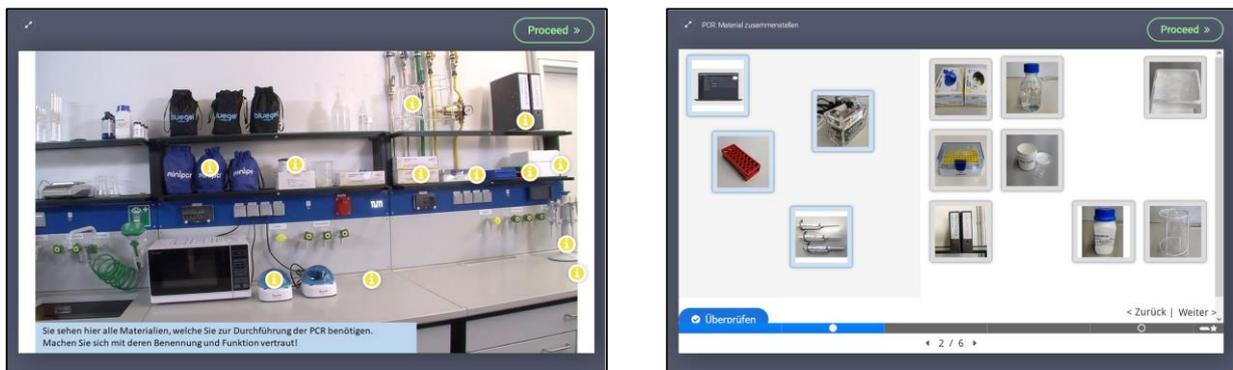


Abbildung 35: Beispiel des „Digitalen Labors“: Vorbereitung der PCR.  
Links: Laborübersicht mit den zur Verfügung stehenden Materialien für die PCR  
Rechts: Drag and Drop Aufgabe der für die PCR notwendigen Materialien

Hierbei können sie jederzeit wieder auf die vorgegebenen Versuchsprotokoll bzw. Versuchsverlaufsschema zugreifen. Durch kurze Fragen (Summary) im Verlauf der Einheit werden sie immer wieder auf wichtige Aspekte zum Umgang bei der praktischen Arbeit hingewiesen und aufgefordert Abläufe nachzuvollziehen. Die Schüler\*innen lernen im Anschluss daran, durch Videos (Interactive Video) und Zuordnungsaufgaben (Drag and Drop, Summary ...) die Funktionsweise einer Mikroliterpipette kennen (s. Abbildung 36), bevor sie die Proben für die PCR „pipettieren“ (s. Abbildung 37). Anschließend können sie sich den Umgang mit einer Zentrifuge erarbeiten, bevor die PCR-Proben in den Thermocycler gestellt werden.

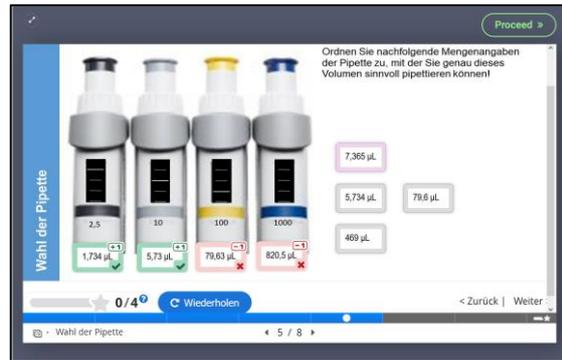


Abbildung 36: Beispiel des „Digitalen Labors“: Nutzung von Mikroliterpipetten  
 Links: Erklärung wesentlicher Einstellungen an einer Mikroliterpipette  
 Rechts: Drag and Drop Aufgabe zur Wahl der richtigen Pipette

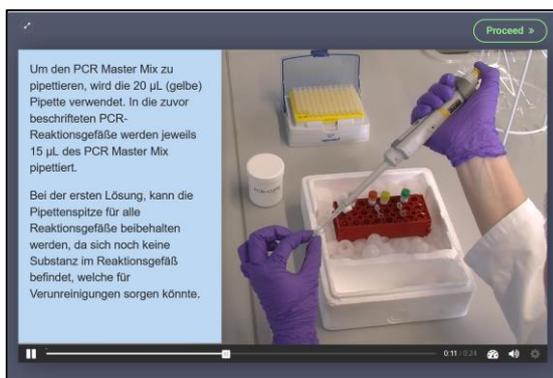


Abbildung 37: Beispiel des „Digitalen Labors“: Pipettieren der PCR-Proben  
 Links: Video zum Pipettieren der PCR-Proben  
 Rechts: PCR-Proben werden in den Thermocycler überführt

Nachdem das Programm für den Thermocycler durch die Schüler\*innen richtig eingestellt wurde (Course Presentation, Drag and Drop), kann der Temperaturverlauf in Abhängigkeit von der Zeit für die gesamte PCR in Zeitraffer beobachtet werden (Video). Zudem erscheinen die PCR-Schritte *Denaturation*, *Annealing* und *Elongation* auf Teilchenebene je nach aktuell vorherrschender Temperatur (s. Abbildung 38).

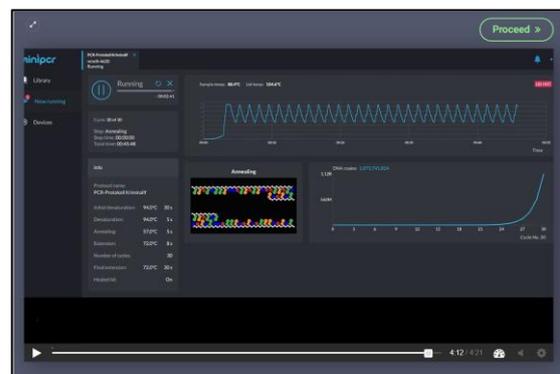
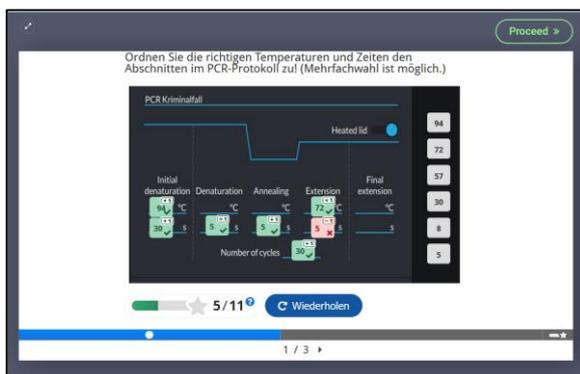


Abbildung 38: Beispiel des „Digitalen Labors“: PCR-Protokoll  
 Links: PCR-Protokoll per Drag and Drop erstellen  
 Rechts: Zeitrafferaufnahme PCR-Verlauf

Analog zur realen zeitlichen praktischen Durchführung im Labor wird das Agarosegel zur Trennung der DNA-Proben hergestellt. Die Schüler\*innen entscheiden sich erneut für die dazu benötigten Materialien (Course Presentation, Drag and Drop) und überprüfen ihr Wissen zur praktischen Herstellung des Agarosegels (Summary), bevor ihnen die Vorgehensweise visuell nahegebracht wird (Interactive Video) (s. Abbildung 39).

Das hergestellte Gel wird in die Agarose-Gelelektrophoresekammer überführt (Interactive Video), nachdem der Aufbau der Agarose-Gelelektrophoresekammer erklärt wurde und anschließend mit TBE-Puffer überschichtet.

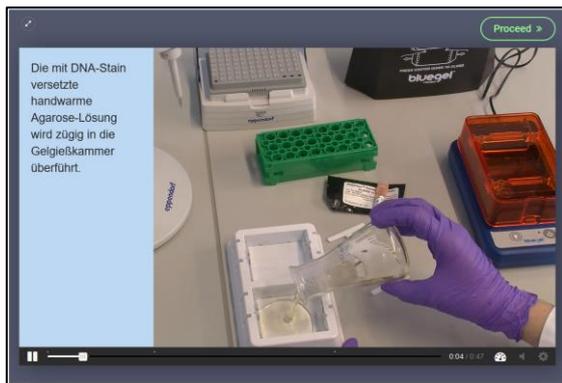


Abbildung 39: Beispiel des „Digitalen Labors“: Agarosegel-Herstellung; Links: Gelgießen; Rechts: Überführung des gehärteten Agarosegels in die Gelelektrophoresekammer

Bevor die PCR-Proben auf das Gel aufgetragen werden (Interactive Video), wird notiert, welche DNA-Probe in welcher Tasche aufgetragen wird. Zudem wird die Bedeutung des 1. und 2. Druckpunktes der Pipette für die Auftragung der DNA-Proben am Gel erklärt (Course Presentation, Video). Anschließend wird die Gelelektrophoresekammer geschlossen und die Agarose-Gelelektrophorese gestartet (s. Abbildung 40).

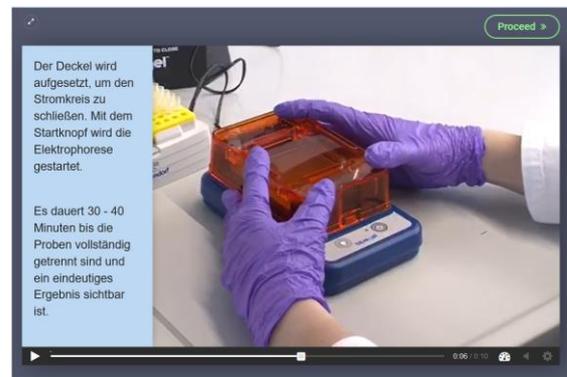
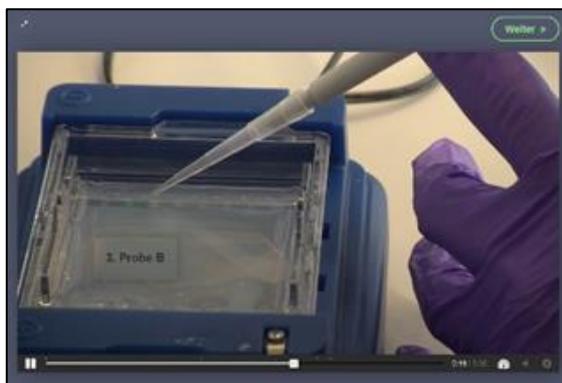


Abbildung 40: Beispiel des „Digitalen Labors“: Agarose-Gelelektrophorese  
Links: Auftrag der DNA-Proben auf das Agarose-Gel am Beispiel von Probe B  
Rechts: Aktivierung der Agarose-Gelelektrophorese

In einer Zeitrafferaufnahme kann anschließend die Auftrennung der Proben im Gel betrachtet werden (Video). Das Ergebnis wird von den Schülern\*innen gedeutet (Summary), bevor eine Erklärung (Image Hotspots) der einzelnen Banden zur Verfügung steht (s. Abbildung 41).

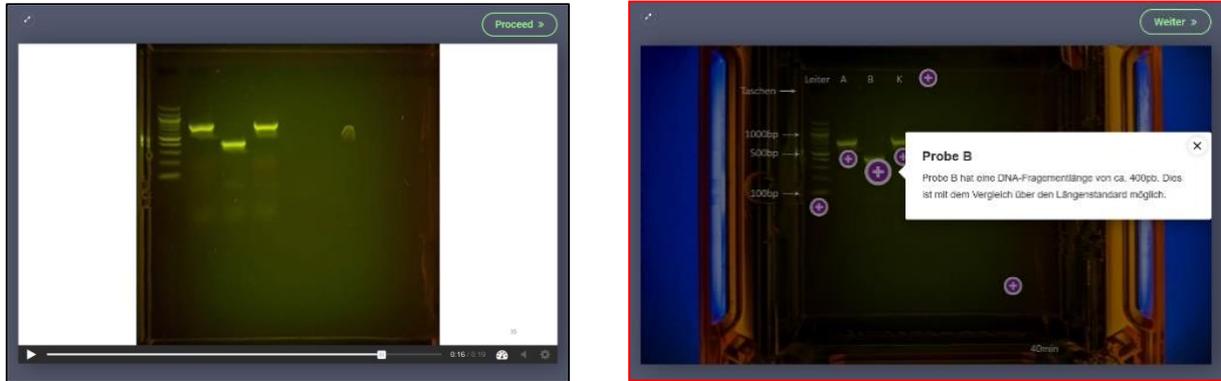


Abbildung 41: Beispiel des „Digitalen Labors“: Ergebnis des „Kriminalfalls“

Links: Bild der DNA-Banden während der Agarose-Gelelektrophorese

Rechts: Bild der DNA-Banden der Agarose-Gelelektrophorese mit Erklärungen

Hiermit sind der Kriminalfall und seine Durchführung im digitalen Labor abgeschlossen. Die Schüler\*innen erhalten keine Gesamtpunktezahl bzgl. ihrer richtig erfüllten Aufgaben.

## 6.7 Statistische Analysemethoden

Zur Vorbereitung der statistischen Auswertung werden die Daten entsprechend gängiger, vordefinierter Kriterien (Bearbeitungszeit, Vollständigkeit, Ratewahrscheinlichkeit, ...) überprüft. Datensätze, die nicht eindeutig einem Prä- oder Post-Test zugeordnet werden können oder eine massiv unterdurchschnittliche Bearbeitungszeit von weniger als fünf Minuten aufweisen, werden von der Analyse ausgeschlossen. Gleiches gilt für Datensätze, bei denen die Ratewahrscheinlichkeit über alle drei Messzeitpunkte unterschritten wurde. Ebenso wurde das Fachwissensitem Item 09 entfernt, weil es einen Schwierigkeitsindex von  $P_{09} < 20$  über alle drei Testzeitpunkte hinweg aufweist. Dies bedeutet, das Item ist durchgehend zu schwer (Bühl, 2019). Im nächsten Schritt werden die aus dem Fragebogen erhaltenen Daten Raschskaliert. Hierfür wird das eindimensionale Raschmodell verwendet, das in der Item-Response-Theorie (IRT) beschrieben ist (s. 6.7.1 *Das eindimensionale Rasch-Modell*). Mithilfe des Raschmodells können Itemschwierigkeiten und Personenfähigkeiten von Teilnehmer\*innen berechnet werden. Die mit der IRT ermittelten Personenfähigkeiten berücksichtigen die Schwierigkeit einzelner Testitems und bieten dadurch eine differenziertere und präzisere Messung individueller Fähigkeiten als einfache Rohwerte. Durch die Skalierung der Fähigkeiten und die Möglichkeit adaptiver Testung ermöglichen sie einen fairen Vergleich zwischen Personen über verschiedene Testformen und -bedingungen hinweg. Durch die Ankerung der Itemschwierigkeiten über die Fragebögen hinweg werden die Personenfähigkeiten miteinander vergleichbar gemacht. Nach der Raschskalierung der Daten werden die erhaltenen Personenfähigkeiten, anstelle der Raw Scores genutzt, um die Auswertung der Ergebnisse gemäß den Prinzipien der klassischen Testtheorie (s. 6.7.2 *Klassische Testtheorie*) durchzuführen. Die KTT ermöglicht eine umfassende Analyse der Personenfähigkeiten und erlaubt es, Schlussfolgerungen aus den gewonnenen Daten zu ziehen. Die Aufbereitung der Datensätze erfolgte mit IBM SPSS 28 (IBM Corp. Released, 2021) und die Analysen der IRT mit dem Programm Winsteps (Linacre, 2023b). Die Berechnungen der KTT wurden mit IBM SPSS 28 (IBM Corp. Released, 2021) und mit R (R Core Team, 2022), mit den packages *rstatix* (Kassambara, 2023) und *tidyr* (Wickham et al., 2023) und JASP 0.17.3 (JASP Team, 2023), durchgeführt.

### 6.7.1 Das eindimensionale Rasch-Modell

Die Item-Response-Theorie (IRT) ist eine statistische Modellierungsmethode, genauer gesagt ein probabilistisches Modell, die sich mit der Analyse von Antwortmustern in Tests und Fragebögen beschäftigt. Sie postuliert, dass die Wahrscheinlichkeit, wie eine Person auf eine bestimmte Frage oder ein Item antwortet, von den individuellen Fähigkeiten der Person und

den Eigenschaften des Items abhängt. (Goldhammer & Hartig, 2020; Sälzer, 2016) Die Modellgleichung für das Rasch-Modell ist in Formel 1 dargestellt (Strobl, 2015).

$$P(U_{ij} = 1 | \theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \quad (1)$$

Lösungswahrscheinlichkeit  $P$  („Probability“) hängt von der Personen-Fähigkeit  $\theta_i$  und der Aufgaben-Schwierigkeit  $\beta_j$  ab (Strobl, 2015). Auf eine detailliertere Erklärung der Modellgleichungen des verwendeten Rasch-Modells wird verzichtet. An dieser Stelle sei auf bekannte Fachliteratur verwiesen: z.B. Rost; Rost (2004, 2006) und Strobl (2015).

Die in dieser Arbeit genutzt Software „Winsteps“ nutzt den JMLE-Schätzlogarithmus (Joint Maximum Likelihood Estimation) zur Modellierung des Antwortverhaltens (Boone & Staver, 2020). Dieses wird in Form von grafischen Darstellungen, auch als Item Characteristic Curve (ICC) bezeichnet und dargestellt (Boone & Staver, 2020; Goldhammer & Hartig, 2020).

### ***Analyse der Itemschwierigkeiten und Personenfähigkeiten***

Um das Rasch-Modell für die Nutzung der Daten anzuwenden, müssen bestimmte Kriterien erfüllt sein. Der **Itemfit** (vgl. Boone et al., 2014) bezieht sich auf die Anpassung der einzelnen Items an die erwarteten Antworten, sowie an die Fähigkeiten der Personen im Durchschnitt (Infit) und an den Extremen (Outfit). Technisch gesehen sind Outfit und Infit Chi-Quadrat-Statistiken. In einer Rasch-Analyse ist das Kriterium die Assoziation zwischen dem Modell und den Daten, d.h. wie gut die Daten mit dem Rasch-Modell übereinstimmen. Die allgemeine Idee von Outfit besteht darin, dass es sich um eine Fit-Statistik handelt, die empfindlich auf Ausreißer reagiert (z.B. Vermutungen oder gedankenlose Fehler), während der Infit sich weniger auf Ausreißer, sondern mehr auf Antworten in der Nähe einer bestimmten Itemschwierigkeit (oder Personenfähigkeit) konzentriert (Boone et al., 2014). Somit ist bei der Überprüfung der Itemfits dem Outfit-Wert eine höhere Bedeutung zuzuordnen, da er empfindlich auf unerwartetes Verhalten von Personen bei Items reagiert, die weit vom Messniveau der Person entfernt sind (Linacre, 2023a). Der MNSQ (Mean Squared Residual) sollte einen Wert zwischen 0,5 und 1,5 haben, wobei ein Wert von 1,0 optimal wäre (Wright & Linacre, 1994). Werte über 1,5 deuten auf völlig unerwartetes Antwortverhalten hin, während Werte unter 0,5 einen zu perfekt passenden Fit anzeigen. "Underfit"-Items mit Werten unter 0,5 sind wenig ergiebig für die Messung, führen aber auch nicht zu einer Verschlechterung. Bei nicht passenden MNSQ-Werten wird der Z-Standardwert betrachtet, der zwischen -1,96 und 1,96 liegen sollte. Bei nicht passenden Fits werden die Personenantworten genauer analysiert, insbesondere die Z-Residuen der Antworten. Große Z-Residuen zeigen, dass eine Person mit niedriger Personenfähigkeit ein schwieriges Item unerwartet beantwortet hat.

Große negative Z-Residuen zeigen an, dass eine Person mit hoher Personenfähigkeit ein einfaches Item unerwartet falsch beantwortet hat. Dieser Prozess wird wiederholt, bis die Fits zufriedenstellend sind. Dabei darf die Grundgesamtheit der Daten jedoch nicht verändert werden. Sehr unerwartete Antworten mit sehr großen (positiven oder negativen) Z-Residuen können aus dem Fragebogen entfernt werden. Eine weitere Möglichkeit besteht darin, die vollständigen Antworten einer Person auszuschließen. Items, die weiterhin nicht den Fit-Vorgaben entsprechen, müssen gegebenenfalls von der Untersuchung ausgeschlossen werden, um eine genaue Messung zu gewährleisten. (Boone et al., 2014) Dem Rasch-Modell liegt die zentrale Annahme zugrunde, dass die Personen- und Item-Parameter auf einer gemeinsamen latenten Dimension liegen und somit **Eindimensionalität** vorliegt (Strobl, 2015) Dies wird mittels PCAR (Principal Component Analysis of Residuals), der Hauptkomponentenanalyse der Residuen überprüft (Boone & Staver, 2020; Wright & Stone, 1979). Hierbei wird gezielt nach Auffälligkeiten in dem Datenteil gesucht, der nicht mit dem Raschmodell übereinstimmt. Es wird eingehend überprüft, ob bestimmte Itemgruppen ähnliche Muster der Unerwartetheit aufweisen. Sollte dies der Fall sein, so besteht die Wahrscheinlichkeit, dass diese Items auch ein bedeutendes gemeinsames Merkmal besitzen, welches nicht als zufälliges Rauschen gedeutet werden kann. Ist der Eigenwert im 1. Kontrast größer 2, wird mithilfe eines Kontrastdiagramms und der Überprüfung der Items (Wortlaute und Inhalt) nachvollzogen, ob es sich um mehrere Dimensionen handeln kann. Für die Raschanalyse muss zudem eine lokale stochastische **Unabhängigkeit der Items** vorliegen, d.h. dass die Antwort einer Person auf ein Item nicht davon abhängt, wie die gleiche Person andere Items des Tests beantwortet hat (Goldhammer & Hartig, 2020; Moosbrugger & Kelava, 2020). Dazu werden die Residualkorrelationen zwischen den Items untersucht, um die lokale Unabhängigkeit zu überprüfen. Die Korrelation ist gegeben, wenn die Antwort auf ein Item von der Antwort auf ein anderes Item innerhalb desselben Wertes abhängig ist. In einem idealen Raschmodell sind die Items lokal unabhängig (Itemkorrelationen:  $-0,7 < r < 0,7$ ) und nur der latente Faktor erklärt die Beziehung zwischen den Items (Linacre, 2009; Semmelink & Maree, 2023). Ein möglicher **Decken- oder Bodeneffekt**, also das Vorhandensein von Befragten, die sich am oberen (Maximalscore) und unteren (Minimalscore) Ende der Skala befinden, muss überprüft werden. Idealerweise liegt die Anzahl an Teilnehmern unter 5 %, die eine dieser Effekte aufweisen (Boone et al., 2014; Davis & Boone, 2021). Analog zum Itemfit wird auch der **Personenfit** (Misfit der Personenfähigkeiten) überprüft. Ein Misfit führt nicht zwingend zum Ausschluss der Person aus dem Datensatz, da eine Personenmisfit deutlich weniger Gewicht für die Analyse hat als ein Itemmisfit. Das **Targeting** (Passung des Tests zum Fähigkeitsniveau der Stichprobe) kann bewertet werden, indem die Differenz zwischen dem durchschnittlichen Personenmaß und dem durchschnittlichen Itemmaß des Datensatzes betrachtet wird. Allgemein gilt, dass eine Differenz unter 1,00 Logits ein gutes Targeting von

Items und Personen darstellt (Boone & Staver, 2020; Finger et al., 2012). Ein weiterer wichtiger Aspekt ist die **Schrittfolgenfolge** (Step Ordering) der Antwortkategorien (Boone et al., 2014). Hierbei sollten die durchschnittlichen Fähigkeiten der Personen, die eine spezifische Kategorie eines Elements gewählt haben, in aufsteigender Reihenfolge für die Kategorien vorhanden sein. Personen, die Kategorie 1 (richtige Antwort) ausgewählt haben, im Durchschnitt über höhere Personenfähigkeiten verfügen sollten als diejenigen, die Kategorie 0 (falsche Antwort) gewählt haben. Mit den **Itemcharakteristikkurven** (ICC) wird die Schwierigkeit der Items auf der Skala dargestellt und eingeschätzt (Boone & Staver, 2020; Boone et al., 2014; Wu & Adams, 2007). Auf der x-Achse ist die Personenfähigkeit aufgetragen, auf der y-Achse der erwartete Score. Durch diese visuelle Darstellung werden Lücken im Messbereich, die Abstände zwischen den ICC, leicht erkennbar. Die Steigung und die Form der ICC sind im Rasch Modell für dichotome Items identisch, lediglich die Position auf der x-Achse verändert sich. Je positiver dabei der x-Wert des Items ist, desto schwieriger ist das Item (Boone & Staver, 2020; Boone et al., 2014; Linacre, 2012; Moosbrugger & Kelava, 2020)

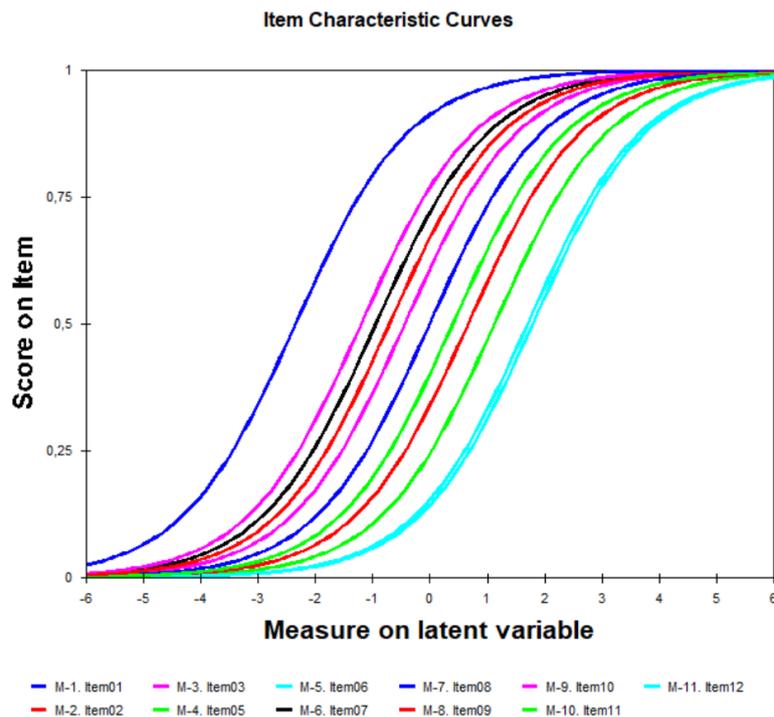


Abbildung 42: ICC eines Beispieldatensatzes in Winsteps erstellt

Eine **Wright Map**, benannt nach ihrem Entwickler Benjamin D. Wright, ist eine grafische Darstellung, die verwendet wird, um die Beziehung zwischen Itemschwierigkeiten und Personenfähigkeiten im Rahmen des Rasch-Modells zu visualisieren (Boone et al., 2014; Wright & Masters, 1982). In der Wright Map, wie in Abbildung 43 zu sehen, werden sowohl Itemschwierigkeiten als auch Personenfähigkeiten entlang einer vertikalen Skala abgetragen.

Diese Skala basiert auf der Einheit "Logits", die zur Messung der Itemschwierigkeiten und Personenfähigkeiten verwendet wird. Schwierige Items werden oben auf der vertikalen Skala der Wright Map platziert. Umgekehrt werden leichte Items unten auf der Skala positioniert, da sie von Personen mit niedrigeren Personenfähigkeiten bewältigt werden können. Personen mit höheren Personenfähigkeiten werden weiter oben auf der vertikalen Skala der Wright Map platziert. Dies spiegelt wider, dass sie die schwierigeren Items erfolgreich bewältigen können. Personen mit niedrigeren Personenfähigkeiten werden weiter unten auf der Skala positioniert, da sie eher in der Lage sind, die leichteren Items zu beantworten.

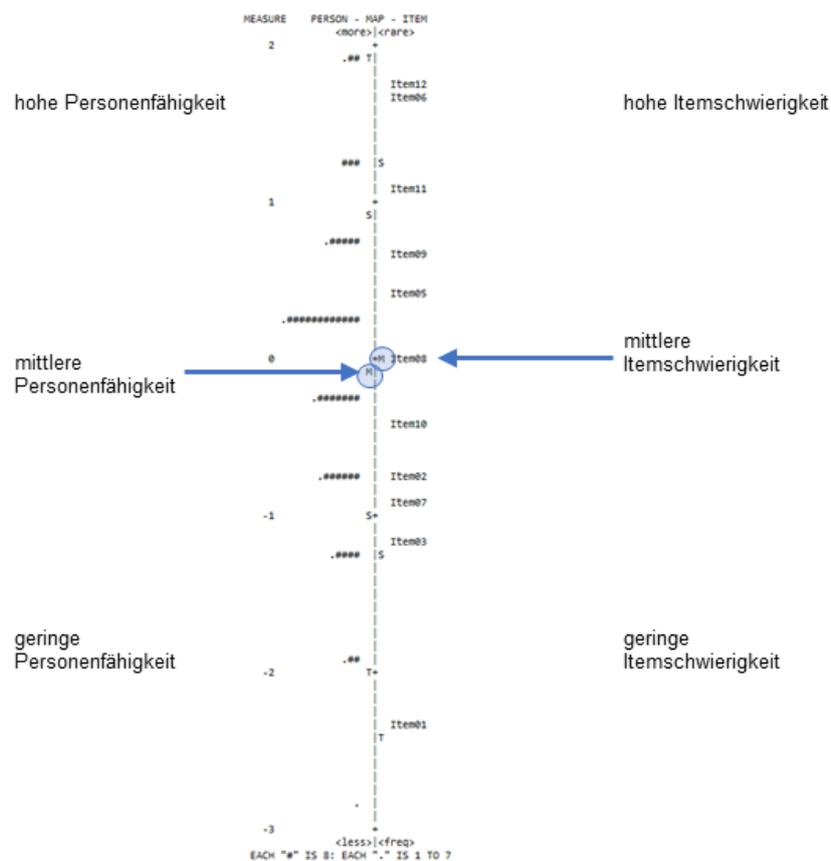


Abbildung 43: Wright Map an einem Beispieldatensatz in Winsteps erstellt

Personen mit einer bestimmten Fähigkeit auf der Wright Map können mit einer Wahrscheinlichkeit von über 50 % die Items richtig beantworten, die auf der Skala unterhalb ihrer eigenen Personenfähigkeit liegen. Bei Items, die oberhalb ihrer eigenen Fähigkeit liegen, ist die Wahrscheinlichkeit, sie korrekt zu beantworten, geringer als 50 %. Durch die Wright Map lässt sich visuell überprüfen, wie gleichmäßig die Items über die Skala verteilt sind. Je breiter und gleichmäßiger die Verteilung ist, desto besser misst das Testinstrument. Boone et al. (2014) nutzen als anschauliche Analogie für das Messinstrument einen "Meterstab". Je gleichmäßiger die Items die Skala (Logits) abdecken, desto feiner ist die Skalierung des "Meterstabes" (cm). Haben Items denselben Schwierigkeitsgrad oder kommt es zu Häufungen

in bestimmten Bereichen, sollten die Items überprüft und bei Redundanz in Bezug auf das untersuchte Merkmal aus dem Fragebogen entfernt werden. Lücken zwischen den Items deuten auf potenzielle Schwächen des Instruments hin (S. Chen et al., 2017). Im nächsten Schritt werden die Items inhaltlich analysiert, um festzustellen, ob die Schwierigkeitshierarchie der Items mit der theoretischen Erwartung übereinstimmt. Es wird untersucht, welche Unterschiede zwischen den leichten und schwierigen Items bestehen. Wenn die Anordnung und Verteilung der Items mit den theoretischen Vorhersagen übereinstimmt, ist dies ein Beleg für die Konstruktvalidität des Instruments (Boone & Staver, 2020). Zusätzlich erfolgt eine Überprüfung der Antwortformate, um die Plausibilität der Distraktoren zu gewährleisten und potenzielle sozial erwünschte Antworten zu identifizieren. Konkret bedeutet dies, dass die durchschnittliche Personenfähigkeit größer sein muss als die Standardabweichung der einzelnen Antwortmöglichkeiten. Falls eine beträchtliche Anzahl der genannten Kriterien nicht erfüllt ist, führt dies zur Ausschlussentscheidung für das betreffende Item. Zum Abschluss der Analyse des Datensatzes werden die Werte zur **Separation und Reliabilität** für die gesamte Stichprobe überprüft. Bei der Rasch-Messung haben wir zwei Reliabilitätswerte (Personenreliabilität, Itemreliabilität) von 0 bis 1 und zwei Separationswerte (Personenseparation, Itemseparation) (Eggert et al., 2017). Letztere stellen eine zusätzliche und angebrachte Ergänzung zur Bewertung der Güte eines Messinstruments dar, da sie Informationen darüber liefern, wie gut die Items in der Lage sind, die Befragten im Hinblick auf das Konstrukt zu differenzieren, sowie darüber, wie gut die Befragten in der Lage sind, die Items zu trennen, die zur Definition des Konstrukts verwendet werden. (Boone et al., 2014; Davis & Boone, 2021; Malec et al., 2007). Die **Personenseparation** wird verwendet, um Personen zu klassifizieren. Wenn die Personenseparation bei einer ausreichend großen Stichprobe gering ist ( $< 2$ ), deutet dies darauf hin, dass das Instrument möglicherweise nicht empfindlich genug ist, um zwischen Personen mit unterschiedlichen Fähigkeiten zu unterscheiden. In solchen Fällen kann es erforderlich sein, mehr Items oder Antwortkategorien hinzuzufügen, um eine bessere Trennschärfe zu erreichen (Boone et al., 2014). Die **Itemseparation** wird zur Überprüfung der Hierarchie der Itemschwierigkeiten verwendet. Wenn die Itemseparation niedrig ist ( $< 4$ ), deutet dies darauf hin, dass die Stichprobe möglicherweise nicht ausreichend groß ist, um die Hierarchie der Item-Schwierigkeiten zu bestimmen und somit die Konstruktvalidität des Instruments zu bestätigen (Boone et al., 2014). Die **Reliabilität** ist ein Maß für die Reproduzierbarkeit der relativen Positionen der Messwerte. Bei einer hohen Reliabilität ist die Wahrscheinlichkeit groß, dass Personen oder Items, die höhere geschätzte Messwerte aufweisen, tatsächlich höhere Messwerte haben im Vergleich zu Personen (oder Elementen) mit niedrigeren Messwerten. Eine präzisere Schätzung wird durch eine größere Stichprobengröße und verringerten Messfehler erreicht. (Boone & Staver, 2020) Die **Personenreliabilität** steigt durch eine Stichprobe mit breitem Fähigkeitsbereich und

einer größeren Testlänge. Auch die Kategorienzahle pro Item und das Stichproben-Item-Targeting spielen eine Rolle. Sie ist unabhängig hingegen vom Stichprobenumfang und weitgehend unbeeinflusst von der Modellgüte. Die Personenreliabilität der Rasch-Analyse entspricht am ehesten der Reliabilität gemäß der klassischen Testtheorie. Generell wird jedoch die durch Cronbachs Alpha gemessene Reliabilität in der klassischen Testtheorie oft höher als diejenige, die durch das Rasch-Modell berechnet wird. (Boone et al., 2014; Linacre, 2023a) Die "model" Personenreliabilität und die "reale" Personenreliabilität dient als Obergrenze („model“) bzw. Untergrenze („real“), wenn Personen auf der Grundlage ihrer Personenfähigkeiten geordnet werden. Die wirkliche Personenreliabilität liegt zwischen der geschätzten Ober- und Untergrenze. Je weniger Rauschen in den Daten vorhanden ist, desto näher liegt die wahre Reliabilität am „model“ Wert. (Linacre, 1997) Neben der Personenreliabilität errechnet Winsteps noch eine Itemreliabilität. Die Item-Reliabilität ist ein Maß, das unabhängig von der Testlänge ist und weitgehend unbeeinflusst von der Modellgüte bleibt. Sie hängt hauptsächlich von zwei Faktoren ab: Der Varianz der Item-Schwierigkeiten und der Größe der Stichprobe. Ein großer Schwierigkeitsbereich der Items und eine größere Stichprobe führen zu einer höheren Item-Reliabilität. Entsprechend der Personenreliabilität wird in Winsteps auch eine „model“ Itemreliabilität und eine „reale“ Itemreliabilität berechnet (Boone et al., 2014; Linacre, 2023a). In der Literatur findet man leicht abweichende Zielwertangaben für die oben beschriebenen Konstrukte. Anbei die für diese Arbeit genutzten Zielwerte nach Boone et al. (2014) und Boone und Staver (2020):

Tabelle 16: Zielwerte zur Überprüfung der Modelleignung (Fit-, Separations- und Reliabilitätswerte)

<i>Allgemeine Kennwerte</i>		
Item Infit MNSQ	0,5 – 1,5	
Item Outfit MNSQ	0,5 – 1,5	
ZSTD Personenfit	-3,0 – 3,0	
Personenreliabilität (Spanne = 0 bis 1,0)	> 0,8	(bei dichotomen Items > 0,5)
Personenseparation	> 2,0	(bei dichotomen Items > 1,5)
Itemreliabilität (Spanne = 0 bis 1,0)	> 0,9	
Itemseparation	> 4,0	

Des Weiteren sind die Items der Fragebögen auf **DIF** (differential item functioning) zu überprüfen. DIF tritt auf, wenn Items für verschiedene Subgruppen einer Analyse unterschiedlich messen (Boone et al., 2014). Um die zeitlich versetzten Fragebögen (Prä-, Post- und Follow-up-Test) vergleichen zu können, sollte hierbei zwischen identischen Items kein DIF auftreten. Hierzu muss vor der DIF-Analyse ein Item geankert werden, um die Skalen miteinander vergleichen zu können (Strobl, 2015). Sollte dennoch ein DIF vorliegen (Wahrscheinlichkeit < 0,01 & DIF Contrast > 0,64 oder < -0,64), ist zu prüfen, ob ein Item

behalten werden kann. Dazu wird dieses zunächst inhaltlich betrachtet, ob es eine begründete Annahme für den DIF bezogen auf die Subgruppe gibt. Anschließend kann das Item entweder von der Analyse ausgeschlossen oder behalten werden. Neben der Möglichkeit den DIF mit Werten anzugeben, lässt sich das Ergebnis auch grafisch in einem Crossplot darstellen und somit einfach deuten, da die Items nicht mehr im Konfidenzintervall eingezeichnet sind. (Boone & Staver, 2020; Boone et al., 2014)

Zum Vergleich der Personenfähigkeiten zwischen Prä- und Posttests durchzuführen, ist es erforderlich, die entsprechenden Itemschwierigkeiten zu ankern. Dies liegt daran, dass Logits nur innerhalb einer Messung („Meterstab“) vergleichbar sind (Boone et al., 2014). Durch das **Ankern** werden die beiden Messungen miteinander verknüpft, und ein Vergleich wird möglich. Dies bedeutet, dass die Itemschwierigkeiten, die im Prä-Test ermittelt wurden, auch für den Posttest verwendet werden. Durch das Verankern der Itemschwierigkeiten werden die berechneten Personenfähigkeiten aus Prä- und Posttest vergleichbar. Nach Berechnung der Personenfähigkeiten in Prä- und Posttest können diese für weitere Berechnungen verwendet werden, auch im Rahmen der klassischen Testtheorie (Boone et al., 2014).

### 6.7.2 Klassische Testtheorie

In der vorliegenden Arbeit wird der Einfluss der unabhängigen Variable der Darbietungsform der Inhalte auf die abhängigen Variablen Veränderung der Personenfähigkeit bezüglich des biotechnologischen Fachwissens und der Naturwissenschaftlichen Arbeitsweise überprüft. Bei allen durchgeführten statistischen Verfahren wurde ein Signifikanzniveau von 5 % ( $p < 0,05$ ) festgelegt (Bühl, 2019). Die aus der IRT erhaltenen Werte werden in der Klassischen Testtheorie (KTT) mit den Verfahren der einfaktoriellen ANCOVA (einfaktorielle Kovarianzanalyse) zur Überprüfung der Gruppenunterschiede zu den Messzeitpunkten und einer ANOVA mit Messwiederholung und Kovariaten zum Vergleich der Messzeitpunkte der Gruppen genutzt (Döring & Bortz, 2015). Ähnlich wie bei der ANOVA liegt auch bei der Kovarianzanalyse (ANCOVA) das Hauptinteresse darin, Unterschiede in den Mittelwerten einer abhängigen Variablen (AV) aufgrund einer oder mehrerer kategorialer unabhängiger Variablen (UV) zu untersuchen. Wenn diese unabhängigen Variablen auch einen Zusammenhang zur betrachteten abhängigen Variablen aufweisen, kann es von Nutzen sein, ihren Einfluss vor der eigentlichen statistischen Analyse zu kontrollieren. Durch diese Vorgehensweise kann die verbleibende Fehlerabweichung verringert werden, was die Wahrscheinlichkeit erhöht, tatsächlich bestehende Effekte aufzudecken (Tabachnick & Fidell, 2014; Wolf & Best, 2010; Wollschläger, 2017). Die Kovarianzanalyse wird genutzt, da bei vorliegender Arbeit vor der Untersuchung existierende Unterschiede zwischen den Personen

in Bezug auf die Schulart und die individuelle nonverbal-figurale kognitive Fähigkeit zum ersten Messzeitpunkt das Untersuchungsergebnis nicht beeinflussen sollen (Bortz & Schuster, 2016). Die Schulart und die individuelle nonverbal-figurale kognitive Fähigkeit zum ersten Messzeitpunkt fließen daher als Kovariaten in die statistischen Modelle ein. Die **Überprüfung der Homogenität der Regressionssteigungen**, welche eine Voraussetzung für die Nutzung der Kovariaten ist, erfolgt durch die Bildung der entsprechenden Interaktionsterme zwischen den Gruppen und den Kovariaten (Schulart und Prä-KFT) (Bortz & Schuster, 2016). Wird keine Interaktion festgestellt, deutet dies darauf hin, dass die Regressionskoeffizienten in den einzelnen Faktoren gleich sind und ein varianzanalytisches Verfahren darf durchgeführt werden. (Tabachnick & Fidell, 2014)

Beide genutzten Verfahren geben einige Voraussetzungen vor, die vor der Durchführung der Berechnungen überprüft werden. Unter anderem muss die Unabhängigkeit der Messungen gegeben sein, die abhängigen Variablen müssen mindestens intervallskaliert und die unabhängige Variable kategorial sein (Bühner & Ziegler, 2017). Die Überprüfung auf univariate **Ausreißer** im Datensatz wird mit einer explorativen Datenanalyse vorgenommen, welche u.a. Box-Plots mit Darstellung der Ausreißer ausgibt. Unter einem leichten Ausreißer (Kreis) versteht man einen Datenwert, der mehr als das 1,5-fache und weniger als das 3-fache des Interquartilsabstands ist. Extreme Ausreißer (mehr als 3-fache des Interquartilsabstands) werden mit einem Stern gekennzeichnet. (Bühl, 2019) Zudem sollten die Gruppen **normalverteilt** sein, was mit dem **Shapiro-Wilk-Test** überprüft wird (Bühl, 2019), welcher vor allem für kleinere Stichproben gute Eigenschaften zeigt (Razali & Wah, 2011). Ist der Wert hier kleiner 0,05, ist die Annahme der Normalverteilung verletzt und die untersuchte Gruppe nicht normalverteilt (Bühl, 2019). Es zeigte sich, dass die ANOVA mit Messwiederholung relativ robust gegenüber der Verletzung der Normalverteilung ist, wodurch es akzeptabel ist, wenn ein oder mehrere Gruppen nicht normalverteilt sind (Wilcox, 2017). Die **Homogenität der Varianzen** wird mittels **Levene-Tests** basierend auf dem Median überprüft. Der Levene-Test ist unabhängig von der Voraussetzung einer Normalverteilung in der Grundgesamtheit und gilt somit als robust. (Jürgen Janssen & Laatz, 2013) Er basiert auf der Nullhypothese, dass die Varianzen der abhängigen Variable über die verschiedenen Gruppen hinweg gleich ist. Der Median wird verwendet, um robustere Schätzungen der Variation in den Gruppen zu erhalten (Backhaus, 2016; Backhaus et al., 2016; Jürgen Janssen & Laatz, 2013). Zeigt die Teststatistik keine Signifikanz, gibt es keine empirische Grundlage für eine Ablehnung der Nullhypothese. Somit liegen keine signifikanten Unterschiede in den Varianzen zwischen den Gruppen vor und somit spricht man von einer Homogenität der Varianzen. (Backhaus et al., 2016; Bortz & Schuster, 2016) Das Ziel der ANCOVA (Kovarianzanalyse) besteht darin, festzustellen, ob es **Unterschiede zwischen den Gruppen** der abhängigen Variable zu den verschiedenen Messzeitpunkten gibt, nachdem die Einflüsse der beiden Kovariaten berücksichtigt wurden.

Dies wurde durch die **Tests der Zwischensubjekteffekte** ermittelt. Ein signifikantes Ergebnis bei der einfaktoriellem ANCOVA weist darauf hin, dass sich mindestens zwei Gruppen in Bezug auf ihre Mittelwerte statistisch signifikant voneinander unterscheiden. (Blanz, 2021; Bortz & Schuster, 2016) Das Ergebnis wird als F-Wert berichtet. Wenn der F-Wert größer 1 ist, gibt es einen Unterschied zwischen den Gruppen. Der zugehörige p-Wert gibt an, ob dieser Unterschied signifikant ist ( $p < 0,05$ ). Die Bewertung der Effektstärke findet über das partielle  $\eta^2$  statt, welches einen Wert zwischen 0 und 1 annehmen kann. Es ist umso größer, je höher der Anteil der erklärten Streuung an der Gesamtstreuung ist. Hierbei wird bei einem Wert von  $\eta^2 > 0,01$  von einem kleinen Effekt, bei  $\eta^2 > 0,06$  von einem mittleren Effekt und bei  $\eta^2 > 0,14$  von einem großen Effekt gesprochen. (Backhaus et al., 2016; Sedlmeier & Renkewitz, 2013) Um festzustellen, welche spezifischen Gruppen sich unterscheiden, wird im Anschluss ein Post-hoc-Test mit Bonferroni-Korrektur durchgeführt. Für die **Vergleiche einer Gruppe über die Messzeitpunkte** hinweg wird durch den **Mauchly-Test** überprüft, ob die Voraussetzung der **Sphärizität** gegeben ist (Döring & Bortz, 2015). Die Sphärizität gibt an, dass die Unterschiede zwischen den unterschiedlichen Interventionsgruppen annähernd gleiche Varianzen haben, d.h. ob Varianzhomogenität für den unabhängigen Faktor vorliegt (Blanz, 2021). Ist der Mauchly-Test nicht signifikant, dann sind die Varianzen der Abstufungen des unabhängigen Faktors homogen und somit wird Sphärizität angenommen (Blanz, 2021; Sedlmeier & Renkewitz, 2013). Wird der Mauchly-Test signifikant, empfiehlt es sich gerade bei kleineren Testgruppen ein Korrekturverfahren zu wählen (Blanz, 2021). Es wird das Vorgehen nach „Huynh-Feldt“ gewählt, welches als Erstes durch die Stichprobendaten schätzt, wie stark die Verletzung der Voraussetzungen ist. Das Maß für die Verletzung der Voraussetzung heißt Epsilon  $\epsilon$ . Dieser Verletzung wird damit begegnet, dass die Zahl der Zähler- und Nennerfreiheitsgrade reduziert wird. Je größer die Verletzung der Annahmen ist, desto mehr werden die Freiheitsgrade für Zähler und Nenner reduziert. Der p-Wert, der aus dem berechneten F-Wert der Studie resultiert, wird dann unter Verwendung der F-Verteilung mit den reduzierten Freiheitsgraden ermittelt. Dadurch ergibt sich ein erhöhter p-Wert im Vergleich zu dem, der mit den nicht reduzierten Freiheitsgraden erzielt würde. Diese Herangehensweise dient dazu, die Wahrscheinlichkeit eines alpha-Fehlers zu minimieren. (Sedlmeier & Renkewitz, 2013) Die Greenhouse-Geisser Korrektur der Freiheitsgrade wäre auch denkbar, welches eine konservativere Korrektur darstellt, aber weniger gebräuchlich ist. (Sedlmeier & Renkewitz, 2013) Zudem sollte das „Huynh-Feldt“ Vorgehen gewählt werden, wenn das „Greenhouse-Geisser-Epsilon“  $\epsilon > 0,75$  ist (Bortz & Schuster, 2016; Girden, 1992). Nach der durchgeführten Kovarianzanalyse bzw. der ANOVA mit Messwiederholung und Kovariaten wird ein **Bonferroni-korrigierter Post-hoc-Test** durchgeführt. Dies ist eine statistische Analyse, um spezifische Paarvergleiche zwischen den Gruppen durchzuführen, also um festzustellen, welche Gruppen signifikant voneinander abweichen. (Backhaus et al., 2016) Die

Bonferroni-Korrektur ist eine Methode, um das Signifikanzniveau anzupassen bzw. die Alpha-Fehler-Kumulierung zu verhindern, d.h. die Wahrscheinlichkeit von Fehlern 1. Art (Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie korrekt ist) zu kontrollieren. (Backhaus et al., 2016; Bortz & Schuster, 2016) Bei einem Bonferroni-korrigierten Post-hoc Test wird das ursprüngliche Alpha-Niveau mit der Anzahl der durchzuführenden Vergleiche multipliziert (Jürgen Janssen & Laatz, 2013). Das korrigierte Alpha-Niveau ist strenger, was bedeutet, dass ein Paarvergleich nur als statistisch signifikant betrachtet wird, wenn der ermittelte p-Wert dieses korrigierte Alpha-Niveau unterschreitet.

## 7 Ergebnisse der Hauptstudie

### 7.1 Analyse der Itemschwierigkeiten und Personenfähigkeiten

Die Qualität der Fachwissens- und NWA-Fragebögen und deren zugehörige Itemschwierigkeiten und daraus folgende Personenfähigkeiten der Testteilnehmer wurden durch Raschanalysen bestimmt. Hierzu werden, wie unter 6.7.1 *Das eindimensionale Rasch-Modell* beschrieben, die entsprechenden Kennwerte überprüft und anschließend diskutiert.

#### 7.1.1 Skala zum Fachwissen Genetik

##### Prä-Test

Die Itemschwierigkeiten des Prä-Tests der Fachwissensskala werden durch das Rasch-Modell berechnet und zunächst der Item-Fit (INFIT MNSQs und OUTFIT MNSQs) betrachtet. Die für den Prä-Test-Fachwissen erhaltene Item-Outfit ist in Abbildung 44 dargestellt.

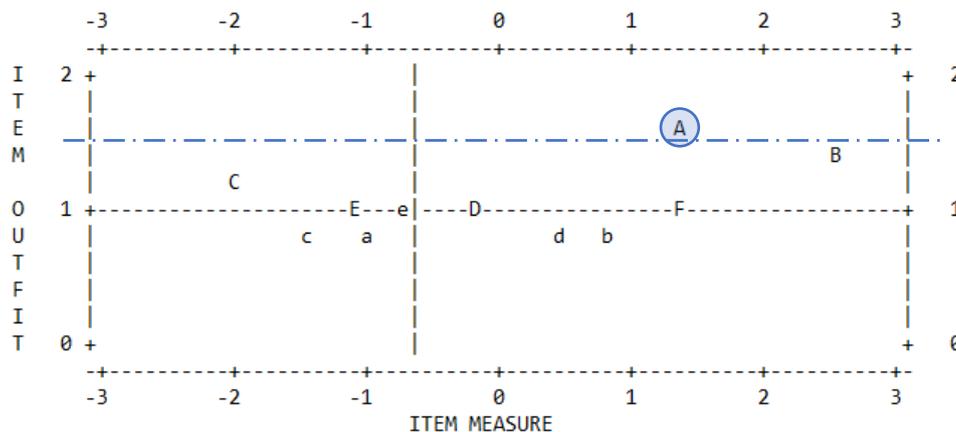


Abbildung 44: Item-Outfit für die Skala Fachwissen im Prä-Test

Die eingezeichnete Linie bei 1,5 stellt den Grenzwert für den OUTFIT MNSQ dar. Die elf Items des Prä-Tests sind hier als Buchstaben A-F und a-e angegeben

Unter Einbezug der Tabelle mit den entsprechenden Zahlenwerten (Anhang 1) ergibt sich für das Item A (PR11\_di) ein OUTFIT MNSQ von 1,53 und somit ein Wert über 1,5. Durch Prüfung der unerwarteten Personenantworten für das Item A (PR11\_di) wurden drei Antworten, mit einem Z-Residual größer 5 aus dem Datensatz entfernt. Die Item-Fits liegen nun im Bereich von 0,5 bis 1,5 (Anhang 2). Der INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD -0,8) und der OUTFIT MNSQ bei 1,02 (ZSTD -0,12). Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft und ist bei allen Items im Erwartungsbereich [-0,7 bis 0,7]. Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,66 im 1. Kontrast (Anhang 3) und liegt somit unter dem Grenzwert von 2. Die Überprüfung auf

Eindimensionalität ist somit abgeschlossen und bestätigt. Betrachtet man die Personenantworten, ergeben sich für die 11 Items ein Maximalscore von 11 und ein Minimalscore von 0. Es liegen keine Personen mit diesen Maximal- bzw. Minimalwerten vor. Somit kann ein Boden- und Deckeneffekt ausgeschlossen werden. Der Personen INFIT MNSQ liegt im Durchschnitt bei 0,99 (ZSTD  $-0,01$ ) und der OUTFIT MNSQ bei 1,02 (ZSTD  $-0,12$ ). Die Überprüfung des INFIT MNSQs und OUTFIT MNSQs der Personen ergibt eine Passung mit einem Misfit von unter 10 %. Die Zuordnung der Scores zu den Personenfähigkeiten im Modell und der Ogive der Skala (s. Abbildung 45) sind in Anhang 4 erfasst.

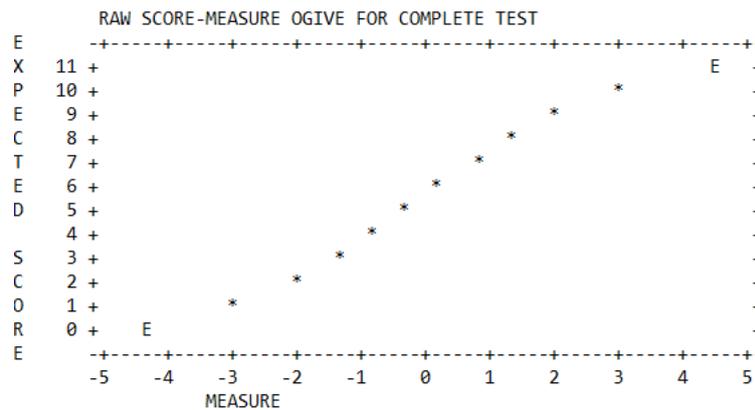


Abbildung 45: Ogive Prä-Test Fachwissen

Darstellung des Zusammenhangs von erreichbaren Punkten im Prä-Test (= EXPECTED SCORE) und den Personenfähigkeiten (=MEASURE)

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten (Anhang 5) sind die Personenfähigkeiten der Itemkategorie 0 durchgehend kleiner als die der Itemkategorie 1 und somit passend. Der Ogive entsprechend lassen sich die Items mittels ICC darstellen. Hierbei wird deutlich, welche Itemschwierigkeiten vorliegen und ob die Itemschwierigkeiten gleichmäßig (ohne Lücken) über die Skala verteilt sind (Abbildung 46).

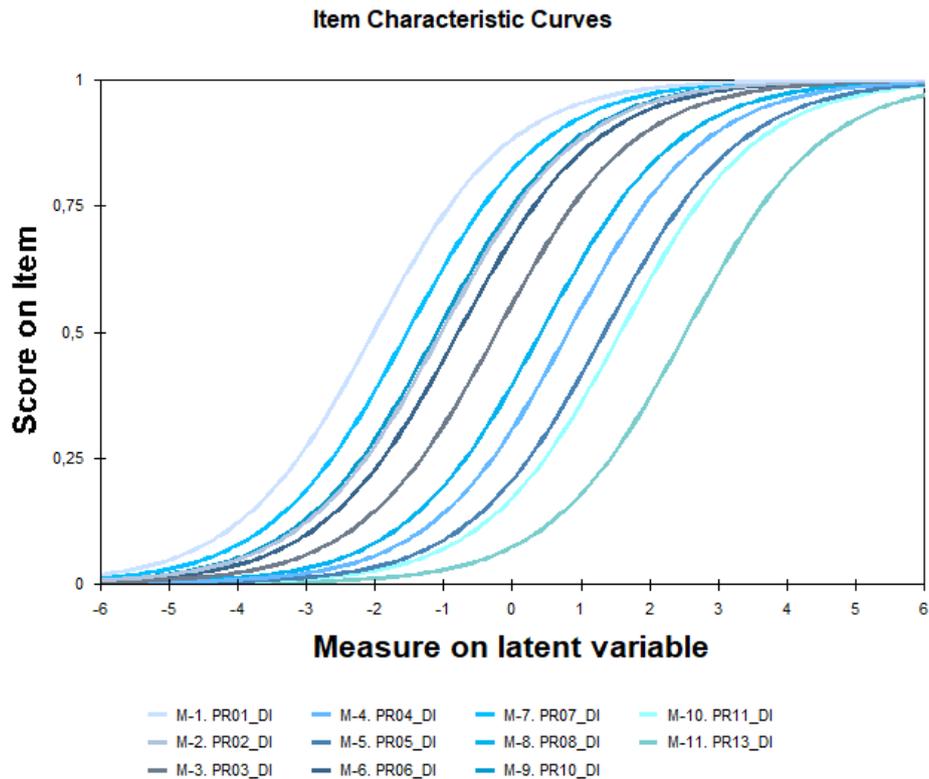


Abbildung 46: ICC für die Skala Fachwissen im Prä-Test  
 Aufgetragen ist die Lösungswahrscheinlichkeit des Items in Abhängigkeit von der Personenfähigkeit (= Measure)

Die Verteilung der ICC kann als grundsätzlich gleichmäßig betrachtet werden. Nur die Kurven des Items PR10\_DI und des Items PR02\_DI liegen sehr nahe beieinander. Ihre Itemschwierigkeiten sind somit ähnlich. Wohingegen zwischen Item PR11\_DI und Item PR13\_DI eine größere Lücke besteht. In Relation zum gesamten Messinstrument ist dieser Bereich größer und trägt somit negativ zur Präzision des Instruments bei.

Für die Darstellung von Personen und Items auf einer Skala wird die Wright Map (Abbildung 47) genutzt. Hierbei werden die Daten entsprechend der Erklärung aus 6.7.1 *Das eindimensionale Rasch-Modell* betrachtet: links die Personenfähigkeiten und rechts die Itemschwierigkeiten.

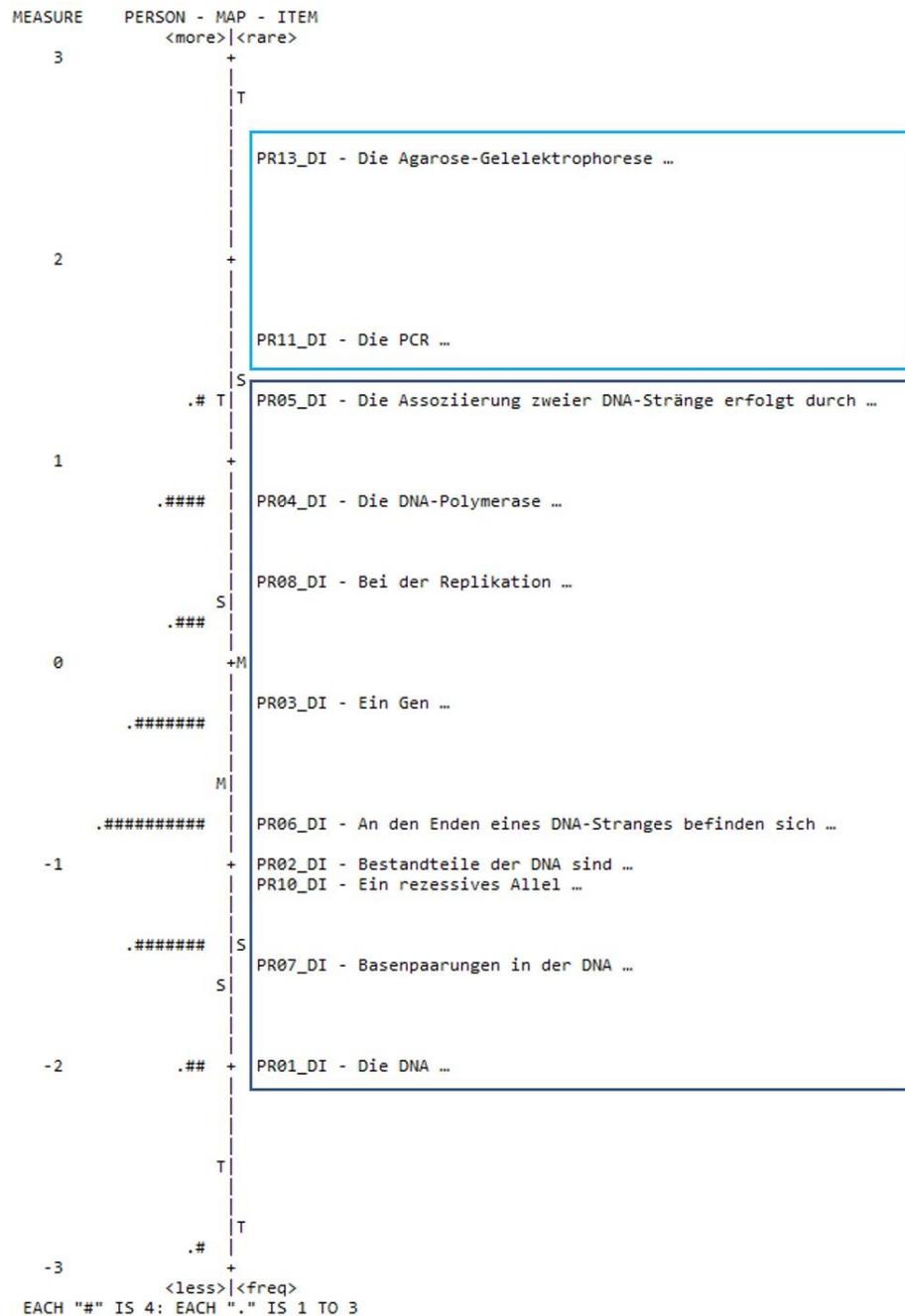


Abbildung 47: Wright Map der Skala Fachwissen im Prä-Test (154 Personen, 11 Items, 2 Kategorien: 0/1). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese

Die Items PR11\_DI und PR13\_DI sind die schwierigsten Items des Fragebogens. Das Fehlen von Personen (#) auf der linken Seite der Wright Map zeigt, dass kaum jemand diese Frage richtig beantwortet hat. Diese Deutung deckt sich mit der Information aus Anhang 5. PR11\_DI wird von 20 Personen richtig beantwortet und PR13\_DI von 9 Personen. Die Items PR05\_DI, PR04\_DI, PR08\_DI und PR03\_DI liegen über dem mittleren Personenfähigkeitswert und sind somit schwerer zu beantworten. Die Items PR01\_DI und PR07\_DI stellen hingegen leichte Items dar. Somit zeigt sich, dass der Fachwissenstest durchaus als anspruchsvoll einzustufen ist. Dies erklärt sich auch mittels der Lösungswahrscheinlichkeit der Aufgaben.

Das Item PR04\_DI mit einer Itemschwierigkeit von 0,81 wird von den nebenstehenden Personen mit entsprechender Personenfähigkeit mit einer Wahrscheinlichkeit von 50 % richtig beantwortet.

Der Mittelwert der Itemschwierigkeiten „M“ auf der rechten Seite der Wright Map wird bei Raschskalierungen auf den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit („M“ linke Seite) ist mit -0,62 Logits geringer. Da die Differenz kleiner 1 ist, kann man von einem noch guten Targeting der Skala sprechen. Um die Aussage über die Qualität der Skala zu vervollständigen, betrachtet man die Real-Werte und Modell-Werte für die Separation und die Reliabilität (Anhang 6). Der wahre Wert liegt zwischen diesen beiden Werten, somit wird der Real-Wert ('REAL') berichtet, da er die Untergrenze darstellt. Die Item-Separation von 6,08 (> 4,00) und die Item-Reliabilität mit 0,97 (> 0,90) der Skala erfüllen die Zielwerte. Die Personen-Separation von 0,61 (> 1,50) und die Personen-Reliabilität von 0,27 (> 0,50) liegen unterhalb der gewünschten Kennwerte.

Bevor die Items im Post-Test und Follow-up-Test mit den Itemschwierigkeiten aus dem Prä-Test geankert und die Skalen überprüft werden, wurde eine DIF Analyse der grundsätzlich ungeankerten Itemschwierigkeiten durchgeführt (unter Nutzung der Itemschwierigkeit des Ankeritems 08 aus dem Prä-Test). Hiermit soll überprüft werden, ob eine Testfairness über die Messzeitpunkte hinweg gegeben ist. Im Vergleich zwischen Prä- und Post-Test zeigte sich DIF bei den Items 01 und 11 (Anhang 7), da die erhaltenen Werte außerhalb des Konfidenzintervalls liegen. Im Vergleich zwischen Prä- und Follow-up-Test (Anhang 8) zeigt sich DIF bei den Items 01 und 13 und im Vergleich zwischen Post- und Follow-up-Test (Anhang 9) DIF zwischen Item 11. Da die Stichprobe klein ist, der DIF nicht in jedem Fragebogen auftritt und sich aus der Literatur heraus keine Grundlage für eine fundierte Argumentation des DIFs auf Itemebene ergibt, werden alle Items beibehalten.

### Post-Test

Um den Prä-Test und den Post-Test vergleichen zu können, werden die Itemschwierigkeiten der Items des Prä-Tests im Post-Test geankert. Anschließend werden die Itemschwierigkeiten der restlichen Post-Test-Items berechnet und der Item-Fit (INFIT MNSQs und OUTFIT MNSQs) betrachtet. Die Abbildung 48 zeigt den Item-OUTFIT für den Post-Test-Fachwissen.

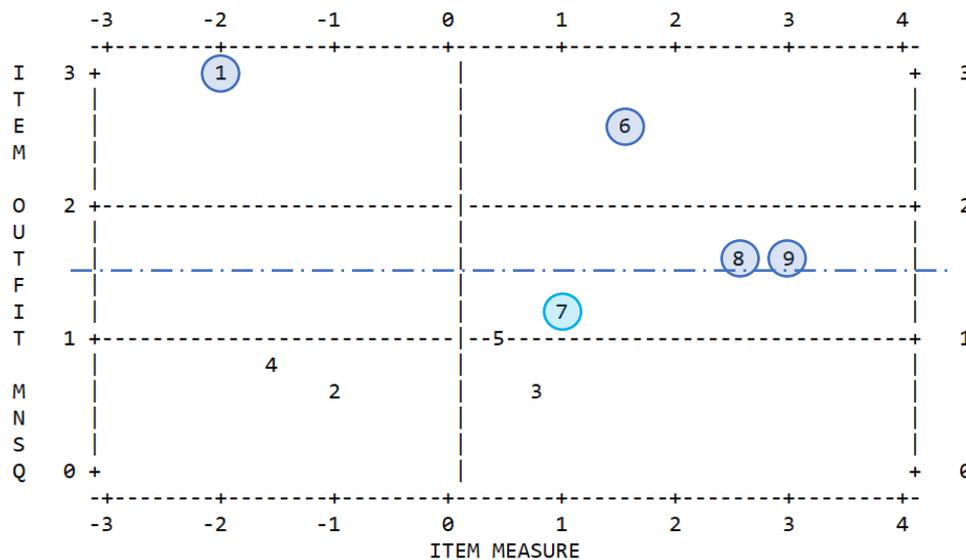


Abbildung 48: Item-OUTfit für die Skala Fachwissen im Post-Test (geankerte Items)

Die eingezeichnete Linie bei 1,5 stellt den Grenzwert für den OUTFIT MNSQ dar. Die neun Items des Post-Tests sind hier als Zahlen 1-9 angegeben.

Betrachtet man die Zahlenwerte Anhang 10 ergeben sich für Item 01 (PO01\_di), Item 06 (PO11\_di), Item 8 (PO13\_di) und Item 9 (PO14\_di) unpassende OUTFIT MNSQ Werte. Die Items 08 und 09 fitten erst durch die Ankerung der Itemschwierigkeiten aus dem Prä-Test nicht mehr (Anhang 11).

Durch Prüfung der unerwarteten Personenantworten wurden bei den entsprechenden Items Antworten ausgeschlossen. Siehe hierzu Tabelle 17. Durch den Ausschluss der Antworten mussten auch bei Item 7 (PO12\_di) zwei Antworten (Z-Residual größer 3 (und kleiner -3)) ausgeschlossen werden, um für alle Items die Item-Fits im Bereich von 0,5 bis 1,5 (Anhang 12) zu erhalten.

Tabelle 17: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Post-Test Fachwissen

#### Übersicht über ausgeschlossene Antworten aufgrund zu großer Z-Residuen

<b>Item 1 (PO01_di)</b>	Ausschluss von 16 Antworten mit Z-Residual (größer 3 und) kleiner -3
<b>Item 6 (PO11_di)</b>	Ausschluss von 5 Antworten mit Z-Residual größer 4 (und kleiner -4)
<b>Item 8 (PO13_di)</b>	Ausschluss von 3 Antworten mit Z-Residual größer 7 (und kleiner -7)
<b>Item 9 (PO14_di)</b>	Ausschluss von 5 Antworten mit Z-Residual größer 5 (und kleiner -5)

Der INFIT MNSQ liegt im Durchschnitt bei 1,09 (ZSTD 0,56) und der OUTFIT MNSQ bei 0,98 (ZSTD -0,05). Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft. Der Erwartungsbereich ist bei allen Items unter 0,7 (Anhang 12).

Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,58 im 1. Kontrast (Anhang 13) und liegt somit unter dem Grenzwert von 2. Betrachtet man die Personenantworten ergeben sich für die 9 Items ein Maximalscore von 9 und ein Minimalscore von 0. Es liegen keine Personen mit diesem Maximalwert vor und 5 Personen mit einem Minimalwert. Ein Boden- und Deckeneffekt ist trotzdem nicht gegeben. Der Personen INFIT MNSQ liegt im Durchschnitt bei 1,06 (ZSTD 0,15) und der OUTFIT MNSQ bei 0,98 (ZSTD 0,20). Die Überprüfung des INFIT MNSQs und OUTFIT MNSQs der Personen ergibt eine Passung mit einem Misfit von unter 10 %. Die Zuordnung der Scores zu den Personenfähigkeiten im Modell sind in Anhang 14 abgebildet.

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten (Anhang 15) sind die Personenfähigkeiten der Itemkategorie 0 durchgehend kleiner als die der Itemkategorie 1 und somit passend. Der Ogive entsprechend lassen sich die Items mittels ICC darstellen (Abbildung 49).

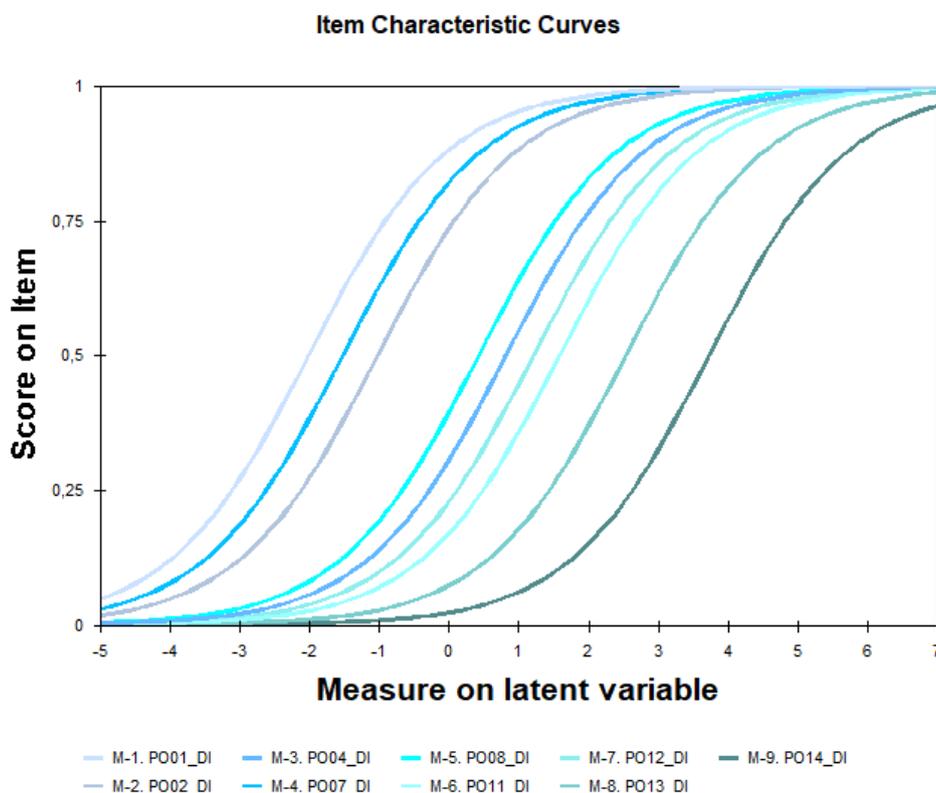


Abbildung 49: ICC der Skala Fachwissen im Post-Test.

Aufgetragen ist die Lösungswahrscheinlichkeit des Items in Abhängigkeit von der Personenfähigkeit (=Measure).

Für die Darstellung von Personen und Items auf einer Skala wird die Wright Map (Abbildung 50) genutzt: links die Personenfähigkeiten und rechts die Itemschwierigkeiten, jeweils in Logits angegeben.

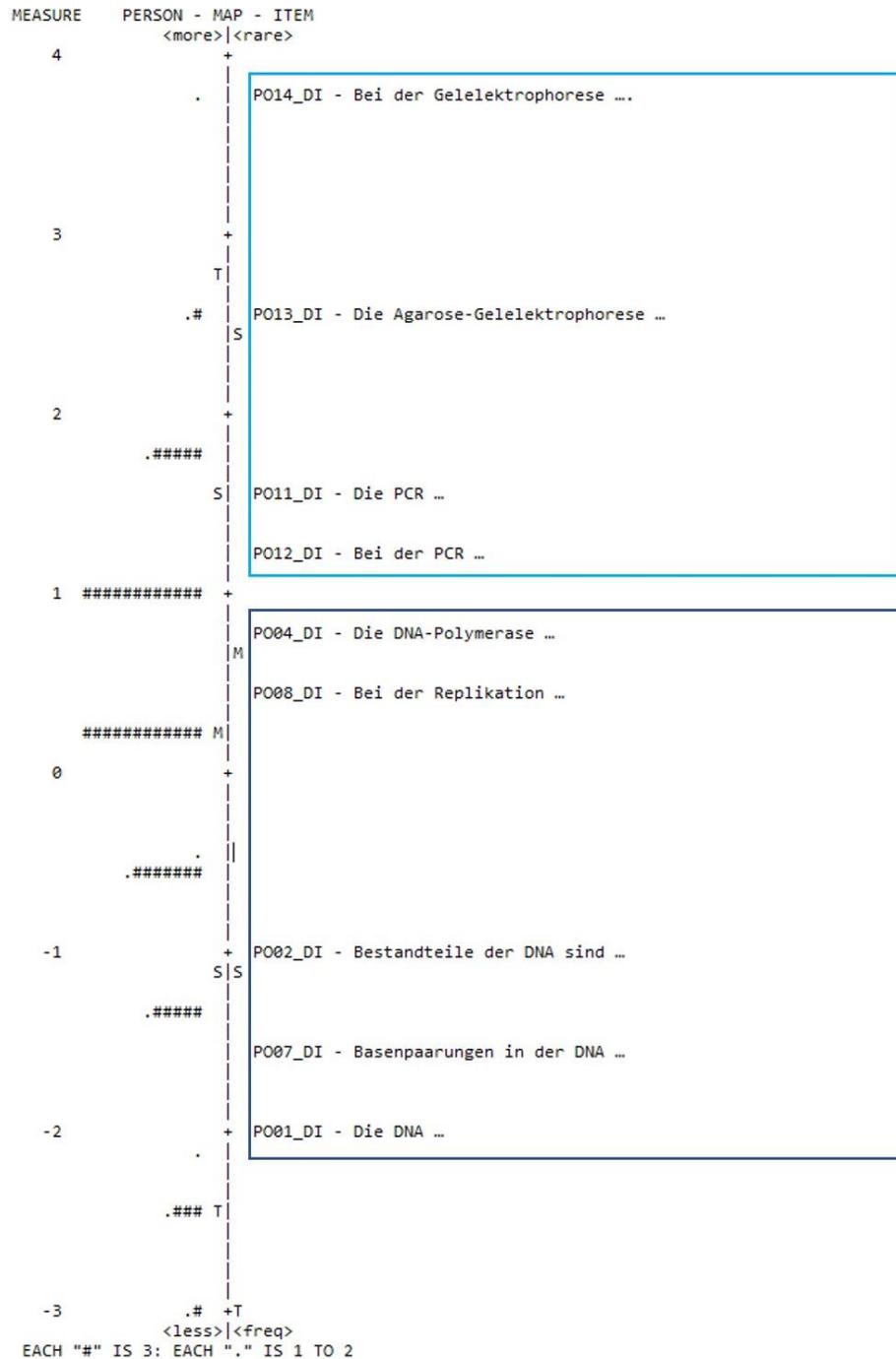


Abbildung 50: Wright Map der Skala Fachwissen im Post-Test (154 Personen, 9 Items, 2 Kategorien: 0/1). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese.

Die Items PO13\_DI und PO14\_DI sind die schwierigsten Items des Fragebogens. Die geringe Anzahl von Personen (# oder .) auf der linken Seite der Wright Map zeigt, dass kaum jemand diese Fragen richtig beantwortet hat. Diese Deutung deckt sich mit der Information aus der Tabelle zu den Anwohnhäufigkeiten je Kategorie (Anhang 15). PO13\_DI wird von 25 Personen richtig beantwortet und PO14\_DI von 8 Personen. Die vier schwersten Items der Skala sind PO12\_DI bis PO14\_DI (hellblauer Kasten). Die Items PO08\_DI, PO04\_DI, PO012\_DI und PO12\_DI liegen über dem mittleren Personenfähigkeitswert und sind somit leichter zu beantworten. Die Items PR01\_DI, PR07\_DI und PO02\_DI stellen hingegen leichte Items dar, die die meisten Personen richtig beantworten konnten. Somit zeigt sich, dass der Fachwissenstest durchaus als anspruchsvoll einzustufen ist. Dies erklärt sich auch durch die Betrachtung der mittleren Personenfähigkeit „M“ bei 0,08 Logits (Anhang 16). Der Mittelwert der Itemschwierigkeiten „M“ auf der rechten Seite der Wright Map wird bei Raschskalierungen auf den Wert 0 Logits festgelegt. Durch die Ankerung der Items aus dem Prä-Test verschiebt sich der Mittelwert der Itemschwierigkeiten im Post-Test auf 0,63 (Anhang 16).

In Abbildung 51 sind die Wright Maps des Prä-Tests (links) und des Post-Tests mit geankerten Itemschwierigkeiten (rechts) im Bereich Fachwissen gezeigt. Die Personenverteilung verschiebt sich im Posttest deutlich nach oben und wird insgesamt breiter. Die mittlere Itemschwierigkeit („M“ rechte Seite) ist im Vergleich zur mittleren Personenfähigkeit („M“ linke Seite) um 0,55 Logits größer, was für ein gutes Targeting spricht. Um die Aussage über die Qualität der Skala zu vervollständigen, betrachtet man die Real-Werte und Modell-Werte für die Separation und die Reliabilität (Anhang 16). Die Item-Separation von 6,86 ( $> 4,00$ ) und die Item-Reliabilität mit 0,98 ( $> 0,90$ ) der Skala erfüllen die Zielwerte. Die Personen-Separation von 0,98 ( $> 1,50$ ) und die Personen-Reliabilität von 0,49 ( $> 0,50$ ) liegen unterhalb der gewünschten Kennwerte.

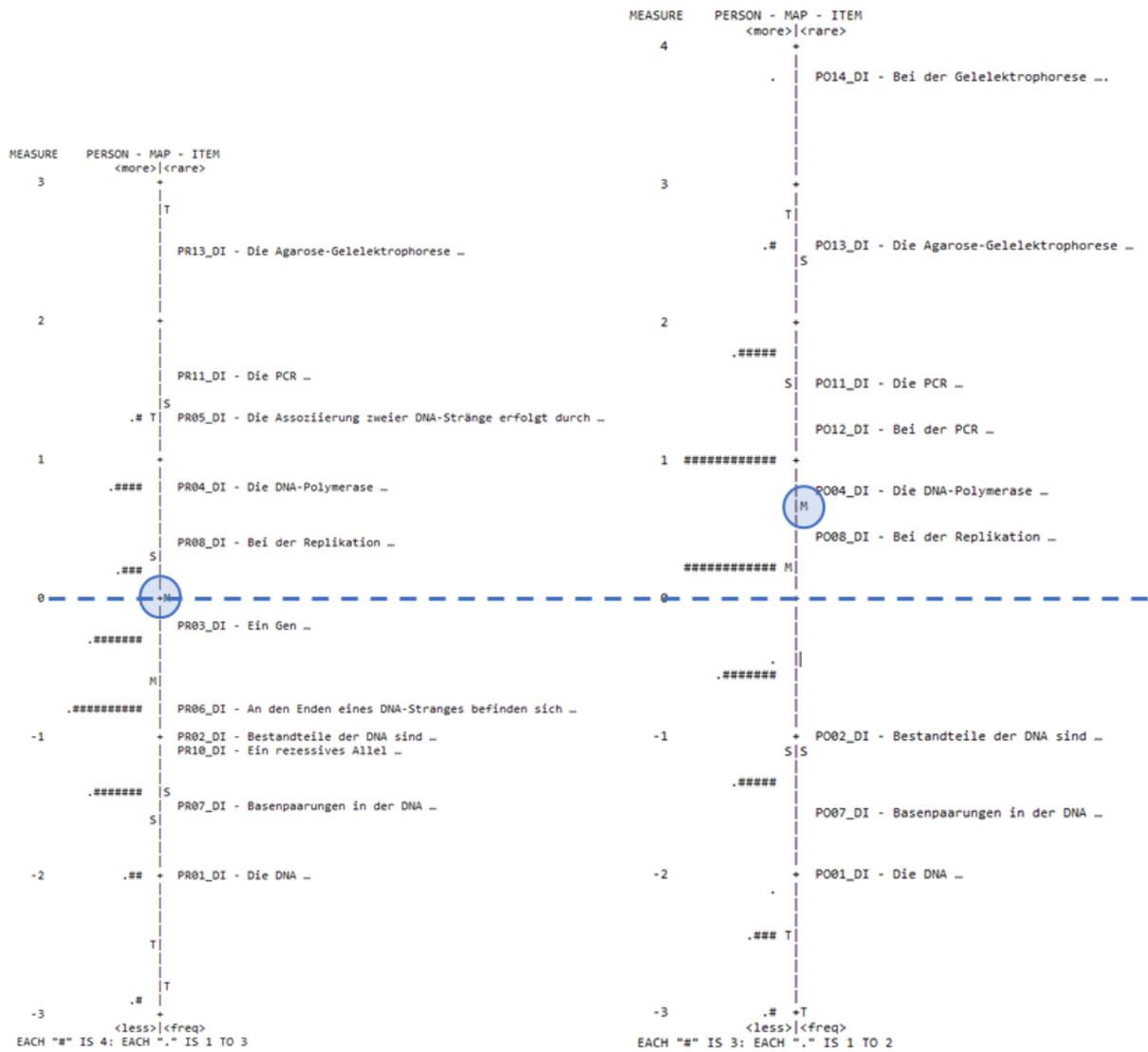


Abbildung 51: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Post-Test mit geankerten Items (rechts). Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit. Hervorgehoben ist die mittlere Itemschwierigkeit des jeweiligen Tests

### Follow-up-Test

Um den Prä-Test und den Follow-up-Test vergleichen zu können, werden die Itemschwierigkeiten der Items des Prä-Tests im Follow-up-Test geankert. Anschließend werden die Itemschwierigkeiten der restlichen Follow-up-Items berechnet und der Item-Fit (INFIT MNSQs und OUTFIT MNSQs) betrachtet. Die Abbildung 52 zeigt den Item-OUTFIT und den Item-INFIT für den Follow-up-Test-Fachwissen.

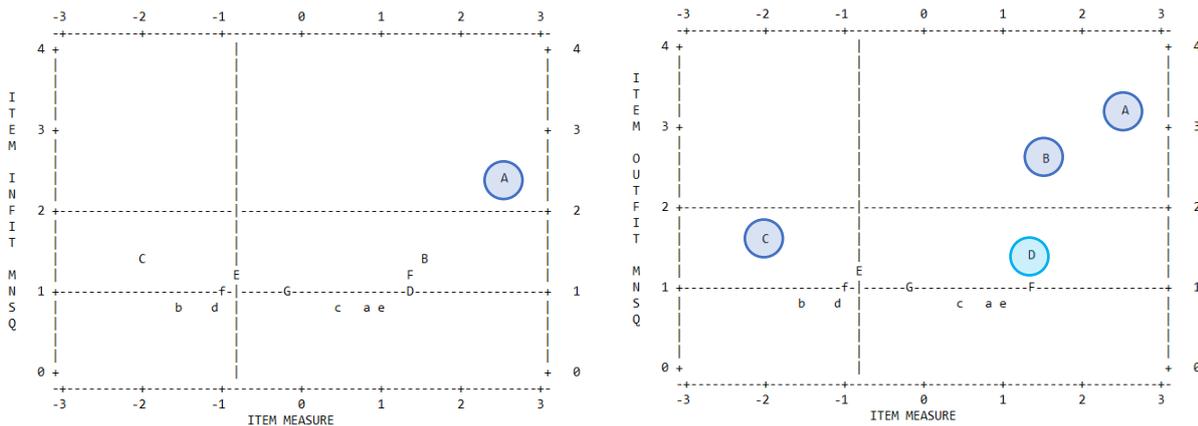


Abbildung 52: Item-Fit für die Skala Fachwissen im Follow-up-Test (geankerte Items)

Die eingezeichnete Linie bei 1,5 stellt den Grenzwert für den OUTFIT MNSQ bzw. INFIT MNSQ dar. Die dreizehn Items des Follow-up-Tests sind hier als Buchstaben A-G und a-f angegeben.

Betrachtet man die Buchstaben (Anhang 18) ergibt sich für den Item-FIT folgendes: Item A (FO13\_DI) ein unpassender INFIT und OUTFIT MNSQ. Für die Items B (FO11\_DI) und C (FO01\_DI) unpassende OUTFIT MNSQ Werte. Die Items fitten erst durch die Ankerung der Itemschwierigkeiten aus dem Prä-Test nicht mehr. Durch Prüfung der unerwarteten Personenantworten wurden bei den entsprechenden Items Antworten ausgeschlossen.

Tabelle 18: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Follow-up-Test Fachwissen

#### Übersicht über ausgeschlossene Antworten aufgrund zu großer Z-Residuen

<b>Item A (FO13_DI)</b>	Ausschluss von 8 Antworten mit Z-Residual (größer 4 und) kleiner -4
<b>Item B (FO11_DI)</b>	Ausschluss von 3 Antworten mit Z-Residual größer 6 (und kleiner -6)
<b>Item C (FO01_DI)</b>	Ausschluss von 3 Antworten mit Z-Residual größer 3 (und kleiner -3)

Das Item D (FO14\_DI) fittet erst durch Löschung der unerwarteten Personenantworten der Items A–C nicht mehr und eine Antwort (Z-Residual größer 5 (und kleiner -5)) muss ausgeschlossen werden. Nun liegt der Item-Fit aller Items im Bereich von 0,5 bis 1,5 (Anhang 19). Der INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD -0,17) und der OUTFIT MNSQ bei 0,97 (ZSTD -0,13). Die Unabhängigkeit der Items ist gegeben (Anhang 19). Die

Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,56 im 1. Kontrast (Anhang 20); Eindimensionalität liegt vor.

Betrachtet man die Personenantworten ergeben sich für die 13 Items ein Maximalscore von 13 und ein Minimalscore von 0. Es liegen keine Personen mit diesem Maximalwert vor und 1 Person mit einem Minimalwert. Ein Boden- und Deckeneffekt ist nicht gegeben.

Der Personen INFIT MNSQ liegt, wie in Anhang 19 zu sehen, im Durchschnitt bei 1,00 (ZSTD 0,64) und der OUTFIT MNSQ bei 0,97 (ZSTD -0,13). Die Überprüfung des INFIT MNSQs und OUTFIT MNSQs der Personen ergibt eine Passung mit einem Misfit von unter 10 %. Die Zuordnung der Scores zu den Personenfähigkeiten im Modell ist in Anhang 21 abgebildet.

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten (Anhang 22) sind die Personenfähigkeiten der Itemkategorie 0 durchgehend kleiner als die der Itemkategorie 1 und somit passend.

Der Ogive entsprechend lassen sich die Items mittels ICC darstellen (Abbildung 53).

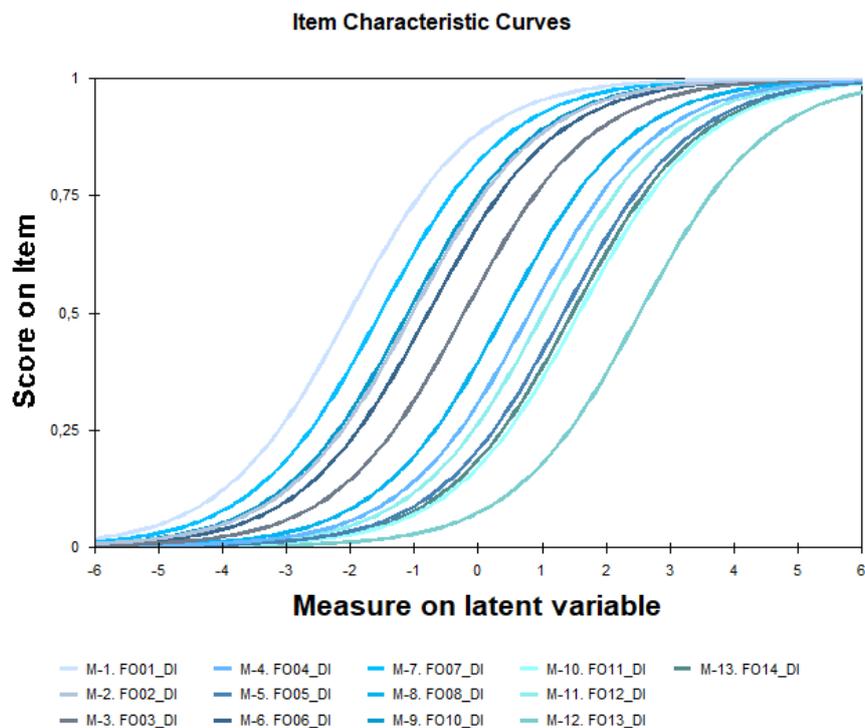


Abbildung 53: ICC Skala Fachwissen Follow-up-Test

Aufgetragen ist Lösungswahrscheinlichkeit des Items in Abhängigkeit von der Personenfähigkeit (=Measure)

Die Verteilung der ICC kann als grundsätzlich gleichmäßig betrachtet werden. Nur die Kurven des Items FO10\_DI und des Items FO02\_DI liegen sehr nahe beieinander. Wohingegen zwischen Item FO11\_DI und Item FO13\_DI eine größere Lücke besteht. Ihre Itemschwierigkeiten sind somit ähnlich bzw. unterschiedlich, was beides auf die Ankerung der Itemschwierigkeiten aus dem Prä-Test zurückzuführen ist.

Für die Darstellung von Personen und Items auf einer Skala wird die Wright Map (Abbildung 54) genutzt.

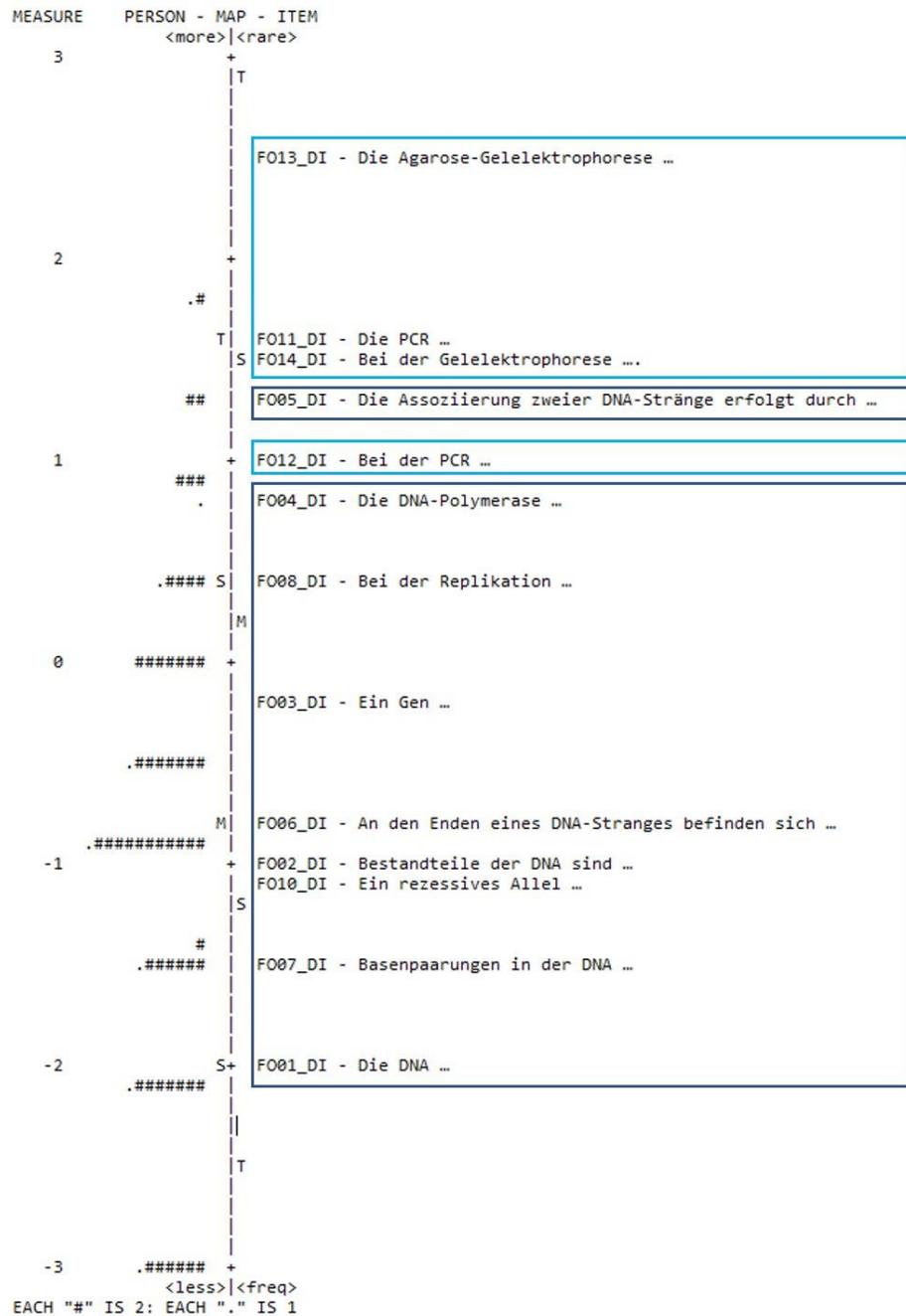


Abbildung 54: Wright Map der Skala Fachwissen im Follow-up-Test (118 Personen, 13 Items, 2 „Kategorien“). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese

Die Items FO14\_DI, FO11\_DI und FO13\_DI sind die schwierigsten Items des Fragebogens. Die geringe Anzahl von Personen (# oder .) auf der linken Seite der Wright Map zeigt, dass kaum jemand diese Frage richtig beantwortet hat. Diese Deutung deckt sich mit der Information über die Antworthäufigkeiten je Item aus Anhang 22. FO13\_DI wird von 8

Personen richtig beantwortet, FO11\_DI von 15 Personen und FO14\_DI von 16 Personen. Die Items FO03\_DI, FO8\_DI, FO04\_DI haben eine mittlere Itemschwierigkeit, stellen aber für die Personen durchaus eine Herausforderung dar. Bei Item FO06\_DI entspricht die Itemschwierigkeit ungefähr der mittleren Personenfähigkeit. Es zeigt sich hier, dass die Personenfähigkeiten gering sind und viele Personen unterhalb des einfachsten Items angesiedelt sind. Somit zeigt sich, dass der Fachwissenstest als Follow-up-Test doch fordernd im Umgang mit dem Wissen der Schüler\*innen ist. Dies bestätigt sich auch durch die Betrachtung (Anhang 23) der Differenz aus der mittleren Itemschwierigkeit 0,19 Logits („M“ rechte Seite) und der mittleren Personenfähigkeit -,86 („M“ linke Seite) welche einen Wert von 1,05 ergibt und somit etwas über dem Grenzwert von 1 Logit liegt, welcher für ein gutes Targeting spricht.

Der Mittelwert der Itemschwierigkeiten „M“ auf der rechten Seite der Wright Map wird bei Raschskalierungen auf den Wert 0 Logits festgelegt. Durch die Ankerung der Items aus dem Prä-Test verschiebt sich der Mittelwert der Itemschwierigkeiten im Follow-up-Test auf 0,19 (Anhang 23). In Abbildung 55 sind die Wright Maps des Prä-Tests (links) und des Follow-up-Tests mit geankerten Itemschwierigkeiten (rechts) im Bereich Fachwissen gezeigt. Die Personenverteilung ähnelt im Follow-up-Test dem Prä-Test, aber die Verteilung der Personen über die Skala wird breiter.

Um die Aussage über die Qualität der Skala zu vervollständigen, betrachtet man die Real-Werte und Modell-Werte für die Separation und die Reliabilität (Anhang 23). Die Item-Separation von 4,81 (> 4,00) und die Item-Reliabilität mit 0,96 (> 0,90) der Skala erfüllen die Zielwerte. Die Personen-Separation von 1,13 (> 1,50) liegt unterhalb und die Personen-Reliabilität von 0,56 (> 0,50) liegt über den gewünschten Kennwerten.

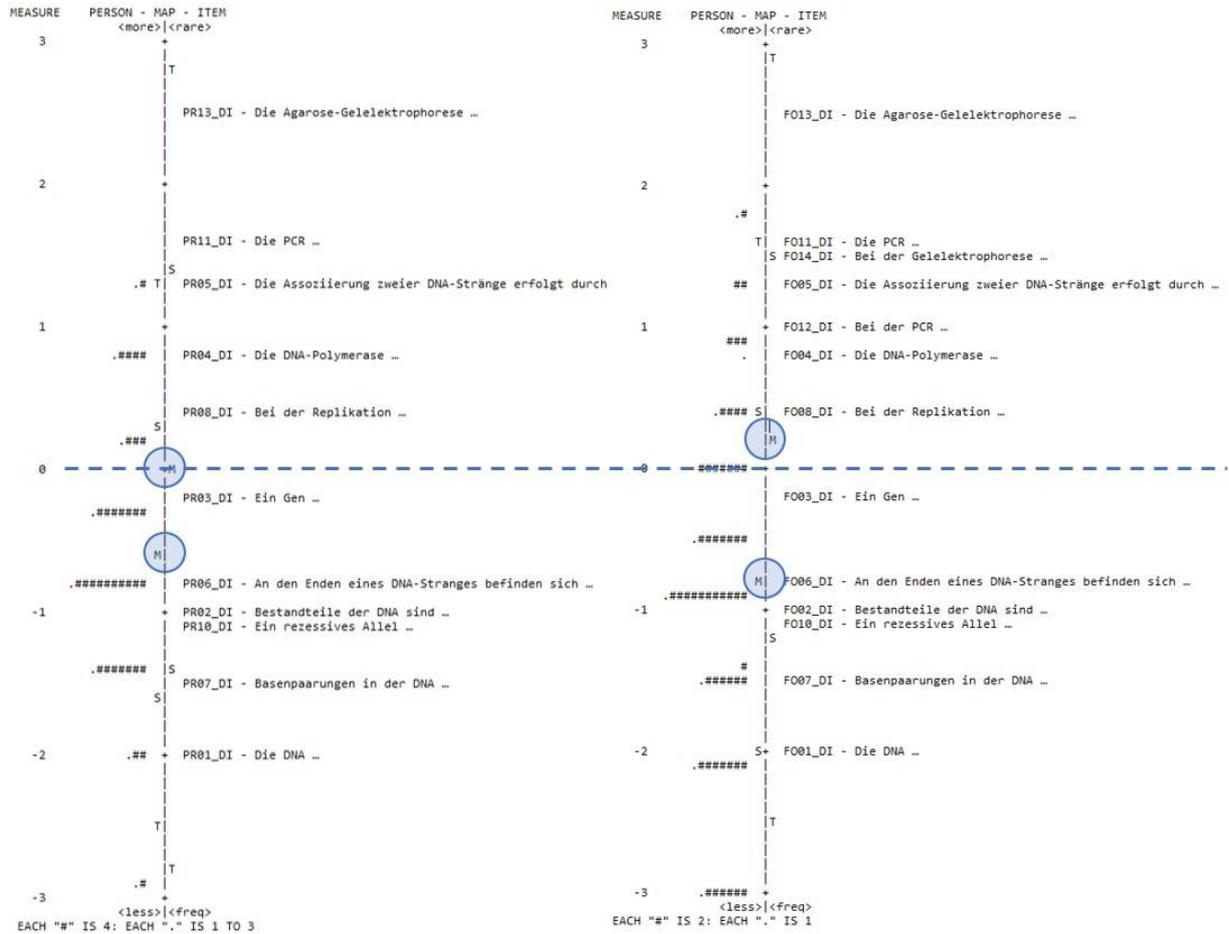


Abbildung 55: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Follow-up-Test (rechts). Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit

## 7.1.2 Zwischendiskussion Fachwissenskala

In Abbildung 56 sind die Wright Maps des Prä-Tests (links), des Post-Tests mit geankerten Itemschwierigkeiten (Itemschwierigkeiten aus dem Prä-Test; mittig) und des Follow-up-Tests mit geankerten Itemschwierigkeiten (Itemschwierigkeiten aus dem Prä-Test; rechts) im Bereich Fachwissen gezeigt, welche nachfolgend als Diskussionsgrundlage dienen.

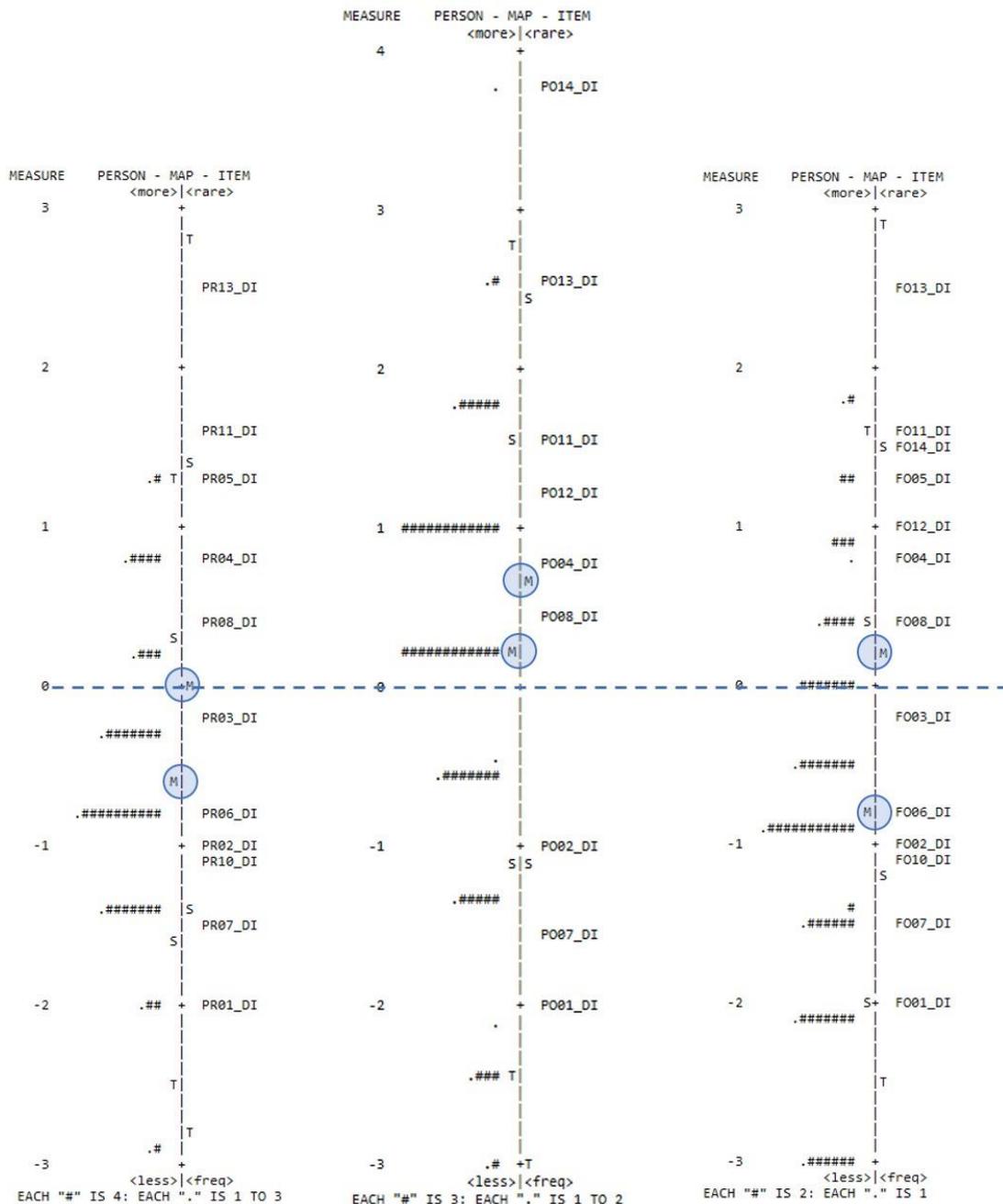


Abbildung 56: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Post-Test (mittig) und Follow-up-Test (rechts). Die Nummerierung der Items ist über alle drei Testzeitpunkte identisch und unterscheiden sich nur über die Vorsilben „pr“ für den Prä-Test, „po“ für den Post-Test und „fo“ für den Follow-up-Test. Die Itemschwierigkeiten des Prä-Tests dienen als Anker für diese im Post- bzw. Follow-up-Test. Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit

Durch die Ankerung der Itemschwierigkeiten ändert sich die Reihung der Items auf der Wright Map über die Testzeitpunkte nicht, wodurch eine einheitliche Skala über die Testzeitpunkte entsteht und die Vergleichbarkeit der Personenfähigkeiten gewährleistet ist. Betrachtet man die Reihung der Items ohne Ankerung im Posttest (Anhang 10) und Follow-up-Test (Anhang 17), so bleibt diese im Vergleich zum Prä-Test weitestgehend erhalten.

Die Reihenfolge auf der Wright Map spiegelt wider, wie verbreitet und verständlich die untersuchten Inhalte im Biologieunterricht sind und wie komplex die kognitiven Anforderungen zur korrekten Beantwortung der Fragen sind. Je spezialisierter und komplexer das Wissen und die Fertigkeiten sind, die zur Beantwortung eines Items erforderlich sind, desto höher ist es auf der Wright Map platziert. Am unteren Ende der Wright Map befinden sich Items, die grundlegendes Wissen über die DNA selbst (01), die Kenntnis der korrekten Basenpaarungen in der DNA (07), das Verständnis eines rezessiven Allels (10), die grundlegenden Bestandteile der DNA (02) und die Enden eines DNA-Stranges (06) abfragen. Diese können als die grundlegendsten Konzepte betrachtet werden, die am ehesten verstanden werden und sind daher im unteren Bereich der Wright Map positioniert. Items die das Konzept eines Gens prüfen (03) oder Items zur Funktion der DNA-Polymerase (04) und zu Replikationsprozessen (08) befinden sich im mittleren Bereich der Skala. Diese Items können als erweiterte Grundlage für den Biologieunterricht an weiterführenden Schulen angesehen werden. Die Items zur Verknüpfung von DNA-Strängen (05), zu molekularbiologischen Techniken wie der PCR (11) und zur Agarose-Gelelektrophorese (13) stellen die höchsten Anforderungen an die Personenfähigkeit der Lernenden. Auch die beiden anderen Items zur PCR (12) und zur Agarose-Gelelektrophorese (14) fallen in diese Kategorie. Dies ist darauf zurückzuführen, dass diese Items ein vertieftes Verständnis der zugrunde liegenden molekularbiologischen Konzepte erfordern.

Die Verteilung der Personen verschiebt sich im Post-Test deutlich nach oben und wird insgesamt breiter. Es zeigt sich also, dass die Personenfähigkeiten im Vergleich zum Prä-Test zugenommen haben. Im Follow-up-Test nähert sich die Verteilung tendenziell wieder der des Prä-Tests an. Dies zeigt sich auch in der Veränderung der mittleren Personenfähigkeit über die Testzeitpunkte (Abbildung 56). Die Breite der Verteilung der Personenfähigkeiten der Schüler\*innen zeigt, dass sich diese im Post-Test im Vergleich zum Prä-Test stärker ausdifferenzieren. Dies spiegelt sich auch in den steigenden Werten der **Personenseparation** über die Fragebögen hinweg wider (Tabelle 19). Dass diese Separationswerte zu allen Testzeitpunkten unter dem gewünschten Kennwert bleiben, liegt unter anderem daran, dass nur Personen der 11. Jahrgangsstufe an der Befragung teilgenommen haben, die grundsätzlich einen ähnlichen Wissensstand aufweisen und es sich somit um eine tendenziell

eher homogene Gruppe an Teilnehmer\*innen handelt. Weitere Erklärungsmöglichkeiten ergeben sich aus zufälligen Antwortmustern, die bei Multiple-Choice-Tests auftreten können und bei denen die Testpersonen raten oder inkonsistent antworten, können zu geringeren Personenseparation führen, da es schwieriger wird, echte Fähigkeitsunterschiede zu identifizieren. Die geringere **Personenreliabilität**, d. h. die geringe Konsistenz der Antworten der Testpersonen auf die verschiedenen Items, zum Prä- und Postzeitpunkt ist auf verschiedene Aspekte zurückzuführen: eine geringe Anzahl von Items (11 bzw. 9 Items) mit nur zwei Antwortkategorien (falsch/richtig) und eine heterogene Testgruppe (Gymnasium/FOS) mit unterschiedlichem Kenntnisstand, was dazu führt, dass die Items die Antworten der Personen möglicherweise nicht ausreichend differenzieren, was zu einer geringen **Personenreliabilität** führt. Darüber hinaus wurden Multiple-Choice-Tests verwendet, die trotz des 2 aus 5-Designs eine gewisse Ratewahrscheinlichkeit aufweisen und somit die Personenreliabilität verringern, da die Testergebnisse dann weniger über das tatsächliche Wissen der Testpersonen aussagen. Um die Personenreliabilität zu verbessern, könnten Maßnahmen wie die Anpassung der Items an das Kompetenzniveau der Zielgruppe, die Erhöhung der Itemanzahl und die Verbesserung der Itemqualität ergriffen werden. Die **Itemseparation** zeigt, dass die Items eine gute Streuung aufweisen, also gut über die Skala verteilt sind und somit unterschiedliche Schwierigkeitsniveaus abdecken. Sie zeigen auch, dass die Tests lang genug sind, um zu prüfen, ohne dabei redundant zu sein. Dies kann zu einer besseren Schätzung der Itemschwierigkeiten führen. Darüber hinaus weisen sie darauf hin, dass die Testitems so gestaltet sind, dass sie die Testpersonen effektiv nach Fähigkeitsniveaus trennen. Items mit einer hohen Trennschärfe tragen dazu bei, Unterschiede zwischen Personen mit unterschiedlichen Fähigkeitsniveaus aufzuzeigen. Die Itemschwierigkeiten werden von den Testpersonen konsistent eingeschätzt, was sich in den entsprechenden **Itemreliabilitäten** widerspiegelt. Eine hohe Itemreliabilität deutet darauf hin, dass die Items mit hoher Wahrscheinlichkeit zuverlässige Indikatoren für das sind, was der Test zu messen vorgibt, und dass die Ergebnisse als zuverlässig angesehen werden können. (Bond & Fox, 2015; Boone & Staver, 2020; Boone et al., 2014; Linacre, 2023a)

Tabelle 19: Reliabilitäten und Separationswerte für Personen und Items

<i>Werte der Reliabilitäten und Separation für Personen und Items je Testzeitpunkt des Fachwissens</i>				
	Kennwerte	Prä	Post	Follow-up
Personenreliabilität (Spanne = 0 bis 1,0)	> 0,5	0,27	0,49	0,56
Personenseparation	> 1,5	0,61	0,98	1,13
Itemreliabilität (Spanne = 0 bis 1,0)	> 0,9	0,97	0,98	0,96
Itemseparation	> 4,0	6,08	6,96	4,81

Die DIF-Analyse der ungeankerten Itemschwierigkeiten (unter Verwendung der Itemschwierigkeit des Ankeritems 08 aus dem Prä-Test) wurde durchgeführt, um die Testfairness, Stabilität und Konsistenz des Fragebogens zwischen den Messzeitpunkten zu überprüfen (Bond & Fox, 2015). Im Vergleich zwischen Prä- und Post-Test wurde für die Items 01 und 11 (Anhang 7) eine DIF festgestellt, da die ermittelten Werte außerhalb des Konfidenzintervalls liegen. Im Vergleich zwischen Prä- und Follow-up-Test (Anhang 8) zeigt sich DIF bei den Items 01 und 13; im Vergleich zwischen Post- und Follow-up-Test (Anhang 9) bei Item 11.

Betrachtet man Item 01, das im Prä-Test das einfachste Item ist und nur einfache grundlegende Informationen zur Lage der DNA abfragt, so wird es im Post-Test und Follow-up-Test schwieriger. Dieses Ergebnis könnte mit dem Problem der Automatizität erklärt werden. Bei Experten kann die Automatizität dazu führen, dass sie Fehler beim Erklären oder Ausführen einfacher Aufgaben machen, weil sie nicht mehr bewusst darüber nachdenken (Willingham, 2002). Ähnliches gilt für die Cognitive Load Theory. Eine Überlastung mit komplexen Informationen, wie sie in der Intervention vorkamen, kann dazu führen, dass grundlegende Informationen vernachlässigt werden oder Fehler in der Anwendung auftreten (Sweller, 2016). Da es sich hier um das leichteste Item beim Prä-Test handelt, könnte es dazu gekommen sein, dass die Schüler\*innen nicht mehr bewusst über die gegebenen Antworten nachgedacht haben. Der DIF von Item 11, das die PCR beschreibt, vom Prä- zum Post-Test und vom Post- zum Follow-up-Test könnte sich dadurch erklären lassen, dass es sich um ein Item handelt, das in der Intervention direkt angesprochen wird. Somit wird es zum Post-Test hin einfacher. Da es sich aber dennoch um ein eher schwieriges Item handelt, ist davon auszugehen, dass die Behaltensleistung möglicherweise zu gering war und das Item auf seine ursprüngliche Schwierigkeit zurückfällt. Dies könnte auf die Bedeutung der Konsolidierung des Lernens hinweisen. Neue Informationen brauchen Zeit, um fest im Langzeitgedächtnis verankert zu werden. Unmittelbar nach der Intervention könnte das Wissen leicht abrufbar sein, aber ohne Konsolidierung durch Wiederholung oder Anwendung könnte diese Information ihre Zugänglichkeit verlieren (Dudai, 2004). Der DIF von Item 13, das sich mit der Agarose-Gelelektrophorese befasst, vom Prä- zum Post-Test könnte durch den Testeffekt erklärt werden, wobei ein einmaliger Abruf von Informationen während des Post-Tests zu einer verbesserten Leistung im Follow-up-Test führen könnte, da der ursprüngliche Abruf die Gedächtnisspur verstärkt (Lipowsky, 2020).

Die DIF-Analyse zeigte, dass der Fragebogen eine Testfairness über die drei Testzeitpunkte zueinander besitzt. Somit werden alle Items im Fragebogen belassen (Bond & Fox, 2015; Boone & Staver, 2020; Davis & Boone, 2021).

Somit zeigt sich, dass der Fachwissenstest in seiner Gesamtheit durchaus als anspruchsvoll einzustufen ist, in sich aber eine gute Zielgenauigkeit aufweist.

### 7.1.3. Skala zur naturwissenschaftlichen Arbeitsweise

Die Fragebögen zur naturwissenschaftlichen Arbeitsweise umfassen Items im Bereich der Methoden der molekularen Genetik mit dem Schwerpunkt der PCR (Item 21 bis 24), der Agarose-Gelelektrophorese (Item 25 bis 28) und der Nutzung von Eppendorf-Pipetten (Item 29 bis 32). Die Skalen der naturwissenschaftlichen Arbeitsweise unterscheiden sich nicht in ihrer Itemzusammensetzung (s. 6.5 *Erhebungsinstrumente*). Die Nummerierung der Items ist bei beiden Testzeitpunkten identisch und unterscheiden sich nur über die Vorsilben „po“ für den Post-Test und „fo“ für den Follow-up-Test.

#### Post-Test

Die Itemschwierigkeiten des Post-Tests der NWA-Skala werden durch das Rasch-Modell berechnet und zunächst der Item-Fit (INFIT MNSQs und OUTFIT MNSQs), wie in Abbildung 57 abgebildet, betrachtet.

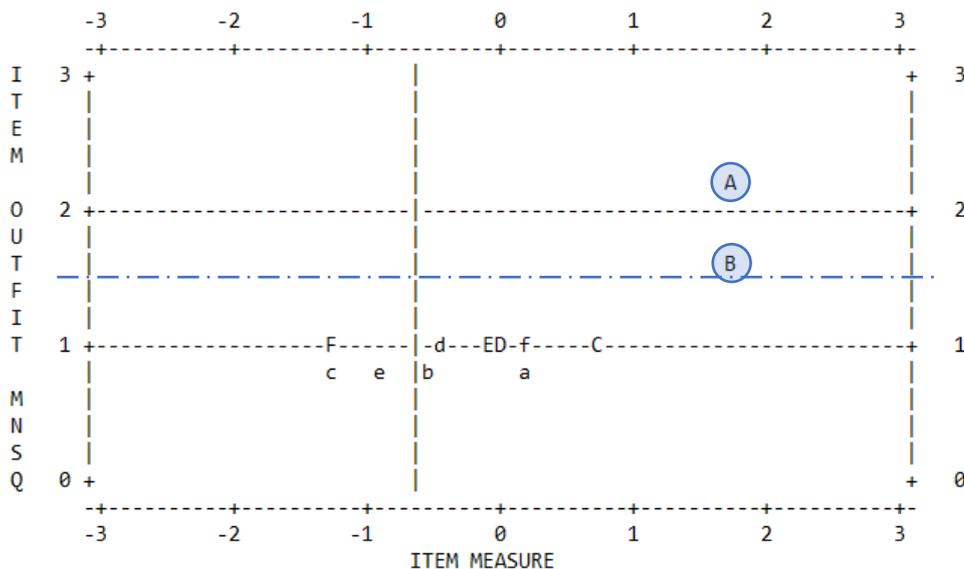


Abbildung 57: Item-Outfit für die Skala NWA im Post-Test

Die eingezeichnete Linie bei 1,5 stellt den Grenzwert für den OUTFIT MNSQ dar. Die elf Items des Post-Tests sind hier als Buchstaben A-F und a-e angegeben.

Unter Einbezug der Tabelle mit den entsprechenden Zahlenwerten (Anhang 24) ergibt sich für das Item A (PO24\_di) ein OUTFIT MNSQ von 2,20 und für das Item B (PO22\_di) ein OUTFIT MNSQ von 1,61. Beide Werte liegen somit über dem Wert von 1,5. Durch Prüfung der unerwarteten Personenantworten für das Item A (PO24\_di) wurden zwei Antworten mit einem Z-Residual von 9 aus dem Datensatz festgestellt und entfernt und bei Item B (PO22\_di) eine Antwort mit einem Z-Residual von 9. Die Item-Fits liegen nun im Bereich von 0,5 bis 1,5 (Anhang 25). Der INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD -0,3) und der OUTFIT MNSQ bei 1,02 (ZSTD -0,3). Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft. Der Erwartungsbereich ist bei allen Items unter 0,7. Die Hauptkomponentenanalyse

der Residuen (PCAR) ergab einen Eigenwert von 1,66 im 1. Kontrast (Anhang 26) und liegt somit unter dem Grenzwert von 2. Eindimensionalität liegt somit vor.

Betrachtet man die Personenantworten ergeben sich für die 12 Items ein Maximalscore von 11 und ein Minimalscore von 0. Es liegen keine Personen mit diesem Maximalscore vor. Der Minimalscore wird von fünf Personen erreicht. Somit kann ein Boden- und Deckeneffekt ausgeschlossen werden. Der Personen INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD – 0,04) und der OUTFIT MNSQ bei 1,02 (ZSTD 0,07). Die Überprüfung des INFIT MNSQs und OUTFIT MNSQs der Personen ergibt eine Passung mit einem Misfit von unter 10 %. Die Zuordnung der Scores zu den Personenfähigkeiten im Modell ist in Anhang 27 einzusehen.

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten (Anhang 28) sind die Personenfähigkeiten der Itemkategorie 0 durchgehend kleiner als die der Itemkategorie 1 und somit passend. Der Ogive entsprechend lassen sich die Items mittels ICC darstellen (Abbildung 58).

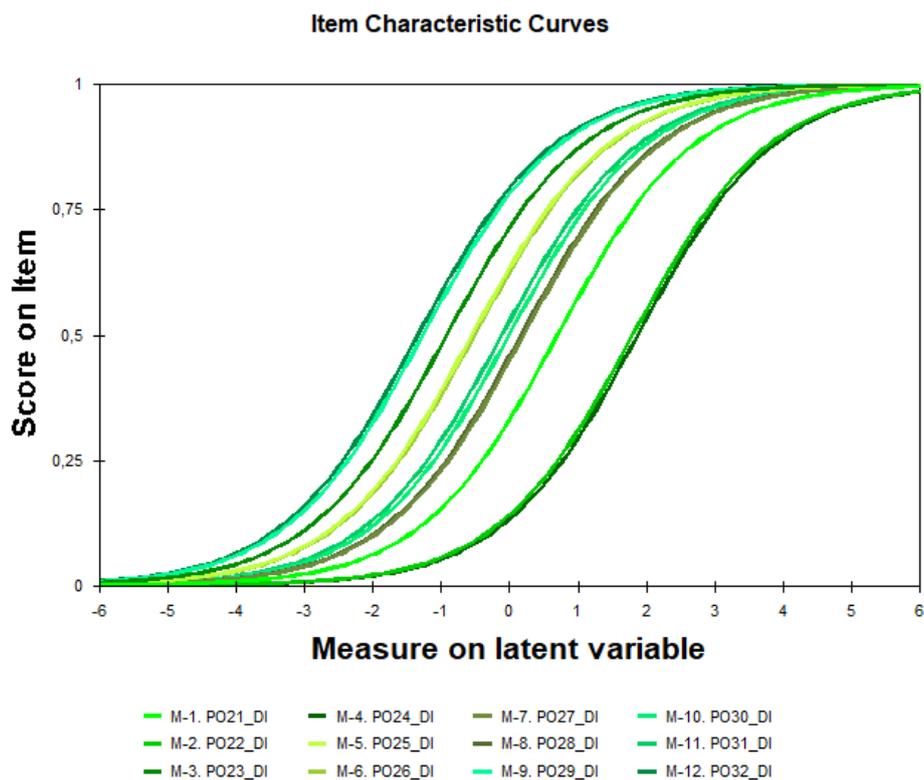


Abbildung 58: Item Characteristic Curves (ICC) NWA im Post-Test

Aufgetragen ist die Lösungswahrscheinlichkeit des Items in Abhängigkeit von der Personenfähigkeit (=Measure).

Die Verteilung der ICC zeigt eine Lücke zwischen Item PO21\_DI und PO22\_DI. Ansonsten ist die Verteilung über die Skala gut abgedeckt. Tendenziell prüfen einige Items dieselbe Schwierigkeit, z.B. Item PO22\_DI und PO24\_DI (ganz rechts), Item PO27\_DI und PO28\_DI (mittig) und Item PO29\_DI und PO32\_DI (links). Somit ergibt sich ein Schwierigkeitsbereich

zwischen den Items, der nicht vollständig durch diese Skala abgedeckt wird und Bereiche, die tendenziell zu stark durch Items abgedeckt sind.

Für die Darstellung von Personen und Items auf einer Skala wird die Wright Map (Abbildung 59) genutzt.

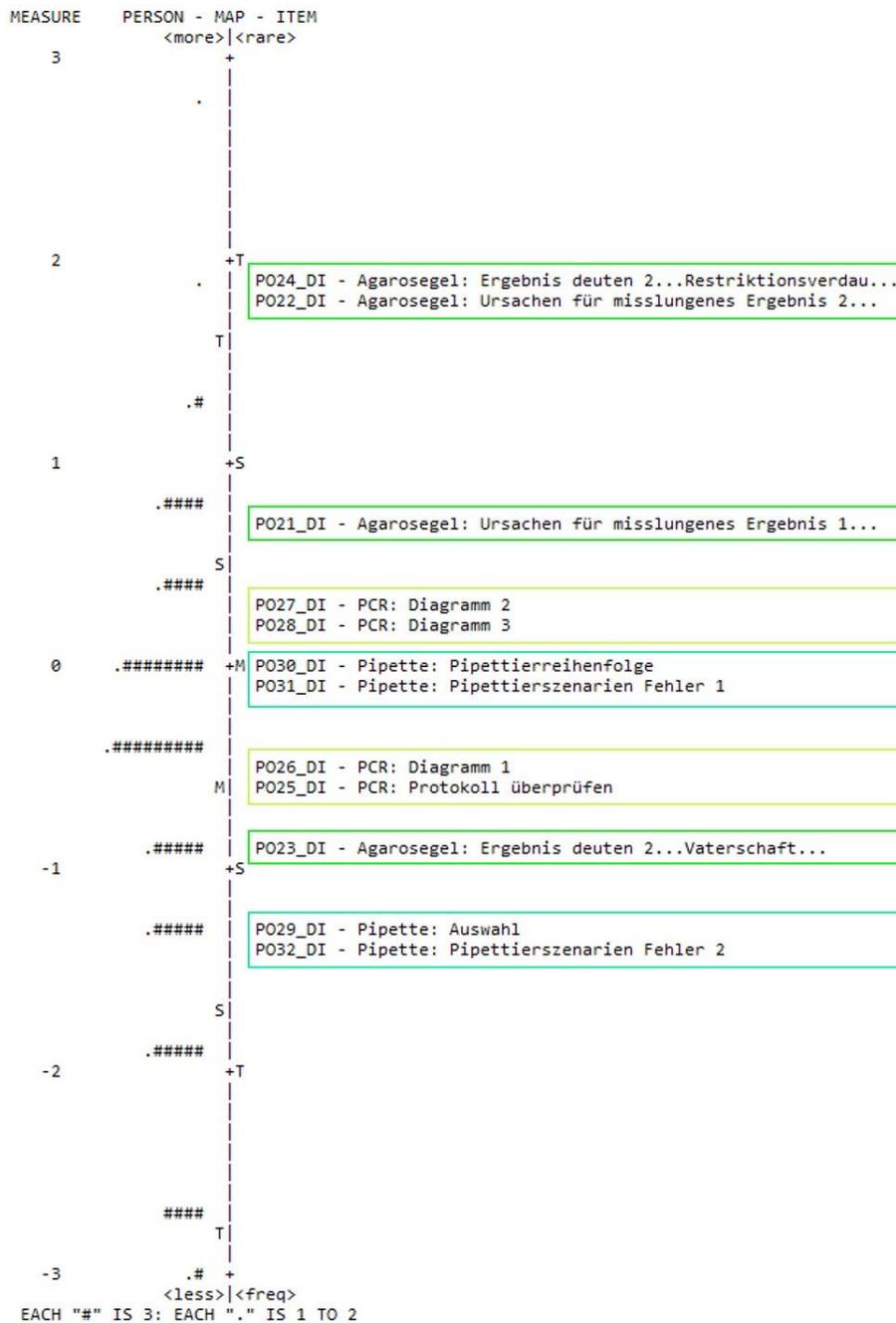


Abbildung 59: Wright Map der Skala NWA im Post-Test (154 Personen, 12 Items, 2 Kategorien: 0/1)  
Inhaltliche Zuordnung der Items: grün: Item 21-24: Agarose-Gelelektrophorese; hellgrün:  
Item 25-28: PRC; türkis: Item 29-32: Pipettieren mit Mikroliterpipette

Die Items PO22\_DI und PO24\_DI sind die schwierigsten Items des Fragebogens. Die Darstellung von lediglich einem „“ auf der linken Seite der Wright Map zeigt, dass nur wenige Teilnehmer diese Frage richtig beantwortet hat. Diese Aussage deckt sich mit der Information aus Anhang 28 zu den Antworthäufigkeiten einzelner Items. Die restlichen Items der Skala befinden sich im mittleren Schwierigkeitsbereich. Zudem zeigt sich, dass im unteren Bereich der Skala (-1,5 bis -3 Logits) keine Items dieser Schwierigkeit zu finden sind, jedoch ein beträchtlicher Anteil an Personen in diesem Bereich ihre Personenfähigkeit haben. Somit zeigt sich, dass der NWA-Test durchaus als anspruchsvoll einzustufen ist. Dies erklärt sich auch mittels der Lösungswahrscheinlichkeit der Aufgaben. Das Item PO30\_DI mit einer Itemschwierigkeit von -,01 wird von den nebenstehenden Personen mit entsprechender Personenfähigkeit mit einer Wahrscheinlichkeit 50 % richtig beantwortet. Der Mittelwert der Itemschwierigkeiten „M“ auf der rechten Seite der Wright Map wird bei Raschskalierungen auf den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit („M“ linke Seite) ist mit -0,69 Logits geringer. Da die Differenz kleiner 1 ist, kann man von einem guten Targeting der Skala sprechen. Um die Aussage über die Qualität der Skala zu vervollständigen, betrachtet man die Real-Werte und Modell-Werte für die Separation und die Reliabilität (Anhang 29). Die Item-Separation von 4,72 (> 4,00) und die Item-Reliabilität mit 0,96 (> 0,90) der Skala erfüllen die Zielwerte. Die Personen-Separation von 1,15 (> 1,50) liegt unterhalb und die Personen-Reliabilität von 0,57 (> 0,50) liegt über den gewünschten Kennwerten.

Bevor die Items im Follow-up-Test mit den Itemschwierigkeiten aus dem Post-Tests geankert und die Skalen überprüft werden, wurde eine DIF Analyse der ungeankerten Itemschwierigkeiten durchgeführt (unter Nutzung der Itemschwierigkeit des Ankeritems 31 aus dem Post-Test). Hiermit soll überprüft werden, ob eine Testfairness über die Messzeitpunkte hinweg gegeben ist. Im Vergleich zwischen Post- und Follow-up Test zeigte sich DIF bei den Items 22, 25, 27 und 29 (Anhang 30), da die erhaltenen Werte außerhalb des Konfidenzintervalls liegen. Da die Stichprobe klein ist und sich aus der Literatur heraus keine Grundlage für eine fundierte Argumentation des DIFs ergibt, werden alle Items beibehalten.

### Follow-up-Test

Um den Post-Test und den Follow-up-Test vergleichen zu können, werden die Itemschwierigkeiten der Items des Post-Tests mit den Items des Follow-up-Test geankert. Anschließend werden die Itemschwierigkeiten der restlichen Follow-up-Items berechnet und der Item-Fit (INFIT MNSQs und OUTFIT MNSQs) betrachtet. Die Abbildung 60 zeigt den Item-OUTFIT und den Item-INFIT für den Follow-up-Test-NWA

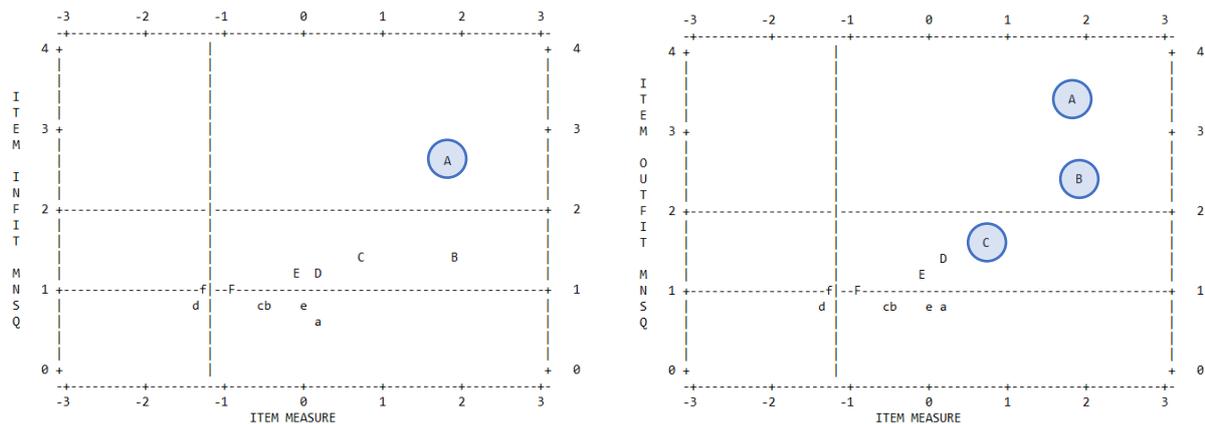


Abbildung 60: Item-Fit für die Skala NWA im Follow-up-Test (geankerte Items)

Die eingezeichnete Linie bei 1,5 stellt den Grenzwert für den OUTFIT MNSQ dar. Die elf Items des Prä-Tests sind hier als Buchstaben A-F und a-e angegeben.

Betrachtet man die Zahlenwerte (Anhang 32) ergibt sich für Item A (FO22\_DI) ein unpassender INFIT und OUTFIT MNSQ und für die Items B (FO24\_DI) und C (FO21\_DI) unpassende OUTFIT MNSQ Werte. Der Misfit der Items FO21\_DI und FO22\_DI entsteht durch die Ankerung der Itemschwierigkeiten aus dem Post-Test. Durch Prüfung der unerwarteten Personenantworten wurden bei den entsprechenden Items Antworten ausgeschlossen, wie in Tabelle 20 zu sehen.

Tabelle 20: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Post-Test NWA

#### Übersicht über ausgeschlossene Antworten aufgrund zu großer Z-Residuen

<b>Item A (FO22_DI)</b>	Ausschluss von 9 Antworten mit Z-Residual größer 3 (und kleiner -3)
<b>Item B (FO24_DI)</b>	Ausschluss von 3 Antworten mit Z-Residual größer 6 (und kleiner -6)
<b>Item C (FO21_DI)</b>	Ausschluss von 1 Antwort mit Z-Residual größer 5 (und kleiner -5)

Nun liegt der Item-Fit aller Items im Bereich von 0,5 bis 1,5. Der INFIT MNSQ liegt im Durchschnitt bei 1,00 (ZSTD -0,17) und der OUTFIT MNSQ bei 0,97 (ZSTD -0,13). Die Unabhängigkeit der Items wird anhand der Korrelationswerte geprüft. Der Erwartungsbereich ist bei allen Items unter 0,7 (Anhang 33). Die Hauptkomponentenanalyse der Residuen (PCAR) ergab einen Eigenwert von 1,56 im 1. Kontrast (Anhang 34) und liegt somit unter dem

Grenzwert von 2; Eindimensionalität liegt vor. Betrachtet man die Personenantworten ergeben sich für die 12 Items ein Maximalscore von 12 und ein Minimalscore von 0. Es liegen keine Personen mit diesem Maximalwert vor und 13 Person mit einem Minimalwert. Ein Deckeneffekt ist nicht gegeben, hingegen aber ein Bodeneffekt. Der Personen INFIT MNSQ liegt, wie in Anhang 19 zu sehen, im Durchschnitt bei 1,00 (ZSTD 0,08) und der OUTFIT MNSQ bei 0,9 (ZSTD 0,19). Die Überprüfung des INFIT MNSQs und OUTFIT MNSQs der Personen ergibt eine Passung mit einem Misfit von unter 10 %. Die Zuordnung der Scores zu den Personenfähigkeiten im Modell ist in Anhang 35 einzusehen.

Betrachtet man die Itemkategorien (0 und 1) und die zugehörigen durchschnittlichen Personenfähigkeiten (Anhang 36), sind die Personenfähigkeiten der Itemkategorie 0 durchgehend kleiner als die der Itemkategorie 1 und somit passend. Der Ogive entsprechend lassen sich die Items mittels ICCs darstellen. Hierbei wird deutlich, welche Itemschwierigkeiten vorliegen und ob die Itemschwierigkeiten gleichmäßig (ohne Lücken) über die Skala verteilt sind (Abbildung 61).

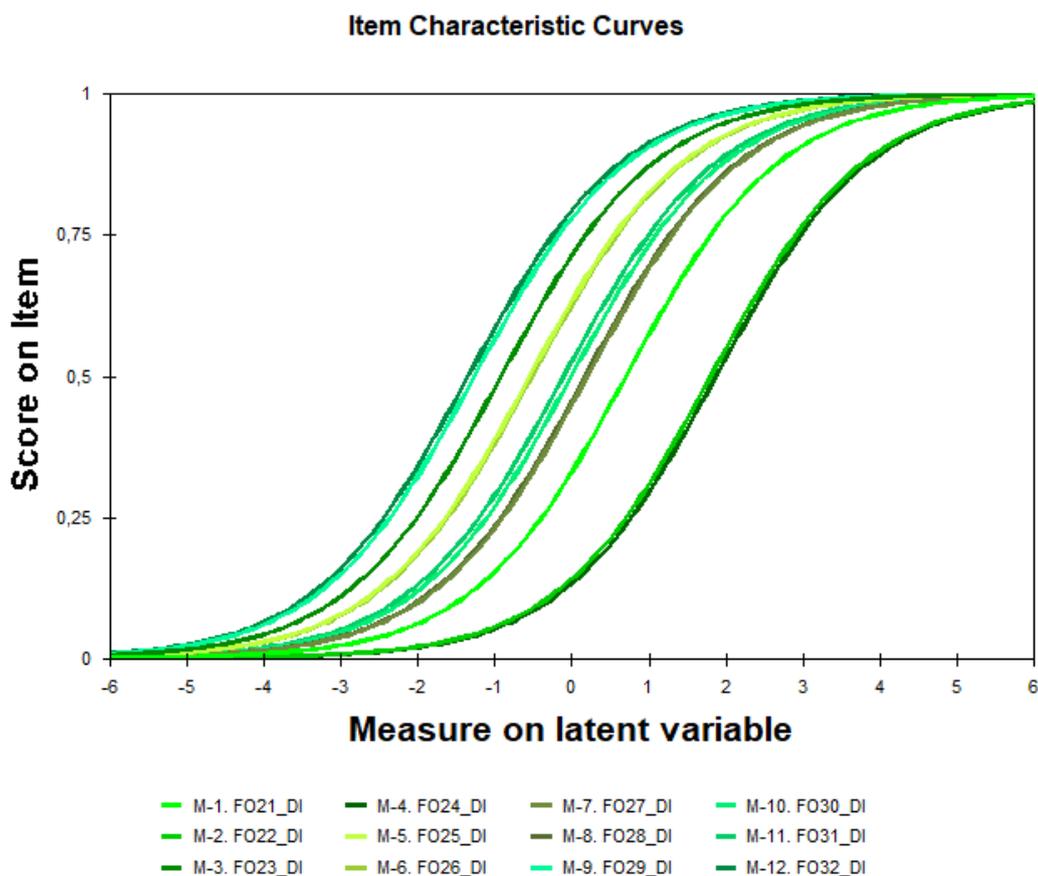


Abbildung 61: ICC der Skala NWA im Follow-up-Test.

Aufgetragen ist die Lösungswahrscheinlichkeit des Items in Abhängigkeit von der Personenfähigkeit (= Measure).

Die ICC und deren Beschreibung entspricht der aus dem Post-Test (s. **Post-Test 7.1.3. Skala zur naturwissenschaftlichen Arbeitsweise**), da die Itemschwierigkeiten aus diesem geankert wurden und somit ihre Lage auf der Skala nicht verändert wird.

Für die Darstellung von Personen und Items auf einer Skala wird die Wright Map (Abbildung 62) genutzt. Hierbei werden die Daten entsprechend der Erklärung aus 6.7.1 *Das eindimensionale Rasch-Modell* betrachtet. Die Verteilung der Items auf der Wright Map ist durch die Ankerung aller Items mit den Itemschwierigkeiten aus dem Post-Test identisch. Somit zeigt sich, dass der NWA-Test durchaus als anspruchsvoll einzustufen ist. Dies erklärt sich auch mittels der Lösungswahrscheinlichkeit der Aufgaben. Das Item FO30\_DI mit einer Itemschwierigkeit von  $-.01$  wird von den nebenstehenden Personen mit entsprechender Personenfähigkeit mit einer Wahrscheinlichkeit 50 % richtig beantwortet. Der Mittelwert der Itemschwierigkeiten „M“ auf der rechten Seite der Wright Map wird bei Raschskalierungen auf den Wert 0 Logits festgelegt. Die mittlere Personenfähigkeit („M“ linke Seite) ist mit  $-1,52$  Logits geringer. Da die Differenz größer 1 ist, ist das Targeting der Skala nicht mehr gut. Somit zeigt sich, dass der NWA-Test als Follow-up Test doch sehr fordernd im Umgang mit den Fähigkeiten der Schüler\*innen ist. Um die Aussage über die Qualität der Skala zu vervollständigen, betrachtet man die Real-Werte und Modell-Werte für die Separation und die Reliabilität (Anhang 37). Der Real-Wert ('REAL') der Item-Separation von  $3,30$  ( $> 4,00$ ) liegt unterhalb und die Item-Reliabilität mit  $0,92$  ( $> 0,90$ ) oberhalb der Zielwerte. Die Personen-Separation von  $0,88$  ( $> 1,50$ ) und die Personen-Reliabilität von  $0,44$  ( $> 0,50$ ) liegen unterhalb der gewünschten Kennwerte.

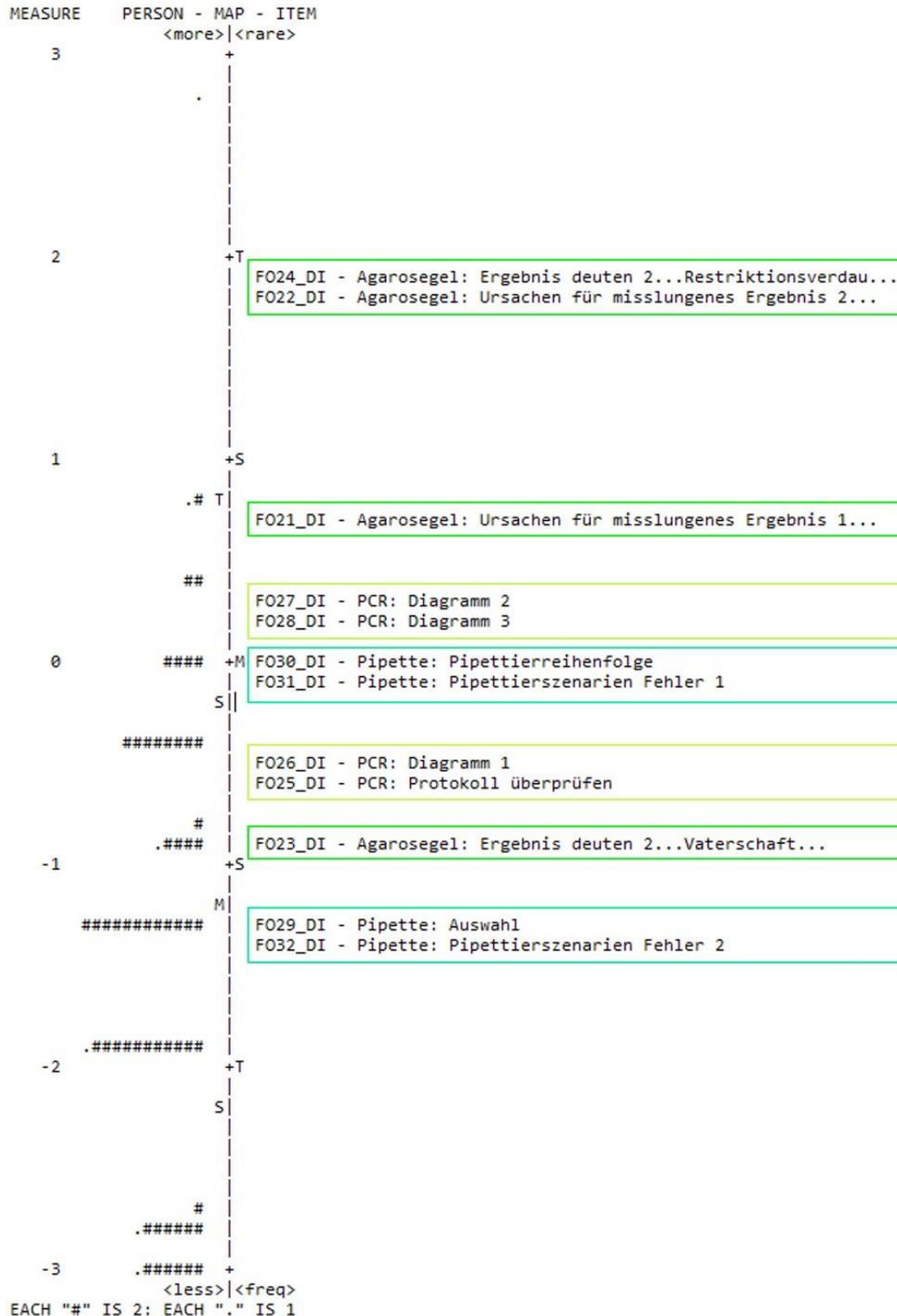


Abbildung 62: Wright Map der Skala NWA im Follow-up-Test (118 Personen, 12 Items, 2 Kategorien: 0/1). Inhaltliche Zuordnung der Items: grün: Item 21-24: Agarose-Gelelektrophorese; hellgrün: Item 25-28: PRC; türkis: Item 29-32: Pipettieren mit Mikroliterpipette

### 7.1.4 Zwischendiskussion NWA-Skala

In Abbildung 63 sind die Wright Maps des Post-Tests (links) und des Follow-up-Tests mit geankerten Itemschwierigkeiten (rechts) im Bereich naturwissenschaftliche Arbeitsweise gezeigt.

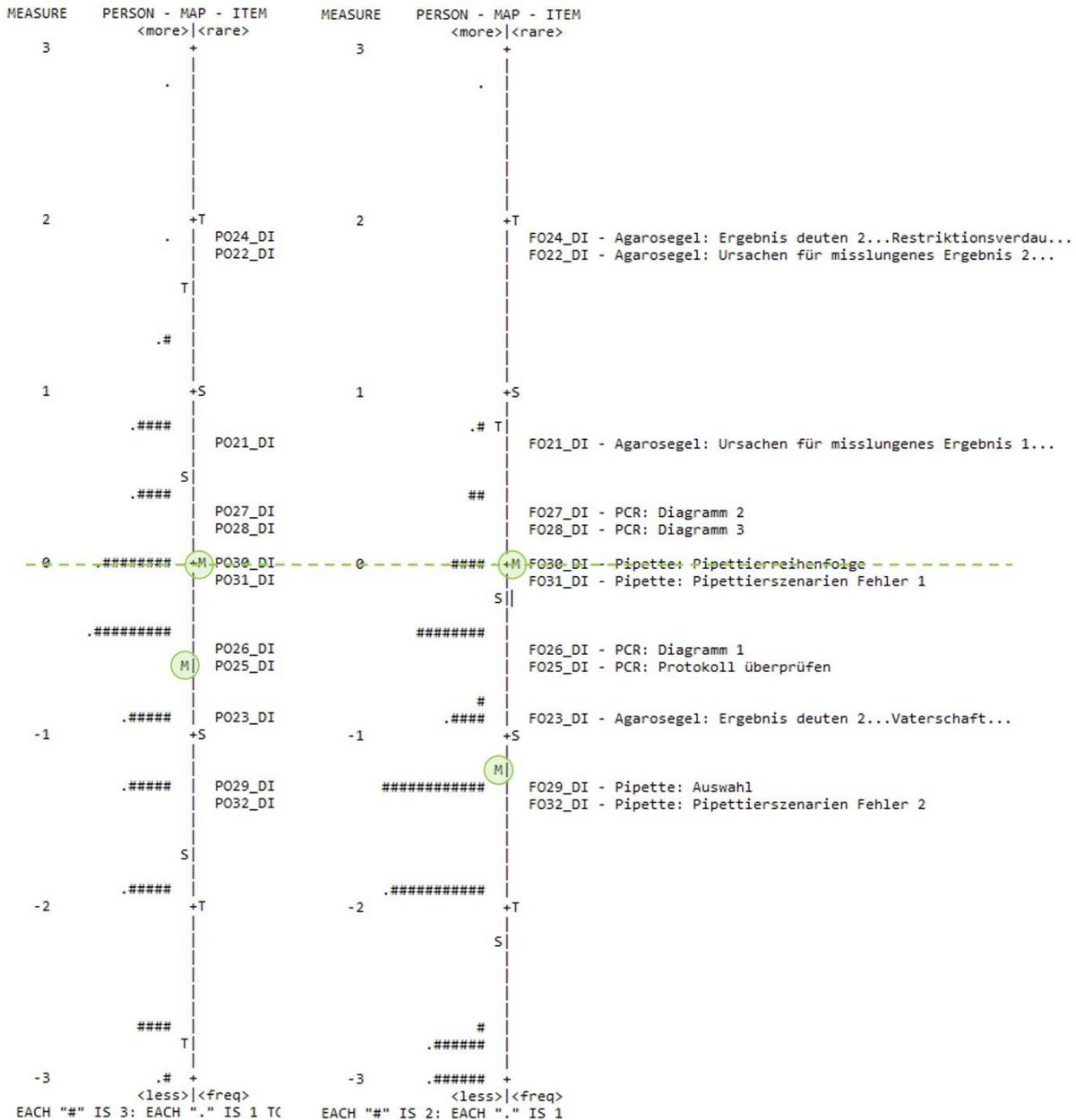


Abbildung 63: Vergleich der Wright Maps in NWA von Post-Test (links) und Follow-up-Test (rechts). Die Nummerierung der Items ist über alle drei Testzeitpunkte identisch und unterscheidet sich nur über die Vorsilben „po“ für den Post-Test und „fo“ für den Follow-up-Test

Durch die Ankerung der Itemschwierigkeiten ändert sich die Reihung der Items auf der Wright Map über die beiden Testzeitpunkte nicht. Betrachtet man die Reihung der Items ohne

Ankerung im Follow-up-Test (Anhang 31), so bleibt diese im Vergleich zum Post-Test weitestgehend erhalten.

Die Verteilung der Items auf der Wright Map deutet darauf hindeuten, dass das Verstehen und Interpretieren von experimentellen Ergebnissen wie Agarosegelen als schwieriger angesehen wird als das Wissen über richtige Laborpraktiken wie Pipettieren. Dies könnte an den kognitiven Fähigkeiten liegen, die benötigt werden, um die Aufgaben zu erfüllen, wobei komplexere Interpretationen und Analysen als anspruchsvoller angesehen werden.

Die untersten Items der Wright Map betreffen das richtige Pipettierverhalten (32) und das Wählen der richtigen Pipetten für bestimmte Volumina (29). Dies sind die grundlegendsten Aufgaben, da sie konkrete und klare Richtlinien für das Pipettieren beinhalten und weniger Interpretation erfordern. Die nächst höher gelegenen Items konzentrieren sich auf das Erkennen von richtigen Aussagen aus der Abbildung eines Gelelektrophoreseschemas eines Vaterschaftstest (23) und die Auswahl von Temperaturen aus PCR-Protokollen (25) und PCR-Diagrammen (26). Sie könnten als grundlegender angesehen werden, da sie mehr auf Erinnerungswissen als auf tieferes Verständnis und kritische Analyse beruhen. Die Items im mittleren Bereich erfordern das Verstehen von Pipettierszenarien (31 und 30) und realen PCR-Diagrammen (28 und 27). Hier ist ein fundiertes Verständnis der Abläufe notwendig, vor allem den beiden PCR-Items. Diese Items (21, 22 und 24) beinhalten das Interpretieren von Agarosegel-Ergebnissen, was ein Verständnis für DNA-Längenstandards, die Anwendung von Spannung auf ein Gel und die Bewertung der Ergebnisse erfordert. Diese sind anspruchsvoller, weil sie ein tieferes Verständnis von den Prinzipien der Agarose-Gelelektrophorese verlangen, was sie zu schwierigeren Aufgaben macht, die ein höheres Fähigkeitslevel erfordern.

Die Verteilung der Personen verschiebt sich im Follow-up-Test deutlich nach unten, wird aber nicht wesentlich schmaler. Es zeigt sich also, dass die Personenfähigkeiten im Vergleich zum Post-Test etwas abgenommen haben. Dies zeigt sich auch in der Veränderung der mittleren Personenfähigkeit über die Testzeitpunkte (Abbildung 63).

Die Breite der Verteilung der Personenfähigkeiten der Schüler\*innen zeigt, dass sich diese im Follow-up-Test im Vergleich zum Post-Test weniger stark ausdifferenzieren. Dies spiegelt sich auch in den sinkenden Werten der **Personenseparation** über die Fragebögen hinweg wider (Tabelle 21). Dass die Werte zu beiden Testzeitpunkten unter dem gewünschten Kennwert bleiben, ist darauf zurückzuführen, dass an der Befragung nur Schüler\*innen der Jahrgangsstufe 11 teilgenommen haben, die grundsätzlich einen ähnlichen Wissensstand aufweisen und somit eine tendenziell homogenere Gruppe an Teilnehmer\*innen ist. Erklärt

werden kann dies auch durch zufällige Antwortmuster, die bei Multiple-Choice-Tests auftreten können und bei denen die Testpersonen raten oder inkonsistent antworten, können zu einer geringeren Personenseparation führen, da es schwieriger wird, tatsächliche Fähigkeitsunterschiede zu identifizieren. Es besteht eine gute **Personenreliabilität** im Post-Test, d.h. die Antworten der Testpersonen auf die verschiedenen Items sind konsistent. Die nicht ausreichende Personenreliabilität im Follow-up-Test lässt sich möglicherweise durch die nur zwei Antwortkategorien (falsch/richtig) im Testdesign und eine heterogene Testgruppe (Gymnasium/FOS) mit sehr unterschiedlichem Wissensstand erklären, wodurch die Items möglicherweise nicht ausreichend zwischen den Antworten der Personen differenzieren. Überdies wurden Multiple-Choice-Tests verwendet, die trotz des 2 aus 5-Designs eine gewisse Ratewahrscheinlichkeit aufweisen und somit die Personenreliabilität verringern, da die Testergebnisse dann weniger über das tatsächliche Wissen der Testpersonen aussagen. Um die Personenreliabilität zu verbessern, könnten Maßnahmen wie die Anpassung der Items an das Kompetenzniveau der Zielgruppe, die Erhöhung der Itemanzahl und die Verbesserung der Itemqualität ergriffen werden. Die **Itemseparation** zeigt, dass die Items im Post-Test eine akzeptable Streuung aufweisen, d.h. gut über die Skala verteilt sind und somit ausreichend unterschiedliche Schwierigkeitsniveaus abdecken. Sie zeigt auch, dass der Post-Test lang genug ist, um zu testen, ohne redundant zu sein. Dies kann zu einer besseren Schätzung der Itemschwierigkeiten führen. Im Follow-up-Test liegt die Itemseparation unter dem Kennwert, was darauf hinweist, dass der untere Bereich der Personenfähigkeiten nicht mehr durch Items abgedeckt wird und der Test in diesem Bereich keine ausreichende Trennschärfe mehr aufweist. Eine kleinere Itemseparation deutet darauf hin, dass der Test für die Teilnehmer\*innen zu schwer oder zu einfach war. Die Itemschwierigkeiten werden von den Testpersonen konsistent eingeschätzt, was sich in den entsprechenden **Itemreliabilitäten** widerspiegelt. Eine hohe Itemreliabilität deutet darauf hin, dass die Items relativ zuverlässige Indikatoren für das sind, was der Test zu messen vorgibt, und dass die Ergebnisse als zuverlässig angesehen werden können. (Bond & Fox, 2015; Boone & Staver, 2020; Boone et al., 2014; Linacre, 2023a)

Tabelle 21: Reliabilitäten und Separationswerte für Personen und Items

*Werte der Reliabilitäten und Separation für Personen und Items je Testzeitpunkt der NWA*

	Kennwerte	Post	Follow-up
Personenreliabilität (Spanne = 0 bis 1,0)	> 0,5	0,57	0,44
Personenseparation	> 1,5	1,15	0,88
Itemreliabilität (Spanne = 0 bis 1,0)	> 0,9	0,96	0,92
Itemseparation	> 4,0	4,72	3,30

Die DIF-Analyse der ungeankerten Itemschwierigkeiten (unter Verwendung der Itemschwierigkeit des Ankeritems 31 aus dem Post-Test) wurde durchgeführt, um die Testfairness, Stabilität und Konsistenz des Fragebogens zwischen den Messzeitpunkten zu überprüfen. Im Vergleich Post-Test zu Follow-up-Test wurden DIFs für die Items 22, 25, 27 und 29 festgestellt.

Der DIF des Items 22, das sich mit Fehlern bei der Durchführung der Agarose-Gelelektrophorese befasst, könnte durch den Testeffekt erklärt werden, da die grundlegende Fehlerbetrachtung des Gels bereits im Post-Test erfolgte. Der Testeffekt besagt, dass ein einmaliger Abruf von Informationen im Post-Test zu einer verbesserten Leistung im Follow-up-Test führen kann, da der erste Abruf die Gedächtnisspur verstärkt (Lipowsky, 2020). Ähnliches kann für den DIF von Item 29 angenommen werden, das die richtige Auswahl einer Mikroliterpipette für eine vorgegebene Flüssigkeitsmenge überprüft.

Der DIF von Item 25, das die Zusammenstellung von PCR-Protokollen überprüft, könnte dadurch erklärt werden, dass es sich um ein Item handelt, das in der Intervention direkt angesprochen wird. Dadurch ist es im Post-Test leichter zu bearbeiten. Da es sich aber dennoch um ein mittelschweres Item handelt, ist davon auszugehen, dass die Behaltensleistung möglicherweise zu gering war und das Item zur Beantwortung eine höhere Personenfähigkeit erfordert. Dies könnte in diesem Fall auf eine mangelnde Festigung des Lerngegenstandes hindeuten. Damit neue Informationen einen dauerhaften Platz im Langzeitgedächtnis finden, sind Zeit und Wiederholungen notwendig. Unmittelbar nach der Intervention könnte das Wissen leicht abrufbar sein, aber ohne Konsolidierung durch Wiederholung oder Anwendung könnte diese Information ihre Zugänglichkeit verlieren (Dudai, 2004). Dies kann auch für Item 27 angenommen werden, das sich auf die Temperaturen bei der PCR in einem Diagramm bezieht.

Die DIF-Analyse zeigte, dass der Fragebogen zwischen den beiden Testzeitpunkten eine Testfairness aufweist. Daher werden alle Items im Fragebogen belassen (Bond & Fox, 2015; Boone & Staver, 2020; Davis & Boone, 2021).

Somit zeigt sich, dass der Fragebogen zu den naturwissenschaftlichen Arbeitsweisen in seiner Gesamtheit als schwierig einzustufen ist, für sich genommen aber eine akzeptable Zielgenauigkeit aufweist.

## 7.2 Fachwissen

Zunächst wird für alle drei Fachwissensskalen (Prä-, Post- und Follow-up-Skala) eine explorative Datenanalyse anhand der Daten der Personenfähigkeiten durchgeführt, um Ausreißer zu identifizieren und die Normalverteilung zu prüfen.

Die Datensätze weisen einige leichte Ausreißer auf (Abbildung 64). Betrachtet man diese, zeigen sich über alle drei Messzeitpunkte hinweg für die Gruppen nachfolgende Anzahl. Die Gruppe Präsenzunterricht (PP) zeigt 2 Ausreißer. In der Gruppe Flipped Classroom (DP) 4 Ausreißer. Und in den Gruppen, Wechselunterricht (PD) und Homeschooling (DD) sind es 6 und 3 Ausreißer. Da es sich bei allen Ausreißern um leichte Ausreißer handelt, werden alle Fälle für die weiteren Analysen genutzt und keine Fälle ausgeschlossen.

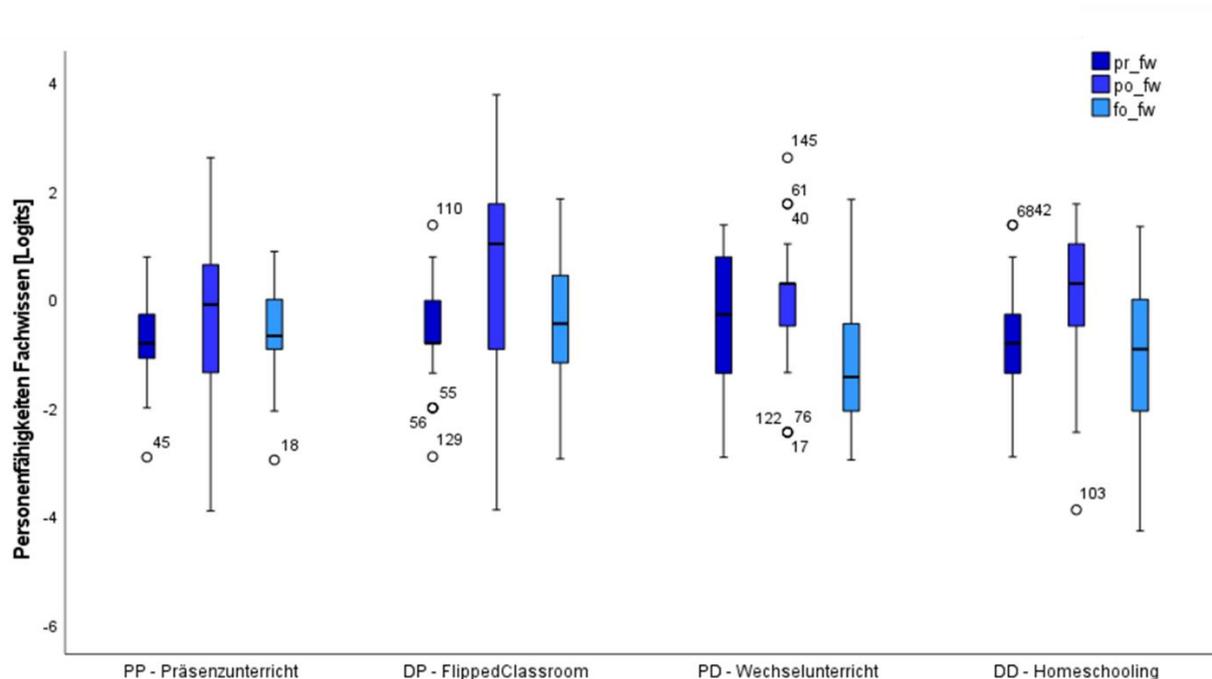


Abbildung 64: Überprüfung der Personenfähigkeiten des Fachwissens auf univariate Ausreißer

Die Testung auf Normalverteilung zeigte gemäß dem Shapiro-Wilk-Test drei nicht normalverteilte Testgruppen ( $p < 0,05$ ) (Anhang 38) über die Testzeitpunkte hinweg. Dies ist bei den Ergebnissen des Prä-Tests in der Gruppe Wechselunterricht ( $p = 0,040$ ) und Homeschooling ( $p = 0,009$ ) der Fall. Ebenso im Post-Test der Gruppe Homeschooling ( $p = 0,002$ ), als auch im Follow-up-Test in der Gruppe Wechselunterricht ( $p = 0,022$ ). Die Überprüfung der Homogenität der Regressionssteigungen erfolgt durch die Bildung der entsprechenden Interaktionsterme zwischen den Gruppen und den Kovariaten (Schulart und Prä-KFT). Im Prä-Test war der Wert der Gruppe\*Prä-KFT Interaktion ( $F(3, 143) = 0,517$ ;  $p = 0,671$ ) und der Gruppe\*Schulart Interaktion ( $F(1, 143) = 0,824$ ;  $p = 0,336$ ) nicht signifikant.

Entsprechendes beim Post-Test mit  $F(3, 143) = 0,400$ ;  $p = 0,753$  für die Gruppe\*Prä-KFT Interaktion und  $F(1, 143) = 0,002$   $p = 0,968$  für die Gruppe\*Schulart Interaktion. Im Follow-up-Test betragen die Werte Gruppe\*Schulart Interaktion  $F(3, 107) = 1,282$ ;  $p = 0,284$  und Gruppe\*Prä-KFT  $F(3, 107) = 1,775$ ;  $p = 0,186$ . Zusammenfassend kann gesagt werden, dass zu allen drei Messzeitpunkten die Interaktionsterme für beide Kovariaten mit der Gruppe nicht signifikant sind und somit die Voraussetzung der Nutzung dieser gegeben ist ( $p > 0,05$ ).

Die Voraussetzung der Homogenität der Varianzen wurde geprüft (Anhang 39) und war für den Prä-Test ( $p = 0,207$ ) und Post-Test ( $p = 0,148$ ) gegeben, wie durch den Levene-Test basierend auf dem Median beurteilt wurde ( $p > 0,05$ ). Der Follow-up-Test zeigte keine Homogenität der Varianzen ( $p = 0,037$ ).

Anschließend wird eine ANOVA mit Messwiederholung und den Kovariaten gerechnet, um Interaktionen zwischen den Messzeitpunkten zu überprüfen (Tabelle 22). Der Interaktionseffekt von Gruppe\*Messzeitpunkt (gruppe\_4:mzp) ist hierbei signifikant  $F(5,68, 211,99) = 2,603$ ;  $p = 0,021$ . Neben dem Interaktionseffekt ist auch die Schulart (schule) signifikant  $F(1,00, 112,00) = 11,987$ ;  $p < 0,001$ .

Tabelle 22: ANOVA Tabelle

<i>ANOVA Table (type III tests)</i>						
Effect	DFn	DFd	F	p	p < 0,05	ges
pr_KFT	1,00	112,00	0,482	0.489000		0,003
schule	1,00	112,00	11,987	0.000759	*	0,062
gruppe_4	3,00	112,00	1,724	0.166000		0,028
mzp	1,89	211,99	2,965	0.057000		0,01
pr_KFT:mzp	1,89	211,99	0,921	0.395000		0,003
schule:mzp	1,89	211,99	2,692	0.073000		0,009
gruppe_4:mzp	5,68	211,99	2,603	0.021000	*	0,026

Die aus der Rasch-Analyse erhaltenen Daten aus dem Prä-, Post- und Follow-up-Fragebogen sind in Tabelle 23 mit ihren Standardabweichungen und Konfidenzintervallen dargestellt.

Tabelle 23: Deskriptive Statistik der Mittelwerte der Fachwissensfragebögen für Prä-, Post- und Follow-up-Test

Messzeitpunkte	Präsenzunterricht (PP)			Flipped Classroom (DP)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Prä	43	-0,785 (0,795)	[-1,030; -0,540]	34	-0,577 (0,943)	[-0,906; -0,248]
Post	43	-0,441 (1,514)	[-0,808; 0,124]	34	0,577 (1,772)	[-0,040; 1,196]
Follow-up	32	-0,630 (0,829)	[-0,928; -0,331]	27	-0,356 (1,176)	[-0,822; 0,108]

Weiterführung Tabelle 23

Messzeitpunkte	Wechselunterricht (PD)			Homeschooling (DD)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Prä	44	-0,316 (1,079)	[-0,645; 0,011]	33	-0,873 (0,888)	[-1,188; -0,558]
Post	44	0,145 (1,145)	[-0,203; 0,493]	33	0,0248 (1,387)	[-0,467; 0,517]
Follow-up	29	-1,328 (1,409)	[-1,864; -0,792]	30	-1,098 (1,335)	[-1,596; -0,599]

Die erhaltenen Daten werden durch die Kovariaten Schulart und den Prä-KFT bereinigt. Die beiden Kovariaten haben Einfluss auf die abhängige Variable und erklären somit einen Teil des Fehlers. Die erhaltenen adjustierten Mittelwerte (Schätzungen) für den Prä-, Post- und Follow-up-Fragebogen des Fachwissens dienen den nachfolgenden Berechnungen und Betrachtungen als Grundlage (s. Tabelle 24).

Die Einbeziehung der Ergebnisse des Prä-Tests als Kovariate zeigte keine Veränderung der Ergebnisse und wurde daher in nachfolgenden Berechnungen nicht berücksichtigt.

Tabelle 24: Schätzungen der Mittelwerte der Fachwissensfragebögen für Prä-, Post- und Follow-up-Test (Ergebnis der Kovarianzanalyse)

Schätzungen der Mittelwerte der Personenfähigkeiten des Fachwissens (Vergleich der Gruppen)

Messzeitpunkte	Präsenzunterricht (PP)			Flipped Classroom (DP)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Prä	43	-0,834 <sup>a</sup> (0,152)	[-1,135; -0,533]	34	-0,637 <sup>a</sup> (0,170)	[-0,973; -0,302]
Post	43	-0,618 <sup>a</sup> (0,227)	[-1,067; -0,169]	34	0,272 <sup>a</sup> (0,253)	[-0,229; 0,772]
Follow-up	32	-0,951 <sup>a</sup> (0,216)	[-1,379; -0,524]	27	-0,715 <sup>a</sup> (0,232)	[-1,175; -0,254]

Messzeitpunkte	Wechselunterricht (PD)			Homeschooling (DD)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Prä	44	-0,269 <sup>a</sup> (0,148)	[-0,561; 0,023]	33	-0,812 <sup>a</sup> (0,180)	[-1,168; -0,456]
Post	44	0,376 <sup>a</sup> (0,22)	[-0,059; 0,812]	33	0,392 <sup>a</sup> (0,269)	[-0,139; 0,922]
Follow-up	29	-1,001 <sup>a</sup> (0,225)	[-1,447; -0,556]	30	-0,749 <sup>a</sup> (0,236)	[-1,217; -0,281]

a. Die Kovariaten in dem Modell werden anhand der folgenden Werte berechnet:

Prä/Post: pr\_KFT = ,6990, Schulart = 1,24; Follow-up: pr\_KFT = ,6786, Schulart = 1,28.

Um die Ergebnisse der ANOVA zu spezifizieren und Aussage darüber treffen zu können, zu welchen Messzeitpunkten welche Gruppen sich wie unterscheiden, werden nachfolgende Berechnungen durchgeführt.

Nachfolgend werden die Verfahren der einfaktoriellen ANCOVA (einfaktorielle Kovarianzanalyse) zur Überprüfung der Gruppenunterschiede zu den jeweiligen Messzeitpunkten (s. 7.2.1 *Vergleich der Gruppen*) und einer ANOVA mit Messwiederholung und Kovariaten zum Vergleich der Messzeitpunkte der einzelnen Gruppen (s. 7.2.2 *Vergleich der Messzeitpunkte*) genutzt.

### 7.2.1 Vergleich der Gruppen

Betrachtet man die Zwischensubjekteffekte der Gruppen je Messzeitpunkt nach Bereinigung um die Schulart und den KFT-Prä-Test zeigt sich, dass zum Messzeitpunkt des Prä-Tests im Bereich des Fachwissens ein signifikanter Unterschied bezüglich der Mittelwerte über die Gruppen zu verzeichnen ist:  $F(3, 148) = 3,085$ ,  $p = 0,029$ ,  $\eta^2 = 0,059$  (Anhang 40). Dies tritt auch beim Post-Test auf, mit  $F(3, 148) = 4,199$ ,  $p = 0,007$ ,  $\eta^2 = 0,078$  (Anhang 42). Beim Messzeitpunkt Follow-up tritt hingegen kein signifikanter Unterschied auf;  $F(3, 112) = 0,426$ ,  $p = 0,735$ ,  $\eta^2 = 0,011$ , wie in der Tabelle Anhang 44 zu entnehmen ist.

Nachfolgend werden die Gruppen mit den aus der Kovarianzanalyse enthaltenen Daten zu den einzelnen Messzeitpunkten verglichen.

Betrachtet man den ersten Testzeitpunkt (**Prä-Fragebogen**, pr\_fw) wie in Abbildung 65 dargestellt (s. Tabelle 24), hat die Personenfähigkeit im Fachwissen der Gruppe Präsenzunterricht (PP) einen Mittelwert von  $-0,834$  ( $SD = 0,152$ ). Die Gruppe Flipped Classroom (DP) liegt mit einer mittleren Personenfähigkeit von  $-0,637$  ( $SD = 0,170$ ), sowie auch die Gruppe Wechselunterricht (PD) mit  $-0,269$  ( $SD = 0,148$ ) etwas höher. Die Gruppe Homeschooling (DD) zeigt hingegen eine mittlere Personenfähigkeit ähnlich der Gruppe Präsenzunterricht mit einem Mittelwert von  $-0,812$  ( $SD = 0,180$ ). Die Schätzungen zur Abbildung 65 sind in Tabelle 24 zu finden. Die Bonferroni-korrigierte Post-hoc-Analyse zeigt für den Prä-Fragebogen keinen signifikanten Unterschied zwischen den Personenfähigkeiten der einzelnen Gruppen (Anhang 41).

Beim zweiten Messzeitpunkt, dem **Post-Fragebogen** (po\_fw), sind die Personenfähigkeiten des Fachwissens, wie in Abbildung 66 dargestellt, verteilt. Die Gruppe Präsenzunterricht (PP) hat einen Mittelwert von  $-0,618$  ( $SD = 0,227$ ) und liegt somit unter allen anderen Mittelwerten. Die Mittelwerte der anderen Gruppen Flipped Classroom (DP)  $0,272$  ( $SD = 0,253$ ), Wechselunterricht (PD)  $0,376$  ( $SD = 0,220$ ) und Homeschooling (DD)  $0,392$  ( $SD = 0,268$ ) zeigen eine höhere Personenfähigkeit. Die Schätzungen zur Abbildung 66 sind in Tabelle 24 zu finden. Die Bonferroni-korrigierte Post-hoc-Analyse zeigt signifikante Unterschiede zwischen der Gruppe Präsenzunterricht (PP) und den anderen drei Gruppen mit entsprechend höheren mittleren Personenfähigkeiten. Zwischen der Gruppe Präsenzunterricht (PP) und der Gruppe Flipped Classroom (DP) mit den Werten  $p = 0,039$ ,  $MDiff = -0,89$ , 95%-KI  $[-1,75; -0,27]$ . Den Gruppen Präsenzunterricht (PP) und Wechselunterricht (PD) mit  $p = 0,018$ ,  $MDiff = -0,99$ , 95%-KI  $[-1,87; -0,12]$  und zwischen den Gruppen Präsenzunterricht (PP) und Homeschooling (DD) mit einem Wert von  $p = 0,048$ ,  $MDiff = -1,01$ , 95%-KI  $[-2,01; 0,00]$ . Zwischen den weiteren Gruppen liegen keine signifikanten Unterschiede in den Personenfähigkeiten vor (Anhang 43).

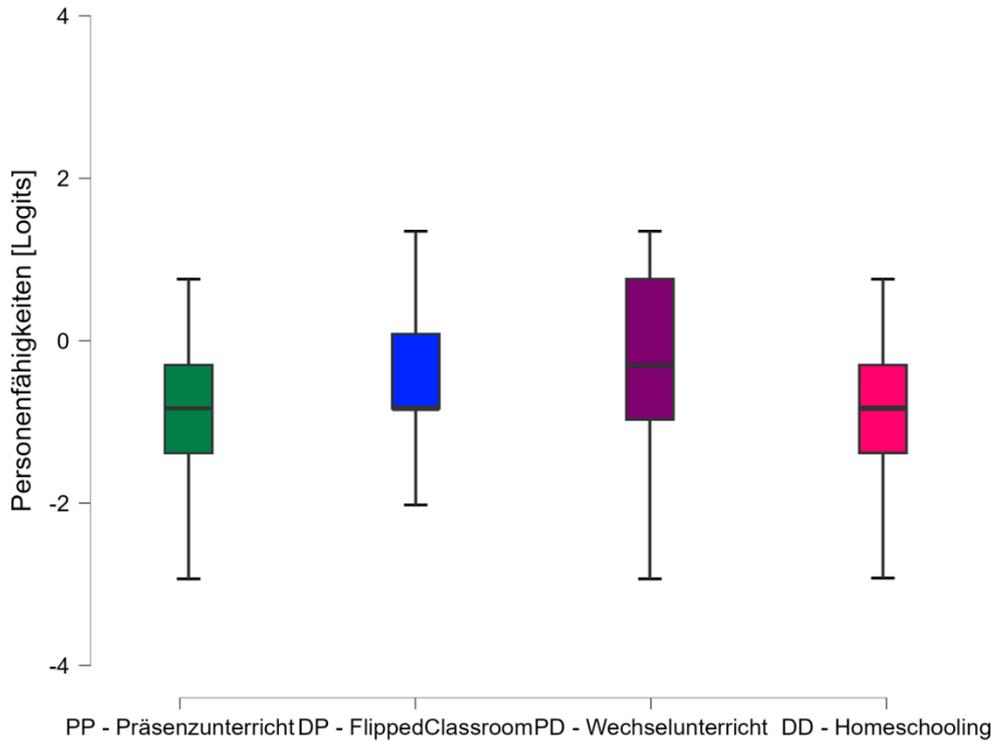


Abbildung 65: Gruppenvergleiche Fachwissen Prä-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,699$ ,  $Schulart = 1,24$ ; Fehlerbalken: 95 % KI;

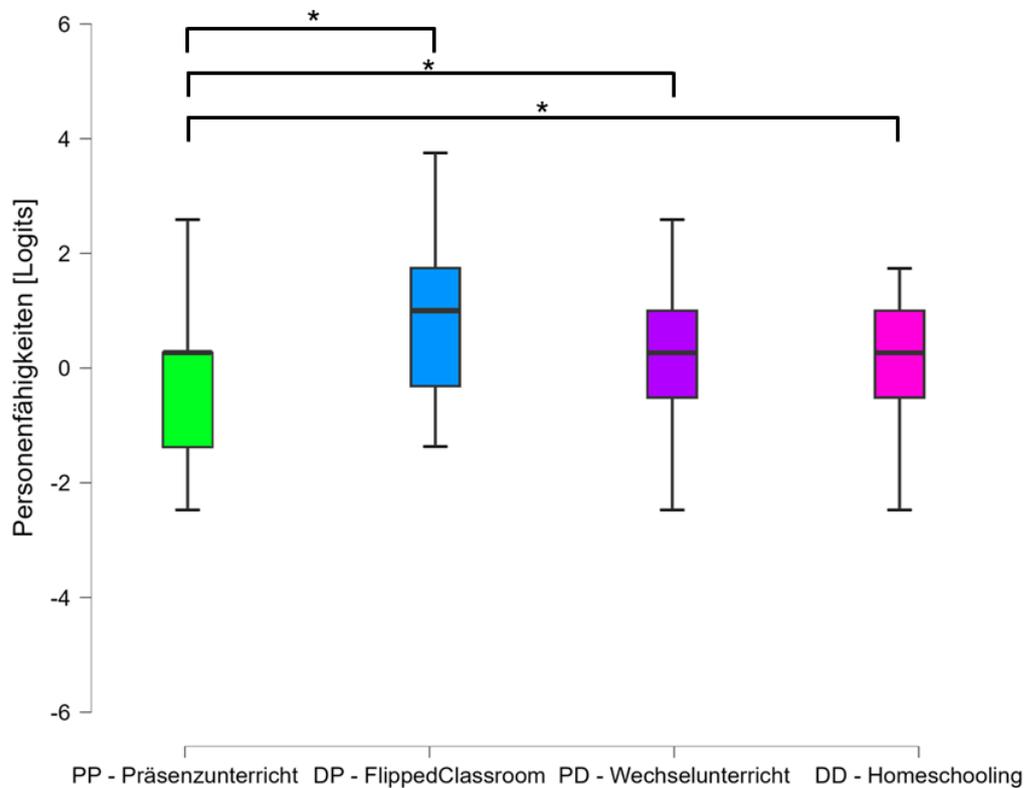


Abbildung 66: Gruppenvergleiche Fachwissen Post-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ ,  $Schulart = 1,24$ . Fehlerbalken: 95 % KI

Beim letzten Testzeitpunkt (**Follow-up-Fragebogen**, fo\_fw), sind die Personenfähigkeiten des Fachwissens, wie in Abbildung 67 dargestellt, verteilt. Die Gruppe Präsenzunterricht (PP) hat einen Mittelwert von -0,951 ( $SD = 0,216$ ) und liegt somit unter allen anderen Mittelwerten. Die Mittelwerte der anderen Gruppen Flipped Classroom (DP) -0,715 ( $SD = 0,232$ ), Wechselunterricht (PD) -1,001 ( $SD = 0,225$ ) und Homeschooling (DD) -0,749 ( $SD = 0,236$ ) zeigen eine höhere Personenfähigkeit. Die Schätzungen zur Abbildung 67 sind in Tabelle 24 zu finden. Die Bonferroni-korrigierte Post-hoc-Analyse zeigt zum dritten Messzeitpunkt keinen signifikanten Unterschied zwischen den Personenfähigkeiten der Gruppen (Anhang 45). In Abbildung 67 sind die Ergebnisse der ANCOVA der vier Testgruppen als Boxplots aufgetragen.

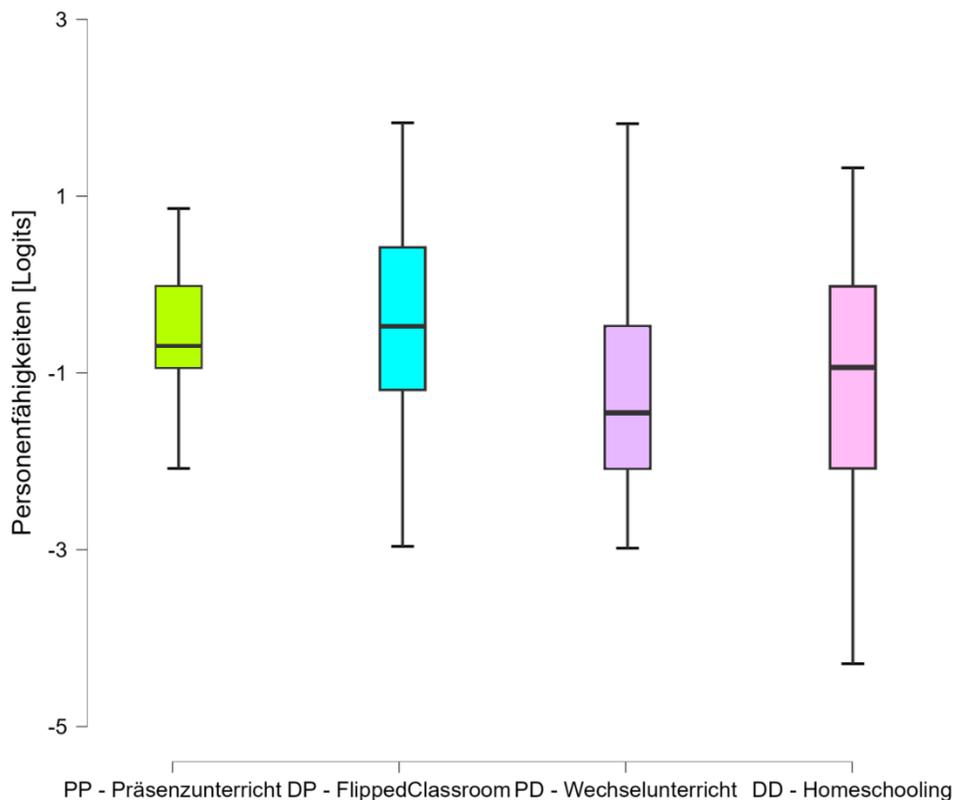


Abbildung 67: Gruppenvergleiche Fachwissen Follow-up-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr\_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI

## 7.2.2 Vergleich der Messzeitpunkte

Zum Vergleich einer Gruppe über die Messzeitpunkte hinweg, werden nur die Daten der Teilnehmer\*innen genutzt, die zu allen Messzeitpunkten an der Erhebung teilgenommen haben (s. Tabelle 25). Bei den Gruppen Präsenzunterricht und Flipped Classroom wurde der Prä-KFT Wert mit aufgenommen. Bei den Gruppen Wechselunterricht und Homeschooling zusätzlich noch die Kovariate Schulart.

Tabelle 25: Schätzungen der Mittelwerte der Fachwissensfragebögen für die vier Gruppen über die Messzeitpunkte hinweg (ANOVA mit Messwiederholung und Kovariaten)

*Schätzungen der Mittelwerte der Personenfähigkeiten des Fachwissens (Vergleich der Messzeitpunkte)*

Gruppeneinteilung	Prä			Post		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Präsenzunterricht	32	-0,892 <sup>a</sup> (1,045)	[-1,258; -0,526]	32	-0,735 <sup>a</sup> (1,604)	[-1,297; -0,173]
Flipped Classroom	27	-0,737 <sup>a</sup> (1,034)	[-1,131; -0,343]	27	0,113 <sup>a</sup> (1,587)	[-0,492; 0,719]
Wechselunterricht	29	-0,232 <sup>a</sup> (1,037)	[-0,613; 0,150]	29	0,258 <sup>a</sup> (1,591)	[-0,328; 0,843]
Homeschooling	30	-0,695 <sup>a</sup> (1,108)	[-1,095; -0,294]	30	0,463 <sup>a</sup> (1,70)	[-0,152; 1,078]

Gruppeneinteilung	Follow-up		
	N	M (SD)	95% KI
Präsenzunterricht	32	-0,951 <sup>a</sup> (0,216)	[-1,379; -0,524]
Flipped Classroom	27	-0,715 <sup>a</sup> (0,232)	[-1,175; -0,254]
Wechselunterricht	29	-1,001 <sup>a</sup> (0,225)	[-1,447; -0,556]
Homeschooling	30	-0,749 <sup>a</sup> (0,236)	[-1,217; -0,281]

a. Die Kovariaten in dem Modell werden anhand der folgenden Werte berechnet: pr\_KFT = 0,6786, Schulart = 1,28.

**Präsenzunterricht** (N = 32)

Der Mauchly-Test auf Sphärizität wurde durchgeführt. Dabei wurde eine Verletzung der Voraussetzung der Sphärizität festgestellt, weshalb eine Korrektur der Freiheitsgrade vorgenommen wurde. Da das Epsilon nach Greenhouse-Geisser  $> 0,75$  ist, wird die Korrektur nach Huynh-Feldt eingesetzt (Anhang 47). Die geschätzten Werte der Gruppe Präsenzunterricht für die drei Messzeitpunkte sind in Abbildung 68 aufgetragen:

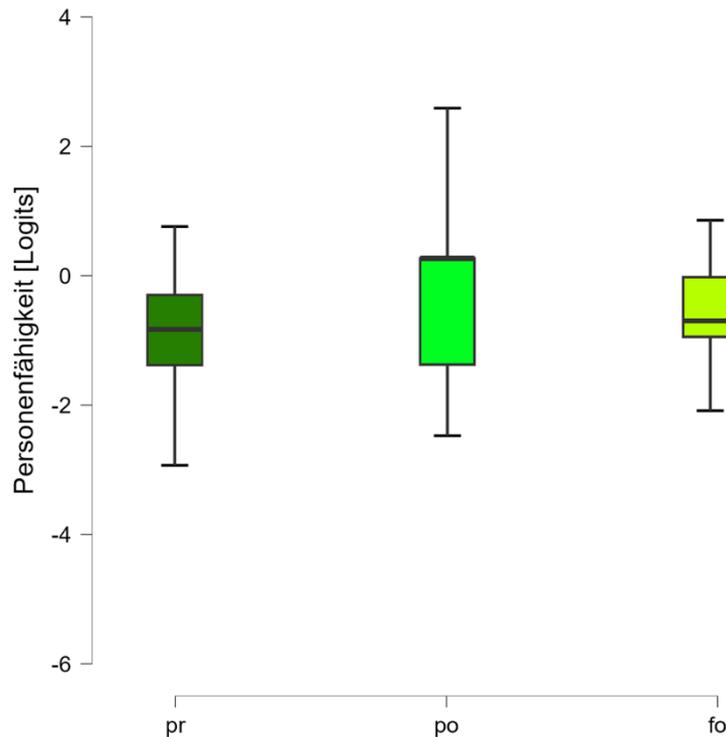


Abbildung 68: Vergleich der Messzeitpunkte - Fachwissen Präsenzunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI

Es gab keinen statistisch signifikanten Unterschied über die Messzeitpunkte hinweg,  $F(1,719, 51,568) = 1,232$ ,  $p = 0,39$ ,  $\eta^2 = 0,039$ . Ein Bonferroni-korrigierter post-hoc-Test zeigte keine signifikanten ( $p < 0,05$ ) Unterschiede zwischen den einzelnen Messzeitpunkten (Anhang 48).

**Flipped Classroom** (N = 27)

Der Mauchly-Test auf Sphärizität wurde durchgeführt. Dabei wurde keine Verletzung der Voraussetzung der Sphärizität festgestellt (Anhang 50). Die geschätzten Werte der Gruppe Flipped Classroom für die drei Messzeitpunkte sind in Abbildung 69 aufgetragen:

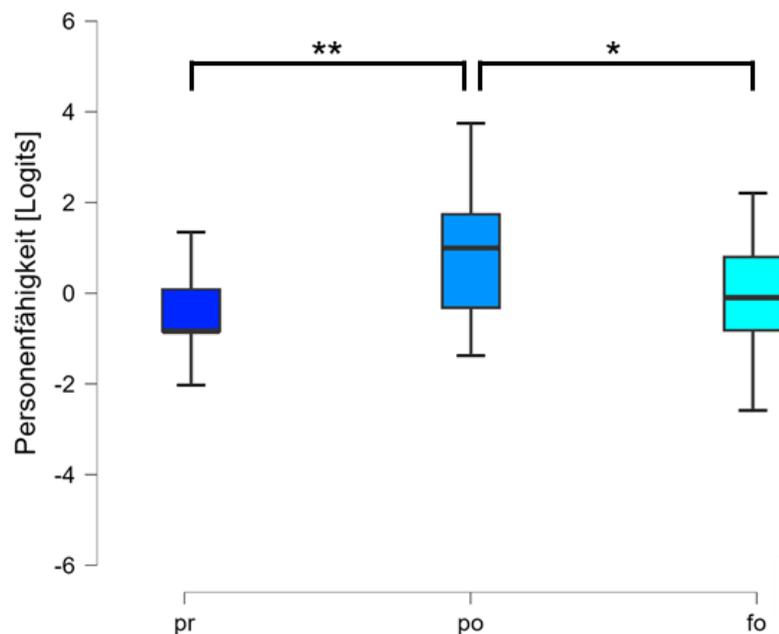


Abbildung 69: Vergleich über die Messzeitpunkte- Fachwissen Flipped Classroom (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI

Es gab keinen statistisch signifikanten Unterschied über die Messzeitpunkte,  $F(2, 50) = 0,116$ ,  $p = 0,891$ ,  $\eta^2 = 0,005$ . Ein Bonferroni-korrigierter post-hoc-Test zeigt signifikante Unterschiede ( $p < 0,05$ ) zwischen zwei Messzeitpunkten (Anhang 51). Zwischen Messzeitpunkt Prä-Test und Post-Test ( $p = 0,005$ ,  $MDiff = -1,10$ , 95%-KI[-1,922; -0,293]) sowie zwischen Post-Test und Follow-up-Test ( $p = 0,012$ ,  $MDiff = 0,85$ , 95%-KI[0,159; -1,539]) liegen signifikante Unterschiede vor.

**Wechselunterricht** (N = 29)

Der Mauchly-Test auf Sphärizität wurde durchgeführt. Dabei wurde keine Verletzung der Voraussetzung der Sphärizität festgestellt (Anhang 53). Die geschätzten Werte der Gruppe Wechselunterricht für die drei Messzeitpunkte sind in Abbildung 70 aufgetragen:

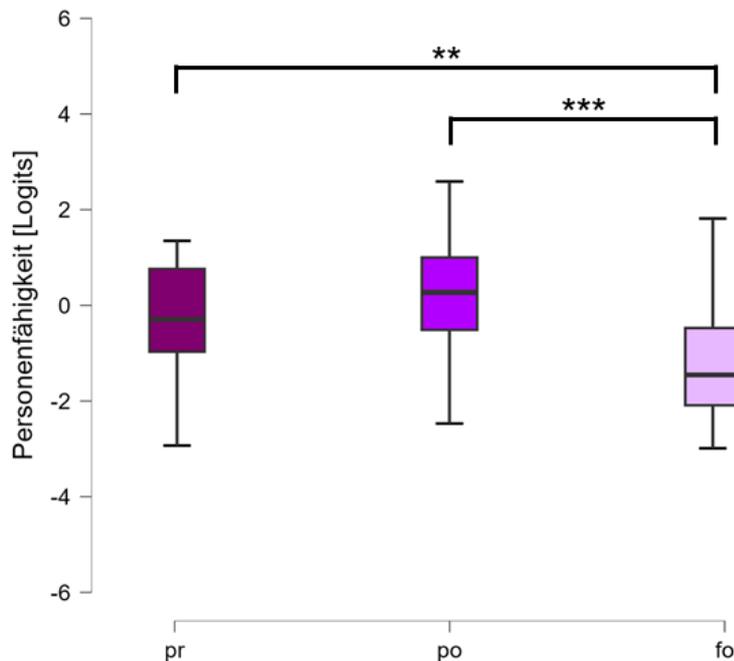


Abbildung 70: Vergleich über die Messzeitpunkte; FW Wechselunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr\_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI

Es gab keinen statistisch signifikanten Unterschied über die drei Messzeitpunkte hinweg,  $F(2, 52) = 0,735$ ,  $p = 0,484$ ,  $\eta^2 = 0,027$ . Ein Bonferroni-korrigierter post-hoc-Test zeigt signifikante ( $p < 0,05$ ) Unterschiede zwischen den Messzeitpunkten (Anhang 54). Zwischen Messzeitpunkt Prä-Test und Follow-up-Test ( $p = 0,002$ ,  $MDiff = 1,01$ , 95%-KI[0,33; -1,68]) sowie zwischen Post-Test und Follow-up-Test ( $p < 0,001$ ,  $MDiff = 1,30$ , 95%-KI[0,59; 2,00]) liegen signifikante Unterschiede vor.

**Homeschooling** (N = 30)

Der Mauchly-Test auf Sphärizität wurde durchgeführt. Dabei wurde keine Verletzung der Voraussetzung der Sphärizität festgestellt (Anhang 56). Die geschätzten Mittelwerte der Gruppe Homeschooling für die drei Messzeitpunkte sind in Abbildung 71 aufgetragen:

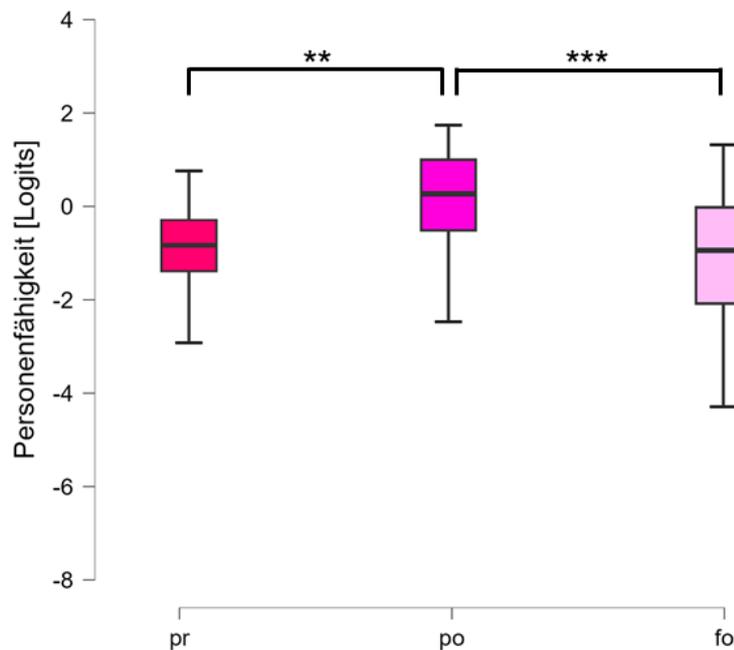


Abbildung 71: Vergleich über die Messzeitpunkte; FW Homeschooling (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr\_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI

Es gab keinen statistisch signifikanten Unterschied über die Messzeitpunkte,  $F(2, 54) = 2,910$ ,  $p = 0,063$ ,  $\eta^2 = 0,097$ . Ein Bonferroni-korrigierter post-hoc-Test zeigt signifikante ( $p < 0,05$ ) Unterschiede zwischen den Messzeitpunkten (Anhang 57). Zwischen Messzeitpunkt Prä-Test und Post-Test ( $p = 0,007$ ,  $MDiff = -0,87$ , 95%-KI[-1,52, -0,206]) sowie zwischen Post-Test und Follow-up-Test ( $p < 0,001$ ,  $MDiff = 1,10$ , 95%-KI[0,56, 1,65]) liegen signifikante Unterschiede vor.

### 7.3 Naturwissenschaftliche Arbeitsweise

Zunächst wird für beide NWA-Skalen (Post und Follow-up) eine explorative Datenanalyse mit den Werten der Personenfähigkeiten aus der Rasch-Analyse durchgeführt, um Ausreißer zu identifizieren und die Normalverteilung zu prüfen. Bei den Daten der Grafik handelt es sich um die Personenfähigkeiten, die sich aus der Rasch-Analyse ergaben.

Die Datensätze weisen einige leichte Ausreißer auf (Abbildung 72). Betrachtet man diese, zeigen sich über alle drei Messzeitpunkte hinweg für die Gruppen die nachfolgende Anzahl. Die Gruppe Präsenzunterricht zeigt sechs Ausreißer. In der Gruppe Flipped Classroom sind es vier Ausreißer und in den Gruppen Wechselunterricht ist es eine Person. Die Gruppe Homeschooling besitzt keine Ausreißer. Da es sich bei allen Ausreißern um leichte Ausreißer handelt, werden alle Fälle für die weiteren Analysen genutzt und keine Fälle ausgeschlossen. Der extreme Ausreißer (mehr als 3-fache des Interquartilsabstands) in der Gruppe Flipped Classroom wird ebenfalls behalten, da er in der Rasch-Analyse unauffällig war.

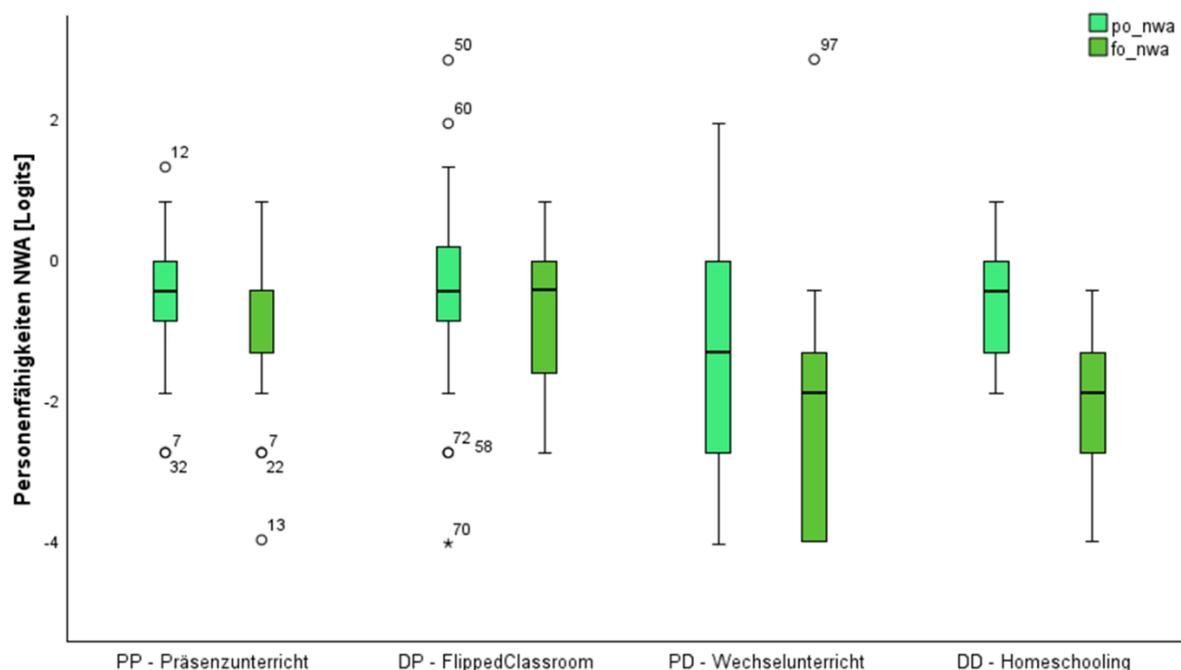


Abbildung 72: Überprüfung auf univariate Ausreißer

Die Testung auf Normalverteilung zeigte gemäß dem Shapiro-Wilk-Test vier nicht normalverteilte Testgruppen ( $p < 0,05$ ) (Anhang 58) über die Testzeitpunkte hinweg. Tritt ein signifikanter Wert bei diesem Test auf, spricht man davon, dass die Annahme der Normalverteilung verletzt wurde. Dies ist bei den Ergebnissen des Post-Tests in der Gruppe Präsenzunterricht ( $p = 0,024$ ) und Homeschooling ( $p = 0,031$ ) der Fall. Ebenso im Follow-up-Test der Gruppen Wechselunterricht ( $p = 0,002$ ) und Homeschooling ( $p = 0,029$ ). Die

Überprüfung der Homogenität der Regressionssteigungen erfolgt durch die Bildung der entsprechenden Interaktionsterme zwischen den Gruppen und den Kovariaten (Schulart und Prä-KFT). Im Post-Test war der Wert der Gruppe\*Prä-KFT Interaktion ( $F(3, 145) = 1,389$ ;  $p = 0,249$ ) und der Gruppe\*Schulart Interaktion ( $F(3, 145) = 2,417$ ;  $p = 0,069$ ) nicht signifikant. Im Follow-up-Test betragen der Wert Gruppe\*Schulart Interaktion  $F(3, 109) = 2,606$ ;  $p = 0,142$  und Gruppe\*Prä-KFT  $F(3, 109) = 1,196$ ;  $p = 0,315$ . Zusammenfassend kann gesagt werden, dass zu allen drei Messzeitpunkten die Interaktionsterme für beide Kovariaten mit der Gruppe nicht signifikant und somit die Voraussetzung der Nutzung dieser gegeben ist ( $p > 0,05$ ). Die Voraussetzung der Homogenität der Varianzen wurde geprüft (Anhang 59) und war für den Post-Test ( $p = 0,012$ ) und den Follow-up-Test ( $p = 0,013$ ) nicht gegeben, wie durch den Levene-Test basierend auf dem Median beurteilt wurde ( $p > 0,05$ ). ANCOVAS sind bei der Verletzung dieser Annahme sehr robust, vor allem wenn die Gruppengrößen ähnlich sind (Huitema, 2011; Ito, 1980; Salkind & Rasmussen, 2007; Shields, 1978).

Anschließend wird eine ANOVA mit Messwiederholung und den Kovariaten gerechnet, um Interaktionen zwischen den Messzeitpunkten zu überprüfen (Tabelle 26). Der Interaktionseffekt von Gruppe\*Messzeitpunkt (gruppe\_4:mzp) ist hierbei signifikant mit  $F(3, 112) = 5,115$ ;  $p = 0,002$ .

Tabelle 26: ANOVA Tabelle

<i>ANOVA Table (type III tests)</i>						
Effect	DFn	DFd	F	p	p < 0,05	ges
pr_KFT	1	112	2,693	0,104		0,018
schule	1	112	1,811	0,181		0,012
gruppe_4	3	112	2,181	0,094		0,042
mzp	1	112	3,169	0,078		0,007
pr_KFT:mzp	1	112	0,516	0,474		0,001
schule:mzp	1	112	3,037	0,084		0,007
gruppe_4:mzp	3	112	5,115	0,002	*	0,033

Die aus der Rasch-Analyse erhaltenen Daten aus dem Post- und Follow-up-Fragebogen sind in Tabelle 27 mit ihren Standardabweichungen und Konfidenzintervallen dargestellt.

Tabelle 27: Deskriptive Statistik der Mittelwerte der NWA-Skalen für Post- und Follow-up-Test

<i>Mittelwerte Personenfähigkeiten der naturwissenschaftlichen Arbeitsweisen (Vergleich der Gruppen)</i>						
Messzeitpunkte	Präsenzunterricht (PP)			Flipped Classroom (DP)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Post	43	-0,580 (0,932)	[-0,867; -0,293]	34	-0,376 (1,352)	[-0,848; 0,095]
Follow-up	32	-1,065 (1,034)	[-1,438; -0,691]	27	-0,840 (0,950)	[-1,216; -0,464]

Weiterführung Tabelle 27

Messzeitpunkte	Wechselunterricht (PD)			Homeschooling (DD)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Post	44	-1,120 (1,577)	[-1,599; -0,640]	33	-0,608 (0,834)	[-0,904; -0,315]
Follow-up	29	-2,140 (1,654)	[-2,770; -1,511]	30	-1,998 (0,986)	[-2,366; -1,629]

Die erhaltenen Daten aus dem Post- und Follow-up-Fragebogen werden durch die Kovariaten Schulart und den Prä-KFT bereinigt. Die beiden Kovariaten haben Einfluss auf die abhängige Variable und erklären somit einen Teil des Fehlers. Die erhaltenen adjustierten Mittelwerte (Schätzungen) für den Post- und Follow-up-Fragebogen der naturwissenschaftlichen Arbeitsweise dienen den nachfolgenden Berechnungen und Betrachtungen als Grundlage. Diese Mittelwerte sind mit ihren Standardabweichungen und Konfidenzintervallen in Tabelle 28 zusammengefasst:

Tabelle 28: Schätzungen der Mittelwerte der NWA-Skalen

*Schätzungen der Mittelwerte der Personenfähigkeiten der naturwissenschaftlichen Arbeitsweisen (Vergleich der Gruppen)*

Messzeitpunkte	Präsenzunterricht (PP)			Flipped Classroom (DP)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Post	43	-0,297 <sup>a</sup> (0,390)	[-1,068; 0,475]	34	-0,312 <sup>a</sup> (0,474)	[-1,249; 0,625]
Follow-up	32	-0,644 <sup>a</sup> (0,497)	[-1,629; 0,341]	27	-0,862 <sup>a</sup> (0,629)	[-2,109; 0,385]

Messzeitpunkte	Wechselunterricht (PD)			Homeschooling (DD)		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Post	44	-0,811 <sup>a</sup> (0,190)	[-1,186; -0,435]	33	-0,702 <sup>a</sup> (0,229)	[-1,154; -0,250]
Follow-up	29	-1,894 <sup>a</sup> (0,254)	[-2,397; -1,391]	30	-2,083 <sup>a</sup> (0,245)	[-2,569; -1,597]

a. Die Kovariaten in dem Modell werden anhand der folgenden Werte berechnet:

Post: pr\_KFT = ,6990, Schulart = 1,24; Follow-up: pr\_KFT = ,6786, Schulart = 1,28.

Um die Ergebnisse der ANOVA zu spezifizieren und Aussage darüber treffen zu können, zu welchen Messzeitpunkten welche Gruppen sich wie unterscheiden, werden nachfolgende Berechnungen durchgeführt.

### 7.3.1 Vergleich der Gruppen

Betrachtet man die Zwischensubjekteffekte der Gruppen je Messzeitpunkte nach Bereinigung um die Schulart und den KFT-Prä-Test zeigt sich, dass zum Messzeitpunkt des Post-Tests im Bereich des Wissens zur naturwissenschaftlichen Arbeitsweise kein signifikanter Unterschied bezüglich der Mittelwerte über die Gruppen zu verzeichnen ist:  $F(3, 148) = 1,123$ ,  $p = 0,342$ ,  $\eta^2 = 0,022$  (Anhang 60). Beim Messzeitpunkt Follow-up tritt hingegen ein signifikanter Unterschied auf:  $F(3, 112) = 4,138$ ,  $p = 0,008$ ,  $\eta^2 = 0,100$ , wie in der Tabelle Anhang 62 zu entnehmen ist.

Nachfolgend werden die Gruppen mit den aus der Kovarianzanalyse enthaltenen Daten zu den einzelnen Messzeitpunkten verglichen.

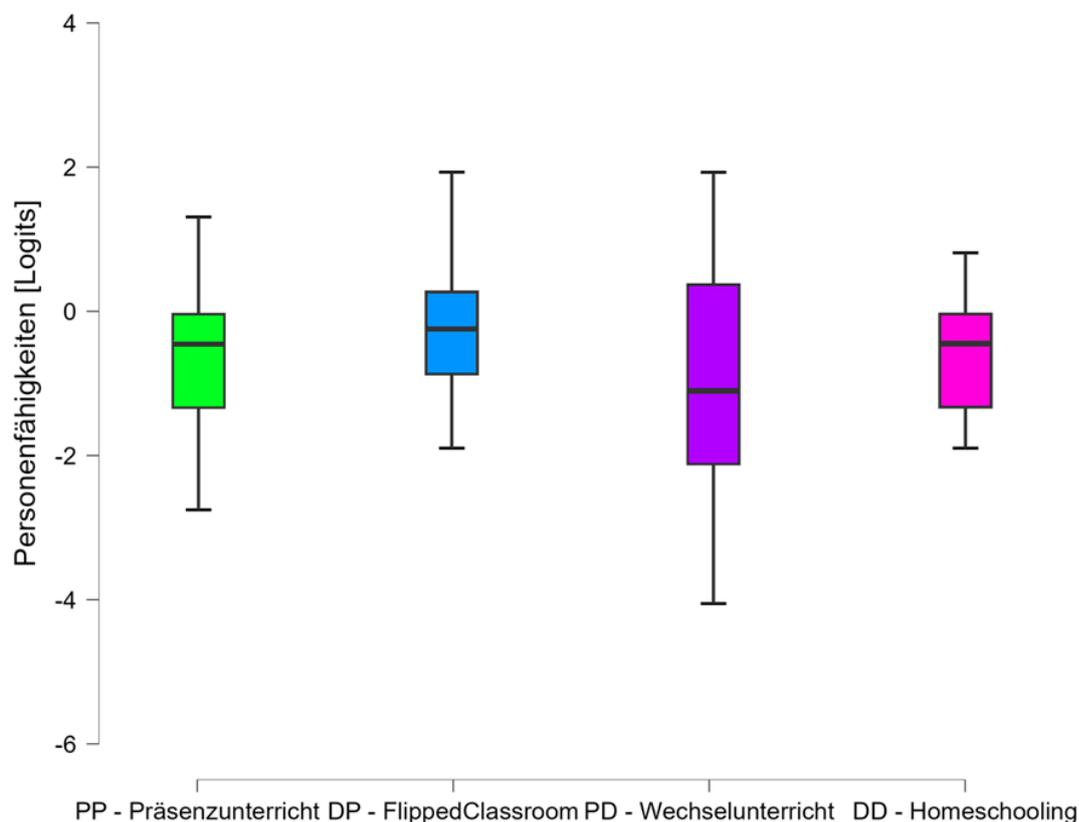


Abbildung 73: Gruppenvergleiche Naturwissenschaftlicher-Arbeitsweise Post-Test. (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ ,  $Schulart = 1,24$ . Fehlerbalken: 95 % KI

Mit Blick auf den ersten Testzeitpunkt (**Post-Fragebogen**,  $po\_nwa$ ), wie in Abbildung 73 dargestellt, hat die Personenfähigkeit bzgl. der naturwissenschaftlichen Arbeitsweise der Gruppe Präsenzunterricht (PP) einen Mittelwert von  $M = -0,297$  ( $SD = 0,390$ ). Die Gruppe Flipped Classroom (DP) liegt mit einer mittleren Personenfähigkeit von  $M = 0,312$  ( $SD = 0,474$ ) ähnlich diesem. Die Gruppe Wechselunterricht (PD) mit  $M = -0,811$  ( $SD = 0,190$ ), sowie auch

die Gruppe Homeschooling (DD) zeigen hingegen eine niedrigere mittlere Personenfähigkeit mit einem Mittelwert von  $M = -0,702$  ( $SD = 0,229$ ). Die Schätzungen zur Abbildung 73 sind in Tabelle 28 zu finden. Die Bonferroni-korrigierte Post-hoc-Analyse zeigt für den Post-Fragebogen keinen signifikanten Unterschied zwischen den Personenfähigkeiten der einzelnen Gruppen (Anhang 61).

Betrachtet man den zweiten Testzeitpunkt (**Follow-up-Fragebogen**, fo\_nwa), wie in Abbildung 74 dargestellt, hat die Personenfähigkeit bzgl. der naturwissenschaftlichen Arbeitsweise der Gruppe Präsenzunterricht (PP) einen Mittelwert von  $M = -0,644$  ( $SD = 0,497$ ). Die Gruppe Flipped Classroom (DP) liegt mit einer mittleren Personenfähigkeit von  $M = -0,862$  ( $SD = 0,629$ ) ähnlich diesem. Die Gruppe Wechselunterricht (PD) mit  $M = -1,894$  ( $SD = 0,254$ ), sowie auch die Gruppe Homeschooling (DD) zeigen hingegen eine niedrigere mittlere Personenfähigkeit mit einem Mittelwert von  $M = -2,083$  ( $SD = 0,245$ ). Die Schätzungen zur Abbildung 74 sind in Tabelle 28 zu finden. Die Bonferroni-korrigierte Post-hoc-Analyse zeigt für den Follow-up-Fragebogen keinen signifikanten Unterschied zwischen den Personenfähigkeiten der einzelnen Gruppen (Anhang 63).

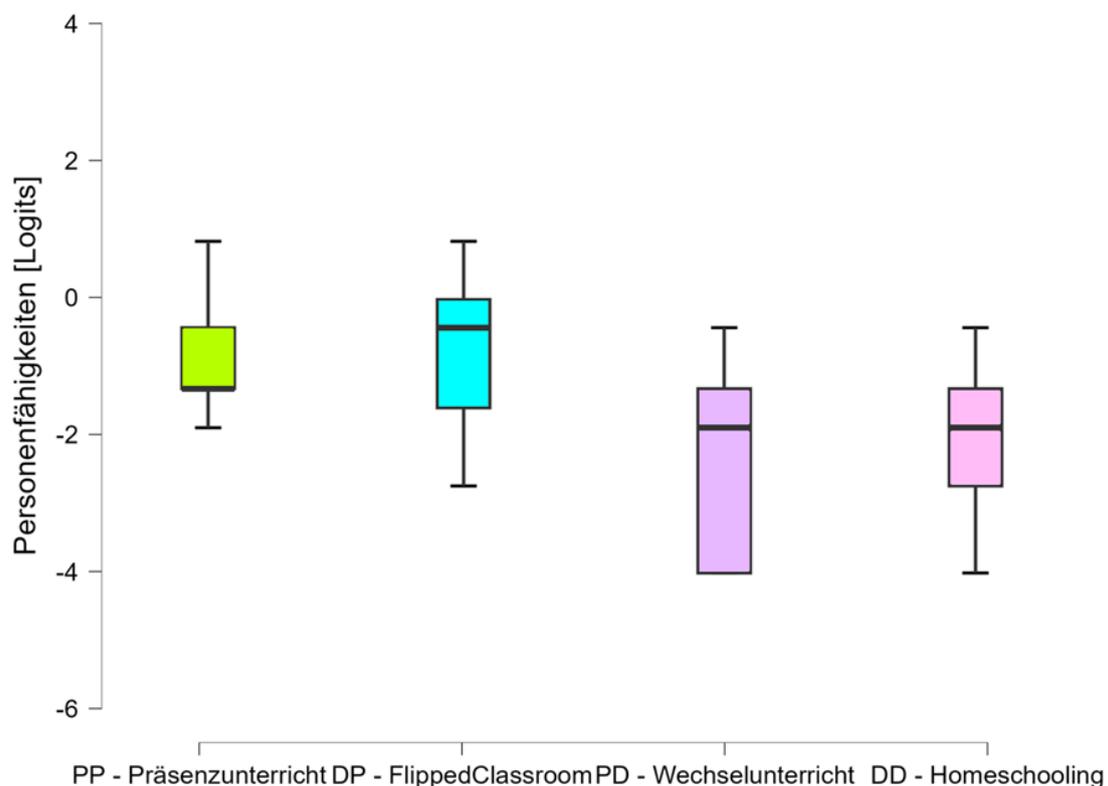


Abbildung 74: Gruppenvergleiche naturwissenschaftlicher Arbeitsweise Follow-up-Test. (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr\_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI

### 7.3.2 Vergleich der Messzeitpunkte

Zum Vergleich einer Gruppe über die Messzeitpunkte hinweg, werden nur die Daten der Teilnehmer\*innen genutzt, die zu beiden Messzeitpunkten an der Erhebung teilgenommen haben. Bei den Gruppen Präsenzunterricht und Flipped Classroom wurde der Prä-KFT Wert als Kovariate in die Berechnung mit aufgenommen. Bei den Gruppen Wechselunterricht und Homeschooling zusätzlich noch die Kovariate Schulart.

Tabelle 29: Schätzungen der Mittelwerte der Personenfähigkeiten aus den Fachwissensfragebögen für die vier Gruppen über die Messzeitpunkte hinweg (ANOVA mit Messwiederholung und Kovariaten)

*Schätzungen der Mittelwerte der Personenfähigkeiten der naturwissenschaftlichen Arbeitsweisen (Vergleich der Messzeitpunkte)*

Gruppeneinteilung	Post			Follow-up		
	N	M (SD)	95% KI	N	M (SD)	95% KI
Präsenzunterricht	32	-0,756 <sup>a</sup> (1,285)	[-1,206; -0,306]	32	-0,644 <sup>a</sup> (0,497)	[-1,629; 0,341]
Flipped Classroom	27	-0,630 <sup>a</sup> (1,272)	[-1,115; -0,145]	27	-0,862 <sup>a</sup> (0,629)	[-2,109; 0,385]
Wechselunterricht	29	-1,065 <sup>a</sup> (1,275)	[-1,534; -0,596]	29	-1,894 <sup>a</sup> (0,254)	[-2,397; -1,391]
Homeschooling	30	-0,337 <sup>a</sup> (1,362)	[-0,830; 0,156]	30	-2,083 <sup>a</sup> (0,245)	[-2,569; -1,597]

a. Die Kovariaten in dem Modell werden anhand der folgenden Werte berechnet: pr\_KFT = ,6786, Schulart = 1,28.

**Präsenzunterricht** (N = 32)

Die geschätzten Werte der Gruppe Präsenzunterricht (PP) für die zwei Messzeitpunkte sind in Abbildung 75 aufgetragen. Hierbei ist der Mittelwert  $M = -0,756$  ( $SD = 1,285$ ) für den Post-Test und  $M = -0,644$  ( $SD = 0,497$ ) für den Follow-up-Test.. Ein Bonferroni-korrigierter post-hoc-Test zeigt signifikante Unterschiede ( $p < 0,05$ ) zwischen den beiden Messzeitpunkten  $p = 0,001$ ,  $MDiff = 0,544$ , 95%-KI [-869; -0,173] (Anhang 65).

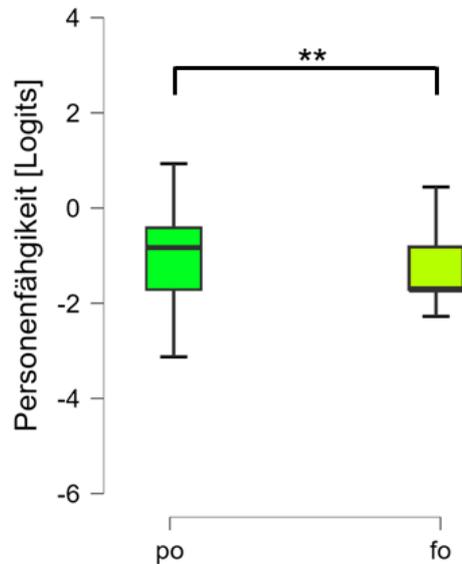


Abbildung 75: Vergleich über die Messzeitpunkte; NWA Präsenzunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI

**Flipped Classroom** (N = 27)

In Abbildung 76 sind die für die Gruppe Flipped Classroom (DP) geschätzten Werte für die beiden Messzeitpunkte (Post-Test und Follow-up-Test) aufgetragen. Hierbei ist der Mittelwert  $M = -0,630^a$  (SD = 1,272) für den Post-Test und  $M = -0,862$  (SD = 0,629) für den Follow-up-Test. Ein Bonferroni-korrigierter post-hoc-Test zeigt signifikante Unterschiede ( $p < 0,05$ ) zwischen den beiden Messzeitpunkten mit  $p = 0,045$ ,  $MDiff = 0,493$ , 95%-KI [0,011; 0,974] (Anhang 67).

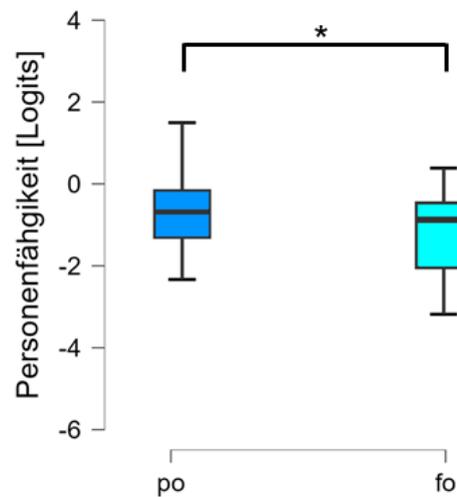


Abbildung 76: Vergleich über die Messzeitpunkte; NWA Flipped Classroom (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI

**Wechselunterricht** (N = 29)

Die geschätzten Werte der Gruppe Wechselunterricht (PD) für die zwei Messzeitpunkte sind in Abbildung 77 aufgetragen. Hierbei ist der Mittelwert  $M = -1,065$  ( $SD = 1,275$ ) für den Post-Test und  $M = -1,894$  ( $SD = 0,254$ ) für den Follow-up-Test. Zwischen den beiden Messzeitpunkten zeigt ein Bonferroni-korrigierter post-hoc-Test signifikante Unterschiede mit  $p = 0,007$ ,  $MDiff = 0,785$ , 95%-KI[0,236; 01,334] (Anhang 69).

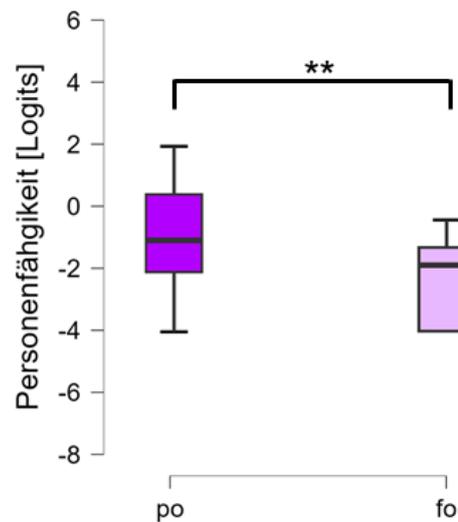


Abbildung 77: Vergleich über die Messzeitpunkte; NWA Wechselunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ ,  $Schulart = 1,24$ . Fehlerbalken: 95 % KI

**Homeschooling** (N = 30)

In Abbildung 78 sind die für die Gruppe Homeschooling (PD) geschätzten Werte für die beiden Messzeitpunkte (Post-Test und Follow-up-Test) aufgetragen. Hierbei ist der Mittelwert  $M = -0,337$  (1,362) für den Post-Test und  $M = -2,083$  (0,245) für den Follow-up-Test. Zwischen den beiden Messzeitpunkten zeigt ein Bonferroni-korrigierter post-hoc-Test signifikante Unterschiede:  $p < 0,001$ ,  $MDiff = 0,785$ , 95%-KI[0,236; 01,334] (Anhang 71).

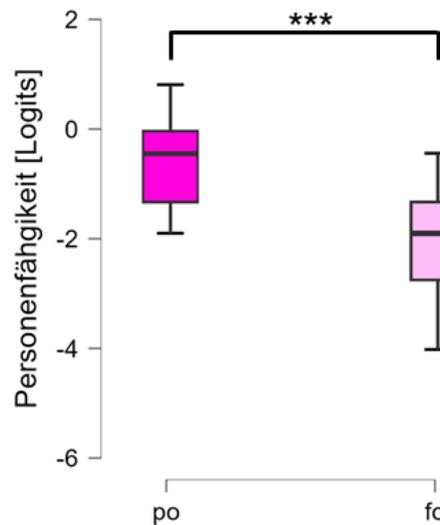


Abbildung 78: Vergleich über die Messzeitpunkte; NWA Homeschooling (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet:  $pr\_KFT = 0,0699$ ,  $Schulart = 1,24$ . Fehlerbalken: 95 % KI

## 8 Diskussion

### 8.1 Methodische Diskussion

Bei der methodischen Reflexion der vorliegenden Arbeit sind verschiedene Einschränkungen und Herausforderungen zu berücksichtigen, die im Folgenden diskutiert werden.

Die Teilnehmerzahl der Studie hätte idealerweise größer sein können, um eine höhere Generalisierbarkeit der Ergebnisse zu gewährleisten. Aufgrund von Genehmigungsanträgen (Kultusministerium, Schulleitung, Fachbetreuer\*innen, Elternbeirat, Eltern, Schüler\*innen) war es jedoch herausfordernd, teilnehmende Schulen zu akquirieren. Zudem war der Erhebungszeitraum durch den vorgegebenen Lehrplan begrenzt, was die zeitliche Flexibilität der Studiendurchführung einschränkte. Kurzfristige Erweiterungen des Teilnehmerkreises waren daher nicht möglich.

Ein weiterer limitierender Faktor war die Verfügbarkeit einiger Probanden zu allen Testzeitpunkten. Die Eingabe der Personencodes stellte für einige Schüler\*innen Problem dar, was zusätzliche Unstetigkeiten in den Grunddatensatz eingefügt hat und berücksichtigt werden musste. Die Stichprobengrößen der einzelnen Gruppen waren nicht ideal, aber ausreichend, um valide Aussagen treffen zu können. Die Zellenbesetzung war im Vergleich noch akzeptabel, was eine gute Vergleichbarkeit der verschiedenen Lernumgebungen untereinander ermöglicht.

Die Verteilung der Schüler\*innen auf die Gruppen bezogen auf ihre Schulzugehörigkeit (Gymnasium und Fachoberschule (FOS)) konnte in der Durchführung der Studie aufgrund des praktischen Schülersversuchs nicht anders umgesetzt werden, wodurch dies in der statistischen Auswertung berücksichtigt werden musste. Es zeigte sich im Verlauf der Studienauswertung, dass die Schüler\*innen der FOS im Vergleich zu den gymnasialen Schüler\*innen durchschnittlich eine geringere Vorwissensleistung aufwiesen. Dieser Befund kann auf den geringeren Stellenwert von Genetik und Gentechnik im Lehrplan der FOS zurückgeführt werden. Erfahrungen aus Schülerkursen am außerschulischen Lernort „Werner-Siemens-Schülerforschungszentrum“ der TU München in den Vorjahren zeigten aber, dass die FOS-Schüler\*innen nicht weniger Verständnis für die Thematik der Genetik und Gentechnik bei der Durchführung des Kriminalfalls aufwiesen als Schüler\*innen des Gymnasiums.

Eine weitere Herausforderung stellte die Entwicklung eines reliablen und validen Messinstruments dar. Zum Zeitpunkt der Erhebung gab es in der fachdidaktischen Forschung kein quantitatives Messinstrument, das die naturwissenschaftlichen Arbeitsweisen im Bereich der Biotechnologie und das entsprechende Fachwissen erfasst. Der erstellte NWA-Prä-Test zeigte, dass das allgemeine Verständnis der Schüler\*innen zu naturwissenschaftlichen

Arbeitsweisen geringer war als erhofft, wodurch dieser nach der Pilotierung verworfen werden musste.

Die Wahl des Multiple-Choice-Designs mit 2 aus 5 Antwortmöglichkeiten, um die Ratewahrscheinlichkeit der Teilnehmer\*innen zu reduzieren, könnte für einige Schüler\*innen eine Herausforderung dargestellt haben, ihr Wissen auf die auswählbaren Antworten zu übertragen.

Der Zeitrahmen für die Bearbeitung der Lernumgebungen zum Fachwissen (Klassenunterricht und Interactive Book) war mit 45 Minuten recht knapp bemessen. Die Durchführung in dieser Zeit war allerdings möglich, wie sich in der Durchführung der Studie gezeigt hat. Die zeitliche Beschränkung war der Verfügbarkeit eines Schulvormittags für die Studie geschuldet. Ein größerer Zeitrahmen könnte die Qualität der Lernumgebungen zur Vermittlung des theoretischen Wissens und damit die Validität der Ergebnisse positiv beeinflussen.

Der gewählte Durchführungszeitpunkt der Studie vor den Sommerferien könnte sich ungünstig auf das Lernverhalten der Schüler\*innen ausgewirkt haben, da das Phänomen des Lernverlustes während der Sommerferien bekannt ist. Das Phänomen beschreibt den Verlust an Wissen und schulischen Fähigkeiten, den einige Schüler\*innen während der Sommermonate erleiden. Die zeitliche Einordnung in das Schuljahr erfolgte jedoch nach dem Lehrplan und ist somit wiederum repräsentativ für die verwendete Thematik (Cooper et al., 1996; Downey et al., 2004).

Ein weiterer Erhebungszeitpunkt zur Überprüfung des Fachwissens hätte unmittelbar nach der ersten Lernumgebung durchgeführt werden können. Damit wäre noch deutlicher geworden, ob die zweite Lerneinheit mit dem Schwerpunkt der naturwissenschaftlichen Arbeitsweisen auch einen zusätzlichen Einfluss auf das Fachwissen hat. Schwierig war dies jedoch mit Blick darauf, die Schüler\*innen nicht an einem Vormittag mit Fragebögen zu überlasten und aus Gründen des Fragebogendesigns (Redundanz der Aufgaben). Eine Lösung wäre, den Studientag zu verlängern und auf den Nachmittag auszudehnen, um den Zeitfaktor zu reduzieren und die Schüler\*innen zu entlasten.

## 8.2 Inhaltliche Diskussion

### F 1 Welche Auswirkung hat die Gestaltung der Lernumgebung auf die Förderung des Fachwissens bei Schüler\*innen im Bereich der Biotechnologie?

Betrachtet man die Ergebnisse, so zeigt sich, dass das Ausgangsniveau in Bezug auf das fachliche Vorwissen bei allen vier Gruppen ähnlich ist. Es gibt keine signifikanten Unterschiede im Vorwissen der Gruppen (s. 7.2.1 *Vergleich der Gruppen*). Inwieweit das Design der Lernumgebung einen Einfluss auf die Verbesserung des biotechnologischen Fachwissens der Schüler\*innen hat, zeigt sich nach der Bearbeitung der Lernumgebung zum Fachwissen. Hier lag die mittlere individuelle Fähigkeit der Präsenzgruppe signifikant unter den mittleren Personenfähigkeiten der anderen drei Gruppen. Zwischen den anderen Gruppen gab es hingegen keine signifikanten Unterschiede. Das bedeutet, dass die Präsenzgruppe nach der Intervention im Vergleich zu den anderen Gruppen einen geringeren Zuwachs der Personenfähigkeit im Fachwissen aufwies. Betrachtet man die Entwicklung innerhalb der einzelnen Gruppen, so zeigt sich zudem, dass nur die Personenfähigkeiten der Gruppen Flipped Classroom und Homeschooling signifikant ansteigen (vgl. 7.2.2 *Vergleich der Messzeitpunkte*), also die beiden Gruppen, die das Fachwissen in digitaler Form präsentiert bekommen haben.

Basierend auf den bisherigen Ergebnissen kann argumentiert werden, dass eine Steigerung der Personenfähigkeit durch das Erlernen von Fachwissen in einer virtuellen bzw. digitalen Lernumgebung erreicht werden kann. Die von Lodge und Horvath (2017) nachgewiesene Möglichkeit, im eigenen Lerntempo zu arbeiten, bot den Lernenden möglicherweise die Chance, ihre Kompetenzen entsprechend ihrer individuellen Bedürfnisse und Fähigkeiten zu entwickeln. Innerhalb eines Zeitfensters von 45 Minuten konnten die Lernenden nicht nur selbstbestimmt agieren, sondern auch ihr Lerntempo eigenständig anpassen, was ein hohes Maß an Selbstmanagement forderte und somit dem Anschein nach eine Steigerung des Lernzugewinns implizierte (S. Zhang et al., 2022). Diese selbstregulierte Art des Lernens, bei der die Lernenden, aufgrund des begrenzten Internets, möglicherweise die Kontrolle über die Auswahl und Systematisierung ihres Lernmaterials behielten, könnte das Prinzip der Lernerautonomie gestärkt haben und könnte förderlich für die Selbstständigkeit gewesen sein (Scheiter, 2022). Die Bereitstellung von Quizfragen und kleineren Spielen innerhalb der Lernumgebung könnte bei den Lernenden zur Steigerung des Interesses und der Motivation geführt haben (Mayrose, 2012).

Die Wirksamkeit solcher interaktiven Werkzeuge zur Förderung der aktiven Auseinandersetzung mit dem Lernmaterial sind bestätigt, was wiederum die Erkenntnis unterstützt, dass solche Methoden nicht nur das Engagement erhöhen, sondern auch die

Kompetenzen der Lernenden erweitern, indem die Lernenden eine aktive Rolle im Lernprozess übernehmen (Deterding et al., 2011; H. Meyer, 2020). Schließlich unterstreicht die positive Reaktion auf diese spielerischen Elemente, wie die Belohnungen in Form von Sternen bei richtigen Ergebnissen, den Mehrwert solcher Methoden für die Steigerung der intrinsischen Motivation. Somit kann angenommen werden, dass sich Lernende tiefer mit dem Lernstoff auseinandersetzen (Ab Rahman et al., 2019; Rahman et al., 2018). Die Verbesserung der Personenfähigkeit durch das Lernen in dieser virtuellen Lernumgebung wurde auch durch die interaktiven Elemente dieser Lernformate unterstützt, die ein unmittelbares Feedback über die Richtigkeit der Lösungen bot und dadurch das Lernen verstärken konnte (Thisgaard & Makransky, 2017). Dieses direkt automatisch erstellte Feedback könnte für das Lernen von zentraler Bedeutung gewesen sein, da es die Lernenden in die Lage versetzte, unmittelbar aus ihren Fehlern zu lernen. Damit wurde im besten Falle das Verständnis in Echtzeit gefördert, was in der Literatur als kritischer Aspekt für den Erwerb und die Vertiefung von Fachkenntnissen in einer solchen Lernumgebung angesehen wird (M. Brown, 2013; X. Chen & DeBoer, 2015; Thisgaard & Makransky, 2017).

Die Gruppen Präsenzunterricht und Wechselunterricht mit der in Präsenz dargebotenen Lernumgebung zum Fachwissen zeigten keinen signifikanten Anstieg in den Personenfähigkeiten der Schüler\*innen. Betrachtet man die Mittelwerte der Gruppen, so weisen diese zwar auf eine Verbesserung der Personenfähigkeit hin, jedoch ist die Varianz innerhalb der Gruppen so groß, dass dies nicht eindeutig festgestellt werden kann.

Mögliche Erklärungsansätze hierfür sind, dass im traditionellen Klassenunterricht das Lerntempo der Schüler\*innen durch den zeitlich vorgegebenen Rahmen des Unterrichts und die Unterrichtsstruktur eingeschränkt wurde. Hier bestimmte die Lehrkraft das Lerntempo für alle Schüler\*innen, sodass auf individuelle Bedürfnisse und Lerngeschwindigkeiten der Schüler\*innen nicht immer Rücksicht genommen werden konnte (Graber & Scholz, 2013). Darüber hinaus kam es sehr wahrscheinlich zu einer geringen Interaktivität im Klassenunterricht, da die Interaktionsformen auf Partnerarbeit und Lehrer-Schüler-Gespräch beschränkt waren, was sich auf den Wissenserwerb der Schüler\*innen ausgewirkt haben könnte, da es zu wenig Möglichkeiten gab, generative Lernstrategien wie Selbsterklärungen anzuwenden, die laut Chi (2022) und Fiorella und Mayer (2022a) das Lernen verbessern. Das Potenzial der Schüler\*innen von aktiven Lernmethoden wie Selbsterklärungen und generativen Lernaktivitäten zu profitieren, konnte in einer großen Klassenkonstellation von 25 bis 30 Schüler\*innen nicht sicher gestellt werden, was zu einer geringeren Lernmotivation und damit zu einer geringeren Selbstwirksamkeit führen könnte (Schulmeister, 2016).

Folglich kann aus diesen Ergebnissen geschlossen werden, dass die **Hypothese H<sub>1.1</sub>** bestätigt wird: Eine virtuelle Lernumgebung für das Fachgebiet Biotechnologie erweist sich hinsichtlich des Lernerfolgs bei Schüler\*innen als überlegen.

Die Studie zeigte, dass nur in den beiden Gruppen, in denen das Fachwissen in einer digitalen Lernumgebung gelernt wurde, eine signifikante Verbesserung der mittleren Personenfähigkeit zu verzeichnen war. Vergleicht man diese Situation mit der Behaltensleistung der anderen Gruppen, so zeigt sich, dass nach den 8 Wochen das Ausgangsniveau der Personenfähigkeiten weitgehend wieder erreicht wurde. Signifikante Unterschiede zwischen den Gruppen bestehen zu diesem Erhebungszeitpunkt nicht mehr (s. *7.2.1 Vergleich der Gruppen*). Betrachtet man die Entwicklung innerhalb der einzelnen Gruppen, so zeigt sich zudem, dass die mittleren Personenfähigkeiten der drei Gruppen, die direkt nach der Intervention signifikant bessere Personenfähigkeiten aufwiesen als der Präsenzunterricht, sich nun alle drei wieder signifikant verschlechtert haben (s. *7.2.2 Vergleich der Messzeitpunkte*). Die Lernenden im Präsenzunterricht zeigte keine signifikante Veränderung in ihrer Personenfähigkeit.

Betrachtet man also die Behaltensleistung, ausgedrückt durch die Veränderung der Personenfähigkeit, so kann diese als weitestgehend unabhängig von der Darbietungsform interpretiert werden. Insgesamt deuten diese Ergebnisse darauf hin, dass der Zuwachs im Bereich des Fachwissens sowohl in der virtuellen als auch in der Präsenzlernumgebung nicht langfristig erhalten werden konnte. Die oben beschriebenen Vorteile, bezogen auf die Lernwirksamkeit virtueller Lernumgebungen, können somit nicht auf den langfristigen Lernzuwachs übertragen werden. Da die Personenfähigkeiten in allen Gruppen mehr oder weniger stark abnehmen, sollten auch externe Gründe genauer betrachtet werden.

Zwischen den beiden Erhebungszeitpunkten lagen die Sommerferien der bayerischen Schulen, was wahrscheinlich zu einem zusätzlichen Wissensverlust der Lernenden geführt haben könnte, da sich die Schüler\*innen nicht weiter mit der grundlegenden Thematik der Genetik beschäftigt haben und insgesamt keine schulisch orientierten Leistungen in dieser Zeit erbrachten (Cooper et al., 1996; Downey et al., 2004). Überdies könnte die Möglichkeit bestehen, dass aufgrund der Position des Themas am Ende des Schuljahres dieses von den Schüler\*innen insgesamt nicht mehr als „so ernst“ in Bezug auf das exakte Lernen wahrgenommen wurde. Das Bewusstsein, dass der Stoff in absehbarer Zeit nicht mehr prüfungsrelevant ist, erzeugte bei den Schüler\*innen möglicherweise keine extrinsische Motivation und kein fremdgesteuertes Handeln, wie es durch Anreizsysteme wie Noten für die Schüler\*innen der Fall ist. (Götz & Nett, 2017)

Betrachtet man das Ergebnis der Gruppe Wechselunterricht, was bezogen auf die Behaltensleistung einen signifikant negativen Effekt in der Personenfähigkeit über die gesamte Messung hinweg aufweist, sollte dieses hinterfragt werden. Das Ergebnis lässt vermuten, dass noch weitere äußere Faktoren hierbei eine Rolle gespielt haben.

Die **Hypothese H<sub>1,2</sub>** (Eine virtuelle Lernumgebung zum Fachwissen im Bereich der Biotechnologie zeigt bei Schüler\*innen eine bessere Behaltensleistung) kann durch dieses Ergebnis nicht bestätigt werden.

## **F 2 Inwiefern hat die Gestaltung der Lernumgebung Auswirkung auf die Nachhaltigkeit des Lernens der naturwissenschaftlichen Arbeitsweise im Bereich der Biotechnologie?**

Analysiert man die Ergebnisse der Studie dahingehend, ob die Gestaltung der Lernumgebung eine entscheidende Rolle für die nachhaltige Verankerung der naturwissenschaftlichen Arbeitsweisen bei den Lernenden im Bereich der Biotechnologie spielt, so zeigt sich, dass die vier Testgruppen über die Messzeitpunkte hinweg keine signifikanten Unterschiede in ihrer durchschnittlichen Personenfähigkeit aufweisen (s. 7.2.1 *Vergleich der Gruppen*), wenngleich die Gruppen des Wechselunterrichts und des Homeschoolings in ihrer durchschnittlichen Personenfähigkeit deutlicher abnehmen als die beiden Gruppen, die die Lernumgebung zur naturwissenschaftlichen Arbeitsweise in Präsenz durchgeführt haben. Betrachtet man ergänzend die Veränderung der jeweiligen Gruppen hinsichtlich der Nachhaltigkeit ihres Lernens über die Zeit, so bestätigt sich dieses Bild (s. 7.2.2 *Vergleich der Messzeitpunkte*). In allen vier Gruppen nimmt die durchschnittliche Personenfähigkeit ab, was bedeutet, dass nicht wirklich nachhaltig gelernt wird, da es jeweils signifikante Unterschiede zwischen den Messzeitpunkten gibt. Dabei ist zu beobachten, dass die signifikante Veränderung der Personenfähigkeit in Bezug auf das Wissen über naturwissenschaftliche Arbeitsweisen in den beiden virtuellen Lernumgebungen zusammen betrachtet größer ist als in den Präsenzgruppen.

Der etwas größere Nachhaltigkeitseffekt der Gruppen mit in Präsenz dargebotener Lernumgebung zu den naturwissenschaftlichen Arbeitsweisen kann wahrscheinlich damit erklärt werden, dass die haptische Erfahrung das Gedächtnis und das Verständnis gefördert hat (Tanhua-Piironen et al., 2010). Die von den Schüler\*innen genutzten fachspezifischen Arbeitsmethoden, zusammen mit der praktischen Laborerfahrung, könnten somit zur Vertiefung und Erweiterung des Verständnisses der Schüler\*innen für das naturwissenschaftliche Wissen beigetragen haben (*America's Lab Report*, 2006; le Plat et al., 2018). Von Bedeutung war hierbei vermutlich auch die Interaktion zwischen Lehrkraft und Schüler\*innen, um den Lernenden die Ziele und die Bedeutung der Laborarbeit zu vermitteln (Högström et al., 2010). Während der Durchführung ergab sich für die Schüler\*innen jederzeit die Möglichkeit des direkten personalisierten Feedbacks durch die Lehrkraft, was gerade bei den komplizierten Verfahren, die im Versuch genutzt wurden, ein besseres Verständnis

ermöglichte (Mottarella et al., 2005; A. Y. Wang & Newlin, 2000). Als weiterer wesentlicher Aspekt für das nachhaltigere Lernen im Präsenzlabor könnte der von Eckert (2020) und Trendel und Lübeck (2018) beschriebene Effekt des Erfahrungslernens angesehen werden. Durch das Erleben und die anschließende Reflexion ihrer Ergebnisse erzielten die Schüler\*innen scheinbar ein tieferes und nachhaltigeres Verständnis des Lernmaterials. Die Verwendung der doch besonderen Materialien (PCR-Cycler, Agarose-Gelelektrophoresekammer und Mikroliterpipetten), die im alltäglichen schulischen Kontext nicht vorkommen, konnte wahrscheinlich eine zusätzliche Motivation erzeugen und so die Lernbereitschaft sowie die kognitive Aktivität der Schüler\*innen anregen (Eckert, 2020).

Die Studie zeigte, dass es zwar einen gruppeninternen Unterschied zwischen dem virtuellen und dem Präsenzversuch gab, bei dem die Nachhaltigkeit des Lernens naturwissenschaftlicher Arbeitsweisen im virtuellen Labor geringer war. Im Vergleich der vier Gruppen zu den Messzeitpunkten konnten diese Effekte aber nicht nachgewiesen werden. Somit könnte davon ausgegangen werden, dass auch das virtuelle Labor eine gewisse Nachhaltigkeit aufwies.

Die virtuelle Lernumgebung war entsprechend der didaktischen Gestaltungsprinzipien aufgebaut und enthielt die gleichen Inhalte wie das Präsenzformat. Das haptische Erleben konnte in diesem Kontext nicht ermöglicht werden, dafür ergaben sich durch den virtuellen Charakter andere Möglichkeiten. Das digitale Labor ermöglichte es den Lernenden, unbeobachtbare und beobachtbare Phänomene, die in diesem Versuch an mehreren Stellen auftraten, miteinander zu verknüpfen (de Jong et al., 2013) und so z.B. auch die bekannte Problematik des Genotyp-Phänotyp-Unterschieds zu reduzieren. Zudem ermöglichte die virtuelle Lernumgebung den Schüler\*innen, in ihrem eigenen Lerntempo zu arbeiten, d.h. auch Schritte im Prozess zu wiederholen, was das Verständnis nachweislich fördert (Govender, 2023). Das Feedback, das bei jedem falschen Entscheidungsschritt erscheint, könnte das Verständnis der Schüler\*innen für Konzepte und Prozesse fördern (Thisgaard & Makransky, 2017) und für den Erwerb praktischer Kompetenzen förderlich sein, da es auf den Handlungen und Entscheidungen der Schüler\*innen basierte und somit Missverständnisse klären und den Lernprozess unterstützen konnte (van Joolingen & de Jong, 1997).

In Bezug auf die **Hypothese H<sub>2</sub>** (In Präsenz dargebotene Inhalte im Bereich des praktischen Arbeitens in der Biotechnologie zeigen bei Schüler\*innen einen nachhaltigeren Lernerfolg) kann somit gesagt werden, dass die Hypothese in Bezug auf den zeitlichen Verlauf der einzelnen Gruppen bestätigt werden kann, jedoch in Bezug auf die Gruppenunterschiede nicht zutrifft.

### **F 3 Welchen Einfluss hat die Kombination der Darbietungsform der Lernumgebungen zum Fachwissen und naturwissenschaftlichen Arbeiten auf die Nachhaltigkeit des Lernens bei Schüler\*innen?**

Aus den Hypothesen der Forschungsfragen 1 und 2 ergibt sich die Kernfrage, inwieweit die Kombination der Lernumgebungen einen Einfluss auf die Nachhaltigkeit des Wissenserwerbs bei den Lernenden hat. Durch die Bestätigung der Hypothesen 1.1 und 2 und die Widerlegung der Hypothese 1.2 ergibt sich für diese Studie die Annahme, dass sich das hybride Format des Flipped Classroom am nachhaltigsten auf das Lernergebnis der Schüler\*innen auswirkt.

Vergleicht man die Veränderung der Personenfähigkeiten in den hybriden und virtuellen Lernumgebungen der jeweiligen Gruppen mit dem **Präsenzunterricht**, der als Kontrollgruppe herangezogen wird, da er die gängige Unterrichtsform in unserer Schullandschaft darstellt, so lassen sich folgende Beobachtungen ableiten.

Das hybride Format, das digitales Lernen von Fachwissen mit in Präsenz durchgeführtem Schülerversuch kombiniert, wird in dieser Studie als **Flipped Classroom** bezeichnet. Die Ergebnisse zeigen, dass dieses Format in Bezug auf die Personenfähigkeiten in dieser Studie wahrscheinlich das lernwirksamste Format darstellt.

Das virtuell aufbereitete Fachwissen wird von den Schüler\*innen eigenständig bearbeitet. Die von Lodge und Horvath (2017) nachgewiesene Möglichkeit, im eigenen Lerntempo zu arbeiten, bietet den Lernenden die Chance, ihre Kompetenzen entsprechend ihrer individuellen Bedürfnisse und Fähigkeiten zu entwickeln. Innerhalb eines Zeitfensters von 45 Minuten konnten die Lernenden nicht nur selbstbestimmt agieren, sondern auch ihr Lerntempo eigenständig anpassen, was ein hohes Maß an Selbstmanagement erforderte und somit möglicherweise einen erhöhten Lernzuwachs implizierte (S. Zhang et al., 2022). Die zur Verfügung stehende Zeit war für einige Lernende knapp bemessen, was möglicherweise auch auf eine zu gering ausgeprägte Fähigkeit zum selbstgesteuerten Lernen zurückzuführen ist. Spielerische Elemente in Lernumgebungen steigerten vermutlich das Interesse und die Motivation der Lernenden (Mayrose, 2012) und bestätigen wahrscheinlich die Wirksamkeit interaktiver Werkzeuge zur aktiven Auseinandersetzung mit Lerninhalten (Deterding et al., 2011; H. Meyer, 2020). Interaktive Elemente, die unmittelbares Feedback geben, verstärken den Lernerfolg und ermöglichen ein unmittelbares Lernen aus Fehlern, was für die Vertiefung von Fachkenntnissen essenziell ist (M. Brown, 2013; X. Chen & DeBoer, 2015; Thisgaard & Makransky, 2017). Die klare Strukturiertheit der Lernumgebung sorgte wahrscheinlich für einen sehr guten Überblick und ermöglichte es den Schüler\*innen somit, das Fachwissen zu vertiefen. Nach dieser Phase ohne wirkliche soziale Interaktion folgte der Präsenzversuch, welcher, da es in Partnerarbeit durchgeführt wird, eine ständige soziale Interaktion erzeugt (H. Meyer, 2020; L. Meyer & Stein, 2022). Die Schüler\*innen hatten zudem die Möglichkeit, direkt

mit der Lehrperson zu sprechen und so Fragen, die sich bei der Aneignung des Fachwissens ergeben haben, direkt zu formulieren. Dies sollte für den Erwerb praktischer Kompetenzen förderlich sein, da Missverständnisse geklärt und Lernprozesse unterstützt werden können (van Joolingen & de Jong, 1997). Der Einsatz von speziellen Materialien im Schülerversuch kann die Motivation und Bereitschaft sowie die kognitive Aktivität der Lernenden erhöhen (Eckert, 2020). Das haptische Erlebnis, den Versuch eigenständig durchzuführen und dabei neue Methoden kennenzulernen, könnte zudem das Verständnis der Schüler\*innen für naturwissenschaftliche Arbeitsweisen gefördert haben (Tanhua-Piironen et al., 2010). Auch das Erfahrungslernen bei der Durchführung des Schülerversuchs, bei dem die Schüler\*innen durch eigenes Erleben und Reflexion ein tieferes Verständnis erzeugt haben könnten (Eckert, 2020; Trendel & Lübeck, 2018).

Die Annahme ist, dass dieses hybrides Lernkonzept, bei dem Schüler\*innen zuerst online eigenständig Stoff erarbeiten und anschließend das Gelernte im Schülerversuch praktisch anwenden, zu einem tieferen Verständnis führt. Dadurch wird die Lerneffektivität verbessert und die Selbstständigkeit unterstützt, sowie die Umsetzung des Erlernten, was zu einem nachhaltigeren Lerneffekt führen kann (Lutz et al., 2023; Ulbrich, 2021)..

Die Kombination aus in Präsenz durchgeführtem Klassenunterricht zur Vermittlung des Fachwissens und der digitalen Bearbeitung des virtuellen Schülerversuchs wird als **Wechselunterricht** in dieser Studie bezeichnet. Aus den Ergebnissen zeigt sich, dass dieser bezogen auf die Lernwirksamkeit ein ungünstigeres Format darstellt. Die Ergebnisse zeigen, dass die Nachhaltigkeit des Lernens, gemessen durch die Personenfähigkeiten, sowohl für das Fachwissen, als auch für die naturwissenschaftliche Arbeitsweise nur bedingt gegeben sind.

Die Schüler\*innen lernten im Klassenunterricht weniger als in der entsprechenden virtuellen Lernumgebung, da durch seinen festen zeitlichen Rahmen das individuelle Lerntempo der Schüler\*innen beschränkt gewesen ist; zumal die Lehrkräfte, wie üblich und notwendig, das Tempo vorgaben und deswegen seltener auf individuelle Lernbedürfnisse eingehen können (Graber & Scholz, 2013). Die etwas zu geringe Interaktivität, welche wiederum dem zeitlichen Faktor geschuldet war, ist auf die Dominanz von Partner- und Lehrer-Schüler-Interaktionen zurückzuführen. Somit beschränkte dies den Einsatz von generativen Lernstrategien wie Selbsterklärungen, die für ein effektiveres Lernen bedeutsam sind (Chi, 2022; Fiorella & Mayer, 2022a). In den großen Klassen mit 25 bis 30 Schüler\*innen konnte das Potenzial für aktives Lernen und damit für gesteigerte Motivation und Selbstwirksamkeit sicherlich nicht voll ausgeschöpft werden und der Prozess des selbstgesteuerten Lernens war bei den Schüler\*innen möglicherweise noch nicht ausreichend entwickelt (Hattie, 2009; Schulmeister, 2016). Mit daraus ggf. entstandenen Wissenslücken oder Fehlvorstellungen wurden die

Schüler\*innen an die virtuelle Lerneinheit übergeben. Hier erhielten sie zwar stetiges Feedback durch die automatisch eingestellten Reaktionen, was der Lernwirksamkeit zu Gute kommen sollte (Narciss, 2018; Thisgaard & Makransky, 2017) und für den Erwerb praktischer Kompetenzen förderlich sein sollte, da es auf den Handlungen und Entscheidungen der Schüler\*innen basierte und somit Missverständnisse klären und den Lernprozess unterstützen könnte (van Joolingen & de Jong, 1997). Das automatische Feedback könnte jedoch bei einigen Schüler\*innen eine eher unspezifische Antwort auf ihre eigentlichen Fragen darstellen. Das Lernen an der Beispielaufgabe des Kriminalfalls mit einer Problemstellung und einer schrittweise dargebotenen Lösung des Problems sollte nach Ayres und Sweller (2022) gerade bei solch komplexen Fragestellungen hilfreich sein. Auch sollte die kognitive Belastung (Sweller et al., 2019) durch die kleinschrittigen Lerneinheiten eines solch komplexen Themas gesenkt werden und somit lernwirksamer sein. Es könnte aber auch sein, dass das redundante Auftreten von Feedback und damit verbundenen ähnlichen Antworten zu einem gegenteiligen Effekt (Redundanzeffekt) bei den Lernenden führen kann und dadurch die kognitive Belastung erhöht wird (Kalyuga & Sweller, 2022). Die Verknüpfung verschiedener Ebenen (z.B. Phänotyp vs. Genotyp), welche in virtuellen Lernumgebungen besser gelernt werden sollte, als in Präsenz dargebotenen Lernangeboten (de Jong et al., 2013), könnte möglicherweise gar nicht richtig wirken, da vielen der Schüler\*innen möglicherweise ihr Problem hierbei nicht bekannt war bzw. die Notwendigkeit eines Konzeptwechsels (conceptual change) in den Augen der Lernenden gar nicht vorlag (Kattmann, 2022). Allerdings ermöglichte die virtuelle Lernumgebung den Schüler\*innen, in ihrem eigenen Lerntempo zu arbeiten; d.h., auch Schritte im Prozess zu wiederholen, was das Verständnis für die naturwissenschaftliche Arbeitsweise nachweislich fördern sollte (Govender, 2023), und auch die zur Verfügung gestellte Zeit war ausreichend.

Durch die Verwendung eines rein virtuellen Lernformats in beiden Lernumgebungen ergibt sich das Konzept des **Homeschooling** in dieser Studie. Die Ergebnisse zeigen, dass dies in Bezug auf die Personenfähigkeiten in dieser Studie eine etwas geringere Lernwirksamkeit aufweist als der Präsenzunterricht.

Die kohärente Abstimmung der beiden Formate virtueller Lernumgebungen sollte den Schüler\*innen das Erlernen von Fachwissen und naturwissenschaftlichen Arbeitsweisen erleichtern (R. E. Mayer & Moreno, 2003). Der gewissenhaft gewählte Einsatz von interaktiven Tools und Methoden sollte dazu beitragen, dass sich die Schüler\*innen aktiv beteiligen und somit lernwirksamer arbeiten (Deterding et al., 2011; H. Meyer, 2020). Die Lernwirksamkeit der gut strukturierten Lernumgebung im Fachwissen (s. *Flipped Classroom*) könnte genutzt worden sein, um anschließend die virtuelle Lernumgebung des Schülerversuchs durchzuführen. Hierbei war vermutlich das automatische Feedback für einige Schüler\*innen

zu unspezifisch für ihre komplexeren Probleme oder aber die Form des Feedbacks (s. *Wechselunterricht*) war für die Schüler\*innen redundant (Kalyuga & Sweller, 2022; Paas & Sweller, 2022). Ein weiterer wesentlicher Aspekt hierbei könnte sein, dass das selbstgesteuerte Lernen auch während bzw. nach der COVID-19 Pandemie weiterhin nicht wirklich durch die Schüler\*innen verinnerlicht wurde (S. Zhang et al., 2022; Zimmerman, 2002). Abschließend ist anzumerken, dass bei der Umsetzung des Homeschooling-Prinzips eine geringe soziale Interaktion über einen sehr langen Zeitraum wirkte. Zwar konnten die Schüler\*innen selbst gewählte Pausen einlegen und in diesen mit ihren Mitschüler\*innen interagieren, dennoch könnte dies einen wesentlichen Punkt darstellen. Für den insgesamt eher geringen nachhaltigen Lerngewinn spielen vermutlich alle diese Überlegungen zusammen.

Auf Basis der Daten dieser Studie und dieses Forschungsdesigns lässt sich die Hypothese H<sub>3</sub> mit Vorsicht bestätigen (Die selbstständige Erarbeitung des Fachwissens in einer virtuellen Lernumgebung in Kombination mit in Präsenz dargebotenen Inhalte im Bereich des praktischen Arbeitens in der Biotechnologie zeigen, bei den Schüler\*innen einen nachhaltigen Lernerfolg). Die Entwicklung innerhalb der Gruppen über die Zeit zeigt signifikante Ergebnisse, jedoch können diese Effekte im direkten Vergleich der Gruppen zu den entsprechenden Messzeitpunkten nicht dargestellt werden.

Zusammenfassend könnte aus dieser Studie abgeleitet werden, dass der Zugang zu adäquat gestalteten digitalen Lernumgebungen für Lehrkräfte während der COVID-19 Pandemie möglicherweise, neben sicherlich einigen weiteren Faktoren, dazu beigetragen hätte, den beobachteten Kompetenzrückgang bei den Schüler\*innen zu mindern.

## 9 Ausblick und Relevanz

Die vorliegende Arbeit befasst sich mit der Untersuchung der Wirksamkeit von hybriden und virtuellen Lernumgebungen für das naturwissenschaftliche Arbeiten in der Biotechnologie, die für Lernende während der COVID-19 Pandemie in Deutschland relevant waren. Die Planung und Entwicklung dieser Studie fand im gleichen Zeitraum statt und berücksichtigte die Tatsache, dass wissenschaftliches Arbeiten auch in Zeiten einer Pandemie von großer Bedeutung ist. In dieser herausfordernden Periode wurden die in der Studie genutzten virtuellen Lernumgebungen unter Verwendung der den Lehrkräften zu dieser Zeit zur Verfügung stehenden Ressourcen wie H5P gestaltet und den Schüler\*innen über LMS wie mebis zur Verfügung gestellt.

Die Relevanz der erhobenen Daten ergibt sich insbesondere daraus, dass sie zum Ende der Pandemie an Schüler\*innen gewonnen wurden. Sie ermöglichen Vergleiche, die in dieser Form noch nicht untersucht wurden und rücken das naturwissenschaftliche Arbeiten in den Fokus, ein Bereich, der in der wissenschaftlichen Betrachtung in der Kombination von Präsenzlernangeboten und virtuellen Lernangeboten bisher zu wenig Beachtung gefunden hat. Dabei wurde anvisiert, dass die Präsenz- und virtuellen Lernumgebungen inhaltsgleich erstellt wurden, um eine konsistente Datengrundlage zu gewährleisten und konkrete Szenarien der Pandemie auf wissenschaftlicher Basis zu untersuchen.

Zu den Einschränkungen der Studie gehört der Neuigkeitseffekt digitaler Lernumgebungen, der zum Zeitpunkt der Untersuchung noch vorhanden war, inzwischen aber an Bedeutung verloren hat. Ein weiterer Faktor ist das Erhebungsinstrument, welches an den Leistungsverlust der Schüler\*innen, welcher u.a. durch die COVID-19 Pandemie verstärkt wurde, angepasst werden müsste. Es sei auch darauf hingewiesen, dass der Vergleich der Unterrichtsformen und die Ergebnisse dieser Studie weiterer Untersuchungen bedürfen und nicht als endgültig angesehen werden können. Unter den genannten Voraussetzungen sind die aufgezeigten Effekte eingetreten.

Als mögliche Erweiterung der Arbeit könnte die Gamification der virtuellen Lernumgebungen durch den Einsatz neuerer Verfahren verstärkt werden (Deterding et al., 2011; Sailer & Homner, 2020), um auch hier eine Art Erlebniseffekt, analog zu den in Präsenz dargebotenen Lernumgebungen, zu erzeugen. Ein Tracking der Schülerhandlungen während der Durchführung könnte wertvolle Einblicke darüber bieten, wie und in welchem Ausmaß die Schüler\*innen das Format nutzen. Dies würde auch für in Präsenz durchgeführte Lernumgebungen gelten, wobei darauf zu achten wäre, dass das Tracking das Handeln der Schüler\*innen nicht beeinträchtigt.

Es wird deutlich, dass praktisches Arbeiten für den naturwissenschaftlichen Unterricht unerlässlich ist und durch virtuelle Lernumgebungen effektiv vorentlastet werden kann, was dann ein möglicherweise zielgerichteteres, kognitiv weniger überforderndes und gefahrloseres Arbeiten in der realen Welt ermöglichen könnte. Darüber hinaus bieten digitale Lernumgebungen adaptive Feedbackmöglichkeiten, wobei die Reaktion auf Fehler derzeit noch fest vorprogrammiert ist. In Zukunft wird künstliche Intelligenz eine Option sein, um die adaptiven Fähigkeiten des Feedbacks weiter zu verbessern und es so zu gestalten, dass es der Reaktion einer realen Person (Lehrkraft) entspricht und auch mit unkonkreteren Fragestellungen durch Schüler\*innen arbeiten kann. Die Forschung in diesem Bereich ist immens und bezieht sich auch auf den naturwissenschaftlichen Erkenntnisweg, wie Bewersdorff et al. (2023) aufzeigen.

Empfehlungen für die Schule in Bezug auf diese Ergebnisse beinhalten die Nutzung des hybriden Formats des Flipped Classroom, das sich als effektiv erwiesen hat, wie auch in der Literatur dokumentiert ist (Al-Samarraie et al., 2020; J. Bishop & Verleger, 2013; Lutz et al., 2023). Um einen echten Mehrwert zu generieren, müsste jedoch wahrscheinlich das selbstgesteuerte Lernen der Schüler\*innen zunächst stärker gefördert werden, damit die Ergebnisse aus dem selbstständigen Lernen im schulischen Kontext zielführender verwendet werden können. Die praktische Durchführung von Versuchen und Experimenten im naturwissenschaftlichen Unterricht bleibt unbestritten wichtig. Eine digitale Lernumgebung wie ein digitales Labor kann aber insbesondere dann sinnvoll sein, wenn Versuche oder Experimente aus Zeit-, Kosten- oder Aufwandsgründen in der Schule oder als Exkursion zu einem außerschulischen Lernort sonst nicht möglich wären. Dazu müsste allerdings das Angebot an kompletten digitalen Lernumgebungen für den naturwissenschaftlichen Erkenntnisweg, die für Schulen verfügbar und bezahlbar sind, deutlich erhöht werden.

## 10 Verzeichnisse

### 10.1 Abbildungsverzeichnis

Abbildung 1: Darstellung des vereinfachten Angebots-Nutzungs-Modells .....	5
Abbildung 2: Das MAIN-TEACH-Modell (nach Charalambous und Praetorius (2020)) .....	7
Abbildung 3: Schematischer Ablauf eines Zyklus der Polymerasekettenreaktion (PCR)....	29
Abbildung 4: Schematische Darstellung des CFTR-Gens und Proteins.....	32
Abbildung 5: Ausschnitt aus der AS-Sequenz des <i>CFTR</i> -Gens (Die AS 1–400 und 461–1480 sind nicht dargestellt). Blau hervorgehoben ist die AS 508 Phenylalanin, welche bei der $\Delta F508$ Mutation fehlt. Übernommen und verändert nach (UniProt, 2007).....	33
Abbildung 6: Normale CFTR-Protein-Bildung und die Mutationsklassen I bis VI.....	33
Abbildung 7: Zusammenhänge von Genotyp und Phänotyp und deren Verknüpfung über die drei verschiedenen Analyseschritte DNA-Extraktion, PCR und Agarose-Gelelektrophorese. Verändert nach Schöppner et al., 2022, S. 81). .....	37
Abbildung 8: Design der Pilotierung der Fragebögen und der digitalen Lernumgebungen. An den Prä-Test schließt sich die Durchführung der digital dargebotenen Lernumgebungen an. Abgeschlossen wird die Pilotierung mit einem Post-Test. ....	49
Abbildung 9: Item 04 im Bereich Fachwissen .....	51
Abbildung 10: Item 22 im Bereich naturwissenschaftlicher Erkenntnisgewinn Agarose- .....	52
Abbildung 11: Wright Map der Skala zur Erfassung des Fachwissens im Prä-Test (59 Personen, 12 .....	60
Abbildung 12: Wright Map der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test (59 Personen, 6 Items, 2 Kategorien: 0/1) .....	62
Abbildung 13: Wright Map der Skala zur Erfassung des Fachwissens im Post-Test (38 Personen, 9 Items, 2 Kategorien: 0/1) .....	64
Abbildung 14: Wright Map der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Arbeitsweise im Post-Test (38 Personen, 6 Items, 2 Kategorien: 0/1) .....	66
Abbildung 15: Wright Map der Skala zur Erfassung des naturwissenschaftlichen Erkenntnisgewinns in der Biotechnologie im Post-Test (38 Personen, 6 Items, 2 Kategorien: 0/1) .....	68
Abbildung 16: Design der Studie. An den Prä-Test schließt sich die Durchführung der in Präsenz und/oder digitalen dargebotenen Lernumgebungen an.	

Abgeschlossen wird der Studientag mit einem Post-Test, bevor acht Wochen später ein Follow-up-Test genutzt wird. Abhängige Variable Wissen (Fachwissen – dunkelblau; Naturwissenschaftliche Arbeitsweise – hellblau); unabhängige Variable Darbietungsform (digital – orange; in Präsenz – hellgrün).....	71
Abbildung 17: Ablauf der Datenerhebung am Studientag.....	73
Abbildung 18: Ablauf der Gesamtstudie .....	73
Abbildung 19: Beispielaufgabe aus dem N2A-Test. Logische Erweiterung der vorgegebenen Abfolge. (eigene Darstellung nach Heller und Perleth (2000); Beispielaufgabe zu Test N2A). .....	79
Abbildung 20: Klassenunterricht im Klassenzimmer zum Fachwissen in Präsenz. ....	81
Abbildung 21: Räumlichkeiten zur praktischen Durchführung des Schülerversuchs. ....	82
Abbildung 22: Zeitungsartikel zum Verschwinden von Clara Müller.....	83
Abbildung 23: Schüler*innen beim Pipettieren der PCR-Proben in der Lernumgebung Labor .....	84
Abbildung 24: Durchführung des Schülerversuchs in Präsenz. ....	85
Abbildung 25: Durchführung der Agarose-Gelelektrophorese durch Schüler*innen.....	86
Abbildung 26: Ergebnis der Gelelektrophorese. ....	87
Abbildung 27: Startseite des mebis-Kurses „Ein Ausflug in die Forensik“ .....	88
Abbildung 28: Beispiel des "Interactive Books" zur Theorie: Die DNA .....	89
Abbildung 29: Beispiel des "Interactive Books" zur Theorie: Polymerase .....	90
Abbildung 30: Beispiel des "Interactive Books" zur Theorie: PCR .....	90
Abbildung 31: Beispiel des "Interactive Book" zur Theorie: Agarose-Gelelektrophorese.....	91
Abbildung 32: Beispiel des "Interactive Books": Anwendungsbereiche der PCR und Gelelektrophorese .....	91
Abbildung 33: Beispiel des „Digitalen Labors“: Links: Willkommenseite; Rechts: Entscheidungsfragen .....	92
Abbildung 34: Beispiel des „Digitalen Labors“: Entscheidungssituation zu Beginn der Fallbearbeitung.....	93
Abbildung 35: Beispiel des „Digitalen Labors“: Vorbereitung der PCR.....	93
Abbildung 36: Beispiel des „Digitalen Labors“: Nutzung von Mikroliterpipetten .....	94
Abbildung 37: Beispiel des „Digitalen Labors“: Pipettieren der PCR-Proben .....	94
Abbildung 38: Beispiel des „Digitalen Labors“: PCR-Protokoll .....	94
Abbildung 39: Beispiel des „Digitalen Labors“: Agarosegel-Herstellung; Links: Gelgießen; .	95
Abbildung 40: Beispiel des „Digitalen Labors“: Agarose-Gelelektrophorese .....	95
Abbildung 41: Beispiel des „Digitalen Labors“: Ergebnis des „Kriminalfalls“ .....	96
Abbildung 42: ICC eines Beispieldatensatzes in Winsteps erstellt.....	100

Abbildung 43: Wright Map an einem Beispieldatensatz in Winsteps erstellt .....	101
Abbildung 44: Item-Outfit für die Skala Fachwissen im Prä-Test .....	108
Abbildung 45: Ogive Prä-Test Fachwissen.....	109
Abbildung 46: ICC für die Skala Fachwissen im Prä-Test.....	110
Abbildung 47: Wright Map der Skala Fachwissen im Prä-Test (154 Personen, 11 Items, 2 Kategorien: 0/1). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese .....	111
Abbildung 48: Item-Outfit für die Skala Fachwissen im Post-Test (geankerte Items) .....	113
Abbildung 49: ICC der Skala Fachwissen im Post-Test.....	114
Abbildung 50: Wright Map der Skala Fachwissen im Post-Test (154 Personen, 9 Items, 2 Kategorien: 0/1). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese. ....	115
Abbildung 51: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Post-Test mit geankerten Items (rechts). Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit. Hervorgehoben ist die mittlere Itemschwierigkeit des jeweiligen Tests .....	117
Abbildung 52: Item-Fit für die Skala Fachwissen im Follow-up-Test (geankerte Items) .....	118
Abbildung 53: ICC Skala Fachwissen Follow-up-Test .....	119
Abbildung 54: Wright Map der Skala Fachwissen im Follow-up-Test (118 Personen, 13 Items, 2 „Kategorien“). Dunkelblau: Grundlegende Items aus dem Bereich der Genetik. Hellblau: Items aus dem Bereich der PCR und Agarose-Gelelektrophorese .....	120
Abbildung 55: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Follow-up-Test (rechts). Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit.....	122
Abbildung 56: Vergleich der Wright Maps im Fachwissen von Prä-Test (links) und Post-Test (mittig) und Follow-up-Test (rechts). Die Nummerierung der Items ist über alle drei Testzeitpunkte identisch und unterscheiden sich nur über die Vorsilben „pr“ für den Prä-Test, „po“ für den Post-Test und „fo“ für den Follow-up-Test. Die Itemschwierigkeiten des Prä-Tests dienen als Anker für diese im Post- bzw. Follow-up-Test. Die blaue gestrichelte Linie kennzeichnet die mittlere Itemschwierigkeit des Prä-Tests und den Wert von 0 Logit.....	123
Abbildung 57: Item-Outfit für die Skala NWA im Post-Test.....	127

Abbildung 58: Item Characteristic Curves (ICC) NWA im Post-Test .....	128
Abbildung 59: Wright Map der Skala NWA im Post-Test (154 Personen, 12 Items, 2 Kategorien: 0/1) .....	129
Abbildung 60: Item-Fit für die Skala NWA im Follow-up-Test (geankerte Items).....	131
Abbildung 61: ICC der Skala NWA im Follow-up-Test.....	132
Abbildung 62: Wright Map der Skala NWA im Follow-up-Test (118 Personen, 12 Items, 2 Kategorien: 0/1). Inhaltliche Zuordnung der Items: grün: Item 21-24: Agarose-Gelelektrophorese; hellgrün: Item 25-28: PRC; türkis: Item 29-32: Pipettieren mit Mikroliterpipette.....	134
Abbildung 63: Vergleich der Wright Maps in NWA von Post-Test (links) und Follow-up- Test (rechts). Die Nummerierung der Items ist über alle drei Testzeitpunkte identisch und unterscheidet sich nur über die Vorsilben „po“ für den Post- Test und „fo“ für den Follow-up-Test.....	135
Abbildung 64: Überprüfung der Personenfähigkeiten des Fachwissens auf univariate Ausreißer .....	139
Abbildung 65: Gruppenvergleiche Fachwissen Prä-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr_KFT = 0,699, Schulart = 1,24; Fehlerbalken: 95 % KI; .....	143
Abbildung 66: Gruppenvergleiche Fachwissen Post-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI .....	143
Abbildung 67: Gruppenvergleiche Fachwissen Follow-up-Test (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr_KFT = 0,0699, Schulart = 1,24. Fehlerbalken: 95 % KI .....	144
Abbildung 68: Vergleich der Messzeitpunkte - Fachwissen Präsenzunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr_KFT = 0,0699. Fehlerbalken: 95 % KI.....	146
Abbildung 69: Vergleich über die Messzeitpunkte- Fachwissen Flipped Classroom (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: pr_KFT = 0,0699. Fehlerbalken: 95 % KI .....	147
Abbildung 70: Vergleich über die Messzeitpunkte; FW Wechselunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden	

	anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	148
Abbildung 71:	Vergleich über die Messzeitpunkte; FW Homeschooling (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	149
Abbildung 72:	Überprüfung auf univariate Ausreißer .....	150
Abbildung 73:	Gruppenvergleiche Naturwissenschaftlicher-Arbeitsweise Post-Test. (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	153
Abbildung 74:	Gruppenvergleiche naturwissenschaftlicher Arbeitsweise Follow-up-Test. (Ergebnis der Kovarianzanalyse). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	154
Abbildung 75:	Vergleich über die Messzeitpunkte; NWA Präsenzunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI.....	156
Abbildung 76:	Vergleich über die Messzeitpunkte; NWA Flipped Classroom (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ . Fehlerbalken: 95 % KI.....	157
Abbildung 77:	Vergleich über die Messzeitpunkte; NWA Wechselunterricht (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	158
Abbildung 78:	Vergleich über die Messzeitpunkte; NWA Homeschooling (Ergebnis der ANCOVA mit Messwiederholung). Die Kovariaten im Modell werden anhand der folgenden Werte berechnet: $pr\_KFT = 0,0699$ , $Schulart = 1,24$ . Fehlerbalken: 95 % KI .....	159

## 10.2 Tabellenverzeichnis

Tabelle 1: Skala zum Fachwissen in der Biotechnologie (Prä-Test-Pilotierung) .....	53
Tabelle 2: Skala zur allgemeinen naturwissenschaftlichen Arbeitsweise (Prä-Test-Pilotierung) .....	54
Tabelle 3: Skala zum Fachwissen in der Biotechnologie (Post-Test-Pilotierung) .....	55
Tabelle 4: Skala zur allgemeinen naturwissenschaftlichen Arbeitsweise (Post-Test-Pilotierung) .....	56
Tabelle 5: Skala zur naturwissenschaftlichen Arbeitsweise biotech. (Post-Test-Pilotierung) .....	57
Tabelle 6: Übersichtstabelle der Skala zur Erfassung des Fachwissens im Prä-Test; sortiert nach Item-Reihenfolge .....	59
Tabelle 7: Übersichtstabelle der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Prä-Test; sortiert nach Item-Reihenfolge .....	61
Tabelle 8: Übersichtstabelle der Skala zur Erfassung des Fachwissens im Post-Test; sortiert nach Item-Reihenfolge .....	63
Tabelle 9: Übersichtstabelle der Skala zur Erfassung des allgemeinen naturwissenschaftlichen Erkenntnisgewinns im Post-Test; sortiert nach Item-Reihenfolge .....	65
Tabelle 10: Übersichtstabelle der Skala zur Erfassung des naturwissenschaftlichen Erkenntnisgewinns in der Biotechnologie im Post-Test; sortiert nach Item-Reihenfolge .....	67
Tabelle 11: Zusammensetzung der Testgruppen nach Schulart und Testzeitpunkt .....	72
Tabelle 12: Übersicht über die genutzten Items je Erhebungszeitpunkt .....	79
Tabelle 13: Pipettierschema für die vier Reaktionsansätze des Kriminalfalls .....	84
Tabelle 14: PCR-Protokoll zum Kriminalfall .....	85
Tabelle 15: Zusammensetzung 1,6%-iges Agarosegel .....	86
Tabelle 16: Zielwerte zur Überprüfung der Modelleignung (Fit-, Separations- und Reliabilitätswerte) .....	103
Tabelle 17: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Post-Test Fachwissen .....	113
Tabelle 18: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Follow-up-Test Fachwissen .....	118
Tabelle 19: Reliabilitäten und Separationswerte für Personen und Items .....	125
Tabelle 20: Anzahl an ausgeschlossenen unerwarteten Antworten je Item im Post-Test NWA .....	131

---

Tabelle 21: Reliabilitäten und Separationswerte für Personen und Items.....	137
Tabelle 22: ANOVA Tabelle.....	140
Tabelle 23: Deskriptive Statistik der Mittelwerte der Fachwissensfragebögen für Prä-, Post- und Follow-up-Test .....	140
Tabelle 24: Schätzungen der Mittelwerte der Fachwissensfragebögen für Prä-, Post- und Follow-up-Test (Ergebnis der Kovarianzanalyse) .....	141
Tabelle 25: Schätzungen der Mittelwerte der Fachwissensfragebögen für die vier Gruppen über die Messzeitpunkte hinweg (ANOVA mit Messwiederholung und Kovariaten).....	145
Tabelle 26: ANOVA Tabelle.....	151
Tabelle 27: Deskriptive Statistik der Mittelwerte der NWA-Skalen für Post- und Follow-up- Test.....	151
Tabelle 28: Schätzungen der Mittelwerte der NWA-Skalen.....	152
Tabelle 29: Schätzungen der Mittelwerte der Personenfähigkeiten aus den Fachwissensfragebögen für die vier Gruppen über die Messzeitpunkte hinweg (ANOVA mit Messwiederholung und Kovariaten) .....	155

### 10.3 Literaturverzeichnis

- Ab Rahman, R., Ahmad, S. & Hashim, U. R. (2019). A Study on Gamification for Higher Education Students' Engagement Towards Education 4.0, 67, 491–502. [https://doi.org/10.1007/978-981-13-6031-2\\_5](https://doi.org/10.1007/978-981-13-6031-2_5)
- Ableitner, O. (2018). *Einführung in die Molekularbiologie*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-20624-6>
- Adesope, O. O., Nesbit, J. C. & Sundararajan, N. (2022). The Mapping Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., 351-359). Cambridge University Press. <https://doi.org/10.1017/9781108894333.037>
- Adkins, S. & Burmeister, M. (1996). Visualization of DNA in agarose gels as migrating colored bands: applications for preparative gels and educational demonstrations. *Analytical biochemistry*, 240(1), 17–23. <https://doi.org/10.1006/abio.1996.0325>
- Alberts, B., Johnson, A., Lewis, J [Julian], Morgan, D., Raff, M., Roberts, K. & Walter, P. (2017). *Molekularbiologie der Zelle* (U. Schäfer, Hg.) (6. Auflage). Wiley-VCH Verlag GmbH & Co. KGaA.
- Al-Samarraie, H., Shamsuddin, A. & Alzahrani, A. I. (2020). A flipped classroom model in higher education: a review of the evidence across disciplines. *Educational Technology Research and Development*, 68(3), 1017–1051. <https://doi.org/10.1007/s11423-019-09718-8>
- America's Lab Report: Laboratory Experiences and Student Learning*. (2006). National Academies Press. <https://nap.nationalacademies.org/read/11311/chapter/5>  
<https://doi.org/10.17226/11311>
- Anderson, C. (1992). FBI gives in on genetics. *Nature*(355), 663. <https://doi.org/10.1038/355663a0>
- Armstrong, J. A. & Schulz, J. R. (2015). Agarose Gel Electrophoresis. *Current Protocols Essential Laboratory Techniques*, 10(1), 7.2.1-7.2.22. <https://doi.org/10.1002/9780470089941.et0702s10>
- Atawodi, S., Atawodi, J. C. & Dzikwi, A. A. (2010). Polymerase Chain Reaction: Theory, Practice and Application: A Review. *Sahel Medical Journal*, 13(2), 54–63. <https://dspace.unijos.edu.ng/jspui/handle/123456789/2607>
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Hrsg.), *Psychology of Learning and Motivation. The psychology of learning and motivation: Advances in research and theory* (Bd. 2, S. 89–195). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)

- Attié, E., Guibert, J. & Polle, C. (2022). Promoting Student Self-Regulation and Motivation Through Active Learning. In J. Keengwe (Hrsg.), *Advances in higher education and professional development (AHEPD) book series. Handbook of research on active learning and student engagement in higher education* (S. 202–226). IGI Global. <https://www.semanticscholar.org/paper/Promoting-Student-Self-Regulation-and-Motivation-Atti%C3%A9-Guibert/fd41260e10f65945ec8d9a97e557e987b4873409>
- Ayres, P. & Sweller, J. (2022). The Split-Attention Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S.199–211). Cambridge University Press. <https://doi.org/10.1017/9781108894333.020>
- Bachhiesl, C. W. (Hrsg.). (2011). *Austria: Forschung und Wissenschaft. Interdisziplinär: Bd. 7. Kriminologische Theorie und Praxis: Geistes- und naturwissenschaftliche Annäherungen an die Kriminalwissenschaft*. LIT Verlag.
- Backhaus (Hrsg.). (2016). *Multivariate Analysemethoden*. Springer. <https://doi.org/10.1007/978-3-662-46076-4>
- Backhaus, K., Erichson, B., Weiber, R. & Plinke, W. (2016). Varianzanalyse. In Backhaus (Hrsg.), *Multivariate Analysemethoden* (S. 173–213). Springer. [https://doi.org/10.1007/978-3-662-46076-4\\_4](https://doi.org/10.1007/978-3-662-46076-4_4)
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. <https://doi.org/10.1126/science.1736359>
- Baepler, P., Walker, J. D. & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education*, 78, 227–236. <https://doi.org/10.1016/j.compedu.2014.06.006>
- Bayrhuber, H. & Kull, U. (Hrsg.). (2005). *Linder Biologie* (22. Auflage). Schroedel Schulbuchverlag.
- Benbunan-Fich, R. & Hiltz, S. R. (1999). Impacts of Asynchronous Learning Networks on Individual and Group Problem Solving: A Field Experiment. *Group Decision and Negotiation*, 8(5), 409–426. <https://doi.org/10.1023/A:1008669710763>
- Berg, J. M., Tymoczko, J. L., Gatto, G. J. & Stryer, L. (2018). *Stryer Biochemie*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-54620-8>
- Bergmann, J. & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day* (1. ed.). ISTE/ASCD.
- Berkaoui, D., Castermans, K., Chandra, R. L., Nacken, H. & Querl, P. (2023, 14. Juni). *Virtuelle Labore*. [https://www.eteaching.org/didaktik/gestaltung/virtuelles\\_Labor#:~:text=Labore%20sind%20spezialisierte%20Orte%2C%20R%C3%A4ume,erm%C3%B6glichen%20die%20Erprobung%20komplexer](https://www.eteaching.org/didaktik/gestaltung/virtuelles_Labor#:~:text=Labore%20sind%20spezialisierte%20Orte%2C%20R%C3%A4ume,erm%C3%B6glichen%20die%20Erprobung%20komplexer)

- Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E. & Nerdel, C. (2023). Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Computers and Education: Artificial Intelligence*, 5, 100177. <https://doi.org/10.1016/j.caeai.2023.100177>
- Bezold, G., Krähn, G. & Peter, R. U. (2000). Falsch-negative Ergebnisse bei diagnostischer PCR wegen PCR-Inhibition: Überwachung mit internen Kontrollen. In A. Plettenberg, W. N. Meigel & I. Moll (Hrsg.), *Dermatologie an der Schwelle zum neuen Jahrtausend* (S. 49–52). Springer. [https://doi.org/10.1007/978-3-642-57191-6\\_15](https://doi.org/10.1007/978-3-642-57191-6_15)
- BioRender. (2023). *Scientific Image and Illustration Software*. <https://www.biorender.com/>
- Bishop, J. & Verleger, M. A. (2013). The Flipped Classroom: A Survey of the Research. *2013 ASEE Annual Conference & Exposition*, 23.1200.1-23.1200.18. <https://peer.asee.org/the-flipped-classroom-a-survey-of-the-research>
- Bishop, K. & Denley, P. (1997). The fundamental role of subject matter knowledge in the teaching of science. *The School science review*, 79, 65–71. <https://www.semanticscholar.org/paper/The-fundamental-role-of-subject-matter-knowledge-in-Bishop-Denley/ec402a049cfe44448f103fb4cbf71be9b5e6cdde>
- Blanz, M. (2021). *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen* (2. Auflage). Kohlhammer Verlag. <https://elibrary.kohlhammer.de/book/10.17433/978-3-17-039819-1>
- Bochter, R., Hofman, H., Hupfer, K. & Sinclair, S. (2009). *Nautilus: Biologie 11 : Ausgabe B* (1. Aufl.). Bayerischer Schulbuch-Verlag.
- Bokolo, A., JR, Adzhar, K., Awanis, R., Anis Farihan Mat, R., Danakorn Nincarean A. L. Eh, Phon, Aziman, A. & Gan Leong, M. (2022). Blended Learning Adoption and Implementation in Higher Education: A Theoretical and Systematic Review. *Technology, Knowledge and Learning*, 27(2), 531–578. <https://doi.org/10.1007/s10758-020-09477-z>
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge/Taylor & Francis Group. <https://psycnet.apa.org/record/2015-33472-000>
- Boone, W. J. & Staver, J. R. (2020). *Advances in Rasch Analyses in the Human Sciences* (1st ed. 2020). Springer. <https://doi.org/10.1007/978-3-030-43420-5>
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Bortz, J. & Schuster, C. (2016). *Statistik: Für Human- und Sozialwissenschaftler. Limitierte Sonderausgabe* (Limitierte Sonderausgabe, 7., vollständig überarbeitete und erweiterte Auflage). Springer. <https://doi.org/10.1007/978-3-642-12770-0>

- Boyle, M. P. & Boeck, K. de (2013). A new era in the treatment of cystic fibrosis: correction of the underlying CFTR defect. *The Lancet. Respiratory medicine*, 1(2), 158–163. [https://doi.org/10.1016/S2213-2600\(12\)70057-7](https://doi.org/10.1016/S2213-2600(12)70057-7)
- Brata, W. W. W., Wibowo, F. C. & Rahmadina, N. (2021). Implementation of discovery learning in a digital class and its effect on student learning outcomes and learning independence level. *F1000Research*, 10, 386. <https://doi.org/10.12688/F1000RESEARCH.51763.1>
- Breuer, J., Vogelsang, C. & Reinhold, P. (2020). Implementation und Nutzung von Unterrichtsmaterialien im schulischen Unterricht - Eine Bestandsaufnahme der mathematisch-naturwissenschaftlichen Fächer. *PhyDid A - Physik und Didaktik in Schule und Hochschule*, 19, 12–22. [https://www.researchgate.net/publication/341576253\\_Implementation\\_und\\_Nutzung\\_von\\_Unterrichtsmaterialien\\_im\\_schulischen\\_Unterricht\\_-\\_Eine\\_Bestandsaufnahme\\_der\\_mathematisch-naturwissenschaftlichen\\_Facher](https://www.researchgate.net/publication/341576253_Implementation_und_Nutzung_von_Unterrichtsmaterialien_im_schulischen_Unterricht_-_Eine_Bestandsaufnahme_der_mathematisch-naturwissenschaftlichen_Facher)
- Brown, M. (2013). From Classroom to Digital Arena in Seeking Higher-level Learning: Student Experience. *Seminar.net*, 9(1). <https://doi.org/10.7577/seminar.2390>
- Brown, T. A. (2011). *Gentechnologie für Einsteiger* (6. Auflage). Spektrum Akademischer Verlag.
- Brown, T. A. (2021). *Gene cloning and DNA analysis: An introduction* (8th ed.). Wiley-Blackwell.
- Buckley, B. C., Gobert, J. D., Kindfield, A. C. H., Horwitz, P., Tinker, R. F., Gerlits, B., Wilensky, U., Dede, C. & Willett, J. (2004). Model-Based Teaching and Learning with BioLogica™: What Do They Learn? How Do They Learn? How Do We Know? *Journal of Science Education and Technology*, 13(1), 23–41. <https://doi.org/10.1023/B:JOST.0000019636.06814.e3>
- Bühl, A. (2019). *SPSS: Einführung in die moderne Datenanalyse ab SPSS 25* (16., aktualisierte Auflage). *Pearson Studium - Scientific Tools*. Pearson. <https://elibrary.pearson.de/book/99.150005/9783863268718>
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2., aktualisierte und erweiterte Auflage). *Studium Psychologie*. Pearson Deutschland GmbH. <https://elibrary.pearson.de/book/99.150005/9783863268091>
- Bybee, R. W. (2002). Scientific Literacy — Mythos oder Realität? In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 21–43). Leske + Budrich. [https://doi.org/10.1007/978-3-322-80863-9\\_2](https://doi.org/10.1007/978-3-322-80863-9_2)
- Calderhead, J. & Miller, E. (1986). *The Integration of Subject Matter Knowledge in Student Teachers' Classroom Practice*. *Research Monograph Series Paper 1*.

- <https://www.semanticscholar.org/paper/The-Integration-of-Subject-Matter-Knowledge-in-1.-Calderhead-Miller/1ef5f7fe011587504e4a513d70ed65d9428a99da>
- Carson, S., Miller, H. B., Srougi, M. C. & Witherow, D. S. (2020). *Molecular biology techniques: A classroom laboratory manual* (Fourth edition). Academic Press. <https://doi.org/10.1016/C2017-0-03102-6>
- Castro-Rodríguez, M., Marín-Suelves, D., López-Gómez, S. & Rodríguez-Rodríguez, J. (2021). Mapping of Scientific Production on Blended Learning in Higher Education. *Education Sciences*, 11(9), 494. <https://doi.org/10.3390/educsci11090494>
- Chandler, P. & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), 293–332. [https://doi.org/10.1207/s1532690xci0804\\_2](https://doi.org/10.1207/s1532690xci0804_2)
- Chang, K.-E., Chen, Y.-L., Lin, H.-Y. & Sung, Y.-T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education*, 51(4), 1486–1498. <https://doi.org/10.1016/j.compedu.2008.01.007>
- Charalambous, C. Y. & Praetorius, A.-K. (2020). Creating a forum for researching teaching and its quality more synergistically. *Studies in Educational Evaluation*, 67, 100894. <https://doi.org/10.1016/j.stueduc.2020.100894>
- Chen, S., Zhu, X. & Kang, M. (2017). Development and validation of an energy-balance knowledge test for fourth- and fifth-grade students. *Journal of sports sciences*, 35(10), 1004–1011. <https://doi.org/10.1080/02640414.2016.1208837>
- Chen, X. & DeBoer, J. (2015). Checkable answers: Understanding student behaviors with instant feedback in a blended learning class. *2015 IEEE Frontiers in Education Conference (FIE)*. <https://www.semanticscholar.org/paper/Checkable-answers%3A-Understanding-student-behaviors-ChenDeBoer/5c6cf707acab4ff9869edabc6aaf6f55f33df90b>
- Chi, M. T. H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H. (2022). The Self-Explanation Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 381–393). Cambridge University Press. <https://doi.org/10.1017/9781108894333.040>
- Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D. & Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive science*(42), 1777–1832. <https://doi.org/10.1111/cogs.12626>

- Chi, M. T. H. & Menekse, M. (2015). Dialogue patterns that promote learning. In Resnick L. B., C. Asterhan & Clarke S. N. (Hrsg.), *Socializing intelligence through academic talk and dialogue* (S. 263–274). AERA.
- Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Chong, T. S. (2005). Recent Advances in Cognitive Load Theory Research: Implications for Instructional Designers. *Malaysian Online Journal of Instructional Technology*, 2(3), 106–117.
- Clark, R. C. & Mayer, R. E. (2016). *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (1. Aufl.). Wiley online library. Wiley. <http://onlinelibrary.wiley.com/book/10.1002/9781119239086>  
<https://doi.org/10.1002/9781119239086>
- Committee AMCTB (2013). *Analytical methods: advancing methods and applications*, 6(2), 333–336. <https://doi.org/10.1039/c3ay90101g>
- Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research*, 66(3), 227–268.  
<https://doi.org/10.3102/00346543066003227>
- Cronach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Daniel, J. Sir (2020). Education and the COVID-19 pandemic. *Prospects*, 49(1-2), 91–96. <https://doi.org/10.1007/s11125-020-09464-3>
- Davis, D. R. & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, 2, 100054. <https://doi.org/10.1016/j.ijedro.2021.100054>
- de Jong, T. (2006). Computer simulations. Technological advances in inquiry learning. *Science*, 312(5773), 532–533. <https://doi.org/10.1126/science.1127750>
- de Jong, T. (2022). The Guided Inquiry Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 394–402). Cambridge University Press.
- de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K. & Zacharia, Z. C. (2023). Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review*, 39, 100536. <https://doi.org/10.1016/j.edurev.2023.100536>

- de Jong, T., Linn, M. C. & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340, 3015–3208.  
<https://doi.org/10.1126/science.1231022>
- Degrelle, S. A., Hennequet-Antier, C., Chiapello, H., Piot-Kaminski, K., Piumi, F., Robin, S., Renard, J.-P. & Hue, I. (2008). Amplification biases: possible differences among deviating gene expressions. *BMC genomics*, 9, 46. <https://doi.org/10.1186/1471-2164-9-46>
- Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011). From game design elements to gamefulness. In A. Lugmayr, H. Franssila, C. Safran & I. Hammouda (Hrsg.), *ACM Other conferences, Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments* (S. 9–15). ACM.  
<https://doi.org/10.1145/2181037.2181040>
- Domingues, L. (Hrsg.). (2017). *Methods in molecular biology: Bd. 1620. PCR: Methods and Protocols* (1st edition 2017). Springer. <https://doi.org/10.1007/978-1-4939-7060-5>
- Döring, N. & Bortz, J. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften: In den Sozial- und Humanwissenschaften* (5 Aufl.). Springer-Lehrbuch. Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Downey, D. B., Hippel, P. T. von & Broh, B. A. (2004). Are Schools the Great Equalizer? Cognitive Inequality during the Summer Months and the School Year. *American Sociological Review*, 69(5), 613–635. <https://doi.org/10.1177/000312240406900501>
- Du, K., Sharma, M. & Lukacs, G. L. (2005). The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nature structural & molecular biology*, 12(1), 17–25. <https://doi.org/10.1038/nsmb882>
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual review of psychology*, 55, 51–86.  
<https://doi.org/10.1146/annurev.psych.55.090902.142050>
- Eckert, M. (2020). Kolbs Erfahrungslernen. In M. Eckert (Hrsg.), *essentials. Online-Lehre mit System: Wie man in der digitalen Lehre passgenaue Lernimpulse setzt und neue Lernerfahrungen ermöglicht* (S. 15–24). Springer Gabler. [https://doi.org/10.1007/978-3-658-32670-8\\_4](https://doi.org/10.1007/978-3-658-32670-8_4)
- Eggert, S., Nitsch, A., Boone, W. J., Nückles, M. & Bögeholz, S. (2017). Supporting Students' Learning and Socioscientific Reasoning About Climate Change—the Effect of Computer-Based Concept Mapping Scaffolds. *Research in Science Education*, 47(1), 137–159. <https://doi.org/10.1007/s11165-015-9493-7>
- Elborn, J. S. (2016). Cystic fibrosis. *Lancet*, 388(10059), 2519–2531.  
[https://doi.org/10.1016/S0140-6736\(16\)00576-6](https://doi.org/10.1016/S0140-6736(16)00576-6)

- Ensinck, M., Keersmaecker, L. de, Heylen, L., Ramalho, A. S., Gijbers, R., Farré, R., Boeck, K. de, Christ, F., Debyser, Z. & Carlon, M. S. (2020). Phenotyping of Rare CFTR Mutations Reveals Distinct Trafficking and Functional Defects. *Cells*, 9(3). <https://doi.org/10.3390/cells9030754>
- The European Business Review (Hrsg.). (2021). *Benefits of Virtual Labs for Education*. <https://www.europeanbusinessreview.com/benefits-of-virtual-labs-for-education/#:~:text=Affordability%0AL263%3A%20%0AL264%3A%20First%20off%2C%20virtual,maintenance%20costs%20and%20other%20expenses>
- Fend, H. (1981). *Theorie der Schule* (2., durchges. Aufl.). *U-&-S-Pädagogik*. Urban & Schwarzenberg.
- Finger, R. P., Fenwick, E., Pesudovs, K., Marella, M., Lamoureux, E. L. & Holz, F. G. (2012). Rasch analysis reveals problems with multiplicative scoring in the macular disease quality of life questionnaire. *Ophthalmology*, 119(11), 2351–2357. <https://doi.org/10.1016/j.ophtha.2012.05.031>
- Fiorella, L. & Mayer, R. E. (2022a). The Generative Activity Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., 339-350). Cambridge University Press. <https://doi.org/10.1017/9781108894333.036>
- Fiorella, L. & Mayer, R. E. (2022b). Principles Based on Social Cues in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 277–285). Cambridge University Press. <https://doi.org/10.1017/9781108894333.029>
- Fiorella, L. & Mayer, R. E. (2022c). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 185–198). Cambridge University Press. <https://doi.org/10.1017/9781108894333.019>
- Fleissner-Martin, J., Bogner, F. X. & Paul, J. (2023). Digital vs. Analog Learning—Two Content-Similar Interventions and Learning Outcomes. *Forests*, 14(9), 1807. <https://doi.org/10.3390/f14091807>
- Forsa Politik- und Sozialforschung GmbH. (2020). *Das Deutsche Schulbarometer Spezial Corona-Krise: Ergebnisse einer Befragung von Lehrerinnen und Lehrern an allgemeinbildenden Schulen* [Im Auftrag der Robert Bosch Stiftung in Kooperation mit der ZEIT] [https://www.vbe.de/fileadmin/user\\_upload/VBE/Service/Meinungsumfragen/2020-06-09\\_-\\_forsa-VBE\\_Corona\\_Schuloeffnungsphase.pdf](https://www.vbe.de/fileadmin/user_upload/VBE/Service/Meinungsumfragen/2020-06-09_-_forsa-VBE_Corona_Schuloeffnungsphase.pdf)

- Frailich, M., Kesner, M. & Hofstein, A. (2009). Enhancing students' understanding of the concept of chemical bonding by using activities provided on an interactive website. *Journal of Research in Science Teaching*, 46(3), 289–310. <https://doi.org/10.1002/tea.20278>
- Gadsby, D. C., Vergani, P. & Csanády, L. (2006). The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature*, 440(7083), 477–483. <https://doi.org/10.1038/nature04712>
- Gey, M. (2015). *Instrumentelle Analytik und Bioanalytik: Biosubstanzen, Trennmethoden, Strukturanalytik, Applikationen* (3. Auflage). Springer-Lehrbuch. Springer Spektrum. <https://doi.org/10.1007/978-3-622-46255-3>
- Gey, M. (2021). *Instrumentelle Analytik und Bioanalytik: Biosubstanzen, Trennmethoden, Strukturanalytik, Applikationen* (4. Auflage). Lehrbuch. Springer Spektrum.
- Gilhespy, I. (2011). Evaluating the effectiveness of digital learning objects: student perspectives and issues of design. <https://www.semanticscholar.org/paper/Evaluating-the-effectiveness-of-digital-learning-of-Gilhespy/b65bcc113140e557f7a219a4bce7330840172e4f>
- Girden, E. R. (1992). *ANOVA: repeated measures* (1. print). Sage University papers / Quantitative applications in the social sciences: Bd. 84. SAGE. <https://doi.org/10.4135/9781412983419>
- Goldhammer, F. & Hartig, J. (2020). Testwertinterpretation, Testnormen und Testeichung. In H. Moosbrugger & A. Kelava (Hrsg.), *Lehrbuch. Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage). Springer.
- Götz, T. & Nett, U. E. (2017). Selbstreguliertes Lernen. In T. Götz (Hrsg.), *UTB - Pädagogische Psychologie, Schulpädagogik: Bd. 3481. Emotion, Motivation und selbstreguliertes Lernen* (2., akt. Auflage, S. 143–184). Schöningh.
- Govender, R. (2023). Teaching and learning using virtual labs: Investigating the effects on students' self-regulation. *Cogent Education*, 10(1). <https://doi.org/10.1080/2331186X.2023.2172308>
- Graber, U. & Scholz, M. (2013). *Vor- und Nachteile von Online- und Präsenzphasen – Digital lehren*. [https://wiki.llz.uni-halle.de/Vor-\\_und\\_Nachteile\\_von\\_Online-\\_und\\_Pr%C3%A4senzphasen#:~:text=Lernende%20k%C3%B6nnen%20sich%20leichter%20entziehen,Didaktik.%20Zuletzt%20ge%C3%A4ndert](https://wiki.llz.uni-halle.de/Vor-_und_Nachteile_von_Online-_und_Pr%C3%A4senzphasen#:~:text=Lernende%20k%C3%B6nnen%20sich%20leichter%20entziehen,Didaktik.%20Zuletzt%20ge%C3%A4ndert)
- Gräber, W. & Nentwig, P. (2002). Scientific Literacy — Naturwissenschaftliche Grundbildung in der Diskussion. In W. Gräber, P. Nentwig, T. Koballa & R. Evans (Hrsg.), *Scientific*

- Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 7–20). Leske + Budrich. [https://doi.org/10.1007/978-3-322-80863-9\\_1](https://doi.org/10.1007/978-3-322-80863-9_1)
- Green, M. R. & Sambrook, J. (2019). Analysis of DNA by Agarose Gel Electrophoresis. *Cold Spring Harbor Protocols*, 2019(1), pdb.top100388. <https://doi.org/10.1101/pdb.top100388>
- Gropengießer, H. (2018). Biologielehrerinnen und Biologielehrer. In H. Gropengießer, U. Harms, U. Kattmann, S. Bögeholz & D. Eschenhagen (Hrsg.), *Unterricht Biologie. Fachdidaktik Biologie: Die Biologiedidaktik* (11. Auflage, S. 212–218). Aulis.
- Gropengießer, H. & Kattmann, U. (2018). Arbeiten mit Schülervorstellungen. In H. Gropengießer, U. Harms, U. Kattmann, S. Bögeholz & D. Eschenhagen (Hrsg.), *Unterricht Biologie. Fachdidaktik Biologie: Die Biologiedidaktik* (11. Auflage, S. 12–15). Aulis.
- Gunbas, N. (2015). Students' mathematics word problem-solving achievement in a computer-based story. *Journal of Computer Assisted Learning*, 31(1), 78–95. <https://doi.org/10.1111/jcal.12067>
- H5P.org. (2023a). *Branching Scenario*. <https://h5p.org/branching-scenario>
- H5P.org. (2023b). *H5P*. <https://h5p.org/>
- H5P.org. (2023c). *Interactive Book*. <https://h5p.org/content-types/interactive-book>
- Haeck, C. & Lefebvre, P. (2020). Pandemic School Closures May Increase Inequality in Test Scores. *Working Papers*(03), 1–10. <https://ideas.repec.org/p/grc/wpaper/20-03.html>
- Hafeez, M. & Akhter, Y. (2021). Effects of Blended Learning in Comparison of Traditional Learning to Provide Safer Learning Environment- A Comparative Review. *International Journal of Educational Research & Social Sciences*, 1604–1615. <https://www.semanticscholar.org/paper/Effects-of-Blended-Learning-in-Comparison-of-to-A-Hafeez-Akhter/c34edfd8bac161c5c2d11a6f3d0f5de70a163c9d>
- Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *APPLIED MEASUREMENT IN EDUCATION*, 15(3), 309–334. [https://doi.org/10.1207/S15324818AME1503\\_5](https://doi.org/10.1207/S15324818AME1503_5)
- Hammann, M. & Asshoff, R. (2023). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten* (5. Auflage). *Unterricht Biologie*. Klett/Kallmeyer.
- Hammerstein, S., König, C., Dreisörner, T. & Frey, A. (2021). Effects of COVID-19-Related School Closures on Student Achievement - A Systematic Review. *Frontiers in psychology*, 12, 1–8. <https://doi.org/10.3389/fpsyg.2021.746289>
- Harms, U. (2018). Fachwissen kennen und anwenden. In H. Gropengießer, U. Harms, U. Kattmann, S. Bögeholz & D. Eschenhagen (Hrsg.), *Unterricht Biologie. Fachdidaktik Biologie: Die Biologiedidaktik* (11. Auflage, S. 51–55). Aulis.

- Harms, U. & Kattmann, U. (2018). Sprache. In H. Gropengießer, U. Harms, U. Kattmann, S. Bögeholz & D. Eschenhagen (Hrsg.), *Unterricht Biologie. Fachdidaktik Biologie: Die Biologiedidaktik* (11. Auflage, S. 378–389). Aulis.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprinted.). Routledge.
- Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S. & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, 4(12), 1303–1312. <https://doi.org/10.1038/s41562-020-01009-0>
- Heller, K. A. & Perleth, C. (2000). *KFT 4-12+ R. Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* (2. Auflage). Beltz.
- Helmke, A. (2003). *Unterrichtsqualität: Erfassen. Bewerten. Verbessern*. Kallmeyer.
- Helmke, A. (2007). *Was wissen wir über guten Unterricht: Wissenschaftliche Erkenntnisse zur Unterrichtsforschung und Konsequenzen für die Unterrichtsentwicklung*, Bertelsmann.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung: Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung* (1. Auflage). Klett Kallmeyer. <https://elibrary.utb.de/doi/book/10.5555/9783772716850>
- Herget, T., Charton, J. E. & Hildemann, S. (2020). Bioethik in der Stammzellforschung und der Genom-Editierung am Beispiel eines Wissenschafts- und Technologieunternehmens. In S. Gerke, J. Taupitz, C. Wiesemann, C. Kopetzki & H. Zimmermann (Hrsg.), *Veröffentlichungen des Instituts Für Deutsches, Europäisches und Internationales Medizinrecht, Gesundheitsrecht und Bioethik der Universitäten Heidelberg und Mannheim Ser: v.48. Die klinische Anwendung von humanen induzierten pluripotenten Stammzellen: Ein Stakeholder-Sammelband* (1. Aufl. 2020, Bd. 48, S. 143–156). Springer. [https://doi.org/10.1007/978-3-662-59052-2\\_4](https://doi.org/10.1007/978-3-662-59052-2_4)
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hodges, C., Moore, S., Lockee, B., Turst, T. & Bond, A. (2020). *The Difference Between Emergency Remote Teaching and Online Learning*. EDUCAUSE REVIEW. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Högström, P., Ottander, C. & Benckert, S. (2010). Lab Work and Learning in Secondary School Chemistry: The Importance of Teacher and Student Interaction. *Research in Science Education*(40), 505–523. <https://doi.org/10.1007/s11165-009-9131-3>

- Horneber, M. & Möller, C. (2021). Kundenorientiert, digital und agil! Gesundheitsversorgung 2030. In B. Simon & N. Krämer (Hrsg.), *Zukunft der Gesundheitsversorgung: Vorschläge und Konzepte aus Perspektive* (S. 41–51). GABLER. [https://doi.org/10.1007/978-3-658-33007-1\\_4](https://doi.org/10.1007/978-3-658-33007-1_4)
- Huitema, B. E. (2011). *Analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies* (2nd ed (Online-Ausg.)). *Wiley series in probability and statistics*. Wiley. <https://doi.org/10.1002/9781118067475>
- IBM Corp. Released. (2021). *IBM SPSS Statistics for Windows* (Version Version 28.0) [Computer software]. IBM Corp.
- Institut für die Qualitätsentwicklung im Bildungswesen (IQB) in Berlin. (2022). *KERMIT IM KOHORTENVERGLEICH: Vergleich der Ergebnisse von KERMIT 3, 5 und 7 aus dem Jahr 2021 mit früheren Erhebungen*. <https://www.hamburg.de/contentblob/15935050/19a8ab3b185aa080dcc894322563776f/data/ergebnisbericht-kermit.pdf>
- Ireland, J., Johnson, N., Adams, D., Eboh, W. & Mowatt, E. (2009). Blended learning in education: effects on knowledge and attitude. *British journal of nursing*, 18(2), 124–130.  
<https://www.semanticscholar.org/paper/Blended-learning-in-education%3A-effects-on-knowledge-Ireland-Johnson/de4e09dfb4f5e1876b32f8bf31f4cc5e40c5b512>
- Ito, P. K. (1980). Robustness of ANOVA and MANOVA test procedures. In P. R. Krishnaiah & C. R. Rao (Hrsg.), *Handbook of Statistics. Analysis of variance* (Vol 1, p. 199–236). [https://doi.org/10.1016/S0169-7161\(80\)01009-7](https://doi.org/10.1016/S0169-7161(80)01009-7)
- Jansohn, M. & Rothhämel, S. (Hrsg.). (2012). *Gentechnische Methoden*. Spektrum Akademischer Verlag. <https://doi.org/10.1007/978-3-8274-2430-3>
- Janssen, J [Jeroen], Kirschner, F. & Kirschner, P. A. (2022). The Collaboration Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 304–312). Cambridge University Press. <https://doi.org/10.1017/9781108894333.032>
- Janssen, J [Jürgen] & Laatz, W. (2013). *Statistische Datenanalyse mit SPSS*. Springer. <https://doi.org/10.1007/978-3-642-32507-6>
- JASP Team. (2023). *JASP (Version 0.17.3)* [Computer software].
- Jessen, A. (2019). Ganz der Papa...: Vaterschaftsnachweis. *Pflegezeitschrift*, 72(5), 66. <https://doi.org/10.1007/s41906-019-0058-7>
- Johnson, C. I. & Marraffino, M. (2022). The Feedback Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 403–417). Cambridge University Press. <https://doi.org/10.1017/9781108894333.042>

- Jonkisz, E., Moosbrugger, H. & Brandt Holger. (2011). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* (2. Aufl. 2012). Springer.
- Kalyuga, S. & Sweller, J. (2022). The Redundancy Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 212–220). Cambridge University Press. <https://doi.org/10.1017/9781108894333.021>
- Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests [Computer software]*. R package version 0.7.2. <https://rpkgs.datanovia.com/rstatix/>
- Kattmann, U. (2005). Lernen mit anthropomorphen Vorstellungen? – Ergebnisse von Untersuchungen zur Didaktischen Rekonstruktion in der Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*, 11, 165–174.
- Kattmann, U. (2007). Didaktische Rekonstruktion — eine praktische Theorie. In D. Krüger (Hrsg.), *Springer-Lehrbuch. Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden ; mit 12 Tabellen* (S. 93–104). Springer. [https://doi.org/10.1007/978-3-540-68166-3\\_9](https://doi.org/10.1007/978-3-540-68166-3_9)
- Kattmann, U. (2015). Die Neigung, Fehler zu machen, und wie man mit ihnen lernen kann. *Das Seminar*, 21(4), 84–95.
- Kattmann, U. (2022). *Schüler besser verstehen: Alltagsvorstellungen im Biologieunterricht*. Aulis. <https://elibrary.utb.de/doi/book/10.5555/9783761490099>
- Kattmann, U., Duit, R., Gropengießer, H. & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18.
- Kemp, N. & Grieve, R. (2014). Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. *Frontiers in psychology*, 5, 1278. <https://doi.org/10.3389/fpsyg.2014.01278>
- Kennedy, A. I., Mejía-Rodríguez, A. M. & Strello, A. (2022). Inequality in remote learning quality during COVID-19: student perspectives and mitigating factors. *Large-scale assessments in education*, 10(1), 29. <https://doi.org/10.1186/s40536-022-00143-7>
- Killermann, W., Hiering, P. & Starosta, B. (Hrsg.). (2020). *Immer besser unterrichten. Biologieunterricht heute: Eine moderne Fachdidaktik* (16 Aufl.). Auer.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)

- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In BMBF (Hrsg.), *TIMSS - Impulse für Schule und Unterricht* (S. 43–57). Bundesministerium für Bildung & Forschung.
- Kostić, J. & Ranđelović, K. R. (2022). Digital distractions: learning in multitasking environment. *Psychological Applications and Trends*, 301–304. <https://www.semanticscholar.org/paper/DIGITAL-DISTRACTIONS%3A-LEARNING-IN-MULTITASKING-Kosti%C4%87-Ran%C4%91elovi%C4%87/baa2ae089307becb84114f787cb202eb038273a7>
- Kramer, M., Förtsch, C., Aufleger, M. & Neuhaus, B. J. (2019). Der Einsatz digitaler Medien im gymnasialen Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 131–160. <https://doi.org/10.1007/s40573-019-00096-5>
- Krimsky, S. & Simoncelli, T. (2012). *Genetic justice: DNA data banks, criminal investigations, and civil liberties*. Columbia University Press. <https://swbplus.bsz-bw.de/bsz364234040kla.htm>
- Krüger, L. (2018). Zystische Fibrose – gibt es Aussicht auf Heilung? *manuelletherapie*, 22(02), 61–66. <https://doi.org/10.1055/a-0585-1569>
- Kubinger, K. D., Holocher-Ertl, S., Reif, M., Hohensinn, C. & Frebort, M. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. *International Journal of Selection and Assessment*, 18(1), 111–115. <https://doi.org/10.1111/j.1468-2389.2010.00493.x>
- Kück, U. (Hrsg.). (2005). *SpringerLink Bücher. Praktikum der Molekulargenetik*. Springer. <https://doi.org/10.1007/b137618>
- Kultusministerkonferenz. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Biologie.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf)
- Kultusministerkonferenz. (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife*. Carl Link Verlag. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2020/2020\\_06\\_18-BildungsstandardsAHR\\_Biologie.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Biologie.pdf)
- Kunter, M., Baumert, J. & Blum, W. (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (1. Aufl.). Waxmann Verlag GmbH. Landesbildungsserver Baden-Württemberg. *Versuchsskizzen: PowerPoint-Präsentation*. <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjBt4am8LCCAxXUSPEDHTycDXUQFnoECBMQAQ&url=https%3A%2F%2Fwww.sc-hule-bw.de%2Ffaecher-und-schularten%2Fmathematisch-naturwissenschaftliche->

- faecher%2Fchemie%2Fneuer-index.html%2Fnuetzliches-fuer-die-  
 unterrichtsvorbereitung%2Fskizze%2Fversuchsskizzen.ppt&usg=AOvVaw0Dix-  
 ScHXhF8Nc0\_j-9RMA&opi=89978449.
- le Plat, D., Henkel, R., Hickmann, T. & Tuckermann, R. (2018). Eisen-Chrom-Redox-Flow-  
 Batterie für Schülerversuche. *CHEMKON*, 25(7), 269–277.  
<https://doi.org/10.1002/ckon.201800003>
- Lee, C.-S. B., Das, R. & Reed, R. (2003). Agarose gel separation/isolation of RNA-protein  
 complexes. *Current protocols in molecular biology*, Chapter 27, Unit 27.1.  
<https://doi.org/10.1002/0471142727.mb2701s63>
- Lee, P. Y., Costumbrado, J., Hsu, C.-Y. & Kim, Y. H. (2012). Agarose gel electrophoresis for  
 the separation of DNA fragments. *Journal of visualized experiments : JoVE*, 62.  
<https://doi.org/10.3791/3923>
- Leopold, C. (2022). The Imagination Principle in Multimedia Learning. In R. E. Mayer & L.  
 Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia  
 learning* (3rd ed., S. 370–380). Cambridge University Press.  
<https://doi.org/10.1017/9781108894333.039>
- Lester, P. M. & King, C. M. (2009). Analog vs. Digital Instruction and Learning: Teaching  
 Within First and Second Life Environments. *Journal of Computer-Mediated  
 Communication*, 14(3), 457–483. <https://doi.org/10.1111/j.1083-6101.2009.01449.x>
- Letzel, V. & Pozas, M. (2022). Differenzierung im digital durchgeführten Distanzunterricht.  
 Eine quantitative Studie zu Implementationshäufigkeit und Prädiktoren von inklusiven  
 Unterrichtsmaßnahmen. *DDS – Die Deutsche Schule*, 2022(03), 263–279.  
<https://doi.org/10.31244/dds.2022.03.03>
- Leutner, D. & Schmeck, A. (2022). The Drawing Principle in Multimedia Learning. In R. E.  
 Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of  
 multimedia learning* (3rd ed., S. 360–369). Cambridge University Press.  
<https://doi.org/10.1017/9781108894333.038>
- Lewis, H. A., Buchanan, S. G., Burley, S. K., Connors, K., Dickey, M., Dorwart, M., Fowler, R.,  
 Gao, X., Guggino, W. B., Hendrickson, W. A., Hunt, J. F., Kearins, M. C., Lorimer, D.,  
 Maloney, P. C., Post, K. W., Rajashankar, K. R., Rutter, M. E., Sauder, J. M.,  
 Shriver, S., . . . Emtage, S. (2004). Structure of nucleotide-binding domain 1 of the  
 cystic fibrosis transmembrane conductance regulator. *The EMBO journal*, 23(2), 282–  
 293. <https://doi.org/10.1038/sj.emboj.7600040>
- Lewis, J [Jenny] & Kattmann, U. (2004). Traits, genes, particles and information: re-visiting  
 students' understandings of genetics. *International Journal of Science Education*,  
 26(2), 195–206. <https://doi.org/10.1080/0950069032000072782>

- Lewis, J [Jenny] & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance - do students see any relationship? *International Journal of Science Education*, 22(2), 177–195. <https://doi.org/10.1080/095006900289949>
- Lexikon der Biologie - Spektrum. (2018). *Konturlänge von DNA*.  
<https://www.spektrum.de/lexikon/biologie/konturlaenge-von-dna/36944>
- Linacre, J. M. (1997). *Rasch Measurement Transactions: KR-20 / Cronbach Alpha or Rasch Person Reliability: Which Tells the "Truth"?* 11(3). p. 580-581.  
<https://www.rasch.org/rmt/rmt113l.htm>
- Linacre, J. M. (2009). Local independence and residual covariance: A study of Olympic figure skating ratings, 10(2), 1–13. <https://www.winsteps.com/a/linacre-figure-skating.pdf>
- Linacre, J. M. (2012). *Winsteps Tutorial 2*. <https://www.winsteps.com/a/winsteps-tutorial-2.pdf>
- Linacre, J. M. (2023a). *A User's Guide to W I N S T E P S: Winsteps Help for Rasch Analysis*.  
<https://www.winsteps.com/a/Winsteps-Manual.pdf>
- Linacre, J. M. (2023b). *Winsteps®* (Version Version 5.6.0) [Computer software]. Portland, Oregon: Winsteps.com. Available from <https://www.winsteps.com/>
- Linder, K. E. (2017). Fundamentals of Hybrid Teaching and Learning. *New Directions for Teaching and Learning*, 2017(149), 11–18. <https://doi.org/10.1002/tl.20222>
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Springer-Lehrbuch. Pädagogische psychologie* (S. 73–101). Springer. [https://doi.org/10.1007/978-3-540-88573-3\\_4](https://doi.org/10.1007/978-3-540-88573-3_4)
- Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Hrsg.), *Lehrbuch. Pädagogische Psychologie* (3., vollständig überarbeitete und aktualisierte Auflage, S. 69–118). Springer. [https://doi.org/10.1007/978-3-662-61403-7\\_4](https://doi.org/10.1007/978-3-662-61403-7_4)
- Lodge, J. & Horvath, J. (2017). Science of learning and digital learning environments. <https://www.semanticscholar.org/paper/Science-of-learning-and-digital-learning-Lodge-Horvath/f9f36c94ea188ed31eb34a2bb7bc4b39edb62833>
- Löffler, G. & Petrides, P. E. (2022). *Löffler/Petrides Biochemie und Pathobiochemie* (10., vollständig überarbeitete Auflage). *Lehrbuch*. Springer. <https://doi.org/10.1007/978-3-662-60266-9>
- Lowe, R. K., Schnotz, W. & Boucheix, J.-M. (2022). The Animation Composition Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 313–323). Cambridge University Press. <https://doi.org/10.1017/9781108894333.033>
- Luger, O., Tröstl, A. & Urferer, K. (2017). *Gentechnik geht uns alle an! Ein Überblick über Praxis und Theorie* (2. Aufl. 2017). *SpringerLink Bücher*. Springer VS. <https://doi.org/10.1007/978-3-658-15605-3>
- Lumi Education GbR. (2022). *Lumi*. <https://app.lumi.education/>

- Lutz, W., Elsholz, M., Haase, S., Burde, J.-P., Wilhelm, T. & Trefzger, T. (2023). Flipped Classroom im Physikunterricht der Sekundarstufe I – Auswirkungen auf die Veränderung des individuellen Interesses im Bereich der E-Lehre. In J. Roth, M. Baum, K. Eilerts, G. Hornung & T. Trefzger (Hrsg.), *Die Zukunft des MINT-Lernens: Band 2. Digitale Tools und Methoden für das Lehren und Lernen* (S. 35–49). Springer Spektrum. [https://doi.org/10.1007/978-3-662-66133-8\\_3](https://doi.org/10.1007/978-3-662-66133-8_3)
- Makransky, G. (2022). The Immersion Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., 296-303). Cambridge University Press. <https://doi.org/10.1017/9781108894333.031>
- Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A. & Phatak, V. (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simulation in healthcare : journal of the Society for Simulation in Healthcare*, 2(1), 4–10. <https://doi.org/10.1097/SIH.0b013e31802b68ee>
- Markl, J. (Hrsg.). (2010). *Biologie: Oberstufe* (1. Auflage). Klett.
- Martins, A., Fonseca, M. J., Lemos, M., Lencastre, L. & Tavares, F. (2020). Bioinformatics-Based Activities in High School: Fostering Students' Literacy, Interest, and Attitudes on Gene Regulation, Genomics, and Evolution. *Frontiers in microbiology*, 11, 578099. <https://doi.org/10.3389/fmicb.2020.578099>
- Mayer, J [Jürgen] (2002). Vom Schulversuch zum Forschenden Unterricht: Wissenschaftliches Arbeiten im Biologieunterricht am Beispiel der Fotosynthese (Workshop zum BLK-Programm).
- Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Hrsg.), *Cambridge Handbooks in Psychology. The Cambridge handbook of multimedia learning* (2nd ed., S. 43–71). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.005>
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>
- Mayer, R. E. (2022). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 57–72). Cambridge University Press. <https://doi.org/10.1017/9781108894333.008>
- Mayer, R. E. & Fiorella, L. (Hrsg.). (2022a). *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108894333>

- Mayer, R. E. & Fiorella, L. (2022b). Principles for Managing Essential Processing in Multimedia Learning: Segmenting, Pre-training, and Modality Principles In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 243–260). Cambridge University Press.  
<https://doi.org/10.1017/9781108894333.025>
- Mayer, R. E. & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52.  
[https://doi.org/10.1207/S15326985EP3801\\_6](https://doi.org/10.1207/S15326985EP3801_6)
- Mayrose, J. (2012). Active Learning Through The Use Of Virtual Environments. *American Journal of Engineering Education*, 3(1), 13–18.  
<https://www.semanticscholar.org/paper/Active-Learning-Through-The-Use-Of-Virtual-Mayrose/013b67e1e3ef43970c2fe5286fb1ecbb0c799350>
- McDonell, M. W., Simon, M. N. & Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *Journal of molecular biology*, 110(1), 119–146. [https://doi.org/10.1016/s0022-2836\(77\)80102-2](https://doi.org/10.1016/s0022-2836(77)80102-2)
- McKinsey & Company (2022). How COVID-19 caused a global learning crisis. *McKinsey & Company*. <https://www.mckinsey.com/industries/education/our-insights/how-covid-19-caused-a-global-learning-crisis#/>
- McNamara, D. (1991). Subject Knowledge and its Application: problems and possibilities for teacher educators. *Journal of Education for Teaching*, 17(2), 113-128.  
<https://www.semanticscholar.org/paper/Subject-Knowledge-and-its-Application%3A-problems-and-McNamara/fca52f4149a70eeba86ec86939525beadde3753f>
- McQueen, M. J. (1998). Ethical and legal issues in the procurement, storage and use of DNA. *Clinical chemistry and laboratory medicine*, 36(8), 545–549.  
<https://doi.org/10.1515/CCLM.1998.093>
- mebis-Redaktion. (2023). *Lernplattform - Überblick: mebis - Landesmedienzentrum Bayern*.  
<https://mebis.bycs.de/beitrag/lernplattform-ueberblick>
- Meier, M., Greefrath, G., Hammann, M., Wodzinski, R. & Ziepprecht, K. (2023). Lehr-Lern-Labore und Digitalisierung: Strömungen und Veränderungen in der Lehr-Lern-Labor-Landschaft als Folge des digitalen Wandels. In M. Meier, G. Greefrath, M. Hammann, R. Wodzinski & K. Ziepprecht (Hrsg.), *Lehr-Lern-Labore und Digitalisierung* (S. 1–16). Springer VS, Wiesbaden. [https://doi.org/10.1007/978-3-658-40109-2\\_1](https://doi.org/10.1007/978-3-658-40109-2_1)
- Meier, M. & Mayer, J [Jurgen] (2014). Selbständiges Experimentieren: Entwicklung und Einsatz eines anwendungsbezogenen Aufgabendesigns. *Der mathematische und naturwissenschaftliche Unterricht*, 67, 4–10.

- Merck KGaA. (2023a). *Agarose: Eigenschaften und Forschungsanwendungen*.  
<https://www.sigmaaldrich.com/DE/de/products/chemistry-and-biochemicals/biochemicals/agarose#csr>
- Merck KGaA. (2023b). *Sicherheitsdatenblatt: GelGreen® Nucleic Acid Stain (10,000X Water)*.  
<https://www.sigmaaldrich.com/DE/de/sds/mm/sct125>
- Meyer, H. (2004). *Was ist guter Unterricht?* (15. Aufl.). Cornelsen Verlag GmbH.
- Meyer, H. (2020). *Didaktische Ansprüche an Homeschooling und Fernunterricht*.  
<https://unterrichten.digital/2020/05/07/hilbert-meyer-homeschooling/>
- Meyer, L. & Stein, M. (2022). Schulische Förderung von Peerkontakten und sozialem Lernen während und nach der Pandemie. *Zeitschrift für Schul- und Professionsentwicklung*, 4(5), 189–206. <https://doi.org/10.11576/PFLB-5905>
- Millar, R. (2006). Twenty First Century Science: Insights from the Design and Implementation of a Scientific Literacy Approach in School Science. *International Journal of Science Education*, 28(13), 1499–1521. <https://doi.org/10.1080/09500690600718344>
- miniPCR. (2022). *Missy Baker Missing: Instructor's and Student's Guide*. miniPCR bioTM.  
<https://www.minipcr.com/product/minipcr-crime-lab-missy-baker/>
- miniPCR bio. (2023). *Apps - Apps for thermocycler*. <https://www.minipcr.com/products/apps/>
- Moosbrugger, H. & Kelava, A. (2011). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* (2. Aufl. 2012). Springer.
- Moosbrugger, H. & Kelava, A. (Hrsg.). (2020). *Lehrbuch. Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage). Springer.
- Mottarella, K., Fritzsche, B. & Parrish, T. (2005). Who Learns More? Achievement Scores following Web-Based versus Classroom Instruction in Psychology Courses. *Psychology Learning & Teaching*, 4(1), 51–54. <https://doi.org/10.2304/plat.2004.4.1.51>
- Mueller, P. A. & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: advantages of longhand over laptop note taking. *Psychological science*, 25(6), 1159–1168. <https://doi.org/10.1177/0956797614524581>
- Müller, H.-J. & Prange, D. R. (2016). *PCR - Polymerase-Kettenreaktion* (2. Aufl. 2016). *SpringerLink Bücher*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-48236-0>
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*, 51 Pt 1, 263–273.  
<https://doi.org/10.1101/sqb.1986.051.01.032>

- Mullis, K. B [Kary B.] & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Methods in Enzymology : Recombinant DNA Part F* (Bd. 155, S. 335–350). Academic Press. [https://doi.org/10.1016/0076-6879\(87\)55023-6](https://doi.org/10.1016/0076-6879(87)55023-6)
- Murphy, E. E. (2013). *Legal and Ethical Issues in Forensic DNA Phenotyping*. <https://doi.org/10.2139/ssrn.2288204>
- Narciss, S. (2018). Feedbackstrategien für interaktive Lernaufgaben. In H. Niegemann & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien* (S. 1–24). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-662-54373-3\\_35-1](https://doi.org/10.1007/978-3-662-54373-3_35-1)
- Narita, M. (1993). Polymerase chain reaction for diagnosis of infectious diseases. *Acta paediatrica Japonica : Overseas edition*, 35(2), 89–97. <https://doi.org/10.1111/j.1442-200x.1993.tb03015.x>
- National Academies Press (US). (2003). *Bio 2010: Transforming undergraduate education for future research biologists*. National Academies Press. <https://doi.org/10.17226/10497>
- Nerdel, C. (2017). *Grundlagen der Naturwissenschaftsdidaktik*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-53158-7>
- Neuhaus, B. J. (2021). Unterrichtsqualität aus der Perspektive der Biologiedidaktik. *Unterrichtswissenschaft*, 49(2), 273–283. <https://doi.org/10.1007/s42010-021-00114-x>
- Nissim-Rafinia, M., Linde, L. & Kerem, B. (2006). The CFTR Gene: Structure, Mutations and Specific Therapeutic Approaches. In A. Bush, Alton Eric W. F. W., Davies, Jane C., Griesenbach, Uta & A. Jaffe (Hrsg.), *Progress in respiratory research: Bd. 34. Cystic fibrosis in the 21st century* (S. 2–10). Karger. <https://doi.org/10.1159/000088467>
- Otsetov, A. G. (2020). Importance of Laboratory Techniques. *Journal of Clinical Chemistry and Laboratory Medicine, Vol.2 Iss.3(146)*, 1–2. <https://doi.org/10.35248/clinical-chemistry-laboratory->
- Paas, F. & Sweller, J. (2022). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 73–81). Cambridge University Press. <https://doi.org/10.1017/9781108894333.009>
- Pérez-Juárez, M. Á., González-Ortega, D. & Aguiar-Pérez, J. M. (2023). Digital Distractions from the Point of View of Higher Education Students. *Sustainability*. <https://www.semanticscholar.org/paper/Digital-Distractions-from-the-Point-of-View-of-P%C3%A9rez-Ju%C3%A1rez-Gonz%C3%A1lez-Ortega/51cab69e44e4aea2bb7af52744a022fc3f8cf82a>
- Peter, V. (2019). *Expertiseabhängige Wirksamkeit von instruktional unterstützten digitalen* [Masterarbeit]. TUM School of Education.

- Plass, J. L. & Hovey, C. (2022). The Emotional Design Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., 324-336). Cambridge University Press. <https://doi.org/10.1017/9781108894333.034>
- Praetorius, A.-K. & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? *Unterrichtswissenschaft*, 49(2), 167–188. <https://doi.org/10.1007/s42010-021-00119-6>
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48(3), 409–446. <https://doi.org/10.1007/s42010-020-00082-8>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing [Computer software]*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rahman, M. H. A., Ismail Yusuf Panessai, I., Mohd Noor, N. A. Z. & Mat Salleh, N. S. (2018). Gamification elements and their impacts on teaching and learning - a review. *The International journal of Multimedia & Its Applications*, 10(06), 37–46. <https://doi.org/10.5121/ijma.2018.10604>
- Razali, N. M. & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33. [https://www.nbi.dk/~petersen/teaching/stat2017/power\\_comparisons\\_of\\_shapiro-wilk\\_kolmogorov-smirn.pdf](https://www.nbi.dk/~petersen/teaching/stat2017/power_comparisons_of_shapiro-wilk_kolmogorov-smirn.pdf)
- Renkl, A. (2022). The Worked Example Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., 231-240). Cambridge University Press. <https://doi.org/10.1017/9781108894333.023>
- Renneberg, R., Süßbier, D., Berkling, V. & Loroach, V. (2023). *Biotechnology for beginners* (R. FitzRoy, Übers.) (3rd ed.). Academic Press. <https://dokumen.pub/biotechnology-for-beginners-3nbsped-0323855695-9780323855693.html>
- Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N. & Chou, J. L. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245(4922), 1066–1073. <https://doi.org/10.1126/science.2475911>
- Rogge, T. & Seifert, A. (2023). Belastungserleben als Herausforderung von Abiturient\*innen im digital vermittelten Distanzunterricht. In D. Newiak, J. Romppel & A. Martin (Hrsg.), *Digitale Bildung jetzt!* (S. 115–134). Springer VS, Wiesbaden. [https://doi.org/10.1007/978-3-658-40845-9\\_8](https://doi.org/10.1007/978-3-658-40845-9_8)

- Rost, J. (2004). *Lehrbuch Testtheorie Testkonstruktion* (2., überarb. und erw. Aufl.). *Psychologie-Lehrbuch*. Huber.  
[http://bvbr.bib-bvb.de:8991/F?func=service&doc\\_library=BVB01&doc\\_number=010369310&line\\_number=0001&func\\_code=DB\\_RECORDS&service\\_type=MEDIA](http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&doc_number=010369310&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA)
- Rost, J. (2006). Item-Response-Theorie. In F. Petermann & M. Eid (Hrsg.), *Hogrefe eLibrary: Band 4. Handbuch der Psychologischen Diagnostik (Reihe: Handbuch der Psychologie Bd. 4)* (1. Auflage). Hogrefe Verlag.
- Roth, J. (2015). Lernpfade – Definition, Gestaltungskriterien und Unterrichtseinsatz. In J. Roth, E. Süss-Stepancik & H. Wiesner (Hrsg.), *Medienvielfalt im Mathematikunterricht* (S. 3–25). Springer.
- Roth, J. & Wiesner, H. (2014). Lernpfade – Ein Weg zur selbständigen und sinnvollen Nutzung von digitalen Werkzeugen durch Schüler/innen. In J. Roth & J. Ames (Hrsg.), *Beiträge zum Mathematikunterricht* (S. 1003–1006). WTM-Verlag.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839), 487–491.  
<https://doi.org/10.1126/science.2448875>
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230(4732), 1350–1354.  
<https://doi.org/10.1126/science.2999980>
- Sailer, M. & Homner, L. (2020). The Gamification of Learning: a Meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Salkind, N. J. & Rasmussen, K. (Hrsg.). (2007). *Encyclopedia of measurement and statistics*. SAGE. <http://site.ebrary.com/lib/uniregensburg/Doc?id=10367430>
- Sälzer, C. (2016). Grundzüge des Rasch-Modells. In C. Sälzer (Hrsg.), *Studienbuch Schulleistungsstudien* (S. 43–51). Springer. [https://doi.org/10.1007/978-3-662-45765-8\\_3](https://doi.org/10.1007/978-3-662-45765-8_3)
- Sánchez, J., Lesmes, M., Azpeleta, C. & Gal, B. (2022). Work station learning activities (WSLA) through the ICAP framework: a qualitative study. *BMC Medical Education*, 22(1), 748. <https://doi.org/10.1186/s12909-022-03794-w>
- Scheiter, K. (2022). The Learner Control Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 418–429). Cambridge University Press.  
<https://doi.org/10.1017/9781108894333.043>
- Schmidt, C. & Dietrich, L. (2022). *Chemie im Biologiestudium: Von Grund auf verständlich erklärt* (2. Auflage). *Lehrbuch*. Springer Spektrum.

- Schmidt, H. & Rothhämel, S. (2012). Polymerase-Kettenreaktion (PCR). In M. Jansohn & S. Rothhämel (Hrsg.), *Gentechnische Methoden* (S. 135–171). Spektrum Akademischer Verlag. [https://doi.org/10.1007/978-3-8274-2430-3\\_4](https://doi.org/10.1007/978-3-8274-2430-3_4)
- Schmidt, R., Süß-Stepancik, E., Wiesner, H. & Roth, J. (2015). Konstruktiver Umgang mit Heterogenität – Der Beitrag von Lernpfaden. In J. Roth, E. Süß-Stepancik & H. Wiesner (Hrsg.), *Medienvielfalt im Mathematikunterricht* (S. 117–135). Springer. [https://doi.org/10.1007/978-3-658-06449-5\\_7](https://doi.org/10.1007/978-3-658-06449-5_7)
- Schnieder, L. C. (2015). Motive und Entscheidungswege bei Nicht-Tester\_innen. Diskussion eines Fallbeispiels. In T. Lemke & K. Liebsch (Hrsg.), *Frankfurter Beiträge zur Soziologie und Sozialpsychologie. Die Regierung der Gene: Diskriminierung und Verantwortung im Kontext genetischen Wissens* (S. 155–168). Springer VS. [https://doi.org/10.1007/978-3-658-09651-9\\_7](https://doi.org/10.1007/978-3-658-09651-9_7)
- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schöppner, P., Großbruchhaus, S. & Nerdel, C. (2022). *Biotechnologie praxisorientiert unterrichten: Aktuelle Kontexte für Schule und Lehrerfortbildung* (1. Auflage). Springer. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=7151637>
- Schöppner, P. & Nerdel, C. (2015). *Anwendung von Molekularbiologischen Methoden im naturwissenschaftlichen Unterricht: Skript zur Lehrerfortbildung*. München. Amgen Foundation; Technische Universität München. <https://www.edu.sot.tum.de/fdls/lehrkraeftefortbildungen/molekularbiologie-bioinformatik/fortbildungsmaterialien/>
- Schulmeister, R. (2016). Präsenz und Selbststudium im E-Learning. Indizien für eine besondere Rolle der Präsenz. In Hochschulrektorenkonferenz & Freie Universität Berlin (Vorsitz), *Digitale Lehrformen für ein studienzentriertes und kompetenzorientiertes Studium*.
- Schwanewedel, J. (2006). Vorstellungen zu Gesundheit und Krankheit im Kontext von Genetik und genetischer Diagnostik. *Erkenntnisweg Biologiedidaktik*(5), 157–171.
- Schwanewedel, J., Hößle, C. & Kattmann, U. (2008). Vorstellungen zur Rolle der Gene in Bezug auf die Gesundheit und Krankheit des Menschen. In U. Harms & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik*. (S. 179–195). Studienverlag. [https://www.researchgate.net/publication/266969083\\_Vorstellungen\\_zur\\_Rolle\\_der\\_Gene\\_in\\_Bezug\\_auf\\_die\\_Gesundheit\\_und\\_Krankheit\\_des\\_Menschen](https://www.researchgate.net/publication/266969083_Vorstellungen_zur_Rolle_der_Gene_in_Bezug_auf_die_Gesundheit_und_Krankheit_des_Menschen)
- Sedlmeier, P. & Renkewitz, F. (2013). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (2., aktualisierte Aufl.). *Always learning*. Pearson.

- Seiler, L., Kuhnel, M., Ifenthaler, D. & Honal, A. (2019). Digital Applications as Smart Solutions for Learning and Teaching at Higher Education Institutions. *Utilizing Learning Analytics to Support Study Success*. <https://www.semanticscholar.org/paper/Digital-Applications-as-Smart-Solutions-for-and-at-Seiler-Kuhnel/a5ad55742865c4bd950f928781f2a8e2ac542d0e>
- Semmelink, D. S. & Maree, D. J. (2023). A Rasch analysis of the High Potential Trait Indicator: A South African sample. *African Journal of Psychological Assessment*, 5. <https://doi.org/10.4102/ajopa.v5i0.115>
- Shamoo, A. E., Resnik, D. B. & Masters, B. R. (2007). Responsible Conduct of Research. *Journal of Biomedical Optics*, 12(3), 39901. <https://doi.org/10.1117/1.2749726>
- Shields, J. L. (1978). *An Empirical Investigation of the Effect of Heteroscedasticity and Heterogeneity of Variance on the Analysis of Covariance and the Johnson-Neyman Technique* [U. S. Army Research Institute for the Behavioral and Social Sciences]. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a058205.pdf>
- Spitzer, M. W. H. & Musslick, S. (2021). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in wake of the COVID-19 pandemic. *PLOS ONE*, 16(8), e0255629. <https://doi.org/10.1371/journal.pone.0255629>
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004a). *Lehrplan G8 Gymnasium. Fachprofile Biologie*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26386.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26386.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004b). *Lehrplan G8 Gymnasium. Fachlehrplan Biologie 11*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26196.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26196.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004c). *Lehrplan G8 Gymnasium. Fachlehrplan Biologie 9*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26259.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26259.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004d). *Lehrplan G8 Gymnasium. Fachlehrplan Englisch 11*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26513.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26513.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004e). *Lehrplan G8 Gymnasium. Fachlehrplan Katholische Religionslehre 10*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26208.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26208.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2004f). *Lehrplan G8 Gymnasium. Fachlehrplan Katholische Religionslehre 11*. [https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26175.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26175.html)

- Staatsinstitut für Schulqualität und Bildungsforschung. (2004g). *Lehrplan G8 Gymnasium. Fachprofil Biologie*.  
[https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id\\_26386.html](https://www.gym8-lehrplan.bayern.de/contentserv/3.1.neu/g8.de/id_26386.html)
- Staatsinstitut für Schulqualität und Bildungsforschung. (2023a). *LehrplanPLUS G9 Gymnasium. Fachprofil Biologie*.  
<https://www.lehrplanplus.bayern.de/fachprofil/gymnasium/biologie/10>
- Staatsinstitut für Schulqualität und Bildungsforschung. (2023b). *LehrplanPLUS FOS. Fachlehrplan Biologie 11*.  
<https://www.lehrplanplus.bayern.de/fachlehrplan/fos/11/biologie>
- Staatsinstitut für Schulqualität und Bildungsforschung. (2023c). *LehrplanPLUS Realschule. Fachlehrplan Biologie 10*.  
<https://www.lehrplanplus.bayern.de/fachlehrplan/realschule/10/biologie#65105>
- Stillfried, S. von & Boor, P. (2021). Nachweismethoden von SARS-CoV-2 in Gewebe [Methods of SARS-CoV-2 detection in tissue]. *Der Pathologe*, 42(2), 208–215.  
<https://doi.org/10.1007/s00292-021-00919-8>
- Stratford, F. L. & Bear, C. E. (2006). Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. In A. Bush, Alton Eric W. F. W., Davies, Jane C., Griesenbach, Uta & A. Jaffe (Hrsg.), *Progress in respiratory research: Bd. 34. Cystic fibrosis in the 21st century* (S. 29–37). Karger.
- Strobach, T. & Wendt, M. (2019). Gedächtnis. In Strobach (Hrsg.), *Was ist eigentlich ...? Allgemeine Psychologie* (S. 33–39). Springer. [https://doi.org/10.1007/978-3-662-57570-3\\_5](https://doi.org/10.1007/978-3-662-57570-3_5)
- Strobl, C. (2015). *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis* (3. erweiterte Auflage). *Sozialwissenschaftliche Forschungsmethoden: Bd. 2*. Reiner Hampp Verlag.
- Sweller, J. (2016). Cognitive Load Theory: What We Learn and How We Learn. In Spector & James (Hrsg.), *Learning, Design, and Technology* (S. 1–17) [publisher not identified].  
[https://doi.org/10.1007/978-3-319-17727-4\\_50-1](https://doi.org/10.1007/978-3-319-17727-4_50-1)
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261–292.  
<https://doi.org/10.1007/s10648-019-09465-5>
- Tabachnick, B. G. & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed., International ed.). *Always learning*. Pearson.  
<https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5832234>
- Tanhua-Piironen, E., Pystynen, J. & Raisamo, R. (2010). Haptic applications as physics teaching tools. *2010 IEEE International Symposium on Haptic Audio Visual Environments and Games*. <https://www.semanticscholar.org/paper/Haptic->

- applications-as-physics-teaching-tools-Tanhua-Piironen-Pystynen/19bdc83d09240d47d0b289456e50e4e07adcf08
- Thisgaard, M. & Makransky, G. (2017). Virtual Learning Simulations in High School: Effects on Cognitive and Non-cognitive Outcomes and Implications on the Development of STEM Academic and Career Choice. *Frontiers in psychology*, 8, 805. <https://doi.org/10.3389/fpsyg.2017.00805>
- Trendel, G. & Lübeck, M. (2018). *Die Entwicklung experimenteller Kompetenzen*. Waxmann Verlag GmbH. <https://www.schulentwicklung.nrw.de/sinus/upload/phase05/nw14-00/n14-00.pdf>
- Ulbrich, H. (2021). Prozessmerkmale guten Unterrichts mit Flipped Classroom. *R&E-SOURCE*. Vorab-Onlinepublikation. <https://doi.org/10.53349/resource.2021.i16.a1003>
- UniProt. (2007). P13569 - CFTR\_HUMAN. <https://www.uniprot.org/uniprotkb/P13569/entry#sequences>
- van der Velde, M., Sense, F., Spijkers, R., Meeter, M. & van Rijn, H. (2021). *Lockdown Learning: Changes in Online Study Activity and Performance of Dutch Secondary School Students during the COVID-19 Pandemic*. Center for Open Science. <http://dx.doi.org/10.31234/osf.io/fr2v8>
- van Gog, T. (2022). The Signaling (or Cueing) Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 221–230). Cambridge University Press. <https://doi.org/10.1017/9781108894333.022>
- van Joolingen, W. R. & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, 25(5), 307–346. <https://doi.org/10.1023/A:1002993406499>
- Wagner, R. & Beckmann, B. M. (2022). Protein-Nucleinsäure-Wechselwirkungen. In J. Kurreck, J. W. Engels & F. Lottspeich (Hrsg.), *Bioanalytik* (4. Aufl. 2022, S. 902–925). Springer Berlin Heidelberg.
- Wang, A. Y. & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology Web-based classes. *Journal of Educational Psychology*, 92(1), 137–143. <https://doi.org/10.1037/0022-0663.92.1.137>
- Wang, J., Lin, M., Crenshaw, A., Hutchinson, A., Hicks, B., Yeager, M., Berndt, S., Huang, W.-Y., Hayes, R. B., Chanock, S. J., Jones, R. C. & Ramakrishnan, R. (2009). High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. *BMC genomics*, 10, 561. <https://doi.org/10.1186/1471-2164-10-561>
- Ward, C. L., Omura, S. & Kopito, R. R. (1995). Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell*, 83(1), 121–127. [https://doi.org/10.1016/0092-8674\(95\)90240-6](https://doi.org/10.1016/0092-8674(95)90240-6)

- Waschull, S. B. (2001). The Online Delivery of Psychology Courses: Attrition, Performance, and Evaluation. *Teaching of Psychology*, 28(2), 143–147. [https://doi.org/10.1207/S15328023TOP2802\\_15](https://doi.org/10.1207/S15328023TOP2802_15)
- Wellnitz, N., Hecht, M., Heitmann, P., Kauertz, A., Mayer, J [Jürgen], Sumfleth, E. & Walpuski, M. (2017). Modellierung des Kompetenzteilbereichs naturwissenschaftliche Untersuchungen. *Zeitschrift für Erziehungswissenschaft*, 20(4), 556–584. <https://doi.org/10.1007/s11618-016-0721-3>
- Wellnitz, N. & Mayer, J [Jürgen] (2013). Erkenntnismethoden in der Biologie. Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 315–345. <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?Fld=1040640#verfuegbarkeit>
- Wellnitz, N. & Mayer, J [Jürgen]. (2016). Methoden der Erkenntnisgewinnung im Biologieunterricht. In A. Sandmann & P. Schmiemann (Hrsg.), *Biologie lernen und lehren: Band 1. Biologiedidaktische Forschung: Schwerpunkte und Forschungsstände*. Logos Verlag.
- Westermeier, R. & Görg, A. (2022). Elektrophoretische Verfahren. In J. Kurreck, J. W. Engels & F. Lottspeich (Hrsg.), *Bioanalytik* (4. Auflage, S. 265–298). Springer Spektrum. [https://doi.org/10.1007/978-3-662-61707-6\\_12](https://doi.org/10.1007/978-3-662-61707-6_12)
- Wickham, H., Vaughan, D. & Girlich, M. (2023). *tidyr: Tidy Messy Data [Computer software]*. <https://tidyr.tidyverse.org>, <https://github.com/tidyverse/tidyr>
- Wiggins, B. L., Eddy, S. L., Grunspan, D. Z. & Crowe, A. J. (2017). The ICAP Active Learning Framework Predicts the Learning Gains Observed in Intensely Active Classroom Experiences. *AERA Open*, 3(2), 233285841770856. <https://doi.org/10.1177/2332858417708567>
- Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4rd ed.). *Statistical modeling and decision science*. Elsevier Science & Technology Books. <http://www.sciencedirect.com/science/book/9780123869838>
- Willingham, D. (2002). Ask the Cognitive Scientist. Inflexible Knowledge: The First Step to Expertise. *American Educator*, 26.
- Wilson, S. M. (1989). A Case Concerning Content: Using Case Studies To Teach Subject Matter. Craft Paper 89-1. <https://www.semanticscholar.org/paper/A-Case-Concerning-Content%3A-Using-Case-Studies-To-Wilson/c24eb661e2ff32c8f03e4ee0b2bcb5a9c3820bce>
- Wolf, C. & Best, H. (Hrsg.). (2010). *Handbuch der sozialwissenschaftlichen Datenanalyse*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92038-2>
- Wollschläger. (2017). *Grundlagen der Datenanalyse mit R*. Springer. <https://doi.org/10.1007/978-3-662-53670-4>

- Wößmann, L. (2021). Bildungsverluste durch Corona: Wie lassen sie sich aufholen? [Not Available]. *Wirtschaftsdienst (Hamburg, Germany : 1949)*, 101(3), 150–151. <https://doi.org/10.1007/s10273-021-2859-8>
- Wright, B. D. & Linacre, J. M. (1994). *Rasch Measurement Transactions: Reasonable mean-square fit values*. 8(3), 370. <https://www.rasch.org/rmt/rmt83b.htm>
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis: [Rasch measurement]*. MESA Press.
- Wright, B. D. & Stone, M. (1979). *Best test design*. MESA Press: Chicago, IL. <https://research.acer.edu.au/measurement/1/>
- Wu, M. & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions, Melbourne.
- Zhang, S., Koning, B. de, Agostinho, S., Tindall-Ford, S., Chandler, P. & Paas, F. (2022). The Cognitive Load Self-Management Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (3rd ed., S. 430–436). Cambridge University Press. <https://doi.org/10.1017/9781108894333.044>
- Zhang, Z., Liu, F. & Chen, J. (2018). Molecular structure of the ATP-bound, phosphorylated human CFTR. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50), 12757–12762. <https://doi.org/10.1073/pnas.1815287115>
- Zhonggen, Y. (2015). Blended Learning Over Two Decades | International Journal of Information and Communication Technology Education. *International Journal of Information and Communication Technology Education*, 11(3), 1–19.
- Zielenski, J. (2000). Genotype and phenotype in cystic fibrosis. *Respiration; international review of thoracic diseases*, 67(2), 117–133. <https://doi.org/10.1159/000029497>
- Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 41(2), 64–70. [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)
- Zumbach, J. (2010). *Lernen mit neuen Medien: Instruktionspsychologische Grundlagen* (1. Auflage). Kohlhammer Verlag. <http://nbn-resolving.org/urn:nbn:de:bsz:24-epflicht-1296533>

## **11 Anhang**

### **11.1 Erhebungsinstrumente**

#### **11.1.1 Fragebogen Fachwissen**

Bitte wählen Sie bei nachfolgender Frage 2 aus 5 Antworten aus.

1 Die DNA ...

---

gibt es nur bei Eukaryoten

---

liegt im menschlichen Körper zyklisch vor  
**liegt bei Eukaryoten im Zellkern vor.**  
ist an der Zellmembran assoziiert  
**ist im menschlichen Körper als Chromosomen gelagert**

2 Bestandteile der DNA sind ...

---

Vesikel

**Zucker**

**Phosphatgruppen**

Proteine

Enzyme

3 Ein Gen ...

---

**ist eine transkriptionelle Einheit**

ist ein kleines DNA-Stück, welches sich im Zellkern frei bewegen kann

**beschreibt eine kodierende DNA-Sequenz**

ist die entspiralisierte Form eines Chromosoms

beschreibt Abschnitte auf der tRNA

4 Die DNA-Polymerase

---

**synthetisiert den Komplementärstrang**

ist Bindestelle für Primer

**katalysiert die Polymerisation von Nukleotiden**

spaltet die Doppelhelix in zwei komplementäre Stränge auf

ist ein Primer

5 Die Assoziierung zweier DNA-Stränge erfolgt durch

---

kovalente Bindungen

**Wasserstoffbrücken**

van-der-Waals Kräfte

Ionenbindung

**elektrostatische Anziehung**

6 An den Enden eines DNA-Stranges befinden sich ...

---

**Hydroxygruppen**

Carboxygruppen

**Phosphatgruppen**

Sulfatgruppen

Carbonylgruppen

7 Markieren Sie die richtigen Basenpaarungen in der DNA ...

---

Adenin bildet 3 Wasserstoffbrücken mit Guanin

Guanin bildet 2 Wasserstoffbrücken mit Cytosin

**Adenin bildet 2 Wasserstoffbrücken mit Thymin**

Thymin bildet 2 Wasserstoffbrücken mit Guanin

**Cytosin bildet 3 Wasserstoffbrücken mit Guanin**

8 Bei der Replikation ...

---

**entsteht eine Kopie der DNA**

**muss das Enzym Helikase beteiligt sein**

wird das Genom absichtlich verändert

wird nur ein Strang kopiert

wird Thymin durch Uracil ersetzt

9 Ein SNP ...

---

löst immer eine Krankheit aus.

**steht für single nucleotide polymorphism**

**ist die häufigste Art genetischer Variation beim Menschen**

steht für short nuclear palindrome

tritt nie im kodierenden Bereich der DNA auf

10 Ein rezessives Allel ...

---

setzt sich bei der Vererbung immer durch.

**kommt nur bei homozygoten Merkmalsträgern zur phänotypischen Ausprägung.**

kommt nur bei heterozygoten Merkmalsträgern zur phänotypischen Ausprägung

beruht häufig auf einer Mutation und einer damit verbundenen Funktionseinschränkung

**kommt seltener zur Ausprägung als ein dominantes Allel.**

11 Die PCR ...

---

steht als Abkürzung für Polymerase-Cytosin-Reaktion

kann in vitro ohne Primer erfolgen

**verläuft in der Reihenfolge Denaturierung, Hybridisierung und Elongation**

**ist eine Methode zur Vervielfältigung einer bestimmten DNA-Sequenz**

erfolgt in der Reihenfolge Denaturierung, Elongation und Annealing

12 Bei der PCR ...

---

**wird das Enzym Taq-Polymerase zur Synthese des komplementären DNA-Strangs genutzt.**

kommen Temperaturen von über 100 °C zum Einsatz.

**wird theoretisch nur ein ursprünglicher DNA-Doppelstrang benötigt.**

wird die Primase zur Kennzeichnung der Synthesepunkte genutzt.

werden Primer, Nukleasen, Polymerase und DNA benötigt.

13 Die Agarose-Gelelektrophorese ...

---

**ist ein Verfahren zur Bestimmung der Länge von DNA-Fragmenten.**

**ist ein Verfahren zur Aufreinigung von DNA Proben.**

ist ein immunchemisches Verfahren.

wird vor der PCR durchgeführt.

wird mit Wechselstrom durchgeführt.

14 Bei der Gelelektrophorese ....

---

ist die DNA negativ geladen.

**übernimmt das Gel die Funktion eines Siebs.**

**nutzt man langkettige Zucker um das Gel herzustellen.**

können Proteine und DNA gleichzeitig aufgetrennt werden.

ist die DNA im Gel ohne Hilfsmittel sichtbar.

## 11.1.2 Fragebogen naturwissenschaftliche Arbeitsweise

### **NWA-Test Allgemein**

Bitte wählen Sie bei nachfolgender Frage 2 aus 5 Antworten aus.

15 Universalindikator kommt in flüssiger oder fester Form vor. Ein gängiges Verfahren ist die Nutzung von pH-Papier (s. Abbildung) um den pH-Wert von Stoffen zu bestimmen. Wählen Sie die beiden richtigen Aussagen aus.

[<https://www.shutterstock.com/de/image-photo/litmus-ph-test-paperisolated-on-white-174629882>]

Legt man pH-Papier auf Natriumhydroxid-Plättchen (NaOH(s)), so verfärbt sich das Papier blau.

**Tröpfelt man Zitronensaft auf pH-Papier, so verfärbt sich das Papier rot.**

**Legt man pH-Papier auf die Zunge des Menschen, so verfärbt sich das Papier grün.**

Schüttet man Natronlauge zu Salzsäure, so wird die Mischung am Ende das pH-Papier immer grün färben.

Taucht man pH-Papier in Cola, so verfärbt sich das Papier blau.



16 Ein wichtiges Gerät für chemische Versuche ist der Gasbrenner. Geben Sie an, welche beiden Aussagen zur Nutzung des Gasbrenners richtig sind.

Leuchtende oder rauschende Flamme kann über die Gaszufuhr reguliert werden.

**Beim Erhitzen von Flüssigkeiten in einem Reagenzglas, dieses leicht schräg halten und schwänken.**

Arbeiten mit dem Gasbrenner sind bevorzugt im Sitzen durchzuführen.

**Beim „Anschalten“ des Gasbrenners zuerst die Gaszufuhr öffnen, dann das Gas entzünden und anschließend die Luftzufuhr öffnen.**

Beim „Ausschalten“ des Gasbrenners zuerst die Gaszufuhr und dann die Luftzufuhr schließen. Restflamme auspusten.

17 Beim Mikroskopieren sind einige Aspekte zu beachten. Wählen Sie die beiden richtigen Aussagen aus.

Deckgläschen waagrecht auf die Probe mit Wassertropfen legen um Blasenbildung im Wasser zu vermeiden.

**Schnitte von Präparaten möglichst dünn anfertigen um sie somit durchsichtig zu machen.**

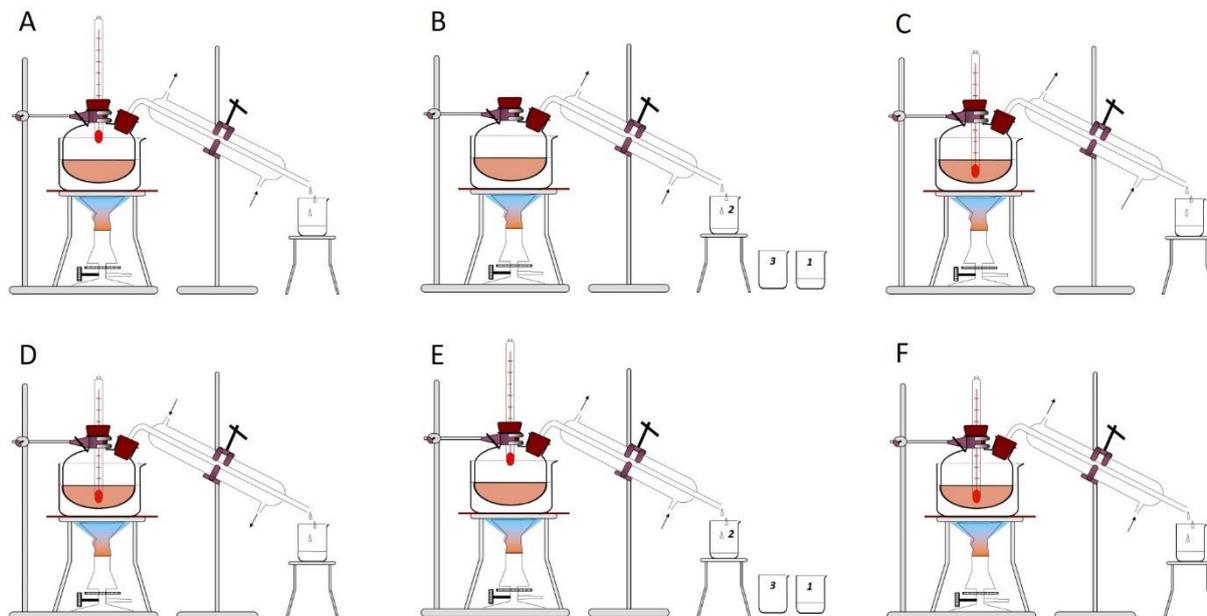
Es ist nicht möglich lebende Organismen unter dem Mikroskop zu betrachten.

**Man beginnt immer mit der geringsten Vergrößerung zu mikroskopieren.**

Beim Mikroskopieren fällt das Licht von oben auf das zu untersuchende Präparat.

18 Nachfolgende Abbildungen zeigen unterschiedliche Aufbauten zu Destillationsapparaturen. Wählen Sie die Aussagen über die Aufbauten aus, die zu einem sinnvollen Ergebnis einer Destillation führen.

[Abbildungen erstellt mit den PowerPoint-Präsentation Versuchsskizzen (Landesbildungsserver Baden-Württemberg)]



- A, nicht B und F  
 E, nicht C und D  
 B, nicht D und E  
 E, nicht A und F  
 D, nicht B und A

19 Chromatographie ist ein wichtiges Trennverfahren was in vielen Bereichen der Analyse seinen Einsatz findet. Versuche mit Farbstoffen lassen sich einfach durchführen. Markieren Sie die beiden Antworten, die sich auch nebenstehender Abbildung bestätigen lassen.

**Der blaue Farbstoff ist besser wasserlöslich als der gelbe, beim grünen (dritter von links) Filzstift.**

Je weiter ein Farbstoff auf dem Papier läuft, desto weniger hydrophil ist er.

Der rote Filzstift (links) besteht nur aus einer Sorte Farbstoff.

**Der orange Farbstoff bindet stärker an die stationäre als an die mobile Phase, als der blaue Farbstoff.**

Die Zeit der Chromatographie ist bei den rechten drei Stiften länger, als die des ersten Stiften (links).

Abbildung aufgrund von  
 Bildnutzungsrechten  
 hier nicht verfügbar.

Abb. Papierchromatographie;  
 Laufmittel Butanol- und Wasser-  
 Gemisch

20 Die Elektrolyse einer Zink-Iodid-Lösung ( $ZnI_2(aq)$ ) sollte Ergebnis A liefern. Bei Ihrem Versuch erhalten Sie allerdings nach einiger Zeit das Ergebnis B (Kabel u.ä. sind nicht dargestellt). Wählen Sie die beiden richtigen angeführten möglichen Erklärungen für dieses fehlerhafte Ergebnis.

[Abbildungen erstellt mit den PowerPoint-Präsentation Versuchsskizzen (Landesbildungsserver Baden-Württemberg)]



Der DNA-Längenstandard (DNA-Leiter) passt nicht zur DNA-Fragmentlänge.

**Der Fluoreszenzfarbstoff (Stain) wurde bei der Gelherstellung vergessen.**

**Die Pole des Stromkreises sind falsch angeschlossen.**

Der Auftragspuffer wurde bei den Proben vergessen.

Es wurden zu große Mengen (Volumina) der Proben in die Taschen aufgetragen.

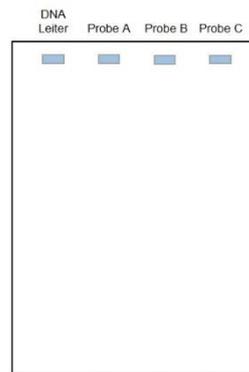


Abb.: Agarose-Gel mit Proben und DNA-Leiter **vor** der Agarose-Gelelektrophorese bei Tageslicht betrachtet.

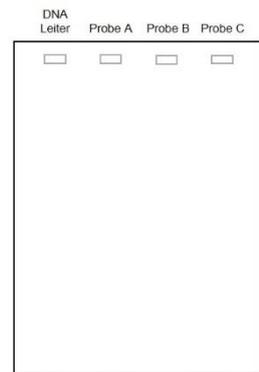


Abb.: Agarose-Gel mit Proben und DNA-Leiter **nach** der Agarose-Gelelektrophorese bei Blau-Licht betrachtet.

23 In der Abbildung sehen Sie das Schema eines Gels nach der Gelelektrophorese. Nachfolgend sind fünf Aussagen über das Gel getätigt. Markieren Sie die beiden richtigen Aussagen die sich aus dem Gel-Bild ableiten lassen.

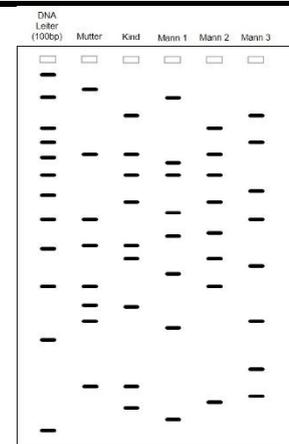
Bei ca. 500 bp ist eine Bande bei allen biologischen Familienmitgliedern zu erkennen.

Mann 2 ist nicht der Vater des Kindes.

**Mann 1 ist nicht der Vater des Kindes.**

**Mutter und Kind stimmen in vier Banden überein.**

Der DNA-Leiter ist eine Mischung aus den DNA-Fragmenten der Väter.



24 Die Gene des Menschen entscheiden darüber ob man den Bitterstoff PTC intensiv (homozygot), moderat (heterozygot) oder nicht schmecken (homozygot) kann. Wenn die Base an Position 785 des besagten Gens (ca. 250 Basenpaare) Cytosin ist, dann kann das Restriktionsenzym *Fnu4HI* dieses Gen an dieser Stelle schneiden. In diesem Falle würde man von „Schmecker“ sprechen.

Nachfolgend ist das Ergebnis nach dem Restriktionsverdau aller Proben (A –C) und der PCR auf einem Agarosegel dargestellt. Entscheiden Sie, welche beiden Aussagen sich aus der Auswertung des Gels treffen lassen.

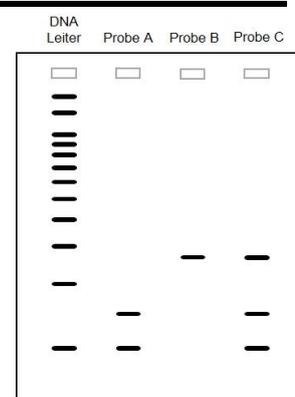
Der DNA-Längenstandard trägt DNA-Fragmente mit derselben Basenabfolge wie das untersuchte Schmecker-Gen.

Person A ist heterozygot bzgl. ihrer DNA für das Schmecker-Gen.

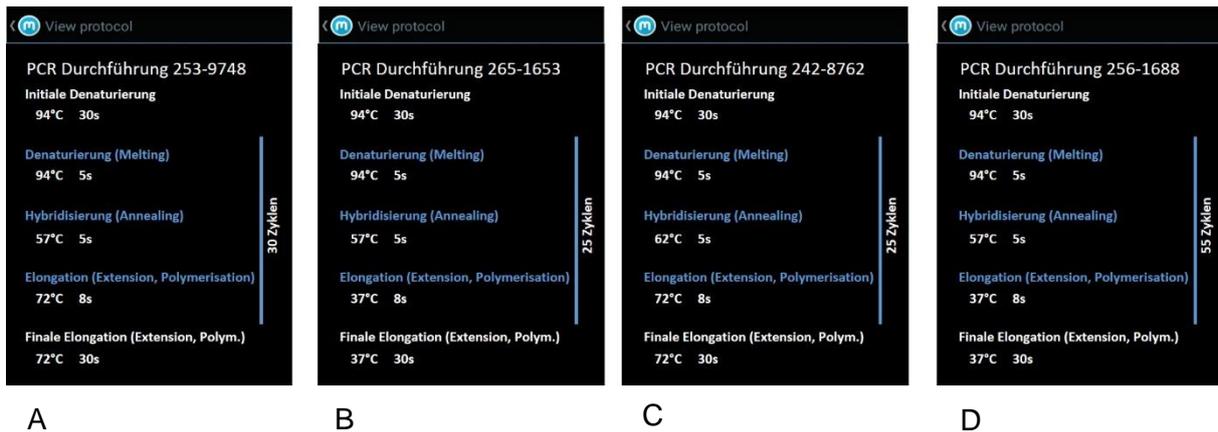
**Person B ist homozygot bzgl. des Schmecker-Gens.**

**Bei Person A zeigt sich, dass das Restriktionsenzym geschnitten hat und diese kann somit als „Schmecker“ bezeichnet werden.**

Person C kann besonders gut bittere Stoffe schmecken, da es drei Banden im Gel aufweist.



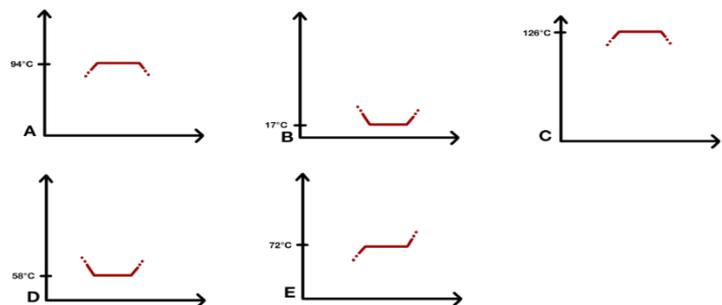
25 Nachfolgend finden Sie vier PCR Protokolle. Wählen Sie aus, welche Protokolle zu einem aussagekräftigen Ergebnis führen würden.



C und D, aber nicht A  
 A und B, aber nicht D  
**A und C, aber nicht B**  
 D und B, aber nicht C  
**C und A, aber nicht D**

26 In der nachfolgend dargestellten Abbildung sind fünf Temperatur-Zeit-Diagramme dargestellt. Entscheiden Sie, welche der nachfolgenden Diagramme Teile eines typischen PCR-Zeit-Temperaturdiagramms sein können.

B und C, aber nicht E.  
**A und D, aber nicht C.**  
 E und B, aber nicht A.  
 B und A, aber nicht C.  
**D und E, aber nicht B.**



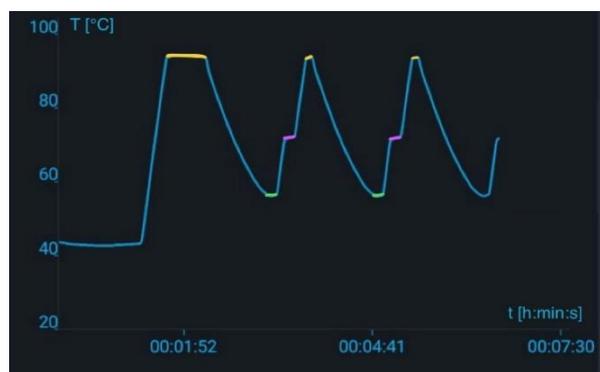
27 Im nachfolgenden Temperatur-Zeit-Diagramm sieht man die verschiedenen Schritte der PCR sehr deutlich. Markieren Sie die beiden richtigen Aussagen zu diesem Diagramm.

Bei der Denaturierung, hier gelb gekennzeichnet, werden die Wasserstoffbrücken der DNA-Doppelstränge gespalten und es entstehen DNA-Einzelstränge.  
 Im hier lila gekennzeichneten Bereich, lagern sich die Primer an den DNA-Einzelsträngen an.  
 Durch Erhöhung der Temperatur auf 94°C wird die optimale Temperatur für die Aktivität der Taq-Polymerase erreicht.  
**Die Neusynthese des komplementären DNA-Strangs ausgehend vom Primer durch die Taq-Polymerase ist der letzte Schritt eines Zyklus.**

Ein PCR-Zyklus hat immer eine Temperaturreihenfolge den Temperaturmarkierungen grün -> lila -> gelb im Diagramm entsprechend.

28 Im nachfolgenden Diagramm ist die Anzahl der DNA-Kopien in Abhängigkeit von der Zyklusanzahl beschrieben. Wählen Sie die beiden Antworten, deren Aussagen hierzu zutreffend sind.

[Abbildung übernommen und verändert nach miniPCR bio (2023)]



[Abbildung übernommen und verändert nach miniPCR bio (2023)]

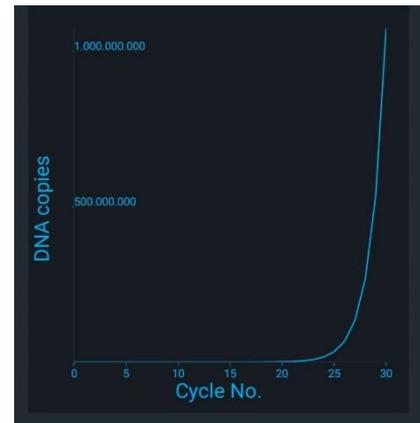
**Pro PCR-Zyklus verdoppelt sich die Anzahl an DNA-Doppelsträngen.**

Nach 30 Zyklen hat man über eine Billion DNA-Kopien.

**Mit der Formel  $2^n$  (zwei hoch n) kann man die Anzahl an DNA-Doppelsträngen ausrechnen die entstehen.**

Nach 15 Zyklen liegen bereits 500.000.000 DNA-Kopien vor.

Für ein Aussagekräftiges Ergebnis sind, wie man in der Grafik sehen kann, 20 Zyklen absolut ausreichend.



29 Für einen Versuch müssen verschiedene flüssige Substanzen pipettiert werden. Hierzu stehen Ihnen vier gängige Pipetten Varianten zur Verfügung. Wählen Sie die beiden Varianten die für das angegebene Volumen die sichersten und sinnvollsten sind.

(A : 0,5-20  $\mu$ l ; B : 2 -20  $\mu$ l ; C : 20 – 100  $\mu$ l ; D : 100 - 1000  $\mu$ l ) – aus Abbildung ablesen – Abb groß genug wählen.

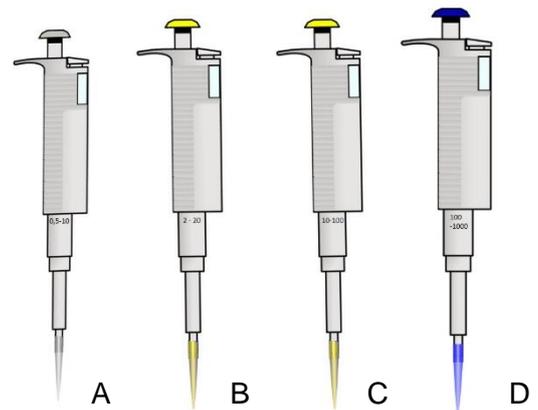
20  $\mu$ l mit Pipette B

**3  $\mu$ l mit Pipette A**

12  $\mu$ l mit Pipette A

**20  $\mu$ l mit Pipette C**

9  $\mu$ l mit Pipette C



30 Folgender Ansatz soll pipettiert werden. Entscheiden Sie welche beiden Reihenfolgen für einen Pipettiervorgang am sinnvollsten sind bei der praktischen Umsetzung.

**C, B, A, D**

**C, B, D, A**

D, A, C, B

D, A, B, C

A, D, C, B

**Zu pipettierende  
Substanzen für Ansatz 26:**

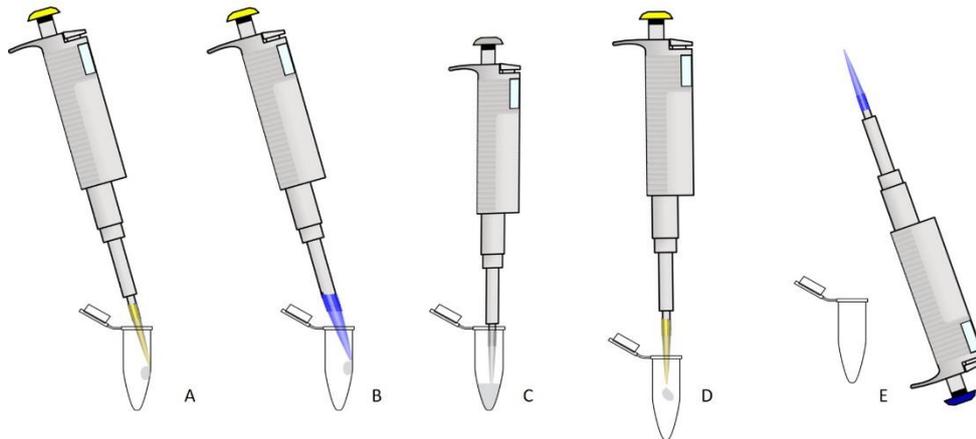
5  $\mu$ l Substanz A

10  $\mu$ l Substanz B

15  $\mu$ l Substanz C

5  $\mu$ l Substanz D

31 Nachfolgend sind 5 Abbildungen angeführt, welche verschiedene Pipettierszenarien darstellen. Entscheiden Sie welche der Szenarien ohne Fehler sind.



**A und C, aber nicht B**

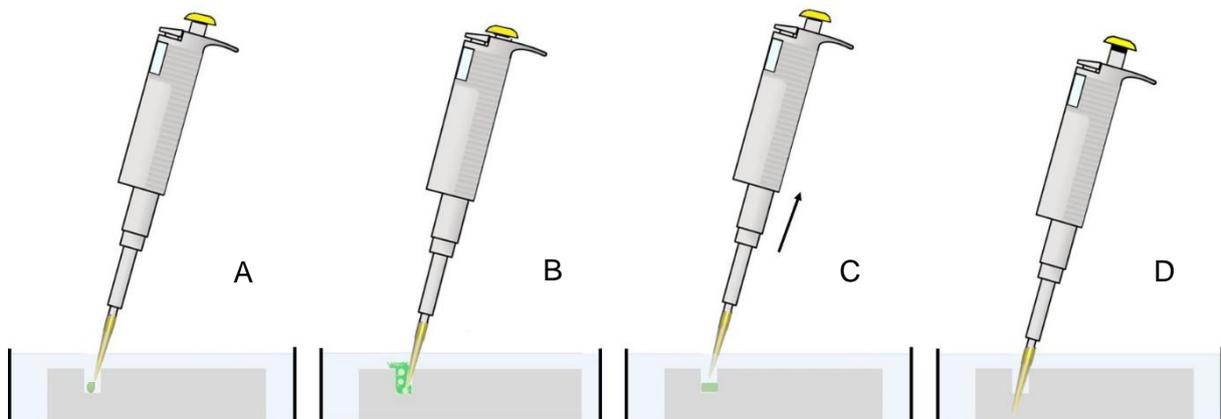
D und A, aber nicht E

**C und A, aber nicht D**

D und E, aber nicht B

E und D, aber nicht A

32 DNA-Proben müssen vor der Gelelektrophorese auf das Agarose-Gel (in der Gelkammer mit Pufferlösung überschichtet) aufgetragen werden. Damit keine Fehler passieren muss man einige Sachen beachten. Wählen Sie die beiden richtigen Antworten aus deren Aussagen richtiges Pipettierverhalten zeigen



**A, aber nicht B**

B, aber nicht D

**A, aber nicht D**

B, aber nicht D

C, aber nicht A

## 11.2 Ergebnisse: Skalendokumentation, Winsteps Datentabellen

### 11.2.1 Prä-Test Fachwissen

#### Anhang 1: Item-Statistik (pr\_fw)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .53 REL.: .22 ... ITEM: REAL SEP.: 6.01 REL.: .97

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	118	154	-1.98	.20	1.13	1.26	1.29	1.62	.16	.32	75.3	77.3	PR01_DI
2	90	154	-1.01	.18	.85	-2.52	.78	-2.54	.54	.38	75.3	67.0	PR02_DI
3	64	154	-.21	.18	1.05	.80	1.08	.89	.33	.39	65.6	68.0	PR03_DI
4	35	154	.82	.21	.89	-.97	.80	-1.24	.47	.35	79.2	78.4	PR04_DI
5	24	154	1.35	.24	1.01	.10	.96	-.10	.31	.31	84.4	84.4	PR05_DI
6	82	154	-.76	.18	.96	-.66	.91	-1.07	.44	.38	65.6	65.6	PR06_DI
7	105	154	-1.50	.19	.95	-.59	.90	-.83	.41	.35	73.4	71.2	PR07_DI
8	45	154	.43	.19	.96	-.43	.90	-.78	.42	.37	74.7	74.4	PR08_DI
9	92	154	-1.07	.18	1.02	.37	1.02	.26	.35	.37	68.8	67.4	PR10_DI
10	23	154	1.41	.24	1.13	.86	1.53	1.96	.12	.31	85.1	85.0	PR11_DI
11	9	154	2.53	.35	1.07	.34	1.36	.86	.08	.21	94.2	94.1	PR13_DI
MEAN	62.5	154.0	.00	.21	1.00	-.13	1.05	-.09			76.5	75.7	
P.SD	35.5	.0	1.35	.05	.09	1.01	.23	1.29			8.4	8.7	

#### Anhang 2: Item-Statistik (pr\_fw) (bereinigt)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .61 REL.: .27 ... ITEM: REAL SEP.: 6.08 REL.: .97

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	118	154	-2.02	.20	1.15	1.37	1.32	1.73	.17	.34	75.3	78.0	PR01_DI
2	90	154	-1.03	.18	.86	-2.38	.78	-2.46	.54	.39	75.3	67.4	PR02_DI
3	64	154	-.22	.18	1.07	.98	1.14	1.42	.32	.40	65.6	68.2	PR03_DI
4	35	154	.81	.21	.90	-.94	.79	-1.17	.46	.35	79.2	78.4	PR04_DI
5	24	154	1.34	.24	1.01	.13	.96	-.08	.31	.31	83.8	84.4	PR05_DI
6	82	154	-.78	.18	.97	-.54	.91	-1.03	.44	.40	65.6	65.9	PR06_DI
7	105	154	-1.53	.19	.96	-.52	.90	-.81	.42	.37	73.4	71.9	PR07_DI
8	45	154	.42	.19	.96	-.37	.90	-.72	.42	.38	74.7	74.6	PR08_DI
9	92	154	-1.10	.18	1.04	.61	1.05	.53	.35	.39	68.8	67.9	PR10_DI
10	20	151	1.58	.25	1.07	.46	1.09	.41	.21	.29	86.8	86.7	PR11_DI
11	9	154	2.53	.35	1.07	.35	1.37	.85	.09	.21	94.2	94.1	PR13_DI
MEAN	62.2	153.7	.00	.21	1.00	-.08	1.02	-.12			76.6	76.1	
P.SD	35.8	.9	1.38	.05	.08	.99	.19	1.19			8.5	8.8	

#### Anhang 3: Hauptkomponentenanalyse der Residuen (PCAR) (pr\_fw)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	16.5877	100.0%	100.0%
Raw variance explained by measures =	5.5877	33.7%	33.3%
Raw variance explained by persons =	1.4927	9.0%	8.9%
Raw Variance explained by items =	4.0950	24.7%	24.4%
Raw unexplained variance (total) =	11.0000	66.3%	100.0%
Unexplned variance in 1st contrast =	1.6565	10.0%	15.1%
Unexplned variance in 2nd contrast =	1.3798	8.3%	12.5%
Unexplned variance in 3rd contrast =	1.2654	7.6%	11.5%
Unexplned variance in 4th contrast =	1.2171	7.3%	11.1%
Unexplned variance in 5th contrast =	1.1004	6.6%	10.0%

## Anhang 4: Personenfähigkeiten (measure) (pr\_fw)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

TABLE OF MEASURES ON TEST OF 11 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-4.28E	1.88	4	-.83	.73	8	1.35	.80
1	-2.94	1.10	5	-.31	.72	9	2.05	.89
2	-2.03	.85	6	.22	.73	10	3.03	1.13
3	-1.38	.77	7	.76	.75	11	4.43E	1.90

CURRENT VALUES, UIMEAN=.0000 USCALE=1.0000

TO SET MEASURE RANGE AS 0-100, UIMEAN=49.1615 USCALE=11.4776

TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UIMEAN=5.4078 USCALE=1.2625

Predicting Score from Measure: Score = Measure \* 1.4233 + 5.4920

Predicting Measure from Score: Measure = Score \* .6889 + -3.7832

Maximum statistically different levels of performance (strata) = 2.3

Wright's Sample-independent Person (Test) Reliability based on maximum strata = .84

## Anhang 5: Step Ordering (pr\_fw)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	S.E. P.SD	INFT MEAN	OUTF MNSQ	PTMA MNSQ	CORR.	ITEM
1	0	0	36	23	-.92	.96	.16	1.3	1.4	-.17	PR01_DI
	1	1	118	77	-.53	.93	.09	1.1	1.1	.17	
2	0	0	64	42	-1.23	.75	.09	.8	.7	-.54	PR02_DI
	1	1	90	58	-.19	.83	.09	.9	.8	.54	
3	0	0	90	58	-.88	.87	.09	1.1	1.1	-.32	PR03_DI
	1	1	64	42	-.26	.93	.12	1.1	1.2	.32	
4	0	0	119	77	-.86	.86	.08	.9	.9	-.46	PR04_DI
	1	1	35	23	.19	.78	.13	.8	.8	.46	
5	0	0	130	84	-.75	.92	.08	1.0	1.0	-.31	PR05_DI
	1	1	24	16	.05	.83	.17	1.0	1.0	.31	
6	0	0	72	47	-1.07	.80	.09	.9	.9	-.44	PR06_DI
	1	1	82	53	-.23	.90	.10	1.0	1.0	.44	
7	0	0	49	32	-1.21	.80	.12	.9	.9	-.42	PR07_DI
	1	1	105	68	-.35	.89	.09	1.0	1.0	.42	
8	0	0	109	71	-.88	.88	.08	1.0	1.0	-.42	PR08_DI
	1	1	45	29	.00	.81	.12	.9	.9	.42	
9	0	0	62	40	-1.03	.86	.11	1.0	1.0	-.35	PR10_DI
	1	1	92	60	-.35	.91	.10	1.0	1.1	.35	
10	0	0	131	87	-.65	.89	.08	1.0	1.0	-.21	PR11_DI
	1	1	20	13	-.08	.84	.19	1.2	1.1	.21	
		MISSING ***		3	2#	-2.92	.00	.00			
11	0	0	145	94	-.65	.95	.08	1.0	1.0	-.09	PR13_DI
	1	1	9	6	-.30	.84	.30	1.5	1.4	.09	

# Missing % includes all categories. Scored % only of scored categories

## Anhang 6: Zusammenfassung Statistik (pr\_fw)

INPUT: 154 PERSON 11 ITEM REPORTED: 154 PERSON 11 ITEM 2 CATS WINSTEPS 5.2.4.0

## SUMMARY OF 154 MEASURED PERSON

	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			S.E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.4	11.0	-.62	.76	.99	-.01	1.02	.12
SEM	.1	.0	.08	.01	.03	.08	.06	.07
P.SD	1.7	.1	.95	.07	.33	.96	.73	.89
S.SD	1.7	.1	.95	.07	.33	.97	.73	.89
MAX.	8.0	11.0	1.35	1.10	1.95	2.33	5.38	2.88
MIN.	1.0	10.0	-2.93	.72	.45	-1.97	.30	-1.44
REAL RMSE	.81	TRUE SD	.49	SEPARATION	.61	PERSON RELIABILITY	.27	
MODEL RMSE	.76	TRUE SD	.56	SEPARATION	.74	PERSON RELIABILITY	.35	
S.E. OF PERSON MEAN = .08								

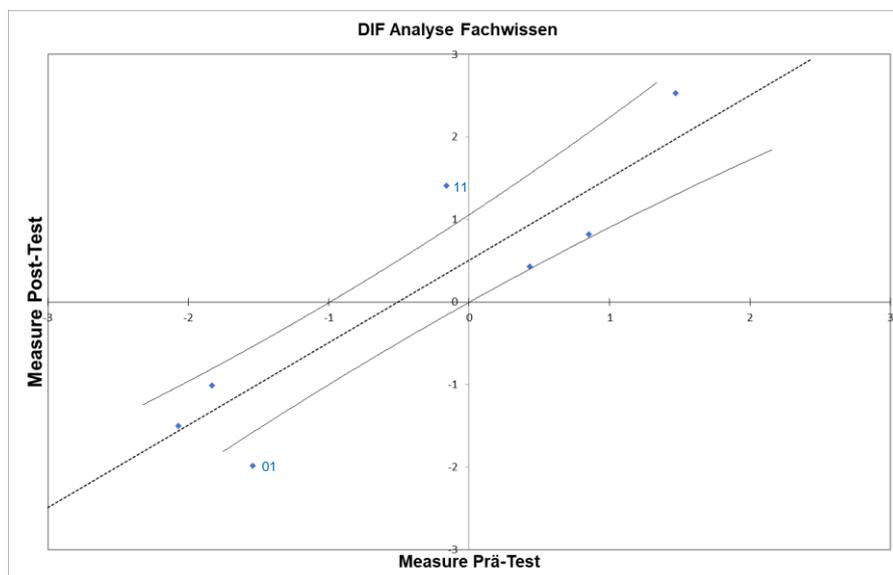
PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00 (approximate due to missing data)  
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .30 SEM = 1.41  
 (approximate due to missing data)  
 STANDARDIZED (50 ITEM) RELIABILITY = .71

## SUMMARY OF 11 MEASURED ITEM

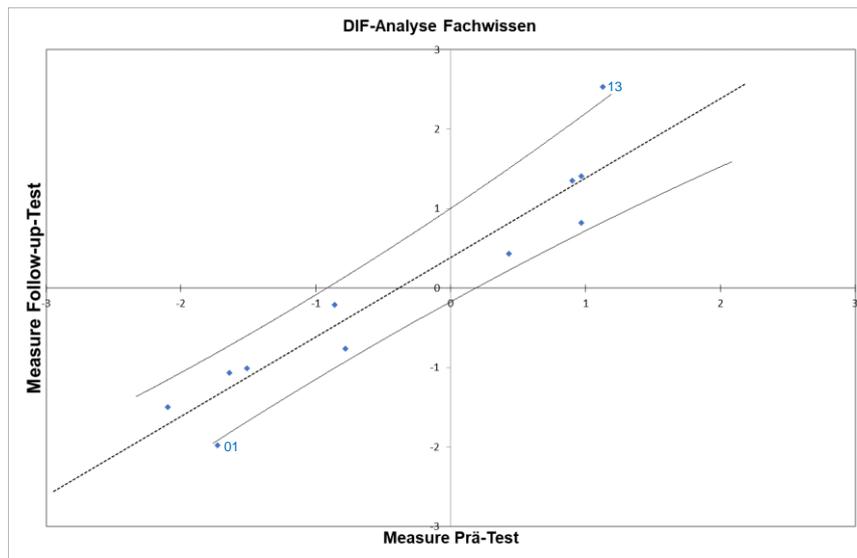
	TOTAL	COUNT	MEASURE	MODEL	INFIT		OUTFIT	
	SCORE			S.E.	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	62.2	153.7	.00	.21	1.00	-.08	1.02	-.12
SEM	11.3	.3	.44	.02	.03	.31	.06	.38
P.SD	35.8	.9	1.38	.05	.08	.99	.19	1.19
S.SD	37.6	.9	1.45	.05	.09	1.04	.20	1.25
MAX.	118.0	154.0	2.53	.35	1.15	1.37	1.37	1.73
MIN.	9.0	151.0	-2.02	.18	.86	-2.38	.78	-2.46
REAL RMSE	.22	TRUE SD	1.36	SEPARATION	6.08	ITEM RELIABILITY	.97	
MODEL RMSE	.22	TRUE SD	1.36	SEPARATION	6.22	ITEM RELIABILITY	.97	
S.E. OF ITEM MEAN = .44								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.99 (approximate due to missing data)  
 Global statistics; please see Table 44.  
 UMEAN=.0000 USCALE=1.0000

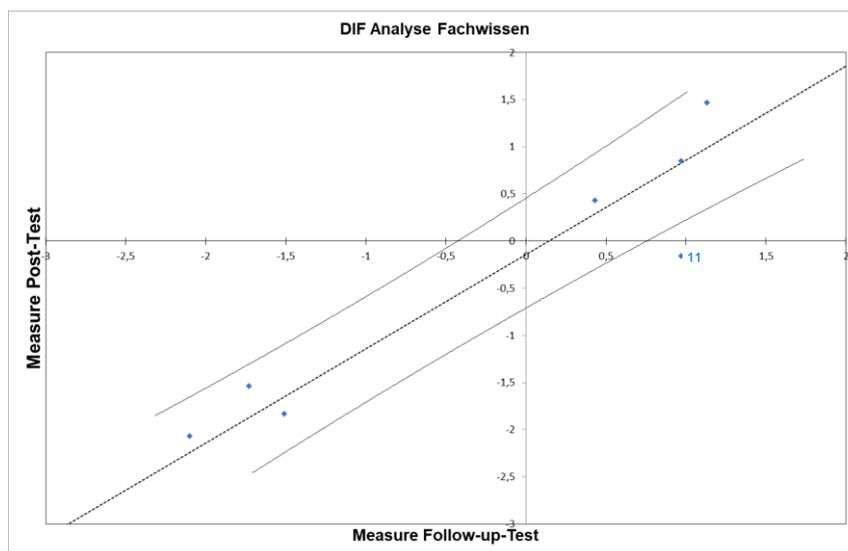
## Anhang 7: DIF-Analyse Prä-Test-Post-Test (fw)



## Anhang 8: DIF-Analyse Prä-Test-Follow-up-Test (fw)



## Anhang 9: DIF-Analyse Post-Test-Follow-up-Test (fw)



## 11.2.2 Post-Test Fachwissen

### Anhang 10: Item-Statistik (po\_fw)

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .66 REL.: .30 ... ITEM: REAL SEP.: 6.52 REL.: .98

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
1	110	154	-1.55	.20	1.14	1.40	1.33	1.89	.31	.44	74.5	76.7	P001_DI
2	117	154	-1.85	.21	.82	-1.62	.73	-1.45	.57	.43	81.0	79.1	P002_DI
3	41	154	.84	.20	.84	-1.81	.70	-1.82	.51	.37	79.7	75.7	P004_DI
4	122	154	-2.08	.22	.97	-.21	.87	-.55	.45	.42	83.0	81.8	P007_DI
5	51	154	.46	.19	.99	-.12	.96	-.21	.41	.40	70.6	71.3	P008_DI
6	70	154	-.17	.18	1.02	.34	1.08	.79	.41	.43	66.7	68.2	P011_DI
7	50	154	.50	.19	1.08	.98	1.17	1.11	.33	.40	67.3	71.8	P012_DI
8	27	154	1.46	.23	.96	-.23	1.05	.28	.34	.33	84.3	83.1	P013_DI
9	13	154	2.39	.30	1.13	.64	1.49	1.13	.11	.25	90.8	91.6	P014_DI
MEAN	66.8	154.0	.00	.21	1.00	-.07	1.04	.13			77.6	77.7	
P.SD	38.2	.0	1.46	.03	.11	1.02	.25	1.17			7.8	6.8	

### Anhang 11: Item-Statistik geankert (po\_fw)

TABLE 14.1 Kurz\_154\_final.sav ZOU103WS.TXT Jun 26 2023 8:44

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .64 REL.: .29 ... ITEM: REAL SEP.: 6.40 REL.: .98

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		DISPLACE	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
1	110	154	-2.02A	.25	1.85	4.56	2.98	4.45	.31	.42	73.9	85.9	.95	P001_DI
2	117	154	-1.03A	.20	.75	-2.60	.67	-2.13	.57	.47	85.6	77.3	-.35	P002_DI
3	41	154	.81A	.19	.78	-3.14	.67	-2.60	.51	.43	77.8	70.7	.57	P004_DI
4	122	154	-1.53A	.22	.96	-.25	.86	-.56	.45	.45	83.0	81.2	-.09	P007_DI
5	51	154	.42A	.18	1.00	.07	.97	-.18	.41	.45	71.2	69.5	.58	P008_DI
6	70	154	1.58A	.21	1.67	5.46	2.53	5.00	.41	.38	64.7	78.1	-1.16	P011_DI
7	50	154	1.04	.19	1.10	1.29	1.24	1.44	.33	.42	67.3	72.2	.00	P012_DI
8	27	154	2.53A	.26	1.35	1.84	1.69	1.66	.34	.30	83.0	88.2	-.49	P013_DI
9	13	154	2.96	.30	1.16	.73	1.63	1.28	.11	.26	90.8	91.6	.00	P014_DI
MEAN	66.8	154.0	.53	.22	1.18	.89	1.47	.93			77.5	79.4	.00	
P.SD	38.2	.0	1.65	.04	.36	2.70	.78	2.48			8.3	7.5	.00	

### Anhang 12: Item-Statistik (po\_fw) (bereinigt)

TABLE 14.1 Kurz\_154\_final.sav ZOU232WS.TXT Jun 26 2023 8:43

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .98 REL.: .49 ... ITEM: REAL SEP.: 6.86 REL.: .98

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		DISPLACE	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
1	110	144	-2.02A	.26	1.28	1.68	1.31	.97	.51	.53	82.0	85.6	.41	P001_DI
2	116	153	-1.03A	.21	.78	-2.11	.69	-1.62	.60	.53	84.5	79.0	-.42	P002_DI
3	41	154	.81A	.19	.83	-2.29	.72	-1.96	.50	.47	76.5	71.4	.68	P004_DI
4	122	154	-1.53A	.23	1.06	.51	.91	-.25	.47	.53	81.9	82.6	-.23	P007_DI
5	51	154	.42A	.19	1.08	1.08	1.12	.92	.41	.49	69.8	70.6	.67	P008_DI
6	65	149	1.58A	.21	1.42	3.73	1.40	1.70	.50	.41	69.2	77.0	-.96	P011_DI
7	48	152	1.21	.20	1.13	1.44	1.04	.29	.38	.44	67.8	73.9	.00	P012_DI
8	25	152	2.53A	.26	1.07	.45	.89	-.18	.42	.34	86.5	86.8	-.25	P013_DI
9	8	149	3.73	.38	1.12	.50	.72	-.31	.20	.24	93.1	94.4	.00	P014_DI
MEAN	65.1	151.2	.63	.24	1.09	.56	.98	-.05			79.0	80.2	.00	
P.SD	39.1	3.2	1.79	.06	.18	1.75	.25	1.13			8.3	7.5	.00	

## Anhang 13: Hauptkomponentenanalyse der Residuen (PCAR) (po\_fw)

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	15.0338	100.0%	100.0%
Raw variance explained by measures =	6.0338	40.1%	44.0%
Raw variance explained by persons =	2.2143	14.7%	16.2%
Raw Variance explained by items =	3.8195	25.4%	27.9%
Raw unexplained variance (total) =	9.0000	59.9%	100.0%
Unexplned variance in 1st contrast =	1.5850	10.5%	17.6%
Unexplned variance in 2nd contrast =	1.3880	9.2%	15.4%
Unexplned variance in 3rd contrast =	1.2418	8.3%	13.8%
Unexplned variance in 4th contrast =	1.1691	7.8%	13.0%
Unexplned variance in 5th contrast =	1.1022	7.3%	12.2%

## Anhang 14: Personenfähigkeiten (measure) (po\_fw)

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

TABLE OF MEASURES ON TEST OF 9 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-3.92E	1.91	4	.27	.86	8	3.75	1.21
1	-2.47	1.17	5	1.00	.85	9	5.27E	1.94
2	-1.37	.97	6	1.75	.88			
3	-.51	.90	7	2.60	.97			

CURRENT VALUES, UIMEAN=.0000 USCALE=1.0000

TO SET MEASURE RANGE AS 0-100, UIMEAN=42.6527 USCALE=10.8773

TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UIMEAN=3.8387 USCALE=.9790

Predicting Score from Measure: Score = Measure \* 1.0624 + 3.8244

Predicting Measure from Score: Measure = Score \* .9311 + -3.5539

Maximum statistically different levels of performance (strata) = 2.0

Wright's Sample-independent Person (Test) Reliability based on maximum strata = .80

## Anhang 15: Step Ordering (po\_fw)

TABLE 14.3 Kurz\_154\_final.sav ZOU232WS.TXT Jun 26 2023 8:43  
 INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA		ABILITY		S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM
			COUNT	%	MEAN	P.SD					
1	0	0	34	24	-1.34	1.52	.26	1.2	1.1	-.51	P001_DI
	1	1	110	76	.36	1.13	.11	1.0	1.0	.51	
	MISSING ***		10	6#	1.87	.84	.28			.32	
2	0	0	37	24	-1.46	1.46	.24	.8	.7	-.60	P002_DI
	1	1	116	76	.59	1.07	.10	.9	.9	.60	
	MISSING ***		1	1#	-2.16	.00				-.12	
3	0	0	113	73	-.37	1.38	.13	1.0	1.0	-.50	P004_DI
	1	1	41	27	1.31	.93	.15	.7	.6	.50	
4	0	0	32	21	-1.27	1.52	.27	1.1	.9	-.47	P007_DI
	1	1	122	79	.43	1.24	.11	1.2	1.2	.47	
5	0	0	103	67	-.35	1.43	.14	1.1	1.2	-.41	P008_DI
	1	1	51	33	.94	1.14	.16	1.0	1.1	.41	
6	0	0	84	56	-.42	1.32	.14	.8	.7	-.50	P011_DI
	1	1	65	44	.97	1.01	.13	1.1	1.0	.50	
	MISSING ***		5	3#	-3.03	.71	.36			-.39	
7	0	0	104	68	-.23	1.43	.14	1.2	1.2	-.38	P012_DI
	1	1	48	32	.91	1.00	.15	1.1	1.0	.38	
	MISSING ***		2	1#	-3.89	.00	.00			-.31	
8	0	0	127	84	-.14	1.35	.12	.8	.8	-.42	P013_DI
	1	1	25	16	1.47	1.05	.21	.9	.7	.42	
	MISSING ***		2	1#	-3.18	.71	.71			-.25	
9	0	0	141	95	.06	1.47	.12	1.3	1.1	-.20	P014_DI
	1	1	8	5	1.38	.56	.21	1.4	.7	.20	
	MISSING ***		5	3#	-1.40	.82	.41			-.18	

# Missing % includes all categories. Scored % only of scored categories

## Anhang 16: Zusammenfassung Statistik (po\_fw)

INPUT: 154 PERSON 9 ITEM REPORTED: 154 PERSON 9 ITEM 2 CATS WINSTEPS 5.2.4.0

## SUMMARY OF 149 MEASURED (NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	3.9	8.9	.21	.92	1.06	.15	.98	.20
SEM	.1	.0	.11	.01	.04	.08	.06	.06
P.SD	1.5	.4	1.30	.09	.46	1.00	.70	.70
S.SD	1.5	.4	1.31	.09	.47	1.01	.70	.70
MAX.	8.0	9.0	3.75	1.24	2.33	2.28	3.13	2.16
MIN.	1.0	7.0	-2.47	.85	.33	-1.68	.18	-.82
REAL RMSE	1.01	TRUE SD	.82	SEPARATION	.81	PERSON RELIABILITY	.40	
MODEL RMSE	.92	TRUE SD	.92	SEPARATION	1.00	PERSON RELIABILITY	.50	
S.E. OF PERSON MEAN = .11								

MINIMUM EXTREME SCORE: 5 PERSON 3.2%

## SUMMARY OF 154 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	3.8	8.8	.08	.95				
SEM	.1	.0	.12	.02				
P.SD	1.7	.4	1.47	.20				
S.SD	1.7	.4	1.48	.20				
MAX.	8.0	9.0	3.75	1.92				
MIN.	.0	7.0	-3.92	.85				
REAL RMSE	1.05	TRUE SD	1.03	SEPARATION	.98	PERSON RELIABILITY	.49	
MODEL RMSE	.97	TRUE SD	1.11	SEPARATION	1.15	PERSON RELIABILITY	.57	
S.E. OF PERSON MEAN = .12								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)  
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .47 SEM = 1.21  
 (approximate due to missing data)  
 STANDARDIZED (50 ITEM) RELIABILITY = .88

## SUMMARY OF 9 MEASURED (NON-EXTREME) ITEM

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	65.1	151.2	.63	.24	1.09	.56	.98	-.05
SEM	13.8	1.1	.63	.02	.07	.62	.09	.40
P.SD	39.1	3.2	1.79	.06	.18	1.75	.25	1.13
S.SD	41.5	3.3	1.90	.06	.20	1.86	.26	1.19
MAX.	122.0	154.0	3.73	.38	1.42	3.73	1.40	1.70
MIN.	8.0	144.0	-2.02	.19	.78	-2.29	.69	-1.96
REAL RMSE	.26	TRUE SD	1.77	SEPARATION	6.86	ITEM RELIABILITY	.98	
MODEL RMSE	.24	TRUE SD	1.78	SEPARATION	7.31	ITEM RELIABILITY	.98	
S.E. OF ITEM MEAN = .63								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.94 (approximate due to missing data)  
 Global statistics: please see Table 44.  
 UMEAN=.0000 USCALE=1.0000

### 11.2.3 Follow-up-Test Fachwissen

#### Anhang 17: Item-Statistik (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: 1.01 REL.: .50 ... ITEM: REAL SEP.: 4.70 REL.: .96

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT  MNSQ ZSTD	OUTFIT  MNSQ ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	73	118	-1.54	.21	1.14 1.50	1.23 1.61	.33	.45	66.1	71.5	F001_DI
2	68	118	-1.32	.21	.97 -.36	.98 -.13	.47	.45	73.7	70.5	F002_DI
3	53	118	-.68	.21	.95 -.61	.85 -1.25	.51	.45	66.1	69.5	F003_DI
4	18	118	1.16	.27	.91 -.49	.91 -.18	.40	.34	83.9	84.9	F004_DI
5	19	118	1.08	.27	1.04 .30	.91 -.17	.32	.34	84.7	84.1	F005_DI
6	51	118	-.59	.21	1.13 1.57	1.17 1.31	.34	.45	66.1	69.6	F006_DI
7	81	118	-1.92	.22	.78 -2.24	.83 -1.00	.59	.43	81.4	74.8	F007_DI
8	25	118	.69	.25	.88 -.90	.88 -.41	.46	.38	84.7	80.2	F008_DI
9	71	118	-1.45	.21	.83 -2.03	.75 -2.00	.59	.45	78.0	71.2	F010_DI
10	18	118	1.16	.27	1.20 1.18	1.97 2.38	.12	.34	83.9	84.9	F011_DI
11	22	118	.88	.26	.90 -.71	.73 -.96	.46	.36	82.2	82.2	F012_DI
12	16	118	1.31	.29	1.17 .95	1.22 .69	.18	.32	85.6	86.5	F013_DI
13	17	118	1.23	.28	1.05 .32	1.32 .95	.25	.33	86.4	85.7	F014_DI
MEAN	40.9	118.0	.00	.24	1.00 -.12	1.06 .07			78.7	78.1	
P.SD	24.6	.0	1.21	.03	.13 1.18	.32 1.21			7.6	6.7	

#### Anhang 18: Item-Statistik geankert (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: 1.01 REL.: .50 ... ITEM: REAL SEP.: 4.33 REL.: .95

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT  MNSQ ZSTD	OUTFIT  MNSQ ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	DISPLACE	ITEM
1	73	118	-2.02A	.23	1.41 3.24	1.69 2.92	.33	.43	64.4	76.6	.61	F001_DI
2	68	118	-1.03A	.21	.96 -.48	1.00 .05	.47	.46	73.7	69.9	-.16	F002_DI
3	53	118	-.22A	.21	1.01 .11	.92 -.54	.51	.45	69.5	71.3	-.31	F003_DI
4	18	118	.81A	.25	.75 -2.06	.71 -1.18	.40	.39	87.3	80.1	.49	F004_DI
5	19	118	1.34A	.28	1.11 .66	.99 .08	.32	.34	84.7	85.2	-.10	F005_DI
6	51	118	-.78A	.21	1.15 1.81	1.18 1.47	.34	.46	64.4	69.6	.33	F006_DI
7	81	118	-1.53A	.22	.75 -2.86	.77 -1.58	.59	.45	81.4	72.7	-.26	F007_DI
8	25	118	.42A	.23	.78 -2.10	.74 -1.32	.46	.41	84.7	76.2	.41	F008_DI
9	71	118	-1.10A	.21	.82 -2.22	.74 -2.26	.59	.46	78.0	70.3	-.22	F010_DI
10	18	118	1.58A	.30	1.39 1.88	2.57 2.93	.12	.32	83.9	87.6	-.26	F011_DI
11	22	118	1.03	.26	.90 -.67	.73 -.92	.46	.37	82.2	82.3	.00	F012_DI
12	16	118	2.53A	.41	2.44 3.30	3.20 2.38	.18	.23	86.4	94.3	-1.00	F013_DI
13	17	118	1.39	.28	1.05 .35	1.35 1.00	.25	.34	86.4	85.7	.00	F014_DI
MEAN	40.9	118.0	.19	.25	1.12 .07	1.28 .23			79.0	78.6	.00	
P.SD	24.6	.0	1.34	.05	.44 1.98	.75 1.69			8.0	7.7	.00	

Anhang 19: Item-Statistik (fo\_fw) (bereinigt)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0  
 PERSON: REAL SEP.: 1.13 REL.: .56 ... ITEM: REAL SEP.: 4.81 REL.: .96

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-AL CORR.	EXP.	EXACT OBS%	MATCH EXP%	DISPLACE	ITEM
1	73	115	-2.02A	.23	1.23	1.91	1.12	.64	.43	.46	68.4	76.0	.40	F001_DI
2	68	118	-1.03A	.21	.95	-.60	.98	-.08	.50	.48	75.2	70.4	-.22	F002_DI
3	53	118	-.22A	.22	1.03	.38	.93	-.40	.52	.47	69.2	72.2	-.35	F003_DI
4	18	118	.81A	.25	.76	-1.90	.71	-1.11	.41	.41	88.0	80.5	-.50	F004_DI
5	19	118	1.34A	.28	1.10	.62	.98	.06	.34	.36	86.3	85.5	-.10	F005_DI
6	51	118	-.78A	.21	1.19	2.12	1.21	1.54	.35	.48	61.5	70.3	-.29	F006_DI
7	81	118	-1.53A	.22	.75	-2.79	.80	-1.27	.59	.48	82.1	73.0	-.35	F007_DI
8	25	118	.42A	.23	.80	-1.82	.89	-.45	.46	.43	83.8	76.8	-.40	F008_DI
9	71	118	-1.10A	.21	.87	-1.53	.82	-1.36	.56	.48	79.5	70.8	-.29	F010_DI
10	15	115	1.58A	.30	1.17	.92	1.20	.59	.22	.34	84.3	87.4	-.01	F011_DI
11	22	118	1.02	.26	.91	-.58	.73	-.88	.46	.39	82.1	82.7	.00	F012_DI
12	8	110	2.53A	.42	1.21	.73	.85	.01	.26	.25	92.7	93.8	-.20	F013_DI
13	16	117	1.48	.29	1.06	.36	1.37	.99	.27	.35	87.1	86.6	.00	F014_DI
MEAN	40.0	116.8	.19	.26	1.00	-.17	.97	-.13			80.0	78.9	.00	
P.SD	25.6	2.2	1.35	.06	.17	1.46	.19	.87			8.6	7.4	.00	

Anhang 20: Hauptkomponentenanalyse der Residuen (PCAR) (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	20.5471	100.0%	100.0%
Raw variance explained by measures =	7.5471	36.7%	35.6%
Raw variance explained by persons =	3.3060	16.1%	15.6%
Raw Variance explained by items =	4.2411	20.6%	20.0%
Raw unexplained variance (total) =	13.0000	63.3%	100.0%
Unexplnd variance in 1st contrast =	1.5607	7.6%	12.0%
Unexplnd variance in 2nd contrast =	1.5343	7.5%	11.8%
Unexplnd variance in 3rd contrast =	1.4399	7.0%	11.1%
Unexplnd variance in 4th contrast =	1.3980	6.8%	10.8%
Unexplnd variance in 5th contrast =	1.1578	5.6%	8.9%

Anhang 21: Personenfähigkeiten (measure) (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

TABLE OF MEASURES ON TEST OF 13 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-4.31E	1.88	5	-.47	.68	10	1.84	.74
1	-2.97	1.09	6	-.02	.66	11	2.45	.84
2	-2.08	.84	7	.42	.66	12	3.34	1.09
3	-1.46	.75	8	.86	.67	13	4.69E	1.88
4	-.94	.70	9	1.33	.69			

CURRENT VALUES, UIMEAN=.0000 USCALE=1.0000  
 TO SET MEASURE RANGE AS 0-100, UIMEAN=47.9195 USCALE=11.1160  
 TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UIMEAN=6.2295 USCALE=1.4451  
 Predicting Score from Measure: Score = Measure \* 1.6701 + 6.1795  
 Predicting Measure from Score: Measure = Score \* .5846 + -3.6078  
 Maximum statistically different levels of performance (strata) = 2.5  
 Wright's Sample-independent Person (Test) Reliability based on maximum strata = .87

## Anhang 22: Step Ordering (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM
1	.	***	36	23#							F001_DI
	0	0	42	37	-1.60	.99	.16	1.1	1.0	-.43	
	1	1	73	63	-.53	1.12	.13	1.1	1.1	.43	
	MISSING	***	3	2#	1.49	.46	.33			.31	
2	.	***	36	23#							F002_DI
	0	0	50	42	-1.58	.98	.14	.8	.8	-.50	
	1	1	68	58	-.32	1.13	.14	1.1	1.2	.50	
3	.	***	36	23#							F003_DI
	0	0	65	55	-1.44	1.09	.14	.9	.8	-.52	
	1	1	53	45	-.14	1.02	.14	1.0	.9	.52	
4	.	***	36	23#							F004_DI
	0	0	100	85	-1.07	1.15	.12	1.0	1.0	-.41	
	1	1	18	15	.33	1.04	.25	.8	.8	.41	
5	.	***	36	23#							F005_DI
	0	0	99	84	-1.04	1.19	.12	1.0	1.0	-.34	
	1	1	19	16	.12	1.00	.24	1.1	.9	.34	
6	.	***	36	23#							F006_DI
	0	0	67	57	-1.23	1.18	.15	1.2	1.3	-.35	
	1	1	51	43	-.36	1.13	.16	1.2	1.2	.35	
7	.	***	36	23#							F007_DI
	0	0	37	31	-1.94	1.12	.19	.8	.9	-.59	
	1	1	81	69	-.36	.94	.10	.8	.8	.59	
8	.	***	36	23#							F008_DI
	0	0	93	79	-1.15	1.09	.11	.9	.9	-.46	
	1	1	25	21	.23	1.15	.24	.9	1.1	.46	
9	.	***	36	23#							F010_DI
	0	0	47	40	-1.71	1.01	.15	.8	.7	-.56	
	1	1	71	60	-.29	1.03	.12	.9	1.0	.56	
10	.	***	36	23#							F011_DI
	0	0	100	87	-.89	1.18	.12	1.2	1.2	-.22	
	1	1	15	13	-.11	.90	.24	1.3	1.2	.22	
	MISSING	***	3	2#	-3.41	.63	.44			-.33	
11	.	***	36	23#							F012_DI
	0	0	96	81	-1.13	1.14	.12	.9	.9	-.46	
	1	1	22	19	.34	.90	.20	.9	.7	.46	
12	.	***	36	23#							F013_DI
	0	0	102	93	-.90	1.25	.12	1.1	1.0	-.26	
	1	1	8	7	.37	.69	.26	1.1	.7	.26	
	MISSING	***	8	5#	-1.48	.50	.19			-.14	
13	.	***	36	23#							F014_DI
	0	0	101	86	-.97	1.18	.12	1.0	1.0	-.27	
	1	1	16	14	.00	1.22	.32	1.2	1.4	.27	
	MISSING	***	1	1#	-2.96	.00				-.16	

# Missing % includes all categories. Scored % only of scored categories

## Anhang 23: Zusammenfassung Statistik (fo\_fw)

INPUT: 154 PERSON 13 ITEM REPORTED: 118 PERSON 13 ITEM 2 CATS WINSTEPS 5.2.4.0

## SUMMARY OF 117 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.4	12.9	-.83	.75	.99	.00	.97	.07
SEM	.2	.0	.11	.01	.03	.08	.06	.08
P.SD	2.3	.3	1.20	.13	.30	.89	.67	.85
S.SD	2.3	.3	1.21	.13	.30	.89	.67	.86
MAX.	10.0	13.0	1.83	1.10	1.84	2.16	4.29	2.02
MIN.	1.0	12.0	-2.98	.66	.40	-2.28	.26	-1.88
REAL RMSE	.80	TRUE SD	.89	SEPARATION	1.11	PERSON RELIABILITY	.55	
MODEL RMSE	.76	TRUE SD	.93	SEPARATION	1.22	PERSON RELIABILITY	.60	
S.E. OF PERSON MEAN = .11								

MINIMUM EXTREME SCORE: 1 PERSON .8%  
LACKING RESPONSES: 36 PERSON

## SUMMARY OF 118 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.4	12.9	-.86	.76				
SEM	.2	.0	.11	.02				
P.SD	2.3	.3	1.24	.16				
S.SD	2.3	.3	1.24	.16				
MAX.	10.0	13.0	1.83	1.87				
MIN.	.0	12.0	-4.29	.66				
REAL RMSE	.82	TRUE SD	.93	SEPARATION	1.13	PERSON RELIABILITY	.56	
MODEL RMSE	.78	TRUE SD	.96	SEPARATION	1.23	PERSON RELIABILITY	.60	
S.E. OF PERSON MEAN = .11								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)  
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .61 SEM = 1.43  
(approximate due to missing data)  
STANDARDIZED (50 ITEM) RELIABILITY = .85

## SUMMARY OF 13 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	40.0	116.8	.19	.26	1.00	-.17	.97	-.13
SEM	7.4	.6	.39	.02	.05	.42	.06	.25
P.SD	25.6	2.2	1.35	.06	.17	1.46	.19	.87
S.SD	26.6	2.3	1.40	.06	.17	1.52	.20	.90
MAX.	81.0	118.0	2.53	.42	1.23	2.12	1.37	1.54
MIN.	8.0	110.0	-2.02	.21	.75	-2.79	.71	-1.36
REAL RMSE	.27	TRUE SD	1.32	SEPARATION	4.81	ITEM RELIABILITY	.96	
MODEL RMSE	.26	TRUE SD	1.32	SEPARATION	5.05	ITEM RELIABILITY	.96	
S.E. OF ITEM MEAN = .39								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.96 (approximate due to missing data)  
Global statistics: please see Table 44.  
UMEAN=.0000 USCALE=1.0000

## 11.2.4 Post-Test naturwissenschaftliche Arbeitsweise (NWA)

### Anhang 24: Item-Statistik (po\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: 1.11 REL.: .55 ... ITEM: REAL SEP.: 4.58 REL.: .95

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	39	154	.71	.20	1.11	1.10	1.05	.31	.32	.39	69.1	76.7	PO21_DI
2	19	154	1.73	.26	1.11	.68	1.61	1.63	.20	.30	85.5	87.9	PO22_DI
3	85	154	-.90	.18	.94	-.75	.89	-1.04	.51	.47	71.7	70.3	PO23_DI
4	19	154	1.73	.26	1.06	.41	2.20	2.75	.18	.30	88.2	87.9	PO24_DI
5	74	154	-.53	.18	.90	-1.51	.85	-1.43	.54	.46	69.7	68.8	PO25_DI
6	73	154	-.50	.18	.93	-1.01	.91	-.85	.51	.46	73.0	68.6	PO26_DI
7	52	154	.21	.19	.86	-1.94	.89	-.77	.52	.42	78.9	71.8	PO27_DI
8	53	154	.18	.19	1.02	.24	1.00	.08	.41	.42	71.7	71.4	PO28_DI
9	95	154	-1.24	.19	1.05	.60	1.04	.38	.43	.47	69.1	72.6	PO29_DI
10	58	154	.00	.19	1.05	.72	1.09	.75	.39	.43	67.1	69.8	PO30_DI
11	61	154	-.10	.18	1.06	.92	1.00	.06	.40	.44	65.8	69.3	PO31_DI
12	97	154	-1.31	.19	.92	-.88	.90	-.78	.52	.47	77.0	72.9	PO32_DI
MEAN	60.4	154.0	.00	.20	1.00	-.12	1.12	.09			73.9	74.0	
P.SD	25.0	.0	.97	.03	.08	.99	.38	1.15			6.8	6.6	

### Anhang 25: Item-Statistik (po\_nwa) (bereinigt)

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: 1.15 REL.: .57 ... ITEM: REAL SEP.: 4.72 REL.: .96

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
1	39	154	.70	.20	1.12	1.17	1.07	.44	.31	.39	68.5	76.4	PO21_DI
2	18	153	1.80	.27	1.09	.57	1.13	.47	.23	.30	85.9	88.3	PO22_DI
3	85	154	-.92	.19	.96	-.56	.90	-.85	.51	.48	71.1	70.2	PO23_DI
4	17	152	1.87	.27	1.02	.15	1.22	.70	.26	.29	89.3	88.9	PO24_DI
5	74	154	-.55	.18	.91	-1.38	.87	-1.28	.53	.46	69.1	68.5	PO25_DI
6	73	154	-.52	.18	.94	-.87	.93	-.65	.50	.46	72.5	68.3	PO26_DI
7	52	154	.20	.19	.86	-1.89	.91	-.66	.51	.42	78.5	71.4	PO27_DI
8	53	154	.17	.19	1.02	.34	1.03	.24	.41	.42	71.1	71.0	PO28_DI
9	95	154	-1.27	.19	1.07	.83	1.07	.57	.44	.49	68.5	72.6	PO29_DI
10	58	154	-.01	.19	1.06	.84	1.12	.95	.39	.44	66.4	69.4	PO30_DI
11	61	154	-.11	.18	1.07	1.05	1.02	.24	.40	.44	65.1	69.0	PO31_DI
12	97	154	-1.35	.19	.94	-.64	.92	-.59	.53	.49	75.2	73.1	PO32_DI
MEAN	60.2	153.8	.00	.20	1.00	-.03	1.02	-.03			73.4	73.9	
P.SD	25.4	.6	1.01	.03	.08	.97	.11	.70			7.3	6.9	

### Anhang 26: Hauptkomponentenanalyse der Residuen (PCAR) (fo\_fw)

TABLE 23.0 Kurz\_154\_final.sav

ZOU455WS.TXT Jul 8 2023 9:29

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations	= 16.7423	100.0%	100.0%
Raw variance explained by measures	= 4.7423	28.3%	28.2%
Raw variance explained by persons	= 1.6672	10.0%	9.9%
Raw Variance explained by items	= 3.0751	18.4%	18.3%
Raw unexplained variance (total)	= 12.0000	71.7%	100.0%
Unexplned variance in 1st contrast	= 1.6435	9.8%	13.7%
Unexplned variance in 2nd contrast	= 1.4751	8.8%	12.3%
Unexplned variance in 3rd contrast	= 1.2856	7.7%	10.7%
Unexplned variance in 4th contrast	= 1.2402	7.4%	10.3%
Unexplned variance in 5th contrast	= 1.1275	6.7%	9.4%

Anhang 27: Personenfähigkeiten (measure) (po\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

TABLE OF MEASURES ON TEST OF 12 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-4.05E	1.86	5	-.45	.64	10	1.93	.84
1	-2.75	1.07	6	-.04	.64	11	2.84	1.10
2	-1.90	.81	7	.37	.65	12	4.19E	1.88
3	-1.33	.71	8	.81	.68			
4	-.87	.66	9	1.31	.74			

CURRENT VALUES, UIMEAN=.0000 USCALE=1.0000  
 TO SET MEASURE RANGE AS 0-100, UIMEAN=49.1864 USCALE=12.1416  
 TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UIMEAN=5.9024 USCALE=1.4570  
 Predicting Score from Measure: Score = Measure \* 1.6866 + 5.9919  
 Predicting Measure from Score: Measure = Score \* .5759 + -3.4504  
 Maximum statistically different levels of performance (strata) = 2.3  
 Wright's Sample-independent Person (Test) Reliability based on maximum strata = .84

Anhang 28: Step Ordering (po\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	S.E. P.SD	INFT MEAN	OUTF MNSQ	PTMA CORR.	ITEM	
1	0	0	115	75	-.92	1.25	.12	1.1	1.1	-.31	P021_DI
	1	1	39	25	-.02	.94	.15	1.1	1.1	.31	
2	0	0	135	88	-.78	1.22	.11	1.2	1.1	-.23	P022_DI
	1	1	18	12	.11	.86	.21	1.1	1.1	.23	
		MISSING ***		1	1#	-4.04	.00			-.22	
3	0	0	69	45	-1.40	1.20	.15	1.0	.9	-.51	P023_DI
	1	1	85	55	-.12	.95	.10	1.0	.9	.51	
4	0	0	135	89	-.76	1.15	.10	1.0	1.0	-.26	P024_DI
	1	1	17	11	.23	1.14	.29	1.1	1.3	.26	
		MISSING ***		2	1#	-4.04	.00	.00		-.31	
5	0	0	80	52	-1.33	1.17	.13	.9	.8	-.53	P025_DI
	1	1	74	48	.00	.92	.11	.9	.9	.53	
6	0	0	81	53	-1.29	1.23	.14	1.0	1.0	-.50	P026_DI
	1	1	73	47	-.03	.87	.10	.9	.9	.50	
7	0	0	102	66	-1.15	1.11	.11	.8	.8	-.51	P027_DI
	1	1	52	34	.20	.99	.14	.9	1.0	.51	
8	0	0	101	66	-1.06	1.22	.12	1.0	1.0	-.41	P028_DI
	1	1	53	34	.01	.96	.13	1.0	1.0	.41	
9	0	0	59	38	-1.39	1.26	.17	1.1	1.1	-.44	P029_DI
	1	1	95	62	-.26	1.02	.10	1.1	1.1	.44	
10	0	0	96	62	-1.07	1.23	.13	1.1	1.0	-.39	P030_DI
	1	1	58	38	-.07	1.00	.13	1.1	1.2	.39	
11	0	0	93	60	-1.10	1.23	.13	1.0	1.0	-.40	P031_DI
	1	1	61	40	-.08	.99	.13	1.1	1.0	.40	
12	0	0	57	37	-1.55	1.21	.16	.9	.9	-.53	P032_DI
	1	1	97	63	-.19	.96	.10	1.0	1.0	.53	

# Missing % includes all categories. Scored % only of scored categories

## Anhang 29: Zusammenfassung Statistik (po\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 154 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

## SUMMARY OF 149 MEASURED (NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.8	12.0	-.58	.72	1.00	.04	1.02	.07
SEM	.2	.0	.09	.01	.02	.07	.05	.07
P.SD	2.3	.0	1.10	.12	.24	.90	.63	.88
S.SD	2.3	.0	1.11	.12	.24	.90	.63	.88
MAX.	11.0	12.0	2.83	1.10	1.69	2.61	4.97	2.63
MIN.	1.0	12.0	-2.75	.64	.49	-2.10	.29	-1.61
REAL RMSE	.76	TRUE SD	.80	SEPARATION	1.05	PERSON RELIABILITY	.53	
MODEL RMSE	.73	TRUE SD	.83	SEPARATION	1.14	PERSON RELIABILITY	.56	
S.E. OF PERSON MEAN = .09								

MINIMUM EXTREME SCORE: 5 PERSON 3.2%

## SUMMARY OF 154 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	4.7	12.0	-.69	.75				
SEM	.2	.0	.10	.02				
P.SD	2.4	.1	1.24	.24				
S.SD	2.4	.1	1.25	.24				
MAX.	11.0	12.0	2.83	1.86				
MIN.	.0	11.0	-4.05	.64				
REAL RMSE	.82	TRUE SD	.94	SEPARATION	1.15	PERSON RELIABILITY	.57	
MODEL RMSE	.79	TRUE SD	.96	SEPARATION	1.22	PERSON RELIABILITY	.60	
S.E. OF PERSON MEAN = .10								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98 (approximate due to missing data)  
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .61 SEM = 1.50  
 (approximate due to missing data)

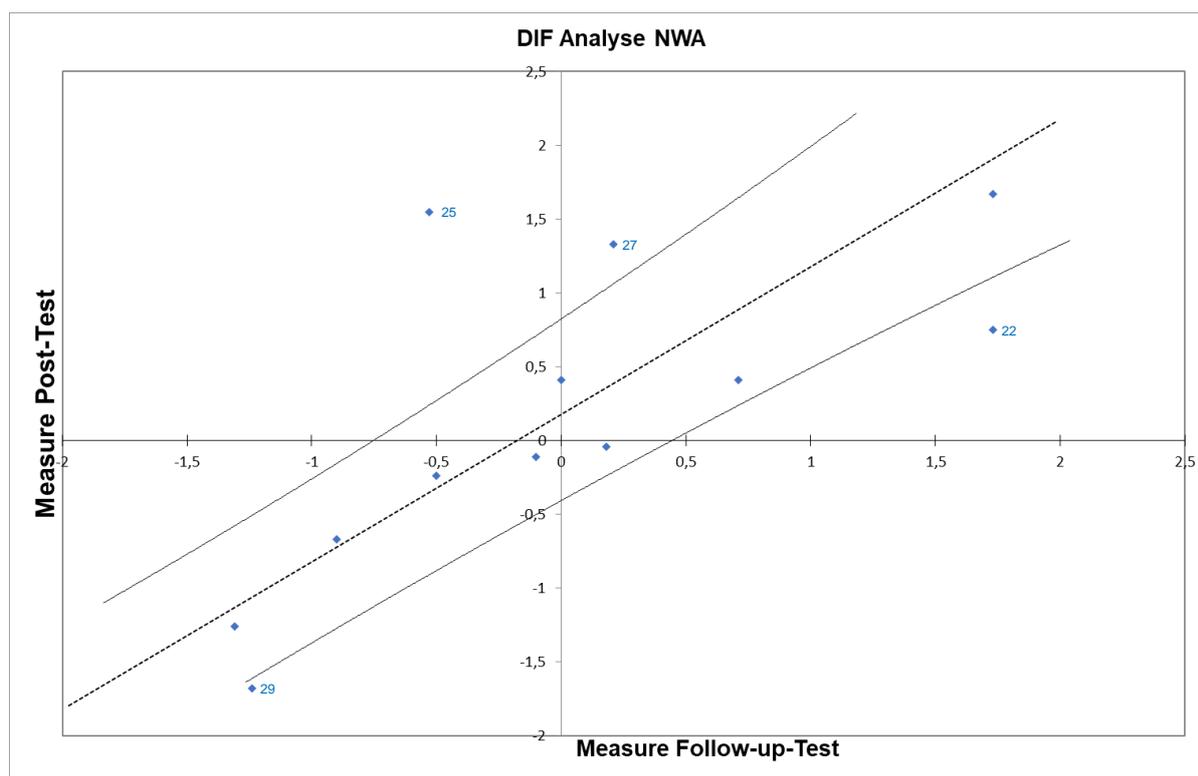
STANDARDIZED (50 ITEM) RELIABILITY = .86

## SUMMARY OF 12 MEASURED (NON-EXTREME) ITEM

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	60.2	153.8	.00	.20	1.00	-.03	1.02	-.03
SEM	7.7	.2	.30	.01	.02	.29	.03	.21
P.SD	25.4	.6	1.01	.03	.08	.97	.11	.70
S.SD	26.6	.6	1.05	.03	.08	1.01	.11	.73
MAX.	97.0	154.0	1.87	.27	1.12	1.17	1.22	.95
MIN.	17.0	152.0	-1.35	.18	.86	-1.89	.87	-1.28
REAL RMSE	.21	TRUE SD	.99	SEPARATION	4.72	ITEM RELIABILITY	.96	
MODEL RMSE	.20	TRUE SD	.99	SEPARATION	4.82	ITEM RELIABILITY	.96	
S.E. OF ITEM MEAN = .30								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.99 (approximate due to missing data)  
 Global statistics: please see Table 44.  
 UMEAN=.0000 USCALE=1.0000

## Anhang 30: DIF-Analyse Post-Test-Follow-up-Test (nwa)



## 11.2.5 Follow-up-Test naturwissenschaftliche Arbeitsweise (NWA)

### Anhang 31: Item-Statistik (fo\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .88 REL.: .44 ... ITEM: REAL SEP.: 3.67 REL.: .93

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM		
1	25	118	.24	.25	1.17	1.29	1.27	1.19	.27	.39	77.1	79.0	F021_DI
2	20	118	.57	.27	1.00	.06	1.17	.69	.36	.37	81.7	82.8	F022_DI
3	46	118	-.85	.21	1.03	.33	.99	-.05	.46	.47	68.8	69.3	F023_DI
4	10	118	1.49	.35	1.09	.44	1.75	1.51	.18	.29	91.7	91.3	F024_DI
5	11	118	1.37	.34	.99	.02	.83	-.30	.32	.30	90.8	90.4	F025_DI
6	37	118	-.42	.22	.87	-1.39	.75	-1.77	.54	.44	77.1	71.6	F026_DI
7	13	118	1.15	.32	.91	-.34	1.30	.85	.33	.32	89.0	88.7	F027_DI
8	33	118	-.22	.23	.97	-.30	1.15	.88	.42	.43	80.7	73.9	F028_DI
9	68	118	-1.85	.22	.94	-.67	.92	-.46	.55	.51	71.6	70.1	F029_DI
10	25	118	.24	.25	.98	-.11	.88	-.48	.41	.39	78.9	79.0	F030_DI
11	34	118	-.27	.23	1.06	.64	1.21	1.24	.37	.43	76.1	73.3	F031_DI
12	59	118	-1.44	.21	.87	-1.60	.88	-.95	.57	.49	74.3	68.5	F032_DI
MEAN	31.7	118.0	.00	.26	.99	-.13	1.09	.20			79.8	78.1	
P.SD	17.7	.0	1.02	.05	.09	.78	.27	.97			7.1	8.1	

### Anhang 32: Item-Statistik geankert (fo\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

PERSON: REAL SEP.: .85 REL.: .42 ... ITEM: REAL SEP.: 2.98 REL.: .90

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	DISPLACE	ITEM		
1	25	118	.70A	.28	1.49	2.58	1.67	2.04	.27	.35	75.2	84.4	-.47	F021_DI
2	20	118	1.80A	.40	2.53	3.58	3.40	3.05	.36	.27	82.6	93.4	-1.22	F022_DI
3	46	118	-.92A	.21	1.01	.11	.97	-.21	.46	.46	68.8	68.7	.06	F023_DI
4	10	118	1.87A	.41	1.44	1.30	2.44	2.08	.19	.27	91.7	93.8	-.40	F024_DI
5	11	118	-.55A	.22	.84	-1.87	.74	-2.02	.32	.44	71.6	70.6	1.93	F025_DI
6	37	118	-.52A	.22	.85	-1.76	.74	-1.97	.54	.44	77.1	70.8	.09	F026_DI
7	13	118	.20A	.25	.67	-2.91	.72	-1.41	.34	.39	89.9	78.7	.93	F027_DI
8	33	118	.17A	.24	1.12	.98	1.45	1.95	.42	.39	77.1	78.3	-.39	F028_DI
9	68	118	-1.27A	.21	.96	-.44	.92	-.65	.54	.48	65.1	67.6	-.58	F029_DI
10	25	118	-.01A	.24	.90	-.88	.79	-1.12	.42	.41	78.9	76.0	.23	F030_DI
11	34	118	-.11A	.23	1.11	.98	1.29	1.57	.37	.41	76.1	74.9	-.17	F031_DI
12	59	118	-1.35A	.21	.87	-1.77	.87	-1.08	.57	.48	74.3	67.4	-.09	F032_DI
MEAN	31.7	118.0	.00	.26	1.15	-.01	1.33	.19			77.4	77.1	.00	
P.SD	17.7	.0	1.01	.07	.48	1.88	.79	1.74			7.4	8.9	.00	

Anhang 33: Item-Statistik (fo\_nwa) (bereinigt)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0  
 PERSON: REAL SEP.: .88 REL.: .44 ... ITEM: REAL SEP.: 3.30 REL.: .92

ITEM STATISTICS: ENTRY ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT O85%	MATCH EXP%	DISPLACE	ITEM
1	24	117	.70A	.28	1.48	2.49	1.45	1.48	.30	.36	75.2	84.4	-.46	F021_DI
2	11	109	1.80A	.42	1.37	1.12	.75	-.32	.44	.28	89.8	93.2	-.54	F022_DI
3	46	118	-.92A	.22	1.03	.40	1.02	-.22	.46	.48	68.6	68.5	-.01	F023_DI
4	7	115	1.87A	.42	1.00	.10	.83	-.16	.27	.27	94.2	93.7	.00	F024_DI
5	11	118	-.55A	.22	.85	-1.74	.76	-1.83	.31	.45	72.4	70.7	1.88	F025_DI
6	37	118	-.52A	.22	.87	-1.38	.77	-1.69	.54	.45	76.2	70.8	.03	F026_DI
7	13	118	.20A	.25	.67	-2.77	.73	-1.32	.34	.40	89.5	78.7	-.89	F027_DI
8	33	118	.17A	.25	1.17	1.27	1.51	2.17	.43	.40	76.2	78.3	-.45	F028_DI
9	68	118	-1.27A	.21	1.01	.15	.98	-.16	.55	.50	61.0	67.2	-.69	F029_DI
10	25	118	-.01A	.24	.92	-.67	.84	-.86	.42	.41	78.1	76.1	.18	F030_DI
11	34	118	-.11A	.24	1.12	1.11	1.32	1.72	.39	.42	75.2	75.0	-.23	F031_DI
12	59	118	-1.35A	.21	.88	-1.57	.88	-.92	.58	.50	73.3	67.0	-.18	F032_DI
MEAN	30.7	116.9	.00	.27	1.03	-.12	.99	-.14			77.5	77.0	.00	
P.SD	18.7	2.5	1.01	.07	.22	1.47	.27	1.27			9.1	8.9	.00	

Anhang 34: Hauptkomponentenanalyse der Residuen (PCAR) (fo\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed	Expected
Total raw variance in observations	15.9603	100.0%	100.0%
Raw variance explained by measures	3.9603	24.8%	24.5%
Raw variance explained by persons	1.2010	7.5%	7.4%
Raw Variance explained by items	2.7593	17.3%	17.1%
Raw unexplained variance (total)	12.0000	75.2%	100.0%
Unexplned variance in 1st contrast	1.6415	10.3%	13.7%
Unexplned variance in 2nd contrast	1.4997	9.4%	12.5%
Unexplned variance in 3rd contrast	1.4571	9.1%	12.1%
Unexplned variance in 4th contrast	1.2563	7.9%	10.5%
Unexplned variance in 5th contrast	1.1581	7.3%	9.7%

Anhang 35: Personenfähigkeiten (measure) (fo\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

TABLE OF MEASURES ON TEST OF 12 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-4.05E	1.86	5	-.45	.64	10	1.93	.84
1	-2.75	1.07	6	-.04	.64	11	2.84	1.10
2	-1.90	.81	7	.37	.65	12	4.19E	1.88
3	-1.33	.71	8	.82	.68			
4	-.87	.66	9	1.32	.74			

CURRENT VALUES, UIMEAN=.0000 USCALE=1.0000  
 TO SET MEASURE RANGE AS 0-100, UIMEAN=49.1800 USCALE=12.1421  
 TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UIMEAN=5.9016 USCALE=1.4571  
 Predicting Score from Measure: Score = Measure \* 1.6867 + 5.9905  
 Predicting Measure from Score: Measure = Score \* .5759 + -3.4495  
 Maximum statistically different levels of performance (strata) = 2.3  
 Wright's Sample-independent Person (Test) Reliability based on maximum strata = .84

## Anhang 36: Step Ordering (fo\_fw)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: CORRELATION ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM
4	.	***	36	23#							F024_DI
	0	0	108	94	-1.56	1.23	.12	.9	.9	-.27	
	1	1	7	6	-.10	1.27	.52	1.1	.8	.27	
	MISSING	***	3	2#	-3.16	.60	.42			-.20	
1	.	***	36	23#							F021_DI
	0	0	93	79	-1.69	1.32	.14	1.3	1.5	-.30	
	1	1	24	21	-.74	.75	.16	1.2	1.1	.30	
	MISSING	***	1	1#	-3.99	.00				-.18	
5	.	***	36	23#							F025_DI
	0	0	107	91	-1.65	1.23	.12	1.2	1.2	-.31	
	1	1	11	9	-.25	1.23	.39	.8	.7	.31	
7	.	***	36	23#							F027_DI
	0	0	105	89	-1.67	1.19	.12	1.0	1.0	-.34	
	1	1	13	11	-.24	1.42	.41	.9	1.1	.34	
11	.	***	36	23#							F031_DI
	0	0	84	71	-1.84	1.20	.13	.9	.9	-.39	
	1	1	34	29	-.71	1.17	.20	1.1	1.3	.39	
10	.	***	36	23#							F030_DI
	0	0	93	79	-1.80	1.21	.13	1.0	1.0	-.42	
	1	1	25	21	-.47	1.06	.22	1.0	.9	.42	
8	.	***	36	23#							F028_DI
	0	0	85	72	-1.87	1.15	.13	.8	.8	-.43	
	1	1	33	28	-.62	1.21	.21	1.1	1.3	.43	
2	.	***	36	23#							F022_DI
	0	0	98	90	-1.67	1.19	.12	.8	.8	-.44	
	1	1	11	10	.23	.91	.29	.8	.5	.44	
	MISSING	***	9	6#	-1.93	1.16	.41			-.09	
3	.	***	36	23#							F023_DI
	0	0	72	61	-2.00	1.24	.15	1.0	1.0	-.46	
	1	1	46	39	-.77	1.00	.15	1.0	1.0	.46	
6	.	***	36	23#							F026_DI
	0	0	81	69	-1.99	1.20	.13	1.0	1.0	-.54	
	1	1	37	31	-.48	.82	.14	.8	.7	.54	
9	.	***	36	23#							F029_DI
	0	0	50	42	-2.35	1.23	.18	.9	.9	-.55	
	1	1	68	58	-.90	.96	.12	1.1	1.1	.55	
12	.	***	36	23#							F032_DI
	0	0	59	50	-2.27	1.15	.15	.8	.8	-.58	
	1	1	59	50	-.76	.95	.13	.9	.9	.58	

# Missing % includes all categories. Scored % only of scored categories

## Anhang 37: Zusammenfassung Statistik (fo\_nwa)

INPUT: 154 PERSON 12 ITEM REPORTED: 118 PERSON 12 ITEM 2 CATS WINSTEPS 5.2.4.0

## SUMMARY OF 105 MEASURED (NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	3.5	11.9	-1.21	.76	1.00	.08	.99	.19
SEM	.2	.0	.10	.01	.02	.07	.04	.07
P.SD	2.0	.3	1.01	.15	.19	.69	.41	.70
S.SD	2.0	.3	1.02	.15	.19	.70	.42	.71
MAX.	11.0	12.0	2.84	1.10	1.53	1.89	2.70	2.28
MIN.	1.0	11.0	-2.75	.64	.58	-2.11	.41	-1.45
REAL RMSE	.81	TRUE SD	.61	SEPARATION	.76	PERSON RELIABILITY	.37	
MODEL RMSE	.78	TRUE SD	.65	SEPARATION	.83	PERSON RELIABILITY	.41	
S.E. OF PERSON MEAN = .10								

MINIMUM EXTREME SCORE: 13 PERSON 11.0%  
LACKING RESPONSES: 36 PERSON

## SUMMARY OF 118 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	3.1	11.9	-1.52	.88				
SEM	.2	.0	.12	.03				
P.SD	2.2	.3	1.30	.36				
S.SD	2.2	.3	1.30	.36				
MAX.	11.0	12.0	2.84	1.84				
MIN.	.0	11.0	-4.02	.64				
REAL RMSE	.97	TRUE SD	.86	SEPARATION	.88	PERSON RELIABILITY	.44	
MODEL RMSE	.95	TRUE SD	.88	SEPARATION	.92	PERSON RELIABILITY	.46	
S.E. OF PERSON MEAN = .12								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .97 (approximate due to missing data)  
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .62 SEM = 1.33  
(approximate due to missing data)  
STANDARDIZED (50 ITEM) RELIABILITY = .78

## SUMMARY OF 12 MEASURED (NON-EXTREME) ITEM

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	30.7	116.9	.00	.27	1.03	-.12	.99	-.14
SEM	5.6	.8	.30	.02	.07	.44	.08	.38
P.SD	18.7	2.5	1.01	.07	.22	1.47	.27	1.27
S.SD	19.6	2.6	1.05	.07	.23	1.53	.28	1.33
MAX.	68.0	118.0	1.87	.42	1.48	2.49	1.51	2.17
MIN.	7.0	109.0	-1.35	.21	.67	-2.77	.73	-1.83
REAL RMSE	.29	TRUE SD	.96	SEPARATION	3.30	ITEM RELIABILITY	.92	
MODEL RMSE	.27	TRUE SD	.97	SEPARATION	3.53	ITEM RELIABILITY	.93	
S.E. OF ITEM MEAN = .30								

ITEM RAW SCORE-TO-MEASURE CORRELATION = -.80 (approximate due to missing data)  
Global statistics: please see Table 44.  
UMEAN=.0000 USCALE=1.0000

## 11.3 Ergebnisse: Fachwissensfragebögen Datentabellen

### 11.3.1 Modellprüfung

Anhang 38: Shapiro-Wilk-Test auf Normalverteilung (fw)

#### Tests auf Normalverteilung

Gruppeneinteilung	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			
	Statistik	df	Signifikanz	Statistik	df	Signifikanz	
pr_fw	PP - Präsenzunterricht	,217	32	<,001	,937	32	,062
	DP - FlippedClassroom	,189	27	,015	,957	27	,323
	PD - Wechselunterricht	,163	29	,047	,925	29	,040
	DD - Homeschooling	,221	30	<,001	,902	30	,009
po_fw	PP - Präsenzunterricht	,160	32	,037	,956	32	,208
	DP - FlippedClassroom	,158	27	,082	,939	27	,117
	PD - Wechselunterricht	,184	29	,013	,932	29	,061
	DD - Homeschooling	,207	30	,002	,875	30	,002
fo_fw	PP - Präsenzunterricht	,167	32	,024	,943	32	,092
	DP - FlippedClassroom	,134	27	,200 <sup>*</sup>	,970	27	,596
	PD - Wechselunterricht	,149	29	,099	,914	29	,022
	DD - Homeschooling	,133	30	,183	,969	30	,504

<sup>\*</sup>. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Anhang 39: Levene-Test auf Homogenität (fw)

#### Test auf Homogenität der Varianz

		Levene-Statistik	df1	df2	Signifikanz
pr_fw	Basiert auf dem Mittelwert	1,630	3	114	,186
	Basiert auf dem Median	1,545	3	114	,207
	Basierend auf dem Median und mit angepaßten df	1,545	3	111,424	,207
	Basiert auf dem getrimmten Mittel	1,550	3	114	,206
po_fw	Basiert auf dem Mittelwert	1,880	3	114	,137
	Basiert auf dem Median	1,815	3	114	,148
	Basierend auf dem Median und mit angepaßten df	1,815	3	101,716	,149
	Basiert auf dem getrimmten Mittel	1,876	3	114	,138
fo_fw	Basiert auf dem Mittelwert	3,467	3	114	,019
	Basiert auf dem Median	2,927	3	114	,037
	Basierend auf dem Median und mit angepaßten df	2,927	3	103,103	,037
	Basiert auf dem getrimmten Mittel	3,268	3	114	,024

## 11.3.2 Vergleich der Gruppen

### Prä-Test Fachwissen

Anhang 40: Tests der Zwischensubjekteffekte (pr\_fw)

#### Tests der Zwischensubjekteffekte

Abhängige Variable: pr_fw						
Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	8,525 <sup>a</sup>	5	1,705	1,937	,092	,061
Konstanter Term	1,253	1	1,253	1,424	,235	,010
pr_KFT	,250	1	,250	,284	,595	,002
schule	,526	1	,526	,598	,441	,004
gruppe_4	8,146	3	2,715	3,085	,029	,059
Fehler	130,279	148	,880			
Gesamt	198,860	154				
Korrigierte Gesamtvariation	138,804	153				

a. R-Quadrat = ,061 (korrigiertes R-Quadrat = ,030)

Anhang 41: Paarweise Vergleiche (pr\_fw)

#### Paarweise Vergleiche

Abhängige Variable: pr_fw						
(I) Gruppeneinteilung	(J) Gruppeneinteilung	Mittelwertdif ferenz (I-J)		Sig. <sup>a</sup>	95% Konfidenzintervall für Differenz <sup>a</sup>	
		Std.-Fehler			Untergrenze	Obergrenze
PP - Präsenzunterricht	DP - FlippedClassroom	-,197	,216	1,000	-,776	,381
	PD - Wechselunterricht	-,565	,220	,068	-1,155	,024
	DD - Homeschooling	-,022	,252	1,000	-,696	,652
DP - FlippedClassroom	PP - Präsenzunterricht	,197	,216	1,000	-,381	,776
	PD - Wechselunterricht	-,368	,235	,713	-,995	,259
	DD - Homeschooling	,175	,261	1,000	-,523	,873
PD - Wechselunterricht	PP - Präsenzunterricht	,565	,220	,068	-,024	1,155
	DP - FlippedClassroom	,368	,235	,713	-,259	,995
	DD - Homeschooling	,543	,223	,097	-,054	1,140
DD - Homeschooling	PP - Präsenzunterricht	,022	,252	1,000	-,652	,696
	DP - FlippedClassroom	-,175	,261	1,000	-,873	,523
	PD - Wechselunterricht	-,543	,223	,097	-1,140	,054

Basiert auf geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

## Post-Test Fachwissen

### Anhang 42: Tests der Zwischensubjekteffekte (po\_fw)

#### Tests der Zwischensubjekteffekte

Abhängige Variable: po_fw						
Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	44,549 <sup>a</sup>	5	8,910	4,549	<,001	,133
Konstanter Term	3,683	1	3,683	1,881	,172	,013
pr_KFT	1,900	1	1,900	,970	,326	,007
schule	19,475	1	19,475	9,943	,002	,063
gruppe_4	24,670	3	8,223	4,199	,007	,078
Fehler	289,873	148	1,959			
Gesamt	335,382	154				
Korrigierte Gesamtvariation	334,422	153				

a. R-Quadrat = ,133 (korrigiertes R-Quadrat = ,104)

### Anhang 43: Paarweise Vergleiche (po\_fw)

#### Paarweise Vergleiche

Abhängige Variable: po_fw						
(I) Gruppeneinteilung	(J) Gruppeneinteilung	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	95% Konfidenzintervall für Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
PP - Präsenzunterricht	DP - FlippedClassroom	-,890 <sup>*</sup>	,323	,039	-1,752	-,027
	PD - Wechselunterricht	-,994 <sup>*</sup>	,329	,018	-1,873	-,115
	DD - Homeschooling	-1,009 <sup>*</sup>	,376	,048	-2,014	-,004
DP - FlippedClassroom	PP - Präsenzunterricht	,890 <sup>*</sup>	,323	,039	,027	1,752
	PD - Wechselunterricht	-,105	,350	1,000	-1,041	,831
	DD - Homeschooling	-,120	,390	1,000	-1,162	,922
PD - Wechselunterricht	PP - Präsenzunterricht	,994 <sup>*</sup>	,329	,018	,115	1,873
	DP - FlippedClassroom	,105	,350	1,000	-,831	1,041
	DD - Homeschooling	-,015	,333	1,000	-,905	,875
DD - Homeschooling	PP - Präsenzunterricht	1,009 <sup>*</sup>	,376	,048	,004	2,014
	DP - FlippedClassroom	,120	,390	1,000	-,922	1,162
	PD - Wechselunterricht	,015	,333	1,000	-,875	,905

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

## Follow-up-Test Fachwissen

### Anhang 44: Tests der Zwischensubjekteffekte (fo\_fw)

#### Tests der Zwischensubjekteffekte

Abhängige Variable: fo_fw						
Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	39,604 <sup>a</sup>	5	7,921	6,266	<,001	,219
Konstanter Term	,043	1	,043	,034	,853	,000
pr_KFT	2,715	1	2,715	2,148	,146	,019
schule	13,133	1	13,133	10,389	,002	,085
gruppe_4	1,614	3	,538	,426	,735	,011
Fehler	141,586	112	1,264			
Gesamt	268,068	118				
Korrigierte Gesamtvariation	181,190	117				

a. R-Quadrat = ,219 (korrigiertes R-Quadrat = ,184)

### Anhang 45: Paarweise Vergleiche (fo\_fw)

#### Paarweise Vergleiche

Abhängige Variable: fo_fw						
(I) Gruppeneinteilung	(J) Gruppeneinteilung	Mittelwertdifferenz (I-J)	Std.-Fehler	Sig. <sup>a</sup>	95% Konfidenzintervall für Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
PP - Präsenzunterricht	DP - FlippedClassroom	-,237	,295	1,000	-1,029	,555
	PD - Wechselunterricht	,050	,326	1,000	-,826	,925
	DD - Homeschooling	-,202	,348	1,000	-1,136	,732
DP - FlippedClassroom	PP - Präsenzunterricht	,237	,295	1,000	-,555	1,029
	PD - Wechselunterricht	,286	,341	1,000	-,631	1,204
	DD - Homeschooling	,034	,355	1,000	-,918	,987
PD - Wechselunterricht	PP - Präsenzunterricht	-,050	,326	1,000	-,925	,826
	DP - FlippedClassroom	-,286	,341	1,000	-1,204	,631
	DD - Homeschooling	-,252	,315	1,000	-1,099	,595
DD - Homeschooling	PP - Präsenzunterricht	,202	,348	1,000	-,732	1,136
	DP - FlippedClassroom	-,034	,355	1,000	-,987	,918
	PD - Wechselunterricht	,252	,315	1,000	-,595	1,099

Basiert auf geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

### 11.3.3 Vergleich der Messzeitpunkte

#### Präsenzunterricht – PP – Fachwissen

##### Anhang 46: Multivariate Tests (PP\_fw)

###### Multivariate Tests<sup>a</sup>

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
Fachwissen_PP	Pillai-Spur	,089	1,411 <sup>b</sup>	2,000	29,000	,260	,089
	Wilks-Lambda	,911	1,411 <sup>b</sup>	2,000	29,000	,260	,089
	Hotelling-Spur	,097	1,411 <sup>b</sup>	2,000	29,000	,260	,089
	Größte charakteristische Wurzel nach Roy	,097	1,411 <sup>b</sup>	2,000	29,000	,260	,089
Fachwissen_PP * pr_KFT	Pillai-Spur	,055	,838 <sup>b</sup>	2,000	29,000	,443	,055
	Wilks-Lambda	,945	,838 <sup>b</sup>	2,000	29,000	,443	,055
	Hotelling-Spur	,058	,838 <sup>b</sup>	2,000	29,000	,443	,055
	Größte charakteristische Wurzel nach Roy	,058	,838 <sup>b</sup>	2,000	29,000	,443	,055

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: Fachwissen\_PP

b. Exakte Statistik

##### Anhang 47: Mauchly-Test auf Sphärizität (PP\_fw)

###### Mauchly-Test auf Sphärizität<sup>a</sup>

Maß: MASS_1		Epsilon <sup>b</sup>					
Innersubjekteffekt	Mauchly-W	Ungefähres Chi-Quadrat	df	Sig.	Greenhouse -Geisser	Huynh-Feldt (HF)	Untergrenze
Fachwissen_PP	,741	8,889	2	,013	,794	,859	,500

Prüft die Nullhypothese, dass sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: Fachwissen\_PP

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

##### Anhang 48: Paarweise Vergleiche (PP\_fw)

###### Paarweise Vergleiche

Maß: MASS_1							95% Konfidenzintervall für Differenz <sup>a</sup>	
(I) Fachwissen_PP	(J) Fachwissen_PP	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>a</sup>	Untergrenze	Obergrenze		
1 pr_fw	2 po_fw	-,399	,206	,188	-,923	,124		
	3 fo_fw	-,133	,166	1,000	-,555	,289		
2 po_fw	1 pr_fw	,399	,206	,188	-,124	,923		
	3 fo_fw	,266	,261	,945	-,394	,927		
3 fo_fw	1 pr_fw	,133	,166	1,000	-,289	,555		
	2 po_fw	-,266	,261	,945	-,927	,394		

Basiert auf geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

## Flipped Classroom – DP – Fachwissen

### Anhang 49: Multivariate Tests (DP\_fw)

#### Multivariate Tests<sup>a</sup>

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
Fachwissen_DP	Pillai-Spur	,017	,208 <sup>b</sup>	2,000	24,000	,813	,017
	Wilks-Lambda	,983	,208 <sup>b</sup>	2,000	24,000	,813	,017
	Hotelling-Spur	,017	,208 <sup>b</sup>	2,000	24,000	,813	,017
	Größte charakteristische Wurzel nach Roy	,017	,208 <sup>b</sup>	2,000	24,000	,813	,017
Fachwissen_DP * pr_KFT	Pillai-Spur	,030	,373 <sup>b</sup>	2,000	24,000	,693	,030
	Wilks-Lambda	,970	,373 <sup>b</sup>	2,000	24,000	,693	,030
	Hotelling-Spur	,031	,373 <sup>b</sup>	2,000	24,000	,693	,030
	Größte charakteristische Wurzel nach Roy	,031	,373 <sup>b</sup>	2,000	24,000	,693	,030

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: Fachwissen\_DP

b. Exakte Statistik

### Anhang 50: Mauchlytest auf Sphärizität (DP\_fw)

#### Mauchly-Test auf Sphärizität<sup>a</sup>

Maß: MASS_1		Epsilon <sup>b</sup>					
Innersubjekteffekt	Mauchly-W	Ungefähres Chi-Quadrat	df	Sig.	Greenhouse -Geisser	Huynh-Feldt (HF)	Untergrenze
Fachwissen_DP	,781	5,933	2	,051	,820	,905	,500

Prüft die Nullhypothese, dass sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: Fachwissen\_DP

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

### Anhang 51: Paarweise Vergleiche (DP\_fw)

#### Paarweise Vergleiche

Maß: MASS_1		95% Konfidenzintervall für Differenz <sup>b</sup>					
(I) Fachwissen_DP	(J) Fachwissen_DP	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	Untergrenze	Obergrenze	
1 pr_fw	2 po_fw	-1,102 <sup>*</sup>	,315	,005	-1,911	-,293	
	3 fo_fw	-,253	,204	,684	-,777	,272	
2 po_fw	1 pr_fw	1,102 <sup>*</sup>	,315	,005	,293	1,911	
	3 fo_fw	,849 <sup>*</sup>	,289	,012	,159	1,539	
3 fo_fw	1 pr_fw	,253	,204	,684	-,272	,777	
	2 po_fw	-,849 <sup>*</sup>	,289	,012	-1,539	-,159	

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

**Wechselunterricht – PD – Fachwissen**

## Anhang 52: Multivariate Tests (PD\_fw)

*Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
Fachwissen_PD	Pillai-Spur	,061	,811 <sup>b</sup>	2,000	25,000	,466	,061
	Wilks-Lambda	,939	,811 <sup>b</sup>	2,000	25,000	,466	,061
	Hotelling-Spur	,065	,811 <sup>b</sup>	2,000	25,000	,466	,061
	Größte charakteristische Wurzel nach Roy	,065	,811 <sup>b</sup>	2,000	25,000	,466	,061
Fachwissen_PD * pr_KFT	Pillai-Spur	,002	,022 <sup>b</sup>	2,000	25,000	,978	,002
	Wilks-Lambda	,998	,022 <sup>b</sup>	2,000	25,000	,978	,002
	Hotelling-Spur	,002	,022 <sup>b</sup>	2,000	25,000	,978	,002
	Größte charakteristische Wurzel nach Roy	,002	,022 <sup>b</sup>	2,000	25,000	,978	,002
Fachwissen_PD * schule	Pillai-Spur	,158	2,340 <sup>b</sup>	2,000	25,000	,117	,158
	Wilks-Lambda	,842	2,340 <sup>b</sup>	2,000	25,000	,117	,158
	Hotelling-Spur	,187	2,340 <sup>b</sup>	2,000	25,000	,117	,158
	Größte charakteristische Wurzel nach Roy	,187	2,340 <sup>b</sup>	2,000	25,000	,117	,158

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: Fachwissen\_PD

b. Exakte Statistik

## Anhang 53: Mauchly-Test auf Sphärizität (PD\_fw)

*Mauchly-Test auf Sphärizität<sup>a</sup>*

Maß: MASS_1		Epsilon <sup>b</sup>					
Innersubjekteffekt	Mauchly-W	Ungefähres Chi-Quadrat	df	Sig.	Greenhouse -Geisser	Huynh-Feldt (HF)	Untergrenze
Fachwissen_PD	,979	,533	2	,766	,979	1,000	,500

Prüft die Nullhypothese, dass sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: Fachwissen\_PD

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

## Anhang 54: Paarweise Vergleiche (PD\_fw)

*Paarweise Vergleiche*

Maß: MASS_1		95% Konfidenzintervall für Differenz <sup>b</sup>					
(I) Fachwissen_PD	(J) Fachwissen_PD	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	Untergrenze	Obergrenze	
1 pr_fw	2 po_fw	-,291	,243	,725	-,911	,330	
	3 fo_fw	1,005 <sup>*</sup>	,263	,002	,333	1,677	
2 po_fw	1 pr_fw	,291	,243	,725	-,330	,911	
	3 fo_fw	1,296 <sup>*</sup>	,275	<,001	,591	2,000	
3 fo_fw	1 pr_fw	-1,005 <sup>*</sup>	,263	,002	-1,677	-,333	
	2 po_fw	-1,296 <sup>*</sup>	,275	<,001	-2,000	-,591	

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

**Homeschooling – DD – Fachwissen**

## Anhang 55: Multivariate Tests (DD\_fw)

*Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
Fachwissen_DD	Pillai-Spur	,214	3,539 <sup>b</sup>	2,000	26,000	,044	,214
	Wilks-Lambda	,786	3,539 <sup>b</sup>	2,000	26,000	,044	,214
	Hotelling-Spur	,272	3,539 <sup>b</sup>	2,000	26,000	,044	,214
	Größte charakteristische Wurzel nach Roy	,272	3,539 <sup>b</sup>	2,000	26,000	,044	,214
Fachwissen_DD * pr_KFT	Pillai-Spur	,181	2,879 <sup>b</sup>	2,000	26,000	,074	,181
	Wilks-Lambda	,819	2,879 <sup>b</sup>	2,000	26,000	,074	,181
	Hotelling-Spur	,221	2,879 <sup>b</sup>	2,000	26,000	,074	,181
	Größte charakteristische Wurzel nach Roy	,221	2,879 <sup>b</sup>	2,000	26,000	,074	,181
Fachwissen_DD * schule	Pillai-Spur	,043	,577 <sup>b</sup>	2,000	26,000	,568	,043
	Wilks-Lambda	,957	,577 <sup>b</sup>	2,000	26,000	,568	,043
	Hotelling-Spur	,044	,577 <sup>b</sup>	2,000	26,000	,568	,043
	Größte charakteristische Wurzel nach Roy	,044	,577 <sup>b</sup>	2,000	26,000	,568	,043

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: Fachwissen\_DD

b. Exakte Statistik

## Anhang 56: Mauchly Test auf Sphärizität (DD\_fw)

*Mauchly-Test auf Sphärizität<sup>a</sup>*

Maß: MASS_1		Epsilon <sup>b</sup>					
Innersubjekteffekt	Mauchly-W	Ungefähres Chi-Quadrat	df	Sig.	Greenhouse -Geisser	Huynh-Feldt (HF)	Untergrenze
Fachwissen_DD	,882	3,273	2	,195	,894	1,000	,500

Prüft die Nullhypothese, dass sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: Fachwissen\_DD

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

## Anhang 57: Paarweise Vergleiche (DD\_fw)

*Paarweise Vergleiche*

Maß: MASS_1		95% Konfidenzintervall für Differenz <sup>b</sup>				
(I) Fachwissen_DD	(J) Fachwissen_DD	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	Untergrenze	Obergrenze
1 pr_fw	2 po_fw	-,865 <sup>*</sup>	,258	,007	-1,524	-,206
	3 fo_fw	,239	,194	,887	-,257	,736
2 po_fw	1 pr_fw	,865 <sup>*</sup>	,258	,007	,206	1,524
	3 fo_fw	1,104 <sup>*</sup>	,214	<,001	,559	1,649
3 fo_fw	1 pr_fw	-,239	,194	,887	-,736	,257
	2 po_fw	-1,104 <sup>*</sup>	,214	<,001	-1,649	-,559

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

## 11.4 Ergebnisse: Naturwissenschaftliche Arbeitsweise

### Datentabellen

#### 11.4.1 Modellprüfung

Anhang 58: Shapiro-Wilk-Test auf Normalverteilung (nwa)

*Tests auf Normalverteilung*

	Gruppeneinteilung	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
po_nwa	PP - Präsenzunterricht	,217	32	<,001	,922	32	,024
	DP - FlippedClassroom	,175	27	,033	,945	27	,159
	PD - Wechselunterricht	,149	29	,100	,957	29	,272
	DD - Homeschooling	,169	30	,028	,922	30	,031
fo_nwa	PP - Präsenzunterricht	,180	32	,010	,945	32	,106
	DP - FlippedClassroom	,182	27	,023	,943	27	,147
	PD - Wechselunterricht	,180	29	,017	,874	29	,002
	DD - Homeschooling	,206	30	,002	,921	30	,029

a. Signifikanzkorrektur nach Lilliefors

Anhang 59: Levene-Test auf Homogenität (nwa)

*Test auf Homogenität der Varianz*

		Levene-Statistik	df1	df2	Signifikanz
po_nwa	Basiert auf dem Mittelwert	3,671	3	114	,014
	Basiert auf dem Median	3,833	3	114	,012
	Basierend auf dem Median und mit angepassten df	3,833	3	91,040	,012
	Basiert auf dem getrimmten Mittel	3,746	3	114	,013
fo_nwa	Basiert auf dem Mittelwert	4,620	3	114	,004
	Basiert auf dem Median	3,722	3	114	,013
	Basierend auf dem Median und mit angepassten df	3,722	3	99,989	,014
	Basiert auf dem getrimmten Mittel	4,877	3	114	,003

## 11.4.2 Vergleich der Gruppen

### ***Post-Test Naturwissenschaftliche Arbeitsweise***

Anhang 60: Tests der Zwischensubjekteffekte (po\_nwa)

#### *Tests der Zwischensubjekteffekte*

Abhängige Variable: po_nwa						
Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	28,728 <sup>a</sup>	5	5,746	4,055	,002	,120
Konstanter Term	5,156	1	5,156	3,638	,058	,024
pr_KFT	9,209	1	9,209	6,499	,012	,042
schule	2,514	1	2,514	1,774	,185	,012
gruppe_4	4,776	3	1,592	1,123	,342	,022
Fehler	209,731	148	1,417			
Gesamt	312,984	154				
Korrigierte Gesamtvariation	238,459	153				

a. R-Quadrat = ,120 (korrigiertes R-Quadrat = ,091)

Anhang 61: Paarweise Vergleiche (po\_nwa)

#### *Paarweise Vergleiche*

Abhängige Variable: po_nwa						
(I) Gruppeneinteilung	(J) Gruppeneinteilung	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>a</sup>	95% Konfidenzintervall für Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
PP - Präsenzunterricht	DP - FlippedClassroom	,016	,560	1,000	-1,481	1,512
	PD - Wechselunterricht	,514	,444	1,000	-,674	1,702
	DD - Homeschooling	,406	,488	1,000	-,901	1,712
DP - FlippedClassroom	PP - Präsenzunterricht	-,016	,560	1,000	-1,512	1,481
	PD - Wechselunterricht	,498	,519	1,000	-,890	1,887
	DD - Homeschooling	,390	,557	1,000	-1,101	1,881
PD - Wechselunterricht	PP - Präsenzunterricht	-,514	,444	1,000	-1,702	,674
	DP - FlippedClassroom	-,498	,519	1,000	-1,887	,890
	DD - Homeschooling	-,109	,289	1,000	-,883	,666
DD - Homeschooling	PP - Präsenzunterricht	-,406	,488	1,000	-1,712	,901
	DP - FlippedClassroom	-,390	,557	1,000	-1,881	1,101
	PD - Wechselunterricht	,109	,289	1,000	-,666	,883

Basiert auf geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

## Follow-up-Test Naturwissenschaftliche Arbeitsweise

### Anhang 62: Tests der Zwischensubjekteffekte (fo\_nwa)

#### Tests der Zwischensubjekteffekte

Abhängige Variable: fo_nwa						
Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	39,479 <sup>a</sup>	5	7,896	5,550	<,001	,199
Konstanter Term	9,340	1	9,340	6,565	,012	,055
pr_KFT	1,812	1	1,812	1,133	,289	,010
schule	,130	1	,130	,092	,763	,001
gruppe_4	17,862	3	5,887	4,138	,008	,100
Fehler	159,333	112	1,423			
Gesamt	489,740	118				
Korrigierte Gesamtvariation	198,813	117				

a. R-Quadrat = ,199 (korrigiertes R-Quadrat = ,163)

### Anhang 63: Paarweise Vergleiche (fo\_nwa)

#### Paarweise Vergleiche

Abhängige Variable: fo_nwa						
(I) Gruppeneinteilung	(J) Gruppeneinteilung	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>a</sup>	95% Konfidenzintervall für Differenz <sup>a</sup>	
					Untergrenze	Obergrenze
PP - Präsenzunterricht	DP - FlippedClassroom	,218	,723	1,000	-1,725	2,160
	PD - Wechselunterricht	1,250	,561	,168	-,259	2,758
	DD - Homeschooling	1,439	,609	,119	-,197	3,075
DP - FlippedClassroom	PP - Präsenzunterricht	-,218	,723	1,000	-2,160	1,725
	PD - Wechselunterricht	1,032	,681	,795	-,798	2,862
	DD - Homeschooling	1,221	,721	,558	-,715	3,158
PD - Wechselunterricht	PP - Präsenzunterricht	-1,250	,561	,168	-2,758	,259
	DP - FlippedClassroom	-1,032	,681	,795	-2,862	,798
	DD - Homeschooling	,189	,350	1,000	-,751	1,130
DD - Homeschooling	PP - Präsenzunterricht	-1,439	,609	,119	-3,075	,197
	DP - FlippedClassroom	-1,221	,721	,558	-3,158	,715
	PD - Wechselunterricht	-,189	,350	1,000	-1,130	,751

Basiert auf geschätzten Randmitteln

a. Anpassung für Mehrfachvergleiche: Bonferroni.

### 11.4.3 Vergleich der Messzeitpunkte

#### **Präsenzunterricht – PP – Naturwissenschaftliche Arbeitsweise**

##### Anhang 64: Multivariate Tests (PP\_nwa)

###### *Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
NWA_PP	Pillai-Spur	,012	,364 <sup>b</sup>	1,000	30,000	,551	,012
	Wilks-Lambda	,988	,364 <sup>b</sup>	1,000	30,000	,551	,012
	Hotelling-Spur	,012	,364 <sup>b</sup>	1,000	30,000	,551	,012
	Größte charakteristische Wurzel nach Roy	,012	,364 <sup>b</sup>	1,000	30,000	,551	,012
NWA_PP * pr_KFT	Pillai-Spur	,000	,015 <sup>b</sup>	1,000	30,000	,905	,000
	Wilks-Lambda	1,000	,015 <sup>b</sup>	1,000	30,000	,905	,000
	Hotelling-Spur	,000	,015 <sup>b</sup>	1,000	30,000	,905	,000
	Größte charakteristische Wurzel nach Roy	,000	,015 <sup>b</sup>	1,000	30,000	,905	,000

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: NWA\_PP

b. Exakte Statistik

##### Anhang 65: Paarweise Vergleiche (PP\_nwa)

###### *Paarweise Vergleiche*

Maß: MASS_1		95% Konfidenzintervall für Differenz <sup>b</sup>				
(I) NWA_PP	(J) NWA_PP	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	Untergrenze	Obergrenze
1 po_nwa	2 fo_nwa	,544 <sup>*</sup>	,152	,001	,233	,856
2 fo_nwa	1 po_nwa	-,544 <sup>*</sup>	,152	,001	-,856	-,233

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

**Flipped Classroom – DP – Naturwissenschaftliche Arbeitsweise**

## Anhang 66: Multivariate Tests (DP\_nwa)

*Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
NWA_DP	Pillai-Spur	,006	,162 <sup>b</sup>	1,000	25,000	,691	,006
	Wilks-Lambda	,994	,162 <sup>b</sup>	1,000	25,000	,691	,006
	Hotelling-Spur	,006	,162 <sup>b</sup>	1,000	25,000	,691	,006
	Größte charakteristische Wurzel nach Roy	,006	,162 <sup>b</sup>	1,000	25,000	,691	,006
NWA_DP * pr_KFT	Pillai-Spur	,000	,001 <sup>b</sup>	1,000	25,000	,971	,000
	Wilks-Lambda	1,000	,001 <sup>b</sup>	1,000	25,000	,971	,000
	Hotelling-Spur	,000	,001 <sup>b</sup>	1,000	25,000	,971	,000
	Größte charakteristische Wurzel nach Roy	,000	,001 <sup>b</sup>	1,000	25,000	,971	,000

a. Design: Konstanter Term + pr\_KFT  
Innersubjektdesign: NWA\_DP

b. Exakte Statistik

## Anhang 67: Paarweise Vergleiche (DP\_nwa)

*Paarweise Vergleiche*

Maß: MASS\_1

(I) NWA_DP	(J) NWA_DP	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	95% Konfidenzintervall für Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
1 po_nwa	2 fo_nwa	,493 <sup>*</sup>	,234	,045	,011	,974
2 fo_nwa	1 po_nwa	-,493 <sup>*</sup>	,234	,045	-,974	-,011

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

**Wechselunterricht – PD – Naturwissenschaftliche Arbeitsweise**

## Anhang 68: Multivariate Tests (PD\_nwa)

*Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
NWA_PD	Pillai-Spur	,001	,023 <sup>b</sup>	1,000	26,000	,880	,001
	Wilks-Lambda	,999	,023 <sup>b</sup>	1,000	26,000	,880	,001
	Hotelling-Spur	,001	,023 <sup>b</sup>	1,000	26,000	,880	,001
	Größte charakteristische Wurzel nach Roy	,001	,023 <sup>b</sup>	1,000	26,000	,880	,001
NWA_PD * pr_KFT	Pillai-Spur	,048	1,261 <sup>b</sup>	1,000	26,000	,272	,048
	Wilks-Lambda	,954	1,261 <sup>b</sup>	1,000	26,000	,272	,048
	Hotelling-Spur	,048	1,261 <sup>b</sup>	1,000	26,000	,272	,048
	Größte charakteristische Wurzel nach Roy	,048	1,261 <sup>b</sup>	1,000	26,000	,272	,048
NWA_PD * schule	Pillai-Spur	,003	,076 <sup>b</sup>	1,000	26,000	,786	,003
	Wilks-Lambda	,997	,076 <sup>b</sup>	1,000	26,000	,786	,003
	Hotelling-Spur	,003	,076 <sup>b</sup>	1,000	26,000	,786	,003
	Größte charakteristische Wurzel nach Roy	,003	,076 <sup>b</sup>	1,000	26,000	,786	,003

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: NWA\_PD

b. Exakte Statistik

## Anhang 69: Paarweise Vergleiche (PD\_nwa)

*Paarweise Vergleiche*

Maß: MASS_1						
(I) NWA_PD	(J) NWA_PD	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	95% Konfidenzintervall für Differenz <sup>b</sup>	
					Untergrenze	Obergrenze
1 po_nwa	2 fo_nwa	,785 <sup>*</sup>	,267	,007	,236	1,334
2 fo_nwa	1 po_nwa	-,785 <sup>*</sup>	,267	,007	-1,334	-,236

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

**Homeschooling – DD – Naturwissenschaftliche Arbeitsweise**

## Anhang 70: Multivariate Tests (DD\_nwa)

*Multivariate Tests<sup>a</sup>*

Effekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat
NWA_DD	Pillai-Spur	,146	4,629 <sup>b</sup>	1,000	27,000	,041	,146
	Wilks-Lambda	,854	4,629 <sup>b</sup>	1,000	27,000	,041	,146
	Hotelling-Spur	,171	4,629 <sup>b</sup>	1,000	27,000	,041	,146
	Größte charakteristische Wurzel nach Roy	,171	4,629 <sup>b</sup>	1,000	27,000	,041	,146
NWA_DD * pr_KFT	Pillai-Spur	,000	,004 <sup>b</sup>	1,000	27,000	,948	,000
	Wilks-Lambda	1,000	,004 <sup>b</sup>	1,000	27,000	,948	,000
	Hotelling-Spur	,000	,004 <sup>b</sup>	1,000	27,000	,948	,000
	Größte charakteristische Wurzel nach Roy	,000	,004 <sup>b</sup>	1,000	27,000	,948	,000
NWA_DD * schule	Pillai-Spur	,082	2,420 <sup>b</sup>	1,000	27,000	,131	,082
	Wilks-Lambda	,918	2,420 <sup>b</sup>	1,000	27,000	,131	,082
	Hotelling-Spur	,090	2,420 <sup>b</sup>	1,000	27,000	,131	,082
	Größte charakteristische Wurzel nach Roy	,090	2,420 <sup>b</sup>	1,000	27,000	,131	,082

a. Design: Konstanter Term + pr\_KFT + schule  
Innersubjektdesign: NWA\_DD

b. Exakte Statistik

## Anhang 71: Paarweise Vergleiche (DD\_nwa)

*Paarweise Vergleiche*

Maß: MASS_1		95% Konfidenzintervall für Differenz <sup>b</sup>					
(I) NWA_DD	(J) NWA_DD	Mittelwertdif ferenz (I-J)	Std.-Fehler	Sig. <sup>b</sup>	Untergrenze	Obergrenze	
1 po_nwa	2 fo_nwa	1,438 <sup>*</sup>	,224	<,001	,977	1,898	
2 fo_nwa	1 po_nwa	-1,438 <sup>*</sup>	,224	<,001	-1,898	-,977	

Basiert auf geschätzten Randmitteln

\*. Die Mittelwertdifferenz ist in Stufe ,05 signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

## 11.5 Aktivität im wissenschaftlichen Umfeld

Stich, J. & Nerdel, C. (09.2023). Vermittlung von Biotechnologie im Unterricht – Verbesserung durch digitale Lernumgebungen? Vortrag. Jahrestagung der Fachsektion Didaktik der Biologie (FDdB), Ludwigsburg.

Stich, J. & Nerdel, C. (08.2023). Biotechnology at school - Digital or analog learning environment? Vortrag. ECER 2023, Glasgow. – keine Konferenzteilnahme

Stich, J. & Nerdel, C. (03.2022). Biotechnologie an der Schule - Digitale oder analoge Lernumgebung? Vortrag. 23. Internationale Frühjahrsschule der Fachsektion Didaktik der Biologie im VBiO, Bielefeld.

Stich, J. & Nerdel, C. (08.2022). Biotechnology at school - Digital or analog learning environment? Vortrag. ECER 2023, Nicosia - Zypern. – keine Konferenzteilnahme

Stich, J. & Nerdel, C. (09.2021). Biotechnologie an der Schule - Digitale oder analoge Lernumgebung. Posterpräsentation. Jahrestagung der Fachsektion Didaktik der Biologie (FDdB), Bonn.

Stich, J. & Nerdel, C. (02.2020). Kompetenzentwicklung von Schüler\*innen in Fachwissen und Kommunikation im Biotechnologischen Schülerlabor. Posterpräsentation. 22. Internationale Frühjahrsschule der Fachsektion Didaktik der Biologie im VBiO, Halle.

Stich, J., Kriehuber, E., Brück, T. & Nerdel, C. (09.2019). Entwicklung von Fachwissen und Methodenkompetenz im biotechnologischen Schülerlabor. Posterpräsentation. Jahrestagung der Fachsektion Didaktik der Biologie (FDdB), Wien.

## Danksagung

An diesem Punkt möchte ich die Gelegenheit ergreifen, um all jenen zu danken, die mich auf dem Weg zu dieser Dissertation begleitet und beigestanden haben.

Mein spezieller Dank richtet sich an meine Erstbetreuerin, Frau Prof. Dr. Claudia Nerdel. Ihre konstruktive Kritik und kontinuierliche Unterstützung während der Ausarbeitung dieser Arbeit. Ihr Engagement war entscheidend für die Vollendung dieses Werkes. Ich danke Prof. Dr. Steffen Schaal für die Zweitbetreuung dieser Arbeit.

Meine Dankbarkeit gilt auch meinen Kolleginnen und Kollegen der FDLS-Arbeitsgruppe. Die Zusammenarbeit, die gemeinsamen Mittagessen und die positive Teamatmosphäre sowohl im Büro als auch außerhalb davon waren eine erhebliche Bereicherung. Des Weiteren bin ich dankbar für die stetige Unterstützung durch Sara Großbruchhaus, die maßgeblich zur Entstehung dieser Dissertation, vor allem in der Coronazeit, beigetragen hat. Mein Dank geht auch an Marie Hornberger, Benjamin Stöger und Patrizia Weidenhiller für ihre beständige und geduldige Unterstützung bei statistischen Herausforderungen. Zudem möchte ich Frau Dr. Patricia Schöppner und Frau Dr. Dagmar Traub danken, für ihr stets offenes Ohr und die damit einhergehende konstruktive Unterstützung.

Ein Dank geht an die Schulleitung des Werner-Heisenberg-Gymnasiums Garching, Herrn OStR Armin Eifertinger und Frau OStRin Christine McAuliffe, deren Unterstützung, auch in herausfordernden Zeiten, die Durchführung meiner Promotion wesentlich erleichtert hat. Ich danke der Biologie-Chemie-Fachschaft des Werner-Heisenberg-Gymnasiums für ihre stete Unterstützung sowohl im privaten als auch im schulischen Bereich.

Zum Schluss möchte ich meiner Familie und meinen Freund\*innen für ihre konstante Unterstützung und Geduld in den vergangenen Jahren danken.

Danke!