

Special Issue

Tasks in Science Education

Research-Based Report of Practice

Wie Schüler:innen Modelle als erdachte Welt begreifen lernen – Eine Vergleichsstudie

Raffael Bachmann¹, Alexis Voutopoulos,² Pascal Amrein³, Katrin Bölsterli Bardy⁴

Received: February 2021 / Accepted: July 2022

Structured Abstract

Hintergrund: Die Kompetenz im Umgang mit Modellen ist für den Naturwissenschaftsunterricht zentral. Bei einer unachtsamen Herangehensweise können Lernendenvorstellungen aufgebaut und automatisiert werden, die einem adäquaten Modellverständnis entgegenwirken. Die gängigsten Lernendenvorstellungen entstehen durch den Transfer von lebensweltlichen Eigenschaften von Stoffen in die Welt der Atome. Des Weiteren ist es wichtig, das Bewusstsein zu fördern, dass ein Modell immer eine Interpretation von Beobachtungen und Messungen ist und somit ein Denkmodell. Obwohl bereits viel in diesem Bereich geforscht wurde, bestehen noch einige Forschungslücken auf der Ebene der Wirksamkeit von einzelnen Aufgaben zum Aufbau eines adäquaten Modellverständnisses.

Absicht: Um diese Forschungslücke zu schmälern, wurden in der vorliegenden Masterarbeit im Sinne einer Pilotstudie zwei konkrete Aufgaben zur Einführung eines möglichst adäquaten Modellverständnisses in Form einer Interventionsstudie ($N=270$) untersucht und miteinander verglichen.

Stichprobe: An der explorativen Interventionsstudie haben elf Zentralschweizer Lehrpersonen der Sekundarstufe mit insgesamt 16 Klassen und 270 Schüler:innen teilgenommen.

Forschungsdesign und Methode: Für den Vergleich der beiden Aufgaben wurden zwei Interventionsgruppen (Blackbox-Interventionsgruppe und Spuren-Interventionsgruppe) sowie eine Kontrollgruppe gebildet. Die Interventionsstudie wurde nach dem Prä-, Post-, Follow-Up-Design durchgeführt.

Resultate: Die Ergebnisse der vorliegenden Untersuchung weisen darauf hin, dass der Lernzuwachs bei den Schüler:innen durch die Blackbox (BB)- Intervention höher ist als durch die Spuren (SP)- Intervention ($p < .001$, $d > 0.81$).

Schlussfolgerung: Die explorative Studie deutet mit ersten Resultaten darauf hin, dass der BB-Ansatz im Unterricht vielversprechend ist. Jedoch führt auch der SP-Ansatz zu einem Lernzuwachs. Zukünftige Forschung könnte in Form einer breit angelegten mehrebenen-analytischen Studie dieser Frage vertieft nachgehen und weitere Fragen, wie z.B. Genderaspekte, klären.

Schlüsselwörter: *Blackbox, Spuren, Tricky Tracks, Denkmodell, Science Education, Nature of Science, de-natured Science, Atommodell*

¹Schule Menzingen, ²Sportschule Kriens, ³Schulhaus Röhrlberg 1, ⁴Pädagogische Hochschule Luzern
✉ katrin.boelsterli@phlu.ch

1 Einleitung

„Models are integral to thinking and working scientifically and it can be argued that science and its explanatory models are inseparable because models are science’s products, methods and its major learning and teaching tools“ (Gilbert, 1991, zitiert nach Harrison & Treagust, 2000, S. 1011).

Wie im einleitenden Zitat von Gilbert aufgezeigt wird, sind Modelle und deren Verständnis ein zentraler Aspekt der naturwissenschaftlichen Bildung (Löffler, 1996). Die Grundlage für wissenschaftliche Erkenntnisse in der Chemie stützt sich weniger auf Gesetze oder Prinzipien, sondern auf Modellvorstellungen, in denen diese Gültigkeit erlangen (Meisert, 2008). In der Chemie ist das Atommodell eines der bedeutendsten Beispiel einer Modellvorstellung.

Beim Lehren und Lernen des atomaren Aufbaus kommt es zu einer Vielzahl an Lernendenvorstellungen (Griffiths & Preston, 1992; Horton, 2009; Kind, 2004). Ein Grossteil der Lernendenvorstellungen entsteht durch einen unreflektierten Umgang mit Modellen im Unterricht oder durch Schulbücher (Horton, 2009; Mikelskis-Seifert & Leisner, o. A.). Um dieser Problematik entgegenzuwirken, wird seit vielen Jahren in der Fachdidaktik geforscht, wie einzelne Modelle und ein adäquates Modellverständnis aufgebaut werden können, also die grundlegende Beschaffenheit von Modellen und deren Konstruktion (Henze et al., 2007; Horton, 2009; Lijnse et al., 1990). Ziel ist es, dass die Schüler:innen Atommodelle als erdachte Welt zur Realität verstehen und akzeptieren lernen (Mikelskis-Seifert & Leisner, o. A.).

Es wurde eine Vielzahl didaktischer Ansätze zum Aufbau eines adäquaten Modellverständnisses erarbeitet, beispielsweise von Horton (2009), Mikelskis-Seifert und Fischler (2003), Lederman und Abd-El-Khalick (2002), Rehm (2009), und Dubach et al. (2022). Von Lederman und Abd-El-Khalick (2002) werden zwei mögliche Herangehensweisen genannt, wie das Verständnis von Atommodellen im Unterricht aufgebaut werden kann. Die zwei Ansätze werden im Folgenden als „Blackbox“ (BB) und „Spuren“ (SP) bezeichnet.

Empirische Daten zur Wirksamkeit der BB und SP Ansätze gibt es jedoch keine. Aus diesem Grund vergleicht die vorliegende Masterarbeit diese zwei Aufgaben von Lederman und Abd-El-Khalick (2002), um das erlangte Modellverständnis der Schüler:innen der Sekundarstufe I durch die zwei Ansätze miteinander zu vergleichen.

2 Theoretischer Hintergrund

2.1 Modelle im Chemieunterricht

Die Chemie sieht sich mit der Schwierigkeit konfrontiert, dass makroskopische Phänomene nicht auf der sichtbaren Ebene erklärt werden können, weil der Aufbau der Atome für niemanden sichtbar ist (Talanquer, 2011). Um sich dennoch den Aufbau der Atome vorstellen zu können, und Erklärungsversuche für Reaktionen machen zu können, entwickelten Chemikerinnen und Chemiker unterschiedliche Modelle, wie das Teilchenmodell, das Bohrmodell oder das Orbitalmodell. All diese Modelle sind Denkmodelle. Modelle der sichtbaren Welt zeichnen sich nach Stachowiak (1973, S. 131ff.) durch Pragmatik, Abbildung und Verkürzung aus. Unter Denkmodellen werden Modelle verstanden, die zusätzlich etwas abbilden, was für niemanden sichtbar ist (Bölsterli Bardy et al., 2021, S. 97). Anders als Anschauungsmodelle, die konkret sind, gehören Denkmodelle zu den mentalen Modellen (Tempel et al., 2018, S. 2). Sie sind laut hypothetischem Realismus auch nicht richtig oder falsch, sondern zweckmässig oder unzweckmässig (Leisner-Bodenthin, 2006, S. 96).

Denkmodelle sind in der Chemie somit ein wichtiges Hilfsmittel, um makroskopische Phänomene auf der submikroskopischen Ebene zu erklären (Talanquer, 2011). Oder anders ausgedrückt: „Denkmodelle helfen, uns Phänomene vorzustellen, die sich nur schwer beobachten und beschreiben lassen.“ (Beerenwinkel et al., 2019, S. 34). Mit Hilfe von Denkmodellen können z.B. Aggregatzustandswechsel oder Molekül-Geometrien erklärt werden (Bindernagel & Eilks, 2008).

2.2 Modellkompetenz im Chemieunterricht

Im Lehrplan 21 (D-EDK, 2013) spielt in der Chemie die Modellkompetenz, beispielsweise in NT3.2, eine grosse Rolle. Beim Aufbau dieser Modellkompetenz sollte, laut Passmore (2015, S. 663), ein tiefgreifendes Modellverständnis entwickelt werden. Ein angemessenes Modellverständnis wird von Leisner-Bodenthin (2006, S. 93f.), unter Einbezug der Arbeiten von Carey et al. (1989) und Driver et al. (1996), zum Wissenschaftsverständnis; und von Justi & Gilbert (2003) und Grosslight et al. (1991) zum Modellverständnis wie folgt definiert:

Die Schüler:innen wissen, dass

- physikalische Modelle vom Menschen geschaffen werden, - wenn die Grenzen der direkten Wahrnehmung erreicht sind - um (in ihrer Gänze) nicht beobachtbare Mechanismen/Objekte zu erklären, vorherzusagen und zu veranschaulichen.
- zur Modellentwicklung Spekulation, Intuition, Annahmen und Abstraktionen notwendig sind.
- Modelle zweckmässig sind und nicht richtig oder falsch.
- physikalische Modelle hypothetisch und vorläufig sind.
- Modelle sich in der Community durchsetzen müssen (Leisner-Bodenthin, 2006, S. 93f.)

Meisert (2008, S. 244f.) gliedert die Modellkompetenz für die Biologiedidaktik in folgende drei Bereiche: „Modellwissen“, d.h. der Kenntnis über Modelle, „Modellarbeit“, der Fähigkeit Modelle zu nutzen, und einem übergeordneten „Modellverständnis“, d.h. dem Verstehen um die Bedeutung von Modellen und der Modellarbeit im Sinne von Nature of Science.

Krell et al. (2015, S. 370) definieren für alle drei Unterrichtsfächer, Biologie, Chemie und Physik, dieselben fünf Aspekte der Modellkompetenz wie Upmeier zu Belzen und Krüger (2010) für die Biologiedidaktik: Eigenschaften von Modellen, alternative Modelle, Zweck von Modellen, Testen von Modellen und Ändern von Modellen. In einer Tabelle vergleichen Krell et al. (2015, S. 371) diese fünf Aspekte mit der Definition der Modellkompetenz anderer Autoren, wie Grosslight et al. (1991); Treagust et al. (2002); Crawford and Cullin (2005); Schwarz and White (2005); Schwarz et al. (2009); Sins et al. (2009). Krüger et al. (2018, S. 149) definieren zu den fünf genannten Aspekten der Modellkompetenz (Upmeier zu Belzen & Krüger, 2010) jeweils drei Niveaustufen, und erstellen so ein empirisch abgestütztes Kompetenzmodell zur Überprüfung der Modellkompetenz von Schüler:innen sowie Studierenden. Bezüglich des Modellverständnisses erstellten Chittleborough et al. (2005, S. 197) einen theoretischen Rahmen, und beschreiben darin, dass das wissenschaftliche Modell und das Lehrmodell einen Einfluss auf das mentale Modell der Lernenden beim Modelllernen und Modellverstehen haben, und sich daraus das effektiv ausgedrückte Modell (expressed model) ergibt.

2.3 Modellkompetenz der Schüler:innen

Millar (1990) beklagte bereits vor 30 Jahren, dass die mentalen Modelle von Kindern sich deutlich von einem wissenschaftlich akzeptierten Modellverständnis unterscheiden. Mikelskis-Seifert und Fischler (2003) haben sich in ihren Untersuchungen mit den Problemen beim Lehren und Lernen der Teilchenstruktur befasst und bestätigen Millar (1990) darin, dass Schüler:innen oftmals erhebliche Schwierigkeiten beim Aufbau von einem angemessenen Verständnis der Teilchenstruktur haben. Auch Begrifflichkeiten wie „Stoff“ und „Atom“ würden häufig nicht trennscharf verwendet (Horton, 2009). Die Erkenntnisse aus der Literaturstudie von Kind (2004, S. 9) deuten darauf hin, dass die Vorstellungen von Teilchen nur unzureichend erfasst werden, da selbst bei Aufforderung etwa 25 % der Schüler:innen gemischten Alters in ihren Antworten nur kontinuierliche Vorstellungen von Materie verwendeten.

Obwohl Chittleborough et al. (2005) feststellen, dass das Modellverständnis im Verlaufe der Schulzeit besser wird, beklagen Ahtee und Varjoli (1998), dass bis zum Ende der Sekundarschule viele Lernendenvorstellungen vorhanden seien und selbst Universitätsstudierende Lernendenvorstellungen aufweisen würden. Zu einem ähnlichen Schluss kamen Mulford und Robinson (2002), die aus diesem Grund sogenannte „Chemical Concepts Inventorys (CCI)“ für erstsemestrige Chemiestudierende zur Erhebung ihrer Lernendenvorstellungen erstellt haben.

Eine erst vor kurzem durchgeführte Studie zum Thema Modelle in allen drei naturwissenschaftlichen Fächern (Chemie, Biologie und Physik) konnte zeigen, dass die Schüler:innen der Sekundarstufe I Atommodelle kennen, und gewisse Lernende von sich aus explizit erwähnen, dass diese Modelle Annahmen über die Struktur der Atome darstellen (Krell et al., 2015, S. 381). Leider ist es jedoch immer noch die Minderheit der Schüler:innen.

2.3.1 Lernendenvorstellungen

In den letzten Jahrzehnten wurde viel über Lernendenvorstellungen zum Modellverständnis bei Schüler:innen und bei Studierenden geforscht (Griffiths & Preston, 1992; Horton, 2009; Kind, 2004). Mikelskis-Seifert und Fischler (2003, S. 76) haben die Lernendenvorstellungen zum Thema Teilchenmodell thematisch in drei Gruppen zusammengefasst: So nehmen Schüler:innen häufig an, erstens, dass Atome, wie jeder andere Körper, eine Farbe oder eine bestimmte Temperatur haben; zweitens, dass der Raum zwischen den Teilchen im Teilchenmodell aus Luft besteht, und drittens, dass die sich ständig bewegenden Teilchen einmal wegen Reibung zur Ruhe kommen.

Kind (2004) teilt in ihrem Review die Lernendenvorstellungen zum Thema Teilchen neben den drei Kategorien von Mikelskis-Seifert und Fischler (2003) in eine vierte Kategorie ein (Kind, 2004, S. 13), nämlich, dass bei gasförmigen Teilchen Bindungen/Kräfte zwischen den Teilchen wirken.

Griffiths und Preston (1992) untersuchten die Lernendenvorstellungen des 12. Schuljahres. Sie konnten insgesamt 52 Lernendenvorstellungen ausmachen. Diese gruppieren sie in 11 Kategorien. Sechs Kategorien beziehen sich auf die Struktur, Aufbau, Grösse, Form, Masse, Bindung und Energie der Moleküle, fünf Kategorien beziehen sich auf die Struktur, Form, Grösse, Masse und animistischen Wahrnehmungen von Atomen.

Einen detaillierten Überblick über mögliche Lernendenvorstellungen zum Thema Atommodelle hat Horton (2009) in einem Review-Artikel zusammengestellt. Er unterscheidet dabei auch zwischen einer Liste relevanter alternativer Lernendenvorstellungen aus Sicht der Forscherinnen und Forscher (Horton, 2009, S. 20ff.) und einer zweiten Liste wichtiger Lernendenvorstellungen aus Sicht der Lehrpersonen (Horton, 2009, S. 24ff.). Wie sich im nächsten Abschnitt zeigen wird, sind diese Lernendenvorstellungen grösstenteils auf ein fehlendes Modellverständnis zurückzuführen.

2.3.2 Herkunft und Stabilität der mangelhaften Modellkompetenz

Die Herkunft der Lernendenvorstellungen von Schüler:innen und Studierenden ist mannigfaltig. Im Folgenden werden einige häufige mögliche Herkünfte genannt:

- Viele Schulbücher fördern unbeabsichtigt neue Lernendenvorstellungen (Horton, 2009, S. 6; Mikelskis-Seifert, 2002)
- Lehrpersonen und deren Unterricht fördern häufig unbeabsichtigt neue Lernendenvorstellungen (Horton, 2009, S. 6; Rehm & Buck, 2009), beispielsweise, indem nur ungenügend zwischen den drei Dimensionen des chemischen Dreiecks unterschieden wird (Johnstone, 1991) (Kapitel 2.5)
- Das Kugel-Stab-Modell fördert die Vorstellung, dass Atome dieselbe Farbe haben, wie der Reinstoff, indem eine gelbe Kugel ein Schwefelatom und eine schwarze Kugel ein Kohlenstoffatom darstellt.
- Periodensysteme, welche bei den Atomen den Reinstoff als Foto abbilden, fördern die Vorstellung, dass das Atom und der dazugehörige Reinstoff dieselben Eigenschaften aufweisen.
- Viele Lernendenvorstellungen werden durch die Schüler:innen selbst konstruiert, indem die Schüler:innen das neu Gelernte mit Alltagserfahrungen zu verknüpfen versuchen.
- Viele Schüler:innen im Alter von 15 Jahren und älter wenden sensorisches Denken über die Materie an, obwohl sie in anderen Bereichen, wie z. B. der Mathematik, im logischen Denken weit fortgeschritten sind. Dies vermutlich, weil die sensorische Erfahrung in den Fällen, in denen die Materie nicht sichtbar ist, dominiert (Kind, 2004, S. 8).
- Fernsehsendungen und Bücher werden von den Schüler:innen selbst neben der Schule als Quelle ihres Wissens über Modelle genannt (Johnstone, 1990b, zitiert nach Rehm & Buck, 2009, S. 31).

Zusammengefasst lässt sich sagen, dass sich der grösste Teil dieser alternativen Lernendenvorstellungen durch die Übertragung von lebensweltlichen Bezügen in die Mikrowelt erklären lassen (Griffiths & Preston, 1992; Mikelskis-Seifert & Fischler, 2003).

Die Frage, warum Lernendenvorstellungen so stabil sind, erklärt Mulford (2002) in Übereinstimmung mit den Herausforderungen der Conceptual Change Theorie (Beerenwinkel, 2006; Duit, 1995) folgendermassen: Erstsemestrige Chemiestudierende würden nur sehr zaghafte ihre Lernendenvorstellungen verändern oder erweitern, selbst wenn wissenschaftlich korrekte Konzepte vermittelt werden, weil ihre Lernendenvorstellungen alltagstauglich sind. Vergleichbar beschreibt Horton (2009, S. 1) erstsemestrige Physikstudierende: „if students encounter new information that contradicts their alternative conceptions it may be difficult for them to accept the new information because it seems wrong.“ Horton (2009) schreibt weiter, dass selbst wenn Studierende an den Prüfungen die neuen Informationen korrekt wiedergeben, sei es gut möglich, dass das auswendiggelernte Wissen schnell wieder vergessen wird, weil diese „anomale“ neue Information keinen Sinn ergibt.

2.4 Modellkompetenz und Nature of Science

Der Gedanke von Nature of Science (NoS) oder Natur der Naturwissenschaften nahm erstmals in den 1950er Jahren in den USA in die Lehrpläne Einzug (McComas, 2020b). Durch den Sputnik-Shock wurde diese Entwicklung weiter beschleunigt und finanziert (McComas, 2020b). Seit dieser Zeit haben sich eine Vielzahl unterschiedlicher Gedankenkonstrukte über NoS entwickelt (McComas, 2020b). Im Versuch einer Begriffsdefinition, grenzen McComas und Clough (2020, S. 5) die Natur der Naturwissenschaften von den Naturwissenschaften ab, indem sie schreiben, dass NoS nicht eine Beschreibung davon ist, wie die natürliche Welt funktioniert (das sind die Naturwissenschaften selbst), sondern viel mehr eine Beschreibung, wie die Naturwissenschaften als Unternehmen funktionieren. Konkreter wird NoS folgendermassen definiert (McComas et al. 1998, zitiert nach McComas & Clough, 2020, S. 5):

The nature of science is a fertile hybrid arena which blends aspects of various social studies of science including the history, sociology, and philosophy of science combined with research from the cognitive sciences such as psychology into a rich description of what science is, how it works, how scientists operate as a social group and how society itself both directs and reacts to scientific endeavors.

Eine etwas kürzere Begriffsdefinition ist die folgende: “The nature of science involves the basic values and beliefs that make up the scientific world view, how scientists go about their work, and the general culture of the scientific enterprise” (AAAS, 2001).

Die Folgefrage lautet, was wir in der Schule unter NoS verstehen möchten, denn NoS ist viel komplexer als der Schulunterricht es abzubilden vermag. Eine diesbezüglich Einigkeit zu erreichen, wird schwierig. Vorschläge zu einem Curriculum für NoS formulierten Lederman (2002), AAAS (1993), Osborne et al. (2003) sowie McComas et al. (2004).

Eine relativ breit akzeptierte Beschreibung von NoS für den Unterricht besteht aus folgenden drei Klustern mit je drei Hauptaspekten von NoS (McComas, 2020a, S. 40):

1. Werkzeuge, Prozesse und Produkte der Naturwissenschaften

- Beweise sind in den Naturwissenschaften unerlässlich
- Gesetze und Theorien sind miteinander verwandt doch unterschiedlich
- Konsensuales methodisches Vorgehen

2. Naturwissenschaften und ihre Grenzen

- Naturwissenschaften unterscheiden sich vom Ingenieurwesen und der Technologie
- Naturwissenschaften sind vorläufig, dauerhaft und selbstkorrigierend
- Naturwissenschaften haben Grenzen

3. Menschliche Aspekte der Naturwissenschaften

- Kreativität ist in den Naturwissenschaften allgegenwärtig
- Subjektivität und Voreingenommenheit sind in den Naturwissenschaften vorhanden
- Gesellschaft und Kultur sind mit den Naturwissenschaften in Wechselwirkung

Welche Bereiche von NoS spielen nun beim Aufbau der Modellkompetenz eine Rolle? Nach Krell (2015) besteht ein enger Zusammenhang zwischen NoS und der Modellkompetenz. Betrachtet man die drei Kluster von McComas (2020a, S. 40), können alle Bereiche von NoS helfen, die Modellkompetenz aufzubauen. So kann zum 1. Cluster „Werkzeuge, Prozesse und Produkte“ beispielsweise ein historischer Überblick über die Herangehensweisen der Erstellung der Atommodelle im Verlaufe der Zeit thematisiert werden, zum 2. Cluster „Naturwissenschaften und ihre Grenzen“ kann durch Atommodelle als Denkmodelle die Grenze der Naturwissenschaften aufgezeigt werden, d.h. es kann gezeigt werden, dass Informationen fehlen, um genaue Modelle zu erstellen. Auch das 3. Cluster „Menschliche Aspekte der Naturwissenschaften“ kann beispielsweise bei der Erstellung des Plumpudding Modells von Thomson oder der unterschiedlichen Interpretation des Goldfolienversuchs durch Thomson und Rutherford (Hansson, et al., 2019) thematisiert werden. Durch diese Beispiele wird offenkundig, wie nahe Modellkompetenz und NoS miteinander verflochten sind. Mikelskis-Seifert (2010, S. 2f.) hebt weiter hervor, dass eine Verknüpfung des Arbeitens mit Modellen und ein Lernen über Modelle notwendig ist, was die Verknüpfung von NoS mit der Modellkompetenz unterstreicht.

2.5 Didaktische Ansätze zum Aufbau der Modellkompetenz

Weil viele Schüler:innen ein inadäquates Verständnis von Modellen und somit eine unzureichende Modellkompetenz besitzen, haben sich in den vergangenen Jahrzehnten viele Fachdidaktikerinnen und Fachdidaktiker Gedanken über didaktische Ansätze zum Aufbau der Modellkompetenz für unterschiedliche Altersklassen gemacht (Horton, 2009; Mikelskis-Seifert & Leisner, o. A.; Peters-Burton & Burton, 2020; Rehm & Buck, 2006).

Dabei fällt auf, dass es unterschiedliche didaktische Vorgehensweisen gibt, mit unterschiedlichen Schwerpunkten. Um die Breite aufzuzeigen, werden im Folgenden zuerst allgemeine Überlegungen zum Aufbau der Modellkompetenz diskutiert und im Anschluss daran mehrere konkrete Herangehensweisen für unterschiedliche Altersklassen skizziert.

2.5.1 Allgemeine Überlegungen zum Aufbau der Modellkompetenz

Stachowiak (1973, S. 184) erläutert, dass anschauliche Modellvorstellungen häufig nur noch wissenschaftsgeschichtliche und didaktische Funktionen inne haben, wie z.B. das nach Analogie von Planetensystemen entworfene Atommodell und schreibt: „Die Anschauung fungiert hier als Denkprothese.“ (1973, S. 184). Damit Atommodelle von den Lernenden tatsächlich als Denkprothesen respektive Denkmodelle betrachtet werden, gilt es laut Johnstone (1991), ein sogenanntes chemisches Dreieck mit drei Dimensionen zu beachten: 1. Die Ebene der Phänomene (das Beobachtbare). 2. Die submikroskopische Ebene (die atomare Denkmodellebene) und 3. die symbolische Ebene (die Formelsprache). Dieses Dreieck wurde immer wieder aufgenommen und teilweise leicht abgeändert. So nennt Rehm (2009) die Ebene der Phänomene „Kontinuum“ und die submikroskopische Welt „Diskontinuum“. Talanquer (2011) spricht bei der Ebene der Phänomene von „Erfahrungen“, bei der submikroskopischen Ebene von „Modellen“ und bei der symbolischen Ebene von „Visualisierungen“.

Johnstone (1991) schreibt, dass viele „Misconceptions“ durch die unsachgemäße Vermischung dieser drei Ebenen resultieren. Im Umkehrschluss bedeutet das für den Unterricht, dass diese Ebenen explizit unterschieden werden müssen. Mehrere Studien bestätigen, dass eine bewusste Thematisierung der drei Ebenen ein besseres Chemieverständnis bei den Schüler:innen hervorbringt (Devetak et al., 2004; Jaber & BouJaoude, 2012). Rehm (2018, S. 200) beschreibt in Übereinstimmung, dass bei seinen Expert:inneninterviews die meisten Chemiedidaktikerinnen und

Chemiedidaktiker sich darüber einig sind, dass das chemische Dreieck das Besondere der Chemie sei. Zudem könne das Experiment diese drei Dimensionen verbinden.

Zum Erlernen der „Modellmethode“, d.h. dem Lernen von und über Modelle im naturwissenschaftlichen Unterricht (Leisner-Bodenthin, 2006, S. 98), erstellte Brademann (1997, S. 207) einen siebenstufigen Ablauf. Dieser bezieht sich nicht nur auf Denkmodelle, sondern Modelle ganz allgemein. Leisner-Bodenthin (2006, S. 98) bezieht sich konkret auf Denkmodelle und beschreibt folgende vier Schritte. Erstens sollen die Lernenden durch das Beobachten eines Phänomens erkennen, dass die direkte Untersuchung des Originals das Phänomen nicht erklären kann, zweitens sollen die Lernenden ein Modell entwickeln beziehungsweise ein Modell auswählen zur Erklärung des Phänomens, drittens soll die Frage/Problem beantwortet, respektive gelöst werden (Modellanwendung). Viertens soll die Zweckmässigkeit und der Erklärungswert des Modells geprüft werden und dabei die Modellnutzung respektive die Modellmethode reflektiert werden. Insbesondere bei Schritt 1 und 4 soll genügend Zeit aufgewendet werden, um die epistemologischen Aspekte zu stärken (Leisner-Bodenthin, 2006, S. 98). Dieses Vorgehen wird als besonders lernwirksam hinsichtlich der Modellkompetenz der Schüler:innen beschrieben (Leisner-Bodenthin, 2006, S. 102). Vorteile sind seine Stringenz und das Hinterfragen des Modells durch die Schüler:innen im vierten Schritt, denn durch diesen vierten Schritt wird es den Lernenden möglich, Modelle als zweckorientierte Konstrukte wahrzunehmen. (Tempel et al., 2018, S. 13).

2.5.2 Konkrete Unterrichtsbeispiele zum Aufbau der Modellkompetenz

Mikelskis-Seifert und Leisner (o. A.) schlagen für den ersten Kontakt mit Denkmodellen ein mehrschrittiges Vorgehen vor. Dabei sind folgende Merkmale entscheidend: 1) die bewusste und dauerhafte Unterscheidung von Erfahrungswelt und Modellwelt. Dabei geht sie auf die ersten zwei Ecken des chemischen Dreiecks von Johnstone (1991) ein; 2) die Betonung des hypothetischen Charakters der konstruierten Modelle; sowie 3) die Nutzung sowie kritische Bewertung alternativer Modellierungen (Mikelskis-Seifert & Leisner, o. A., S. 1). In der Einführung werden die Lernenden durch selbstgebastelte Alltagsmodelle (z.B. Schiff aus Lego) sowie mehrere Blackbox-Experimente (Kapitel 2.6.1 und 3.2.1) an die Modellproblematik herangeführt. Beim Übergang in die Mikrowelt werden Schülervorstellungen im submikroskopischen Bereich gesammelt und Lernenden- und Lehrpersonenäußerungen in den zwei Welten (Erfahrungswelt und Modellwelt) verortet. Durch das Lernen an Stationen werden verschiedenen Phänomene in den beiden Welten untersucht und gedeutet (z.B. Volumenzunahme durch Erwärmen). In der abschliessenden Reflexionsphase wird über die Modellierungen der Phänomene nachgedacht und u.a. diese für die Aggregatzustände angewendet.

Kind (2004, S. 13) schlägt für die Sekundarstufe I vor, dass zu Beginn die von der Primarschule aufgebauten Konzepte über Teilchen und Denkmodelle aufgegriffen werden, und die Kinder die Möglichkeit erhalten sollten, ihre Lernendenvorstellungen in einem „sicheren“ Umfeld zu äussern. Sie hebt auch hervor, dass Lehrpersonen mit den Lernenden geduldig sein sollen beim Aufbau einer adäquaten Modellkompetenz, wenn Wissenschaftlerinnen und Wissenschaftler beinahe 2000 Jahre brauchten, bis die Idee der Atome im frühen 20. Jahrhundert akzeptiert wurde. Konkret möchte Kind (2004) die Lernenden in die Welt der Atome einführen, indem die Kinder in mehreren Schritten erfahren sollen, wie klein Atome sind. Es soll mit mikroskopischen Aufnahmen von Mikroorganismen eingestiegen werden, dann der Begriff „Atomoskop“ eingeführt werden, als ein Mikroskop, welches Atome abbildet. Dann sollen die Kinder aufgefordert werden, sich vorzustellen durch ein „Atomoskop“ zu schauen und ein Atom zu zeichnen. Schliesslich wird das Tunnelrasterelektronenmikroskop als ein echtes Atomoskop präsentiert und Bilder werden gezeigt. Am Ende wird hypothetisch ein Kupferblech auf 50 km vergrössert, und die Lernenden sollen schätzen, wie gross nun ein einzelnes Kupferatom ist. Der Fokus besteht darin, die Winzigkeit der Atome aufzuzeigen.

Rehm und Buck (2006) gehen einen Schritt weiter als Kind (2004). Sie möchten zwar auch die Winzigkeit der Atome anschaulich machen, sehen jedoch, „von den sinnlich wahrnehmbaren Stoffen zu den gedachten Atomen“ (Rehm & Buck, 2006, S. 151) hinzuführen als eine der didaktischen Hauptschwierigkeiten für die Sekundarstufe I, da zwischen der bunten Welt der Stoffe und der eigenschaftsarmen Welt der Atome ein drastischer Unterschied besteht (Rehm & Buck, 2006, S. 151). Dabei beschränken auch sie sich für den Anfangsunterricht auf die zwei ersten Ebenen von Johnstone (1991). Sie schlagen vor, in die Welt der Atome über einen „Sprung“ einzuführen, und somit sich bewusst von einer kontinuierlichen Deduktion zu entfernen (Rehm & Buck, 2006, S. 151), wie das Löffler (1996) vorschlägt. Gestartet wird mit einer Übung „System und Komponente“ (Rehm & Buck, 2006, S. 152ff.), in der es darum geht, dass die Schüler:innen erkennen, dass das gesamte Universum aus ineinander verschachtelten Systemen besteht (z.B. die Stadt als System von Strassen, Parks, Gebäuden usw.). Dabei wird erarbeitet, dass die Komponenten eines Systems andere Eigenschaften besitzen als das System selbst (eine Stadt hat andere Eigenschaften als ein Gebäude dieser Stadt). Weiter sollen die Lernenden erkennen, dass ein System seinerseits wieder eine Komponente eines übergeordneten grösseren Systems ist (die Stadt ist Komponente des System Land, ein Land Komponente eines Erdteils, Erdteile Komponenten der Erde, die Erde Komponente des Sonnensystems... bis hin zum Universum als letztes bekanntes höheres System). Dabei ist es nützlich, zu fragen, was von den abgebildeten Systemen ein Mensch mit dem Auge, dem Fernglas, oder dem Teleskop sehen kann und welche Bilder ausgedachte Bilder, Bilder von Vorstellungen seien (Rehm & Buck, 2006, S. 154). Im Anschluss geht die Reise vom „System Universum“ System für System zurück, bis man wieder beim Mensch als System angelangt ist und merkt, dass auch der Mensch wiederum aus Komponenten besteht, und die Reise in die Winzigkeit weitergeht in Richtung Moleküle und Atome (Rehm & Buck, 2006, S. 154f.). Hier gilt das Augenmerk darauf zu richten, dass Lichtmikroskope Vergrösserungen und somit reelle Bilder sind, das Tunnelrasterelektronenmikroskop im Gegensatz dazu virtuelle Bilder erstellt. D.h. man sieht etwas nicht im Tunnelrastermikroskop, sondern das Tunnelrastermikroskop konstruiert aufgrund von Daten (elektrischen Mikroströmen) ein Bild.

Daraus kann geschlossen werden: „Bilder von Einzelatomen sind immer virtuelle Bilder.“ (Rehm & Buck, 2006, S. 155f.). Weiter kann darauf hingewiesen werden, dass ein Tunnelrasterelektronenmikroskop seine Bilder nicht reproduzieren kann, weil Atome in ständiger Unruhe sind. Mit Hilfe des Tunnelrasterelektronenmikroskops kann diskutiert werden, dass im atomaren System die Eigenschaft Unruhe hinzukommt, während die Eigenschaften, die wir von Stoffen kennen, wie Farbe, Temperatur, im atomaren System fehlen (Rehm & Buck, 2006). Dadurch werden automatisch zwei wichtige Aspekte über Atome vermittelt. 1. Atome haben andere Eigenschaften als die Stoffe, die sie bilden, 2. Es gibt nur virtuelle Bilder von Atomen. Am Ende fassen die Lernenden das Gelernte zusammen. Aus diesen Zusammenfassungen entsteht eine Klassenzusammenfassung, um hervorzuheben, dass Forschende im Team zu ihren Erkenntnissen gelangen (Rehm & Buck, 2006, S. 158).

Für den fortgeschritteneren Unterricht beschreiben Peters-Burton und Burton (2020, S. 190ff.) zur Einführung der Atommodelle einen historisch-genetischen Ansatz von Demokrit bis Schrödinger. Dabei thematisieren sie die einschlägigen, historischen Atommodelle der Reihe nach. Ziel ist es, dass die Schüler:innen durch metakognitive Fragen die Vorgehensweise bei der Erstellung der Atommodelle, die Evidenz, auf der die Modelle erstellt wurden, sowie deren Nutzen durchdringen und reflektieren. Zudem wird reflektiert, warum gewisse Atommodelle verworfen, adaptiert oder noch heute Gültigkeit haben.

Horton (2009, S. 18) schlägt für Studierende an der Universität vor, dass Tests am Anfang und am Ende der Veranstaltung durchgeführt werden, welche alternative Lernendenvorstellungen als falsche Antworten (Distraktoren) enthalten. Denn, wie Kind (2004), ist Horton der Ansicht: „Learning is an active process, and what students do with facts and ideas with which they have been presented depends to a very high degree on what they already think and believe. Being able to recognize and work with these student-held ideas and conceptions is thus a key component of an effective educational strategy.“ (Horton, 2009, S.1). Mehrere solcher formativer Tests wurden bereits für mehrere Niveaus und Themen in der Chemie entwickelt. Weiter schreibt Horton (2009), dass die Dozierenden durch das Bewusstsein über die Lernendenvorstellungen der Studierenden dazu befähigt werden, genauer zuzuhören und die Studierenden passend zu unterstützen. Zwischen den zwei formativen Tests zur Messung des Conceptual Change findet ein „patient guided-inquiry approach“ statt, in welchem die Studierenden mit Experimentresultaten konfrontiert werden, ihre Gedanken dazu äussern sollen (Horton, 2009, S. 4). Dabei sollen die Studierenden durch ihre Argumentationen gelenkt werden, so dass sie zu neuen Interpretationen gelangen. Durch dieses Setting konnten deutlich bessere Studienergebnisse erzielt werden (Horton, 2009, S. 4).

2.6 Didaktische Umsetzung der Interventionsstudie

Um die Schüler:innen mit der Andersweltlichkeit der für niemanden sichtbaren Denkmodellwelt vertraut zu machen, wurde in der vorliegenden Studie eine Unterrichtssequenz durchgeführt, die sich auf mehrere, fachdidaktische Konzepte bezieht. Es wurde konsequent auf die Unterscheidung der drei Dimensionen des chemischen Dreiecks (Johnstone, 1991) geachtet. Auch die Modellmethode von Leisner-Bodenthin (2006, S. 98) mit den vier Schritten wurde beachtet. Als Phänomen zum Start diente in Schritt eins die Elektrolyse von Wasser. Für Schritt zwei wurde, aufgrund der Intervention, die Erstellung von Denkmodellen bearbeitet. In Schritt drei wurden viele verschiedene Modelle thematisiert wie: die Formelschreibweise, ein Legosteinmodell, das Modell von Demokrit, Dalton, Thomson, Rutherford sowie das Schalenmodell. Schliesslich wird in Schritt vier das Schalenmodell eingeführt.

Zur Einführung in die submikroskopische Denkmodellwelt, wurden der Sprung zu den Atomen und das System und Komponenten-Modell in Anlehnung an Rehm und Buck (2006) thematisiert. Ein genetisch historischer Ansatz diente der Einführung unterschiedlicher historischer Modelle von Demokrit bis Rutherford, was somit einer Adaption für die Sek I der Idee von Peters-Burton und Burton (2020) entspricht. Die eigentliche Intervention bezog sich auf zwei bekannte didaktische Vorgehensweisen von Lederman und Abd-El-Khalick (2002). Aus diesem Grund werden diese zwei Vorgehensweisen im Folgenden genauer geschildert.

2.6.1 Blackbox-Aufgabe

Die Blackbox (BB)- Aufgabe (Lederman und Abd-El-Khalick, 2002) (Kapitel 3.2.1) ist eine Möglichkeit, den Lernenden aufzuzeigen, wie Forscherinnen und Forscher bei der Erstellung von Modellen vorgehen. Die Blackbox-Aufgabe ist enaktiv (Bruner, 1966), d.h. die Schüler:innen handeln (Krey & Schwanewedel, 2018, S. 162). Dabei erhalten die Schüler:innen eine verschlossene Box und müssen deren Inhalt versuchen, zu erraten und von ihrer Vorstellung des Boxinhalts eine Skizze zu zeichnen. Gemäss Lederman und Abd-El-Khalick (2002) vermittelt die BB den Schüler:innen ein experimentelles Vorgehen zur Erstellung von Modellen, beispielsweise der Erstellung von Atommodellen. Dabei beobachten die Schüler:innen, sammeln erste Daten und formulieren Hypothesen, um ihre Daten zu erklären. Anschliessend überlegen sich die Schüler:innen Versuche, um ihre Hypothesen zu überprüfen. Aufgrund ihrer Versuche müssen sie entscheiden, ob ihre Hypothesen gültig oder ungültig sind. Letztlich sollen die Schüler:innen aus ihren Erkenntnissen ein Modell ableiten und skizzieren und dessen Gültigkeit überprüfen. BB-Aktivitäten können genutzt werden, um ein angemessenes Verständnis für naturwissenschaftliches Arbeiten zu fördern (Lederman & Abd-El-Khalick, 2002).

Die Schüler:innen sollen folgende Punkte durch das BB-Experiment verstehen (Lederman & Abd-El-Khalick, 2002):

- Eine klare Unterscheidung von Beobachtung und Deutung soll verinnerlicht werden.
- Die Rolle der Kreativität, Vorstellungskraft und die Interpretationsfähigkeit in der Wissenschaft sollen erkannt werden.
- Wissenschaftliche Erkenntnisse werden empirisch erworben. D.h. sie basieren auf Experimenten und Beobachtungen.
- Wissenschaftliche Erkenntnisse (Theorien, Gesetze) sind vorläufig und können sich ändern.
- Wissenschaftliche Modelle (Atome, Gene) sind keine Kopien der Realität. Vielmehr sind es die Modelle, die die Basis bilden, um ein Naturphänomen zu erklären und zu verstehen.

Zusätzlich, zu den oben aufgeführten Punkten, ergibt sich die Gelegenheit, an folgenden Fähigkeiten zu arbeiten:

- Beobachten und Daten sammeln
- Erarbeiten von Strategien, um Hypothesen zu überprüfen
- Modelle konstruieren

2.6.2 Spuren-Aufgabe

Lederman und Abd-El-Khalick (2002) zufolge vermitteln Schulbücher oft ein unangemessenes Bild von den Naturwissenschaften. Die Konzeption im Schulbuch ist so, dass Naturwissenschaftlerinnen und Naturwissenschaftler für jede Frage, die aus der Natur stammt, die korrekte Antwort bereit haben. Dadurch wird den Schüler:innen suggeriert, Naturwissenschaften verhalten sich wie gängige Schulaufgaben, der immer eine eindeutig richtige Lösung zugeordnet werden kann. Diese Erfahrungen widerspiegeln aber keinesfalls die Vorgehensweise in der naturwissenschaftlichen Forschung. Denn es ist selten der Fall, dass es auf die untersuchten Fragen nur eine Antwort gibt. Dies geht darauf zurück, dass wissenschaftliche Erkenntnisse ein Produkt aufbereiteten Wissens und daraus resultierender Deutungen sind. Dabei wirken auch die Vorstellungskraft und die Kreativität mit ein. Dies, obwohl möglichst viele Grundlagen aus empirischen Studien hervorgehen. Daraus folgt, dass naturwissenschaftliche Erkenntnisse nie absolut oder definitiv sind. Das naturwissenschaftliche Wissen, inklusive Theorien und Gesetze, ist vorläufig und kann sich jederzeit ändern (Lederman & Abd-El-Khalick, 2002).

Bei der Spuren-Aufgabe (Tricky tracks) (Kapitel 3.2.2) müssen die Schüler:innen drei Abbildungen betrachten und dazu eine Geschichte schreiben. Auf den Abbildungen sind Spuren von Tieren zu sehen. Welche Tiere es sind und was die Tiere machen, ist nicht sichtbar. Die Schüler:innen haben somit zu wenige Informationen, um genau zu wissen, wie die Spuren entstanden sind. Dadurch entstehen unterschiedliche Geschichten. Den Schüler:innen soll anschliessend die Analogie zur naturwissenschaftlichen Forschung und somit auch zur Entwicklung von Atommodellen aufgezeigt werden, bei welcher Informationen zur Erstellung des Atommodells fehlen und somit ein Interpretationsspielraum vorhanden ist. Lederman et al. (2020, S. 298) schreiben: „Students completing this activity will gain experience in distinguishing between observation and inference and realize that, based on the same evidence (observations, or data), several answers to the same question may be equally valid.“

Die Schüler:innen sollen folgende Punkte durch die SP-Aufgabe verstehen (Lederman & Abd-El-Khalick, 2002):

- Eine klare Unterscheidung von Beobachtung und Schlussfolgerung soll verinnerlicht werden.
- Die Rolle der Kreativität, Vorstellungskraft und die Interpretationsfähigkeit in der Wissenschaft sollen erkannt werden.
- Insbesondere soll klar werden, dass verschiedene Personen, trotz derselben Beobachtungsgrundlage bzw. Daten, verschiedene Erklärungen, Interpretationen und/oder Modelle daraus ableiten. Dabei gilt es zu berücksichtigen, dass alle Erklärungen, Interpretationen und Modelle gleichermaßen gültig sind, solange sie gleich plausibel sind.

2.7 Forschungslücken

Viele Schüler:innen denken, dass es in den Naturwissenschaften immer genau eine Lösung gibt und dass diese Lösung bekannt ist (Lederman & Abd-El-Khalick, 2002). Wird bei der Vermittlung von Atommodellen nicht explizit darauf geachtet, dass es sich hier um Denkmodelle mit Interpretationsspielraum handelt, ist die Gefahr gross, dass die Schüler:innen sich Atome so vorstellen, wie sie im vermittelten Atommodell dargestellt werden (z.B. als Kugel) (Kind, 2004; Rehm, 2009). Deshalb sind sich die Fachdidaktikerinnen und Fachdidaktiker einig, dass NoS eng verknüpft sein muss mit dem Aufbau einer adäquaten Modellkompetenz, und dass dieser Aufbau mehrere Hürden beinhaltet (Leisner-Bodenthin, 2006; McComas, 2020b; Rehm & Buck, 2006). Einige Studien zeigten bereits die Wirksamkeit ihrer didaktischen Vorgehensweise zum Aufbau von Modellkompetenz auf (Horton, 2009; Leisner, 2005; Mikelskis-Seifert &

Fischler, 2003). Weitere Studien untersuchten die Lernwirksamkeit für eine gesamte Unterrichtseinheit (Mikelskis-Seifert & Leisner, o. A.). Da die bekannte Blackbox-Aufgabe (Lederman, 2002) bei Mikelskis-Seifert und Leisner (o. A.) eingebettet in eine gesamte Unterrichtseinheit einen Lernzuwachs hervorrufen konnte, die Blackbox-Aufgabe als einzelne Aufgabe jedoch nicht auf ihre Lernwirksamkeit untersucht wurde, haben wir die Blackbox-Aufgabe in den Blick genommen und stellten sie in unserer Interventionsstudie der Spurenaufgabe (Lederman & Abd-El-Khalick, 2002) gegenüber, die ebenfalls aufzeigt, wie Forschende Denkmodelle erstellen. Bei diesem Vergleich der zwei Aufgaben interessierte uns auch, ob nicht nur eine gesamte Unterrichtseinheit, wie bei Mikelskis-Seifert und Leisner (o. A.), einen Lernzuwachs zur Folge hat, sondern bereits zwei Lektionen, wie sich das auch schon McComas (2020b) gefragt hatte.

2.8 Forschungsfragen und Hypothesen

Aus der Forschungslücke lassen sich folgende Forschungsfragen und Hypothesen ableiten:

1. Führen die beiden Aufgaben (Blackbox und Spuren) zu einem besseren Modellverständnis der Schüler:innen?

H1a: Die Schüler:innen haben zum Messzeitpunkt T1 unabhängig des Unterrichtsansatzes ein signifikant besseres Modellverständnis als zum Messzeitpunkt T0, weil davon ausgegangen wird, dass das Verständnis über Modelle durch das Behandeln von Phänomenen gesteigert werden kann (Séré, 1990).

H1b: Die Schüler:innen haben zum Messzeitpunkt T2 unabhängig des Unterrichtsansatzes ein signifikant besseres Modellverständnis als zum Messzeitpunkt T0, weil davon ausgegangen wird, dass die Schüler:innen trotz Vergessenskurve auch nach zwei Monaten noch mehr wissen über Modelle als vor dem Unterricht.

2. Nimmt das Modellverständnis der beiden Interventionsgruppen (Blackbox-Aufgabe und Spuren-Aufgabe) vom Posttest (T1) zum Follow-up Test (T2) ab?

H2a: Das Modellverständnis der Schüler:innen ist direkt nach der Unterrichtseinheit (T1) signifikant besser als zwei Monate später (T2), weil aufgrund der beschränkten Kapazität des Kurzzeit- und des Arbeitsgedächtnisses altes Wissen durch Neues verdrängt wird (Escher & Messner, 2015). Ein anderer Grund ist, dass das Wissen ggf. zum gewünschten Zeitpunkt nicht abgerufen werden kann. Demnach ist das Wissen noch vorhanden, doch durch fehlende Abrufhinweise oder eine zu schwache Gedächtnisspur blockiert (Escher & Messner, 2015).

3. Welche Intervention (Blackbox-Aufgabe oder Spuren-Aufgabe) ist besser geeignet, um das Modellverständnis bei den Schüler:innen zu fördern?

H3a: Die Schüler:innen der Blackbox-Interventionsgruppe weisen einen grösseren Kompetenzzuwachs von T0 zu T1 auf als die Schüler:innen der Spuren-Interventionsgruppe, weil Lederman und Abd-El-Khalick (2002) davon ausgehen, dass bei der Blackboxaufgabe mehr Verstehensprozesse aufgebaut werden können als bei der Spurenaufgabe und zudem an mehr überfachlichen Fähigkeiten gearbeitet wird. Auch das handelnde Lernen und das unmittelbare Verknüpfen des Denkens mit dem Handeln bei der Blackboxaufgabe könnte vorteilhaft sein für den Verstehensprozess (Vogel, 2014)). Zudem könnte der kooperative Aspekt vorteilhaft sein, wenn man von einer guten Gruppenkonstellation ausgeht (Dubs, 2009; Grunder et al., 2014; Zobrist, 2014).

H3b: Die Schüler:innen der Blackbox-Interventionsgruppe weisen einen grösseren Kompetenzzuwachs von T0 zu T2 auf als die Schüler:innen der Spuren-Interventionsgruppe. Die Begründung ist analog zu H3a.

3 Methoden

3.1 Forschungsdesign

Diekmann (2007) nennt drei Bedingungen, die erfüllt werden müssen, wenn es sich um ein echtes, experimentelles Design handelt.

1. Die Versuchsteilnehmenden müssen in mindestens zwei Gruppen aufgeteilt werden.
2. Diese Zuteilung muss zufällig stattfinden.
3. Die Versuchsbedingungen müssen anhand von mindestens einer unabhängigen Variablen manipuliert werden (Diekmann, 2007).

Durch das zufällige Zuordnen der Klassen zu den drei Gruppen (Kontrollgruppe, SP und BB) und den beiden unterschiedlichen Interventionen, sind die Bedingungen nahezu erfüllt. Weil sich das zufällige Aufteilen auf Gruppen und

nicht einzelne Individuen bezieht, weicht das Design von der zweiten Bedingung leicht ab. Daher handelt sich um eine quantitative Interventionsstudie mit quasi-experimentellem Forschungsdesign.

3.2 Vorgehen

Für den Vergleich der beiden Unterrichtseinheiten wurde eine Interventionsstudie nach dem Prä-, Post-, Follow-Up-Design mit zwei Interventionsgruppen und einer Kontrollgruppe durchgeführt. Zu diesem Zweck wurden die teilnehmenden Schulklassen zufällig in drei Gruppen (Blackbox (BB), Spuren (SP) und Kontrollgruppe) aufgeteilt. Die Lernenden lösten dreimal denselben Fragebogen mit qualitativen und quantitativen Antwortformaten. Die qualitativen Aufgaben wurden durch drei Rater analysiert. Die Punktezahlen der Aufgaben wurden anschliessend inferenzstatistisch ausgewertet.

Die Kontrollgruppe hatte keine Intervention durchgeführt. Sie hatten lediglich die Tests (mit zeitlichem Abstand) zweimal durchgeführt, um ausschliessen zu können, dass eine Verbesserung des Testresultats durch mehrmaliges Ausfüllen erreicht werden kann.

Die Intervention der beiden Interventionsgruppen BB und SP bestand aus einer Unterrichtseinheit im Umfang von acht Lektionen, aufgeteilt auf vier Doppellektionen. Beide Unterrichtseinheiten wurden auf den Lehrplan 21 (D-EDK, 2015) und das LUKAS-Prozessmodell (Wilhelm et al., 2014) abgestimmt. Die kompletten Unterlagen zur Durchführung der Unterrichtseinheiten (Grobplanung, Feinplanungen, Skript, Lösungen, Arbeitsblätter und Test) wurden an 19 Lehrpersonen der Sekundarstufe I verschickt, die sich bereit erklärt hatten, bei der Datenerhebung mitzumachen. In der gesamten Unterrichtseinheit wurden folgende Kompetenzen des Lehrplans 21 (D-EDK, 2015) abgedeckt:

- NT. 3.2. Die Schüler:innen können Stoffumwandlungen einordnen und erklären.
- NT. 2.1. Die Schüler:innen können Stoffe untersuchen, beschreiben und ordnen.

Einzig die Kompetenzstufe NT.2.1.2c „Die Schüler:innen können Unterschiede zwischen Modell und Wirklichkeit aufzeigen.“ (D-EDK, 2015) wurde für die Blackbox- oder Spuren-Aufgabe unterschiedlich unterrichtet. D.h. während sechs Lektionen identisch in beiden Gruppen unterrichtet wurden, unterschied sich die zweite Doppellektion von der Vorgehensweise in der BB und SP Gruppe. In dieser Doppellektion wurde ausschliesslich das Modellverständnis behandelt. Das Modellverständnis im Sinne von Nature of Science wurde somit getrennt von den fachlichen Inhalten der anderen sechs Lektionen behandelt.

3.2.1 Aufbau der zweiten Doppellektion Blackbox-Aufgabe

In der Blackbox-Aufgabe (Lederman & Abd-El-Khalick, 2002) werden die Schüler:innen aufgefordert herauszufinden, was in einer Box ist, ohne hinein schauen zu können. Beim Anschliessenden Öffnen der Box können die Lernenden ihre Ergebnisse mit dem realen Inhalt der Box vergleichen und stellen vermutlich fest, dass trotz der zuvor durchgeführten Experimente gewisse Fehlschlüsse gemacht wurden, und ihre Folgerungen teilweise falsch waren.

Zusätzlich hat die Lehrperson eine weitere Blackbox, die nicht geöffnet wird. Bei dieser Blackbox machen die Schüler:innen die Erfahrung, dass im übertragenen Sinn eine Blackbox in der Realität auch mal nicht geöffnet werden kann, d.h. die Realität ungewiss bleibt. Ein solches Beispiel ist der Aufbau eines Atoms. Als Folgerung kann aufgezeigt werden, dass Atommodelle Denkmodelle bleiben, solange die Realität „unsichtbar“ ist.

3.2.2 Aufbau der zweiten Doppellektion Spuren-Aufgabe

Bei der Spuren-Aufgabe (Tricky tracks) (Lederman & Abd-El-Khalick, 2002) wird mit verschiedenen Tierspuren (vermutlich Vogelspuren) gearbeitet. Die Schüler:innen sollen versuchen herauszufinden, was passiert sein könnte, dass die Spuren in einem solchen Muster hinterlassen worden sind auf drei hintereinander folgenden Bildern. Durch den Austausch unter Peers erkennen die Lernenden, dass eine Spur verschiedene Interpretationen zulässt, und nachdem die Tiere verschwunden sind, die Wahrscheinlichkeit verschwindend klein ist, herauszufinden, was nun wirklich passiert ist. Im übertragenen Sinn soll die Aufgabe den Schüler:innen aufzeigen, dass auch bei Experimenten zu Atomen die Wahrscheinlichkeit verschwindend klein ist, daraus ein Modell zu erstellen, welches der Wirklichkeit entspricht. So können die Lernenden an Denkmodelle herangeführt werden.

3.2.3 Stichprobe

Um ein möglichst aussagekräftiges Resultat zu erhalten, musste eine Vielzahl von Lehrpersonen gefunden werden, die sich bereit erklärten, eine der zwei Unterrichtseinheiten durchzuführen. Daher wurden diverse NT-Lehrpersonen in mehreren Zentralschweizer Kantonen angeschrieben. Glücklicherweise stand der Wechsel zum Lehrplan 21 (D-EDK, 2015) bevor. So hat wohl der Hinweis im Informationsmail, dass die Lehrpersonen eine fertige, nach Lehrplan 21 und LUKAS-Prozessmodell ausgerichtete Unterrichtseinheit erhalten würden, bei vielen Lehrpersonen das Interesse geweckt. Die Nachfrage war so gross, dass eine Beschränkung auf Niveau A-Klassen möglich war. Dadurch konnte die Heterogenität als Störvariable verringert werden.

Ein grosser Unterschied innerhalb der Testgruppe stellt die Jahrgangsstufe dar. In einigen Gemeinden wird das Thema „vom Aufbau der Stoffe“ in der ersten Oberstufe (7. Klasse) behandelt, in anderen erst in der zweiten Oberstufe (8. Klasse). Andererseits gab es Unterschiede bezüglich des Vorwissens. Der prominenteste Unterschied in diesem Zusammenhang war das Thema Wasser. Einigen Schüler:innen war die Zusammensetzung von Wassermolekülen bereits bekannt, anderen nicht.

Im Verlauf dieser Vergleichsstudie zogen drei Lehrpersonen ihre Teilnahme zurück. Die abschliessende Stichprobengrösse (N) kann aus Tabelle 1 entnommen werden. Tabelle 1 gibt weiter Auskunft über die Teilnehmerinnen und Teilnehmer der Studie (ohne Dropouts). Es sind alles Klassen aus der Zentralschweiz.

Tab. 1. Stichprobe.

	NT-LP	Klassen	SuS	m	w
BB	6	8	126	42	84
SP	3	6	113	48	65
Kontrollgruppe	2	2	31	16	15
Total	11	16	270	106	164

Anmerkung: NT-LP= Anzahl Naturwissenschaftslehrpersonen; SuS= Anzahl Schüler:innen; m= männlich; w= weiblich; BB= Blackbox-Interventionsgruppe; SP= Spuren-Interventionsgruppe.

3.3 Forschungsinstrument

Als Forschungsinstrument wurde ein Paper-Pencil-Test zur Unterrichtseinheit eingesetzt. Dieser war bei beiden Interventionsgruppen identisch. Der Test bestand aus neun Aufgaben, wovon Aufgaben 5 und 6 die Kompetenz zum Modellverständnis behandelten. Da nur diese beiden Fragen für die Untersuchung relevant sind, wurden die anderen Aufgaben im Rahmen dieser Studie nicht berücksichtigt. Für die Kontrollgruppe wurde eine gekürzte Version des Tests verwendet, die lediglich die für diese Studie relevanten Aufgaben 5 und 6 enthielt.

Beim Aufbau des Tests wurden die Kriterien eines guten Tests von Froncek und Thielsch (2011) beachtet. Nachfolgend wird auf drei Kriterien genauer eingegangen:

- **Anspruchsniveau variierend:** Es wurde darauf geachtet, dass sowohl einfachere als auch anspruchsvollere Aufgaben unterschiedlicher Taxonomiestufen nach Bloom vorkommen (Bloom & Engelhart, 1976). Ausserdem beinhaltet der Test unterschiedliche Frageformate (Multiple Choice, Offene Fragen, Sortieraufgaben, Berechnungen).
- **Struktur:** Der Test ist so aufgebaut, dass er mit leichteren Fragen beginnt und allmählich schwieriger wird. Ausserdem ist er thematisch und nach Aufgabentyp sortiert (5 Multiple Choice Aufgaben und 4 offene Fragen).
- **Inhaltliche Konsistenz:** Es wurde darauf geachtet, dass die Gewichtung der Fragen und die Aufgabenanzahl ungefähr im Verhältnis zum gelehrteten Inhalt während der Unterrichtseinheit stand.

Die Datenerhebung mit Hilfe des Tests wurde nach dem Pre-Post-Follow-up-Design durchgeführt. Einen Monat vor Beginn der Unterrichtseinheit wurde der Test zum ersten Mal geschrieben. Dadurch konnte das Vorwissen der Schüler:innen erfasst werden. Unmittelbar nach der Durchführung der Unterrichtseinheit wurde der Test zum zweiten Mal gelöst. So konnte der Kompetenzzuwachs über die vier Doppellektionen ermittelt werden. Zwei Monate nach der Unterrichtseinheit wurde der Test ein drittes Mal durchgeführt, um zu überprüfen, wie viel die Lernenden noch wussten.

3.4 Datenauswertung

3.4.1 Auswertung der Daten

Die Datenauswertung wurde von drei Ratern durchgeführt. Um eine möglichst hohe Übereinstimmung der Interpretation der qualitativen Daten zu erreichen, wurde ein Ratingmanual erstellt und zur Validierung die ersten Tests gemeinsam analysiert. In Grenzfällen fand eine gemeinsame Besprechung statt, bevor Punkte vergeben wurden.

Für die vorliegende Studie waren lediglich Aufgabe 5 und 6 relevant. Aufgabe 5 bestand aus vier Multiple Choice Aufgaben (Abb. 1). Aufgrund der quantitativen Natur dieser Aufgabe war ein Lösungsschlüssel ausreichend. Aufgabe 6 bestand aus zwei Fragen mit offenem Antwortformat (Kapitel 4.2). Hierzu wurden in einem Ratingmanual Stichworte festgehalten, die für die Punktevergabe ausschlaggebend waren. Um zu überprüfen, ob alle drei Rater gleich bepunktet haben, wurde die Interraterreliabilität gemessen. Hierfür analysierten alle drei Rater jeweils 127 zufällig

ausgewählte Antworten zu beiden Teilaufgaben von Aufgabe 6 des Posttests (T1) ein zweites Mal. Aufgrund dieser dreifach erhobenen Daten wurde die Intra-Klassen-Korrelation (ICC) berechnet. Die ICC(3,1)-Werte sind für die Aufgabe 6a 75.9% und für die Aufgabe 6b 88.6%. Nach Cicchetti (1994) ist somit für beide Aufgaben eine exzellente Übereinstimmung erreicht. Bei Koo und Li (2016) ist die Bewertung etwas strenger. Die beiden ICC-Werte werden jedoch noch als gut eingestuft. Da während der erneuten Auswertung im Rahmen der Berechnung der Interraterreliabilität kein Austausch zu Grenzfällen stattfand, kann davon ausgegangen werden, dass die ICC-Werte eher konservativ sind.

3.4.2 Quantitative Auswertung der Resultate

Die Resultate der Tests wurden mit IBM SPSS Version 26 ausgewertet. Sechs Items erfassen die einzelnen Punktezahlen der sechs Teilaufgaben (5a, 5b, 5c, 5d, 6a, 6b), zwei Items die beiden Summen der jeweiligen Aufgaben 5 und 6 und ein Item das Punktetotal beider Aufgaben 5 und 6. Somit hat jeder Testzeitpunkt (T0, T1, T2) insgesamt je neun Items. Die Reliabilität wurde mit Crohnbach's Alpha berechnet. Zur Berechnung wurden die Zeitpunkte T1 und T2 berücksichtigt, da zum Zeitpunkt T0 kaum Wissen vorhanden war. Bei den zwei Zeitpunkten T1 und T2 wurden die je 6 Items berücksichtigt, welche die einzelnen Punktezahlen der sechs Teilaufgaben beschreiben. Der Crohnbach's-Alpha Wert der 12 Items beträgt $\alpha = .624$ und gilt laut Bowling (2002) als akzeptabel.

Mittels Q-Q-Diagrammen und dem Kolmogorov-Smirnov-Test wurde die Normalverteilung der Daten überprüft. Anschliessend wurden Abweichungen innerhalb der einzelnen Testgruppen mittels t-Tests für abhängige Stichproben untersucht. Zusätzlich wurden alle t-Tests für abhängige Stichproben mit ANOVA mit Messwiederholung (rm-ANOVA) und post-hoc-Tests (Bonferroni-Korrektur) auf potentielle Ungenauigkeiten hin überprüft, die aufgrund der alpha-kumulierten t-Tests entstehen können. Lag bei den Daten eine Verletzung der Voraussetzung der Sphärizität vor, wurde eine Greenhouse-Geisser Korrektur der Freiheitsgrade vorgenommen (Girden, 1992). Die Unterschiede zwischen den beiden Testgruppen wurden mit t-Tests für unabhängige Stichproben berechnet. Der Levene-Test überprüfte die Gleichheit der Varianzen (Levene, 1960). Bei gleicher Varianz wurde der Student t-Test verwendet, bei ungleicher Varianz der Welch-Test. Des Weiteren wurde mittels t-Tests für unabhängige Stichproben untersucht, inwiefern sich die zwei Interventionsgruppen voneinander unterscheiden. Die Gruppen wurden aufgrund der ungleichen Verteilung der Stichprobengrössen mittels Mann-Whitney-U-Tests miteinander verglichen.

Die Effektstärken wurden abhängig von den verwendeten Tests auf drei verschiedene Arten berechnet: Für t-Tests mit unabhängigen Stichproben wurde Cohen's ds (Cohen, 1988), für t-Tests von abhängigen Stichproben Cohen's dz verwendet (Rosenthal, 1991). Für die Berechnung der Effektstärke beim Mann-Whitney-U-Test wurde eine Formel von Rosenthal (1991) verwendet. Damit die Skalierung der Effektstärken einheitlich ist, wurden die Werte anschliessend auf <https://www.psychometrica.de/effektstaerke.html> in einen Cohen's d-Wert umgerechnet.

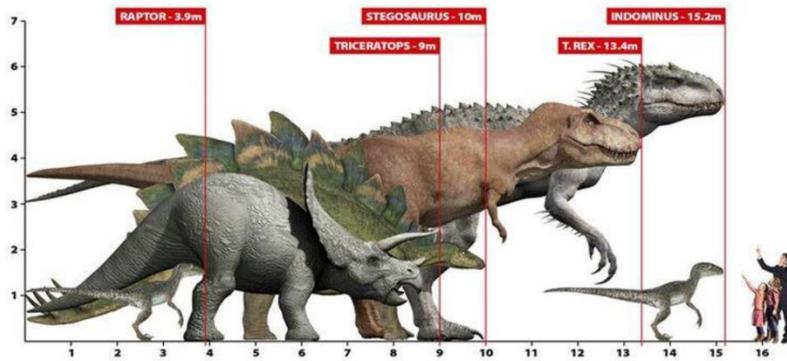
Die Resultate der Kontrollgruppe wurden mittels t-Tests für abhängige Stichproben untersucht.

Um direkte Aussagen über den Kompetenzzuwachs machen zu können, wurden nicht nur die erreichten Punkte, sondern auch die Veränderungen der Mittelwerte der Punktezahlen miteinander verglichen. Für diesen Vergleich wurden neben den je neun Items pro Testzeitpunkt (T0, T1 und T2), welche die erreichten Punktezahlen beschreiben, 27 weitere Items, die deren Veränderung beschreiben, definiert (je neun Items für die Veränderungen von T0 zu T1, T1 zu T2 und T0 zu T2). Diese Auswertungen werden durch Delta (Δ) gekennzeichnet.

4 Ergebnisse

Nachfolgend werden die Ergebnisse der Auswertung präsentiert. Anstelle von absoluten Punkten wird in den Grafiken die prozentuale Erreichung der Gesamtpunktzahl angegeben. Der Grund dafür ist eine unterschiedliche Maximalpunktzahl, die dazu geführt hätte, dass die einzelnen Grafiken nicht miteinander hätten verglichen werden können.

4.1 Aufgabe 5

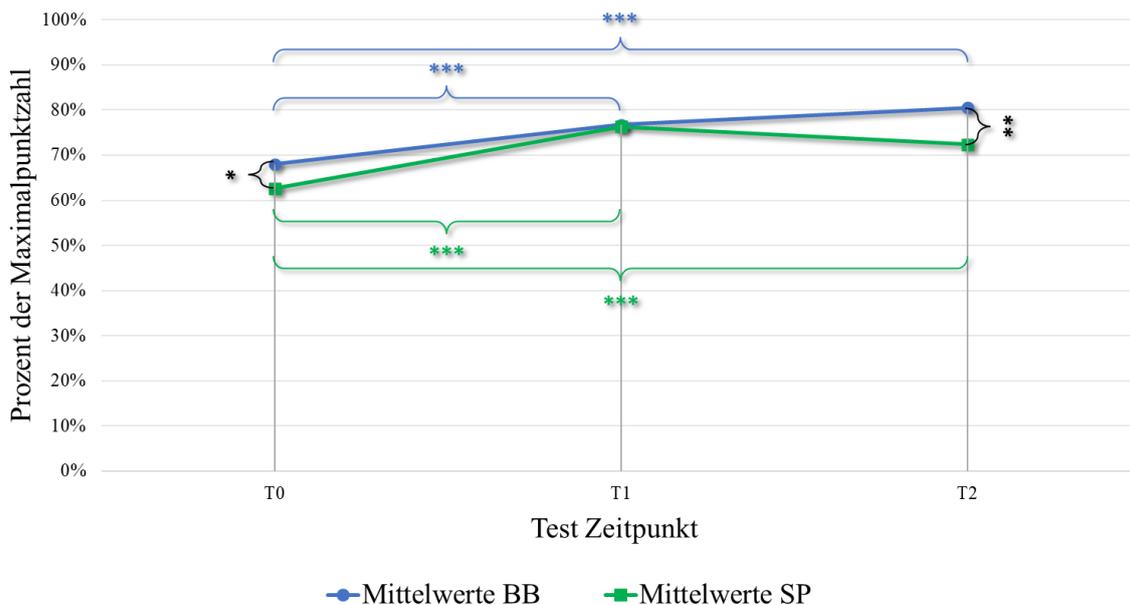


Sind die Aussagen zum Bild oben richtig oder falsch? Kreuze an

	Richtig	Falsch
Bei den Dinos auf dem Bild handelt es sich um eine Abbildung der Realität.	<input type="checkbox"/>	<input type="checkbox"/>
Bei den Menschen auf dem Bild handelt es sich um eine Abbildung der Realität.	<input type="checkbox"/>	<input type="checkbox"/>
Die Größenangaben für die Dinos auf dem Bild entstammen aus einem Denkmodell. Man wird nie wissen, ob der Raptor wirklich kleiner war als der T. Rex.	<input type="checkbox"/>	<input type="checkbox"/>
Die Farben der Dinos entstammen aus einer Abbildung der Realität. Die Dinos hatten genau diese Farben.	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 1. Aufgabe 5.

Bei Aufgabe 5 (Abb. 1) mussten die Schüler:innen verschiedene Aussagen zu einer Abbildung beurteilen. Dabei ging es darum, zu erkennen, welche Aspekte der Abbildung einem Denkmodell oder einer realen Abbildung entspringen. Die Schüler:innen beurteilten die Aussagen mit "richtig" oder "falsch".



Anmerkung: BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe; * $p < .05$, ** $p < .01$, *** $p < .001$.

Abb. 2. Aufgabe 5.

Mittelwertvergleiche innerhalb der Blackbox-Interventionsgruppe (BB) zwischen T0, T1 und T2 sowie innerhalb der SP Interventionsgruppe zwischen T0, T1 und T2. Zudem ist der Mittelwertvergleich zwischen den zwei Gruppen BB und SP zu den drei Zeitpunkten T0, T1 und T2 dargestellt.

Vergleicht man bei Aufgabe 5 die Mittelwerte der Punktezahlen der BB und der SP-Gruppe miteinander, besteht zum Messzeitpunkt T1 kein signifikanter Unterschied (Abb. 2). Im Gegensatz dazu besteht sowohl zum Messzeitpunkt T0 ($p < .05$, $d_s = .22$) als auch zum Messzeitpunkt T2 ($p < .01$, $d_s = .36$) ein signifikanter Unterschied mit jeweils kleinem Effekt (Abb. 2 und Tab. 2) zu Gunsten der BB-Gruppe.

Tab. 2. Aufgabe 5. Mittelwertvergleiche zwischen der BB und SP Gruppe zu den Zeitpunkten T0, T1 und T2.

Beschreibung	Durchgang	M		t	p	d _s
		BB (n = 126)	SP (n = 113)			
Vergleich BB mit SP	T0	1.36	1.25	1.67	<.05*	.22
	T1	1.54	1.53	.14	.443	.02
	T2	1.61	1.45	2.81	<.01**	.36

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, *p < .05, **p < .01, ***p < .001; Effektstärke nach Cohen (1988): |d_s| > .2 ≙ klein, |d_s| > .5 ≙ mittel, |d_s| > .8 ≙ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

Vergleicht man den Kompetenzzuwachs der Schüler:innen jeweils innerhalb der zwei Gruppen, wird erkennbar, dass bei Aufgabe 5 ein signifikanter Kompetenzzuwachs von T0 zu T1 ($p < .001$) und von T0 zu T2 ($p < .001$) sowohl bei den Schüler:innen der BB-Gruppe als auch bei der SP-Gruppe auftritt (Abb. 2 und Tab. 3). Bei beiden Gruppen ist hingegen keine signifikante Zu- oder Abnahme zwischen T1 und T2 auszumachen. Die Effektstärken sind von T0 zu T1 sowohl bei der BB-Gruppe ($d_z = -.34$) als auch bei der SP-Gruppe ($d_z = -.37$) klein (Tab. 3). Von T0 zu T2 sind die Effektstärken bei der BB-Gruppe mit $d_z = -.39$ und bei der SP-Gruppe mit $d_z = -.35$ ebenfalls klein.

Tab. 3. Aufgabe 5. Mittelwertvergleiche innerhalb der Gruppe BB zwischen T0, T1 und T2 sowie innerhalb der Gruppe SP zwischen T0, T1 und T2.

Beschreibung	Durchgang	M			t	p	d _z
		T0	T1	T2			
BB (n = 126)	T0 - T1	1.36	1.54	-	-3.81	<.001***	-.34
	T1 - T2	-	1.54	1.61	-1.58	.941	-.14
	T0 - T2	1.36	-	1.61	-4.36	<.001***	-.39
SP (n = 113)	T0 - T1	1.25	1.53	-	-3.95	<.001***	-.37
	T1 - T2	-	1.53	1.45	1.37	.087	.13
	T0 - T2	1.25	-	1.45	-3.74	<.001***	-.35

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, *p < .05, **p < .01, ***p < .001; Effektstärke nach Cohen (1988): |d_s| > .2 ≙ klein, |d_s| > .5 ≙ mittel, |d_s| > .8 ≙ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe. Die m-ANOVA mit anschliessendem post-hoc Test (Bonferroni-Korrektur) zeigt dieselben Signifikanzen auf.

Um den Kompetenzzuwachs der Lernenden der zwei Interventionsgruppen miteinander zu vergleichen, wurde der Mittelwertunterschied (ΔM) der BB und SP-Gruppe separat berechnet (z.B. $\Delta(T0 - T1)$). Diese Mittelwertdifferenzen wurden anschliessend miteinander verglichen. Der Unterschied des Kompetenzzuwachses zwischen den beiden Interventionsgruppen BB- und SP-Gruppe ist nur von T1 zu T2 ($p < .05$, $d_s = .27$) signifikant mit kleinem Effekt zu Gunsten der BB-Gruppe (Tab. 4).

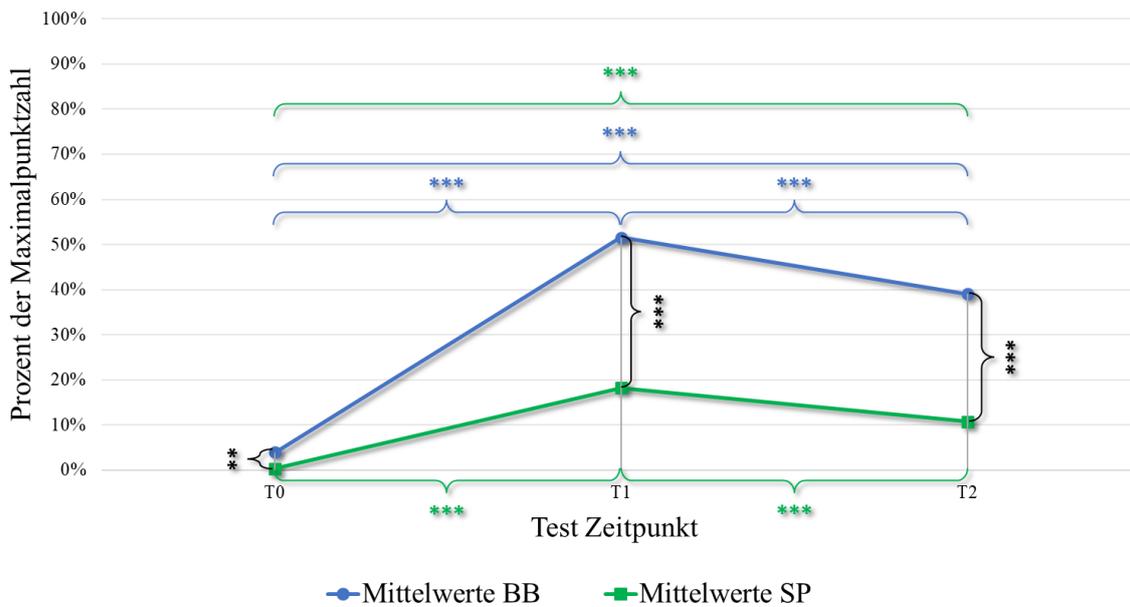
Tab. 4. Aufgabe 5. Vergleich der Mittelwertdifferenzen $\Delta(T0-T1)$, $\Delta(T1-T2)$ und $\Delta(T0-T2)$ zwischen der BB und der SP-Gruppe.

Beschreibung	Durchgang	ΔM		t	p	d_s
		BB ($n = 126$)	SP ($n = 113$)			
Vergleich ΔBB mit ΔSP	$\Delta(T0 - T1)$.17	.27	-1.20	.884	-.16
	$\Delta(T1 - T2)$.08	-.08	2.08	<.05*	.27
	$\Delta(T0 - T2)$.25	.19	.71	.240	.09

Anmerkung: ΔM = Mittelwertdifferenz von jeweils zwei Zeitpunkten; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, * p <.05, ** p <.01, *** p <.001; Effektstärke nach Cohen (1988): $|d_s| >.2 \hat{=}$ klein, $|d_s| >.5 \hat{=}$ mittel, $|d_s| >.8 \hat{=}$ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

4.2 Aufgabe 6

In Aufgabe 6 mussten die Schüler:innen folgende zwei Fragen mit offenem Antwortformat zum Thema Atommodelle beantworten. „A.) Was ist die Schwierigkeit beim Erstellen von Atommodellen? B.) Wie geben Forscherinnen und Forscher beim Erstellen von Atommodellen vor?“



Anmerkung: BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe; * p <.05, ** p <.01, *** p <.001.

Abb. 3. Aufgabe 6. Mittelwertvergleiche innerhalb der Gruppe BB zwischen T0, T1 und T2 sowie innerhalb der Gruppe SP zwischen T0, T1 und T2. Zudem ist der Mittelwertvergleich zwischen den Gruppen BB und SP zu den drei Zeitpunkten T0, T1 und T2 dargestellt.

Vergleicht man bei Aufgabe 6 die Mittelwerte der Punktezahlen der BB und der SP-Gruppe miteinander, besteht zum Messzeitpunkt T0 (p <.05, $d_s = .38$), T1 (p <.001, $d_s = 1.31$) und T2 (p <.001, $d_s = 1.29$) jeweils ein signifikanter Unterschied zu Gunsten der BB-Gruppe (Abb. 3 und Tab. 5). Während der Effekt bei T0 klein ist, ist er bei T1 und T2 gross (Tab. 5).

Tab. 5. Aufgabe 6. Mittelwertvergleiche zwischen der BB und SP Gruppe zu den Zeitpunkten T0, T1 und T2.

Beschreibung	Durchgang	M		t	p	d _s
		BB (n = 126)	SP (n = 113)			
Vergleich BB mit SP	T0	.10	.01	2.97	<.05**	.38
	T1	1.29	.46	10.09	<.001***	1.31
	T2	.98	.27	9.97	<.001***	1.29

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, *p <.05, **p <.01, ***p <.001; Effektstärke nach Cohen (1988): |d_s| >.2 ≙ klein, |d_s| >.5 ≙ mittel, |d_s| >.8 ≙ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

Vergleicht man die Kompetenzen der zwei Gruppen BB und SP separat zu den drei Zeitpunkten T0, T1 und T2, zeigt sich, dass von T0 zu T1 (p <.001) und von T0 zu T2 (p <.001) bei beiden Gruppen ein signifikanter Kompetenzzuwachs zu verzeichnen ist (Abb. 3 und Tab. 6). Von T0 zu T1 ist die Effektstärke sowohl bei der BB-Gruppe (d_z = -1.63) als auch bei der SP-Gruppe (d_z = -.81) gross. Von T0 zu T2 besteht bei der BB-Gruppe (d_z = -1.29) eine grosse und bei der SP-Gruppe (d_z = -.60) eine mittlere Effektstärke. Mit kleinem Effekt kann von T1 zu T2 sowohl bei der BB-Gruppe (p <.001, d_z = .49) als auch bei der SP-Gruppe (p <.001, d_z = .40) eine signifikante Kompetenzabnahme beobachtet werden.

Tab. 6. Aufgabe 6. Mittelwertvergleiche innerhalb der Gruppe BB zwischen T0, T1 und T2 sowie innerhalb der Gruppe SP zwischen T0, T1 und T2.

Beschreibung	Durchgang	M			t	p	d _z
		T0	T1	T2			
BB (n = 126)	T0 - T1	.10	1.29	-	-18.33	<.001***	-1.63
	T1 - T2	-	1.29	.98	5.52	<.001***	.49
	T0 - T2	.10	-	.98	-14.47	<.001***	-1.29
SP (n = 113)	T0 - T1	.01	.46	-	-8.66	<.001***	-.81
	T1 - T2	-	.46	.27	4.26	<.001***	.40
	T0 - T2	.01	-	.27	-6.40	<.001***	-.60

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, *p <.05, **p <.01, ***p <.001; Effektstärke nach Cohen (1988): |d_z| >.2 ≙ klein, |d_z| >.5 ≙ mittel, |d_z| >.8 ≙ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe. Die mr-ANOVA mit anschliessendem post-hoc Test (Bonferroni-Korrektur) zeigt dieselben Signifikanzen auf.

Um den Kompetenzzuwachs der Lernenden der zwei Interventionsgruppen miteinander zu vergleichen, wurde der Mittelwertunterschied (ΔM) der BB und SP-Gruppe separat berechnet (z.B. Δ(T0 - T1)). Diese Mittelwertdifferenzen wurden anschliessend miteinander verglichen. Der Unterschied des Kompetenzzuwachses zwischen den beiden Interventionsgruppen BB- und SP-Gruppe ist von T0 zu T1 mit p <.001, d_s = 1.16 signifikant mit grossem Effekt zu Gunsten der BB-Gruppe (Tab. 7). Ebenfalls ein signifikanter Unterschied des Kompetenzzuwachses zwischen den zwei Gruppen ist zwischen T0 zu T2 (p <.001, d_s = 1.09) mit grossem Effekt sichtbar. Auch hier zu Gunsten der BB-Gruppe. Zwischen T1 und T2 ist die Kompetenzabnahme zwischen den zwei Gruppen nicht signifikant (p = .962) (Tab. 7).

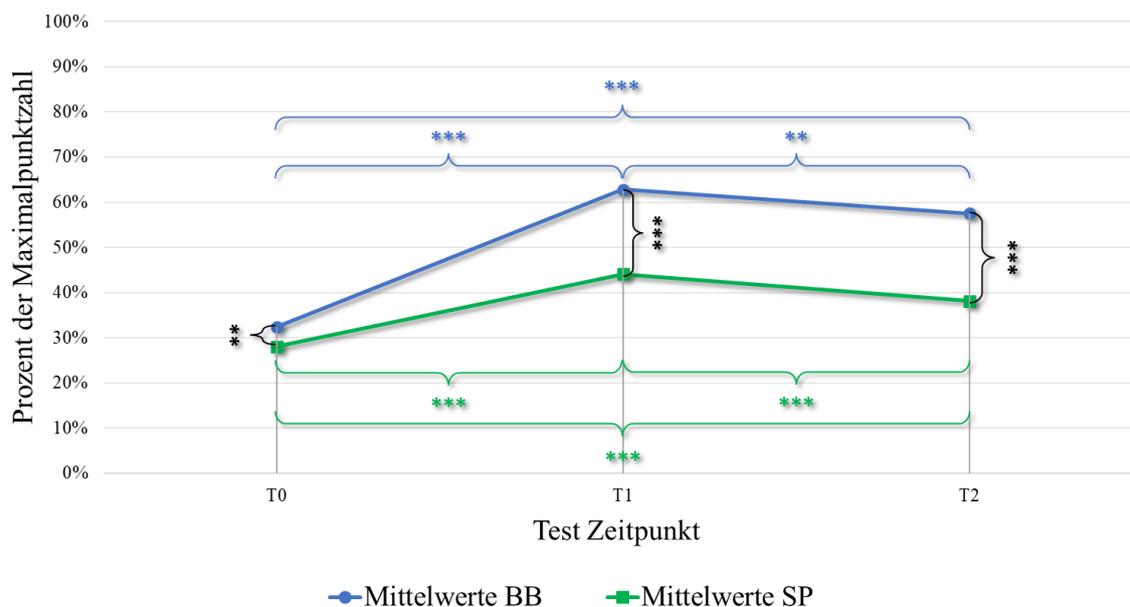
Tab. 7. Aufgabe 6. Vergleich der Mittelwertdifferenzen Δ(T0-T1), Δ(T1-T2) und Δ(T0-T2) zwischen der BB und der SP-Gruppe.

Beschreibung	Durchgang	ΔM		t	p	d _s
		BB (n = 126)	SP (n = 113)			
Vergleich ΔBB mit ΔSP	Δ(T0 - T1)	1.19	.45	8.97	<.001***	1.16
	Δ(T1 - T2)	-.31	-.19	-1.78	.962	-.23
	Δ(T0 - T2)	.88	.26	8.43	<.001***	1.09

Anmerkung: ΔM = Mittelwertdifferenz von jeweils zwei Zeitpunkten; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, *p <.05, **p <.01, ***p <.001; Effektstärke nach Cohen (1988): |d_s| >.2 ≙ klein, |d_s| >.5 ≙ mittel, |d_s| >.8 ≙ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

4.3 Punkttotal

Vergleicht man über alle Teilaufgaben von Aufgabe 5 und Aufgabe 6 die Mittelwerte der Punktezahlen zwischen der BB- und der SP-Gruppe miteinander, haben zu allen drei Messzeitpunkten (T0 ($p < .01$), T1 ($p < .001$) und T2 ($p < .001$)) die Schüler:innen der BB-Gruppe signifikant höhere Punktezahlen (Abb. 4 und Tab. 8). Zu den Messzeitpunkten T1 ($d_s = 1.01$) und T2 ($d_s = 1.20$) mit grossem, zum Zeitpunkt T0 ($d_s = .36$) mit kleinem Effekt.



Anmerkung: BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe; * $p < .05$, ** $p < .01$, *** $p < .001$.

Abb. 4. Aufgabe 5 und Aufgabe 6. Mittelwertvergleiche innerhalb der Gruppe BB zwischen T0, T1 und T2 sowie innerhalb der Gruppe SP zwischen T0, T1 und T2. Zudem ist der Mittelwertvergleich zwischen BB und SP zu den drei Zeitpunkten T0, T1 und T2 dargestellt.

Tab. 8. Aufgabe 6. Vergleich der Mittelwertdifferenzen $\Delta(T0-T1)$, $\Delta(T1-T2)$ und $\Delta(T0-T2)$ zwischen der BB und der SP-Gruppe.

Beschreibung	Durchgang	M		t	p	d _s
		BB (n = 126)	SP (n = 113)			
Vergleich BB mit SP	T0	1.46	1.26	2.75	<.01**	.36
	T1	2.83	1.98	7.80	<.001***	1.01
	T2	2.59	1.72	9.25	<.001***	1.20

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, * $p < .05$, ** $p < .01$, *** $p < .001$; Effektstärke nach Cohen (1988): $|d_s| > .2 \hat{=}$ klein, $|d_s| > .5 \hat{=}$ mittel, $|d_s| > .8 \hat{=}$ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

Über alle Teilaufgaben von Aufgabe 5 und 6 zeigt sich bei beiden Interventionen BB und SP von T0 zu T1 und von T0 zu T2 ein signifikanter Kompetenzzuwachs ($p < .001$) (Abb. 4 und Tab. 9). Bei der BB-Gruppe ist der Effekt des Kompetenzzuwachses sowohl von T0 zu T1 ($d_{\zeta} = -1.55$) als auch von T0 zu T2 ($d_{\zeta} = -1.21$) gross. Bei der SP-Gruppe ist von T0 zu T1 ($d_{\zeta} = -.79$) und von T0 zu T2 ($d_{\zeta} = -.64$) eine mittlere Effektstärke zu verzeichnen. Von T1 zu T2 kann sowohl bei der BB-Gruppe ($p < .01$; $d_{\zeta} = .29$) als auch bei der SP-Gruppe ($p < .001$; $d_{\zeta} = .34$) eine signifikante Kompetenzabnahme beobachtet werden (Abb. 4; Tab. 9).

Tab. 9. Aufgabe 5 und Aufgabe 6. Mittelwertvergleiche innerhalb der Gruppe BB zwischen T0, T1 und T2 sowie innerhalb der Gruppe SP zwischen T0, T1 und T2.

Beschreibung	Durchgang	M			t	p	d_z
		T0	T1	T2			
BB ($n = 126$)	T0 - T1	1.46	2.83	-	-17.38	<.001***	-1.55
	T1 - T2	-	2.83	2.59	3.26	.001**	.29
	T0 - T2	1.46	-	2.59	-13.55	<.001***	-1.21
SP ($n = 113$)	T0 - T1	1.26	1.98	-	-8.42	<.001***	-.79
	T1 - T2	-	1.98	1.72	3.58	<.001***	.34
	T0 - T2	1.26	-	1.72	-6.77	<.001***	-.64

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, * p < .05, ** p < .01, *** p < .001; Effektstärke nach Cohen (1988): $|d_z|$ > .2 $\hat{=}$ klein, $|d_z|$ > .5 $\hat{=}$ mittel, $|d_z|$ > .8 $\hat{=}$ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe. Die m -ANOVA mit anschliessendem post-hoc Test (Bonferroni-Korrektur) zeigt dieselben Signifikanzen auf.

Um den Kompetenzzuwachs der Lernenden der zwei Interventionsgruppen miteinander zu vergleichen, wurde der Mittelwertunterschied (ΔM) der BB und SP-Gruppe separat berechnet (z.B. $\Delta(T0 - T1)$). Diese Mittelwertdifferenzen wurden anschliessend miteinander verglichen.

Beim Vergleich des Kompetenzzuwachses der beiden Gruppen BB- und SP ist sowohl von T0 zu T1 (p < .001; $d_s = .72$) als auch von T0 zu T2 (p < .001; $d_s = .81$) ein signifikanter Unterschied mit mittlerer bis grosser Effektstärke erkennbar zu Gunsten der BB-Gruppe (Tab. 10). Von T1 zu T2 ist kein signifikanter Unterschied ersichtlich.

Tab. 10. Aufgabe 5 und 6. Vergleich der Mittelwertdifferenzen $\Delta(T0-T1)$, $\Delta(T1-T2)$ und $\Delta(T0-T2)$ zwischen der BB und der SP-Gruppe.

Beschreibung	Durchgang	ΔM		t	p	d_s
		BB ($n = 126$)	SP ($n = 113$)			
Vergleich ΔBB mit ΔSP	$\Delta(T0 - T1)$	1.37	.72	5.55	<.001***	.72
	$\Delta(T1 - T2)$	-.24	-.27	.26	.396	.03
	$\Delta(T0 - T2)$	1.13	.46	6.27	<.001***	.81

Anmerkung: ΔM = Mittelwertdifferenz von jeweils zwei Zeitpunkten; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, * p < .05, ** p < .01, *** p < .001; Effektstärke nach Cohen (1988): $|d_s|$ > .2 $\hat{=}$ klein, $|d_s|$ > .5 $\hat{=}$ mittel, $|d_s|$ > .8 $\hat{=}$ gross; BB = Blackbox-Interventionsgruppe; SP = Spuren-Interventionsgruppe.

4.4 Kontrollgruppe

Der Mittelwert des Punktetotals der Kontrollgruppe hat sich von T0 zu T1 nicht signifikant verändert.

Tab. 8. Kontrollgruppe. Mittelwertvergleich innerhalb der Kontrollgruppe T0 und T1.

Beschreibung	Durchgang	M		t	p	d_z
		T0	T1			
Kontrollgruppe ($n = 31$)	T0 - T1	1.58	1.71	-1.91	.244	-.21

Anmerkung: M = Mittelwert; t = t-Wert des t-Tests; p = Wahrscheinlichkeit des t-Tests, * p < .05, ** p < .01, *** p < .001; Effektstärke nach Cohen (1988): $|d_z|$ > .2 $\hat{=}$ klein, $|d_z|$ > .5 $\hat{=}$ mittel, $|d_z|$ > .8 $\hat{=}$ gross.

5 Diskussion / Schlussfolgerung

Die Diskussion ist entlang der drei Forschungsfragen strukturiert, gefolgt durch die Beantwortung der übergeordneten Fragestellung. Weiter werden die praktischen Implikationen, die Methodenkritik und schliesslich das Fazit mit dem Ausblick diskutiert.

5.1.1 Forschungsfrage 1

Die erste Forschungsfrage lautet: *Führen die beiden Aufgaben (Blackbox und Spuren) zu einem besseren Modellverständnis der Schüler:innen?*

In unseren Hypothesen H1a und H1b gingen wir davon aus, dass die Schüler:innen zu den Messzeitpunkten T1 und T2 unabhängig vom Modellansatz ein signifikant besseres Modellverständnis haben als zum Messzeitpunkt T0 (Séré, 1990).

Da bei fast allen Testaufgaben eine signifikante Zunahme des Mittelwerts von T0 zu T1 respektive zu T2 nachgewiesen werden konnte, ist davon auszugehen, dass ein Kompetenzzuwachs stattgefunden hat. Zudem ist bei der Kontrollgruppe keine signifikante Zunahme erkennbar, weshalb weiter anzunehmen ist, dass sich dieser Kompetenzzuwachs auf die beiden Aufgaben (BB und SP) zurückführen lässt. Dieser Kompetenzzuwachs könnte dadurch erklärt werden, dass die Schüler:innen bei beiden Interventionen eigene Erfahrungen über ihre Sinne machen durften und eigene Denkmodelle erfunden haben (Kind, 2004).

Aufgrund der Tatsache, dass beide Null-Hypothesen der Alternativhypothesen H1a und H1b verworfen werden können, kann davon ausgegangen werden, dass sich das Modellverständnis bei den Schüler:innen beider Interventionsgruppen bereits aufgrund der zwei Lektionen andauernden Intervention verbessert hat. Dies gibt Hinweise dazu, dass nicht eine ganze Unterrichtseinheit wie bei Mikelskis-Seifert und Leisner (o. A.) zwingend notwendig ist, um einen signifikanten Lernzuwachs messen zu können, und beantwortet zu einem gewissen Teil die von McComas (2020b) gestellte Frage der Länge einer wirksamen Intervention. Diese Annahme wird zusätzlich durch die Kontrollgruppe gestützt.

5.1.2 Forschungsfrage 2

Die zweite Forschungsfrage lautet: *Nimmt das Modellverständnis der beiden Interventionsgruppen (Blackbox und Spuren) vom Posttest (T1) zum Follow-up Test (T2) ab?*

In unserer Hypothese H2a gingen wir davon aus, dass das Modellverständnis der Schüler:innen direkt nach der Unterrichtseinheit (T1) signifikant besser ist als zwei Monate später (T2) (Escher & Messner, 2015).

Ein Indiz für die Nullhypothese H02a ist die Zunahme der Punkte bei Aufgabe 5. Dies könnte dadurch erklärt werden, dass Aufgabe 5 aus Multiple Choice Aufgaben bestand. Für Schüler:innen ist es einfacher sich zwischen vorgefertigten Antworten zu entscheiden, als selbst eine Antwort auf eine offene Frage zu formulieren. Denn bei den vorgelegten Multiple Choice Aufgaben hatten die Schüler:innen eine 50% Chance auch ohne Vorwissen richtig zu liegen. Dass die Lehrpersonen den Schüler:innen den bepunkteten Test nach dem Messzeitpunkt T1 zurückgegeben haben, erhöhte ihre Chance weiter, beim nächsten Test die richtige Antwort zu finden. Folglich widerspiegeln die Antworten der Schüler:innen nicht zwingend deren effektive Kompetenzausprägung und müssen in ihrer Aussagekraft vermutlich relativiert werden.

Beide Interventionsgruppen zeigen beim Punktetotal der Aufgabe 6 eine signifikante Abnahme von T1 zu T2. Dies erklären wir uns mit den drei verschiedenen Vergessensprozessen nach Wickelgren (1975). Somit kann die Nullhypothese verworfen werden, und es kann davon ausgegangen werden, dass das Modellverständnis der Schüler:innen bei beiden Interventionsgruppen vom Messzeitpunkt T1 zum Messzeitpunkt T2 abnimmt.

5.1.3 Forschungsfrage 3

Die dritte Forschungsfrage lautet: *Welche Aufgabe (Blackbox oder Spuren) ist besser geeignet, um das Modellverständnis bei den Schüler:innen zu fördern?*

In unseren Hypothesen H3a und H3b gingen wir davon aus, dass die Schüler:innen der Blackbox-Interventionsgruppe im Vergleich zu jenen der Spuren-Interventionsgruppe einen grösseren Kompetenzzuwachs (Differenz in der Kompetenzausprägung) von T0 zu T1 und von T0 zu T2 aufweisen (Dubs, 2009; Grunder et al., 2014; Lederman & Abdel-Khalick, 2002; Vogel, 2014; Zobrist, 2014). Zum anderen nahmen kann daraus gefolgert werden, dass die Schüler:innen der Blackbox-Interventionsgruppen eine höhere Kompetenzausprägung zum Messzeitpunkt T1 und T2 aufweisen als die Schüler:innen der Spuren-Interventionsgruppe.

Durch den Vergleich der Veränderungen der Mittelwerte zwischen den einzelnen Messzeitpunkten T0 - T1 und T0 - T2 kann gezeigt werden, dass in der BB-Gruppe ein grösserer Kompetenzzuwachs stattfindet. Der höhere Kompetenzzuwachs aufgrund des BB-Experiments lässt sich verschieden begründen. Mit dem BB-Experiment werden im

Vergleich zum SP-Ansatz drei weitere Verstehensprozesse gefördert (Lederman & Abd-El-Khalick, 2002). Zudem wird vermutet, dass durch den handelnden Zugang, der im BB-Ansatz verfolgt wird, das kognitive Begreifen bzw. das abstrakte Wissen besser aufgebaut wird (Grunder et al., 2014). Eine dritte mögliche Erklärung bietet das forschende Lernen, welches an den handelnden Aspekt anknüpft und bei der die Diskrepanz von Erfahrung und Theorie erkannt werden kann. Zudem spielt kooperatives Lernen bei der BB-Aufgabe eine Rolle (Dubs, 2009; Grunder et al., 2014; Zobrist, 2014). Während bei der BB-Aufgabe in Gruppen gearbeitet wird, arbeiten die Schüler:innen bei der SP-Aufgabe in Einzelarbeit. Kooperatives Lernen lässt individuelle Lösungen zu, und ermöglicht es, in den einzelnen Gruppen zu kommunizieren. Allerdings ist die Lernwirksamkeit stark von der Gruppenkonstellation abhängig (Dubs, 2009; Grunder et al., 2014; Zobrist, 2014).

Eine weitere Erklärung wäre, dass der Test besser auf die BB-Aufgabe abgestimmt war als auf die Spuren-Aufgabe. Dieser Erklärungsansatz halten wir für unwahrscheinlich, da wir beim Erstellen des Tests bewusst darauf geachtet haben, dass alle Teilaufgaben nur Aspekte enthalten, die den Gedanken der Denkmodelle widerspiegeln, ohne Bezug auf die BB- oder Spuren-Aufgabe. Alle diese genannten Aspekte sind Indizien dafür, dass die Nullhypothesen der Alternativhypothesen H3a, H3b, H3c und H3d verworfen werden können.

Gegen diese Interpretation sprechen die Untersuchungen der Teilaufgaben zur Aufgabe 5. Hier schneidet die BB-Gruppe schlechter ab als die SP-Gruppe. Eine mögliche Erklärung könnte sein, dass im Skript der SP-Gruppe mehr Aufgaben zu diesem Aufgabentyp zu finden sind. Folglich hatten die Schüler:innen der SP-Gruppe mehr Übungsmöglichkeiten. Weiter schreiben Funk und Dickson (2011) zu Multiple Choice Fragen, dass sie im Vergleich zu offenen Fragen das Verständnis der Schüler:innen eher überschätzen und dass bei ausschliesslicher Verwendung dieses Aufgabentyps die Fähigkeiten der Schüler:innen nur ungenügend repräsentiert werden. Folglich müssen die Ergebnisse der Aufgabe 5 (Multiple-Choice) isoliert kritisch betrachtet werden.

Da die Untersuchung des Punktetotals aller Aufgaben klar für die Verwerfung der Nullhypothese spricht und die obigen gegenteiligen Indizien relativiert werden konnten, können alle vier Alternativhypothesen angenommen werden. Aufgrund dessen, dass alle vier Nullhypothesen verworfen werden können, kann davon ausgegangen werden, dass sich die BB-Aufgabe besser eignet als die Spurenaufgabe, um das Modellverständnis der Schüler:innen zu fördern.

5.1.4 Beantwortung der übergeordneten Fragestellung

Die vorliegende Arbeit befasst sich mit der übergeordneten Frage, welche Aufgabe besser geeignet ist, damit die Schüler:innen die Welt der Atome als erdachte Welt zur sichtbaren Welt betrachten und akzeptieren, wie das von mehreren Autoren als didaktisch vielversprechender Ansatz proklamiert wird (McComas, 2002; Mikelskis-Seifert & Fischler, 2003; Mikelskis-Seifert & Leisner, o. A.).

Die Untersuchungen dieser Studie geben erste Hinweise darauf, dass bei beiden Interventionen ein Kompetenzzuwachs stattfindet. Es konnte auch gezeigt werden, dass dieser bei der BB-Aufgabe signifikant höher ist als bei der SP-Aufgabe. Diese Ergebnisse decken sich mit den vorgängig erstellten Hypothesen und dem beobachteten Lernzuwachs in der Unterrichtseinheit von Mikelskis-Seifert und Leisner (o. A.), bei der die BB-Aufgabe integriert war. Deswegen kann abschliessend festgehalten werden, dass sich die BB-Aufgabe besser dazu eignet, dass die Schüler:innen die Welt der Atome als erdachte Welt zur sichtbaren Welt betrachten und akzeptieren lernen.

Die Literatur deutet klar darauf hin, dass es sich lohnt, das Modellverständnis sowohl im Anfangsunterricht bei der Einführung des Teilchenmodells und der ersten Atommodelle im Unterricht explizit zu thematisieren (Mikelskis-Seifert, 2009; Rehm, 2007; Vos, 1990) als auch im fortgeschrittenen Unterricht in der Quantenchemie (Passon, Zügge & Grebe-Ellis, 2019). Somit kann die BB-Aufgabe als kleiner Puzzlestein bei einem mehrschrittigen Aufbau der Modellkompetenz betrachtet werden.

5.2 Praktische Implikationen

Die Ergebnisse geben erste Hinweise darauf, dass sich sowohl die BB-Aufgabe als auch die SP-Aufgabe eignen, um das Modellverständnis in der Schule einzuführen. Denn bei beiden Aufgaben deuten die Resultate darauf hin, dass nach der Durchführung der Unterrichtslektion ein Kompetenzzuwachs im Modellverständnis vorhanden war.

Spannend bei der vorliegenden Arbeit ist weiter, dass ein Kompetenzzuwachs bereits durch eine Doppellektion möglich ist, während beispielsweise die Intervention von Mikelski-Seifert und Leisner (o. A.) eine gesamte Unterrichtseinheit beinhaltet.

5.3 Methodenkritik

Die vorliegende explorative Studie ist im Rahmen einer Masterarbeit im Sinne einer Pilotstudie entstanden. Die Resultate deuten somit erste Tendenzen an, die in einer mehrerebenen-analytischen Studie mit einer deutlich grösseren randomisierten Stichprobe und einer Reihe von Kovariaten überprüft werden könnten. Dazu müssten die Stichproben aller Gruppen erhöht werden. Weiter müsste darauf geachtet werden, dass auf die BB-Gruppe und die SP-Gruppe gleich viele Lehrpersonen und gleich viele Klassen kommen. Auch die Wartekontrollgruppe müsste, wie die zwei Interventionsgruppen, drei Messpunkte beinhalten (zusätzlich T2). Weiter müsste das Forschungsinstrument angepasst und im Bereich der zu untersuchenden Kompetenzen zusätzlich erweitert werden, um die Reliabilität weiter zu erhöhen.

5.4 Fazit und Ausblick

Die vorliegende explorative Studie ist eine der ersten empirischen Forschungen, die zwei konkrete Aufgaben zum Aufbau des Modellverständnisses in einer nur zweistündigen Intervention untersucht. Für die Untersuchung wurden die BB-Aufgabe und die SP-Aufgabe gewählt, die von Lederman und Abd-El-Khalik (2002) beschrieben wurden. Die Masterarbeit ging der Frage nach, welcher dieser beiden Ansätze besser geeignet ist, damit die Schüler:innen die Welt der Atome als erdachte Welt zur sichtbaren Welt betrachten und akzeptieren lernen. Die Ergebnisse unserer explorativen Studie geben erste Hinweise darauf, dass die BB-Aufgabe besser geeignet ist, um dieses Ziel zu erreichen. Die didaktische Analyse der beiden Ansätze lässt vermuten, dass das handelnde Lernen das Einführen des Modellverständnisses gewinnbringend beeinflusst. Um diese ersten explorativen Resultate zu erhärten, müsste eine breitangelegte Mehrebenen-analytische Studie durchgeführt werden, die im Rahmen der Masterarbeit nicht möglich war.

In Zukunft wäre spannend, weitere Variablen, wie beispielsweise der Geschlechterunterschied beim Erlernen des Modellverständnisses zu untersuchen. Der fehlende Unterschied zwischen Mädchen und Jungen beim Kompetenzzuwachs in dieser explorativen Studie (Resultate nicht gezeigt), könnte darauf hindeuten, dass beide Ansätze geschlechtsunabhängig das Modellverständnis aufbauen. Eine breitangelegte Mehrebenen-analytische Studie könnte den Genderaspekt jedoch deutlich vertiefter untersuchen.

Weiter weisen mehrere Artikel darauf hin, dass die Lernendenvorstellungen unter anderem durch fehlende fachdidaktische Kompetenzen der Lehrpersonen zustande kommen (Bindernagel & Eilks, 2008; Rehm, 2007). Den Effekt des in dieser Studie gewählten Vorgehens, den Lehrpersonen neben den Arbeitsblättern, Lektionspräparationen und Tests anzubieten und bei Unklarheiten Fragen stellen zu können, könnte ebenfalls mehrebenen-analytisch in einer Interventionsstudie beforcht werden.

References

- Ahtee, M. & Variola, I. (1998). Students' Understanding of Chemical Reactions. *International Journal of Science Education*, 20(3), 305–316.
- American Association for the Advancement of Science [AAAS]. (1993). *Benchmarks for Science Literacy*. Oxford University Press.
- American Association for the Advancement of Science [AAAS]. (2001). *Atlas of science literacy: Mapping K-12 learning and goals*. Washington, DC: Author.
- Beerenwinkel, A., Berset, T., Durrer, K., Herger, H., Iten, M., & Stettler, A. (2019). *Prisma 1 - Natur und Technik Sekundarstufe I Ausgabe für die Schweiz*. Klett und Balmer.
- Bölsterli Bardy, K., Stalder, G., Brun, R., Schmid, M. (2021) *Anorganische Chemie und Atombau*. PH Luzern.
- Dubach, J., Schelleis, N., Bölsterli, K., Cacchione, T., Martarelli, C. S., Probst, M., & Tempelmann, S. (2022). *Unzugängliche Welten für das erfahrungsbasierte Lernen erschließen. Immersive Virtuelle Realität im naturwissenschaftlichen Sachunterricht*. In A. Becher, E. Blumberg, T. Goll, K. Michalik, & C. Tenberge (Eds.), *Probleme und Perspektiven des Sachunterrichts. Sachunterricht in der Informationsgesellschaft*, Vol. 32, (pp. 147-154). Klinkhardt. <https://doi.org/10.25656/01:24211>
- Beerenwinkel, A. (2006). *Fostering conceptual change in chemistry classes using expository texts* [Dissertation]. Universität Wuppertal, Wuppertal. <http://elpub.bib.uni-wuppertal.de/edocs/dokumente/fbg/paedagogik/diss2006/beerenwinkel/dg0602.pdf>
- Bindernagel, J. A. & Eilks, I. (2008). Der "Roadmap-Ansatz" zur Beschreibung des Pedagogical Content Knowledge von Der "Roadmap-Ansatz" zur Beschreibung des Pedagogical Content Knowledge von Lehrkräften zum Teilchenkonzept. In D. Höttecke (Hrsg.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung: [34.] Jahrestagung in Essen [vom 17. bis 20. September] 2007 ; [Gesellschaft für Didaktik der Chemie und Physik, Band 28]*. LIT-Verlag.
- Bowling, A. (2002). *Research Methods in Health: Investigating health and health services*. Open University Press.
- Brademann, T. (1997). Grundsätze und Beispiele zur Arbeit mit physikalischen Denkmodellen im Physik-unterricht der Sekundarstufe I. *Physik in der Schule*, 35(6), 207–210.
- Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.
- Carey, S., Evans, R., Honda, M., Jay, E. & Unger, C. (1989). 'An experiment is when you try it and see if it works': a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11(5), 514–529. <https://doi.org/10.1080/0950069890110504>
- Chittleborough, G. D., Treagust, D. F., Mamiala, T. L. & Mocerino, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science & Technological Education*, 23(2), 195–212. <https://doi.org/10.1080/02635140500266484>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Eds.). Lawrence Erlbaum Associates.

- Crawford, B., & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong & Eijkelhof H. (Hrsg.), *Research and the quality of science education* (309-323). Springer.
- D-EDK. (2013). *Lehrplan 21 Natur, Mensch, Gesellschaft: Konsultationsfassung*. http://projekt.lehrplan.ch/lehrplan/V2/container/31_6_0_0_1_1.pdf
- D-EDK. (2015). *Lehrplan 21: Natur und Technik mit Physik, Chemie, Biologie Kompetenzaufbau 3. Zyklus*. http://projekt.lehrplan.ch/lehrplan/V5/ablage/FS1F_Fachbereich_NT_3.Zyklus.pdf
- Devetak, I., Urbančič, M., Wissiak Grm, K. S. & Krnel, D. (2004). Submicroscopic Representations as a tool for evaluating students' chemical conceptions. *Acta Chim. Slov.*, 51, 799–814.
- Diekmann, A. (2007). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen (rororo Rowohlt's Enzyklopädie)* (17. Aufl. Orig.-Ausg.). Rowohlt-Taschenbuch-Verl.
- Driver, R., Leach, J., Millar, R. & Scott, P. (1996). *Young People's Images of Science*. Open University Press.
- Dubs, R. (2009). *Lehrerverhalten: ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht* (2., vollständig neu bearb. Aufl.). Franz Steiner Verlag.
- Escher, D. & Messner, H. (2015). *Lernen in der Schule*. hep verlag.
- Froncek, B. & Thielsch, M., T. (2011). Merkmale guter schriftlicher Prüfungen (Materialien der Sektion Aus-, Fort- und Weiterbildung in Psychologie im Berufsverband Deutscher Psychologinnen und Psychologen). In M. Krämer & Preiser, S. Brusdeylins-Hammer, K. (Hrsg.), *Psychologiedidaktik und Evaluation VIII (Berichte aus der Psychologie)* (S. 365–372). Shaker.
- Funk, S. C. & Dickson, K. L. (2011). Multiple-Choice and Short-Answer Exam Performance in a College Classroom. *Teaching of Psychology*, 38(4), 273–277.
- Girden, E. R. (1992). *ANOVA: Repeated measures. Sage university papers. Quantitative applications in the social sciences: no. 07-084*. Newbury Park, Calif.: Sage Publications.
- Griffiths, A. K. & Preston, K. R. (1992). Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. *Journal of Research in Science Teaching*, 29(6), 611–628. <https://doi.org/10.1002/tea.3660290609>
- Grosslight, L., Unger, C [C.], Jay, E [E.] & Smith, C. (1991). Understanding Models and their Use in Science: Conceptions of Middle and High School Students and Experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grunder, H.-U., Ruthemann, U., Scherer, S., Singer, P. & Vettiger, H. (Hrsg.). (2014). *Unterricht: verstehen - planen - gestalten - auswerten* (4. unveränd. Aufl.). Schneider Hohengehren.
- Hansson, L., Arvidsson, Å., Heering, P. & Pendrill, A.-M. (2019). Rutherford visits middle school. a case study on how teachers direct attention to the nature of science through a storytelling approach. *Physics Education*, 54(4), 45002. <https://doi.org/10.1088/1361-6552/ab07e7>
- Harrison, A. G. & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026. <https://doi.org/10.1080/095006900416884>
- Henze, I., van Driel, J. H. & Verloop, N. (2007). Science Teachers' Knowledge about Teaching Models and Modelling in the Context of a New Syllabus on Public Understanding of Science. *Research in Science Education*, 37(2), 99–122. <https://doi.org/10.1007/s11165-006-9017-6>
- Horton, C. (2009). *Student Alternative Conceptions in Chemistry: (Published without Appendix 2 and with abridged References in the California Journal of Science Education, 7(2)2007)*. Arizona State University. <http://modeling.asu.edu/Projects-Resources.html>
- Jaber, L. Z. & BouJaoude, S. (2012). A Macro–Micro–Symbolic Teaching to Promote Relational Understanding of Chemical Reactions. *International Journal of Science Education*, 34(7), 973–998. <https://doi.org/10.1080/09500693.2011.569959>
- Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
- Justi, R. & Gilbert, J. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369–1386. <https://doi.org/10.1080/0950069032000070324>
- Kind, V. (2004). *Beyond Appearances: Students' misconceptions about basic chemical ideas: A report prepared for the Royal Society of Chemistry*. https://www.researchgate.net/publication/228799159_Beyond_Appearances_Students%27_Misconceptions_About_Basic_Chemical_Ideas
- Koo, T, K. & Li, M., Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Krell, M., Reinisch, B. & Krüger, D. (2015). Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Research in Science Education*, 45(3), 367–393. <https://doi.org/10.1007/s11165-014-9427-9>
- Krey, O. & Schwanewedel, J. (2018). Lernen mit externen Repräsentationen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 159–176). Springer Berlin Heidelberg.
- Krüger, D., Kauertz, A. & Upmeyer zu Belzen, A. (2018). Modelle und das Modellieren in den Naturwissenschaften. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 141–157). Springer Berlin Heidelberg.
- Lederman, N., & Abd-El-Khalick, F. (2002). Avoiding De-Natured Science: Activities that Promote Understandings of the Nature of Science. In McComas, William, F. (Hrsg.), *Science & Technology Education Library. The Nature of Science in Science Education* (Bd. 5, S. 83–126). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47215-5_5

- Lederman, N. G. (2002). The state of science education: Subject matter without context. *Electronic Journal of Science Education*, 3(2).
- Lederman, Norman, G., Abd-El-Khalick, F. & Lederman, Judith, Sweeney. (2020). Chapter 17 Avoiding De-Natured Science: Integrating Nature of Science into Science Instruction. In McComas, William, F. (Hrsg.), *Nature of Science in Science Instruction* (S. 295–326). Springer International Publishing.
- Leisner, A. (2005). *Entwicklung von Modellkompetenz im Physikunterricht. Eine Evaluationsstudie in der Sekundarstufe I*. Dissertation. Logos Verlag.
- Leisner-Bodenthin, A. (2006). Zur Entwicklung von Modellkompetenz im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 91–109. https://archiv.ipn.uni-kiel.de/zfdn/pdf/006_12.pdf
- Levene, H. (1960). Robust Tests for Equality of Variances. In I. Olkin (Hrsg.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (S. 278–292). University Press.
- Lijnse, P. L., Licht, P., Vos, W. de & Waarlo, A. J. (Hrsg.). (1990). *Relating Macroscopic Phenomena to Microscopic Particles: A central problem in secondary science education*. CD-β Press.
- Löffler, G. (1996). Über die Grundlegung einer Verständigung über den Begriff des Atoms Zur Erinnerung an Martin Wagenscheins 100. Geburtstag. *Chimica et Didacticae*, 22, 296–322.
- McComas, William, F. (Hrsg.). (2002). *Science & Technology Education Library. The Nature of Science in Science Education*. Kluwer Academic Publishers.
- McComas, William, F. (2004). Keys to teaching the nature of science: Focusing on the nature of science in the science classroom. *The Science teacher*, 71(9), 24–27.
- McComas, William, F. (2020a). Chapter 3: Principal Elements of Nature of Science: Informing Science Teaching while Dispelling the Myths. In McComas, William, F. (Hrsg.), *Nature of Science in Science Instruction* (S. 35–66). Springer International Publishing.
- McComas, William, F. (Hrsg.). (2020b). *Nature of Science in Science Instruction*. Springer International Publishing.
- McComas, William, F. & Clough, Michael, P. (2020). Chapter 1: Nature of Science in Science Instruction: Meaning, Advocacy, Rationales, and Recommendations. In McComas, William, F. (Hrsg.), *Nature of Science in Science Instruction* (S. 3–22). Springer International Publishing.
- Meisert, A. (2008). Vom Modellwissen zum Modellverständnis - Elemente einer umfassenden Modellkompetenz und deren Fundierung durch lernerseitige Kriterien zur Klassifikation von Modellen. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243–261.
- Mikelskis-Seifert, S. (2002). *Die Entwicklung von Metakzepten zur Teilchenvorstellung bei Schülern*. Logos Verlag.
- Mikelskis-Seifert, S. (2009). Lernen über Modelle. Am Beispiel der Teilchenstruktur der Materie. *Praxis der Naturwissenschaften - Chemie in der Schule*, 58(7), 15–21.
- Mikelskis-Seifert, S. (2010). Denken in und mit Modellen. *Piko-Brief*(8), 1–7.
- Mikelskis-Seifert, S. & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen - Stand der Forschung und Entwurf einer Unterrichtskonzeption. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 75–88. ftp://ftp.rz.uni-kiel.de/pub/ipn/zfdn/2003/5.Seifert_Fischler_75-88.pdf
- Mikelskis-Seifert, S. & Leisner, A. (o. A.). Das Denken in Modellen fördern ein Unterrichtsbeispiel zur Entwicklung von Teilchenvorstellungen, 1–6. <https://docplayer.org/36570434-Das-denken-in-modellen-foerdern-ein-unterrichtsbeispiel-zur-entwicklung-von-teilchenvorstellungen.html>
- Millar, R. (1990). Making sense: What use are particle ideas to children? In P. Lijnse, P. Licht, W. de Vos, & A. J. Waarlo (Hrsg.), *Relating macroscopic phenomena to microscopic particles* (S. 283–293). CD-[beta] Press]. In P. L. Lijnse, P. Licht, W. de Vos & A. J. Waarlo (Hrsg.), *Relating Macroscopic Phenomena to Microscopic Particles: A central problem in secondary science education* (S. 283–292). CD-β Press. http://www.fi.uu.nl/publicaties/literatuur/1990_lijnse_phenomena_particles.pdf
- Mulford, D. R. & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *Journal of Chemical Education*, 79(6), 739. <https://doi.org/10.1021/ED079P739>
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R. & Duschl, R. (2003). What “ideas-about-science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40(7), 692–720.
- Passmore, C. (2015). Models. In R. Gunstone (Hrsg.), *Encyclopedia of Science Education* (S. 659–663). Springer Netherlands.
- Peters-Burton, Erin, E. & Burton, Stephen, R. (2020). Chapter 9: The Use of Metacognitive Prompts to Foster Nature of Science Learning. In McComas, William, F. (Hrsg.), *Nature of Science in Science Instruction* (S. 179–198). Springer International Publishing.
- Rehm, M. (2007). Crossing the bridge from the macro to the micro world. *NVOX*.
- Rehm, M. (2009). Der Sprung zu den Atomen. *Unterricht Chemie*, 20(114), 5–8.
- Rehm, M. (Hrsg.). (2018). *Unterrichtspraxis: Perspektiven von Expertinnen und Experten: Bd. 2. Wirksamer Chemieunterricht* (1. Aufl.). Schneider Hohengehren.
- Rehm, M. & Buck, P. (2006). Der Teil und das Ganze - ein Lehr-Lern-Arrangement vor der Einführung von Atommodellen. In H. Fischler & C. S. Reiners (Hrsg.), *Die Teilchenstruktur der Materie im Physik- und Chemieunterricht*. Logos Verlag.
- Rehm, M. & Buck, P. (2009). How can atoms, the foundation of chemistry, be taught adequately in the 21st century? *Problems of Education In the 21st Century*, 17, 25–38.
- Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist*, 46(10), 1086–1087. <https://doi.org/10.1037/0003-066X.46.10.1086>

- Schwartz, M. S., Sadler, P. M., Sonnert, G. & Tai, R. H. (2009). Depth versus breadth: How content coverage in high school science courses relates to later success in college science coursework. *Science Education*, 93(5), 798–826. <https://doi.org/10.1002/sce.20328>
- Schwarz, C. & White, B. (2005). Metamodeling knowledge: developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165–205.
- Séré, M. G. (1990). Passing from one model to another: which strategy? (The CD-[beta] series on research in education). In P. L. Lijnse, P. Licht, W. de Vos & A. J. Waarlo (Hrsg.), *Relating Macroscopic Phenomena to Microscopic Particles: A central problem in secondary science education* (S. 50–66). CD-β Press.
- Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R. & van Hout-Wolters, B. H. A. M. (2009). The Relation between Students' Epistemological Understanding of Computer Models and their Cognitive Processing on a Modelling Task. *International Journal of Science Education*, 31(9), 1205–1229. <https://doi.org/10.1080/09500690802192181>
- Stachowiak. (1973). *Allgemeine Modelltheorie*. Springer.
- Talanquer, V. (2011). Macro, Submicro, and Symbolic: The many faces of the chemistry “triplet”. *International Journal of Science Education*, 33(2), 179–195. <https://doi.org/10.1080/09500690903386435>
- Tempel, B. J., Randler, C., Rehm, M. & Wilhelm, M. (2018). Model competences in chemistry and biology lessons - what skills do teachers need? A systematic literature review; Modellkompetenzen im Chemie- und Biologieunterricht - welche Fähigkeiten brauchen Lehrkräfte? Ein systematisches Literaturreview. *Progress in Science Education*, 1(1). <https://doi.org/10.25321/PRISE.2018.471>
- Treagust, D., Chittleborough, G. & Mamiala, T. (2002). Student's understanding of the role of scientific models in learning science. *International Journal of Science Education*, 24(4), 357–368. <https://doi.org/10.1080/09500690110066485>
- Upmeyer zu Belzen, A [A.] & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Vogel, D. (2014). Entdeckendes Lernen. In PH Luzern (Hrsg.), *Grundlagen und Grundformen des Unterrichtens. Studiendband Grundjahr-Mentorat 1. und 2. Semester* (S. 79–86).
- Vos, W. de. (1990). Seven Thoughts on Teaching Molecules. In P. L. Lijnse, P. Licht, W. de Vos & A. J. Waarlo (Hrsg.), *Relating Macroscopic Phenomena to Microscopic Particles: A central problem in secondary science education* (S. 163–176). CD-β Press. http://www.fi.uu.nl/publicaties/literatuur/1990_lijnse_phenomena_particles.pdf
- Wickelgren, W. A. (1975). Alcoholic intoxication and memory storage dynamics. *Memory & Cognition* (3), 385–389.
- Wilhelm, M., Luthiger, H. & Wespi, C. (2014). *Prozessmodell zur Entwicklung von kompetenzorientierten Aufgabensets*. Pädagogische Hochschule Luzern, Luzern.
- Zobrist, B. (2014). Kooperatives Lernen. In PH Luzern (Hrsg.), *Grundlagen und Grundformen des Unterrichtens. Studiendband Grundjahr-Mentorat 1. und 2. Semester* (S. 87–94).