

## **Empirische Struktur von Modellnutzung im Chemieunterricht**

### **Motivation & Problemstellung**

Die allgegenwärtige Verwendung von Modellen in den Naturwissenschaften wird sowohl in den Bildungsstandards als auch in der fachdidaktischen Forschung widerspiegelt. Das breite Verständnis von Modellen führt dabei zu begrifflichen und empirischen Herausforderungen. Während beispielsweise der Rahmenlehrplan für das Fach Chemie in der Sekundarstufe I in Berlin vorsieht, dass Schülerinnen und Schüler (SuS) „Modelle bezüglich ihrer Eignung prüfen [...]“ sollen (Berliner Senatsverwaltung für Bildung, Jugend und Familie, 2016), ist gar nicht expliziert, welche Kriterien zur Eignung vorliegen. Zusammenfassend sollen SuS mithilfe von Modellen eine hypothetisch-deduktive Denkweise erlernen und einüben, ohne dass vollständig geklärt wäre, wie diese Anforderung beschrieben und strukturiert werden soll. Das kann – außerhalb solcher normativen Überlegungen – auch als Desiderat aus dem aktuellen Forschungsstand abgeleitet werden (Krüger, Kauertz & Upmeier zu Belzen, 2018; Gilbert & Justi, 2016). In empirischer Hinsicht sind Forschungsergebnisse zur Nutzung von Modellen als Mittel der Erkenntnisgewinnung und zur Modellierung, d.h. der selbstständigen Konstruktion von Modellen auf Basis eines Phänomens, sehr heterogen (Nicolaou & Constantinou, 2014). Für diese empirische Heterogenität wird ein uneinheitlicher Begriffsrahmen verantwortlich gemacht (ebd., S. 72), der im vorliegenden Projekt auf die nicht explizierten epistemologischen Grundannahmen zurückgeführt wird: Die Annahme, ein Modell sei nützlich, wenn es eine korrekte, strukturanaloge Abbildung eines Originals wäre, setzt bei der Eignungsprüfung voraus, dass das Original überhaupt strukturell zu erkennen wäre. Das ist bei submikroskopischen Teilchen beliebig schwierig. Die Annahme, ein Modell sei hingegen nützlich, wenn es dem Original nur vorbehaltlich strukturell analog wäre und vor allem für die Generierung prüfbarer Hypothesen diene, setzt bei der Eignungsprüfung keinen Original-Modell-Vergleich voraus. Aus den beiden exemplarischen Annahmen leiten sich jeweils andere Kriterien für eine empirische Fähigkeits- und/oder Dimensionalitätsprüfung ab. Aus diesem Grund wird eine Forschungslücke identifiziert, die kontext- und chemiespezifische Aufgaben für das Durchlaufen des hypothetisch-deduktiven Denkens mithilfe von Modellen fordert. Deren Konstruktion und Prüfung soll auf die epistemologischen Grundannahmen zurückgeführt und anschließend daran reflektiert werden, um einen Beitrag zur Begriffsklärung zu leisten.

### **Theoretischer Hintergrund & Fragestellung**

Modelle zum Durchlaufen hypothetisch-deduktiver Denkschritte (Nehring, Nowak, Upmeier zu Belzen & Tiemann, 2015) oder das Modellieren in problemorientierten Umgebungen (Schwarz, Reiser, Davis, Kenyon, Achér, Fortus, Shwartz, Hug & Krajcik, 2009) sind bereits dokumentiert und geprüft worden. Als theoretische Rahmung für eine zweckorientierte Modellnutzung sind der wissenschaftliche Realismus (z.B. Devitt, 2006) und der konstruktive Empirismus (z.B. Van Fraassen, 1980) zielführend. Ersterer geht von der realen Existenz von Atomen aus, die sich wie kleine Kugeln verhalten, weil sie die beste Erklärung für beobachtete Phänomene sind. Zweiterer zweifelt deren Existenz nicht notwendigerweise an, stellt aber die Frage nach der Realität ihrer Eigenschaften in den Hintergrund. Wichtig sei nicht, ob Atome wirklich so existieren, sondern ob ihre vorbehaltlich formulierten Eigenschaften „das Phänomen retten.“ (Van Fraassen, 1980, S. 41 ff., Übers. d. Autors).

Diese Unterscheidung ist für Lehr-/Lernprozesse keineswegs trivial. Das Planetenmodell von Atomen wird im Schulunterricht völlig selbstverständlich neben Lewis-Strukturformeln verwendet. Im ersten Fall bewegen sich die Elektronen in kreisförmigen Bahnen um den Atomkern, im zweiten Fall wird impliziert, dass Elektronen feste räumliche Positionen einnehmen. Gleichzeitig ist beides nicht möglich. Es wird hier angenommen, dass der Realitätsgehalt beider Modelle keine dominante Rolle in Erkenntnisgewinnungsprozessen spielt, sondern viel mehr ihre Zweck- und Situationsgebundenheit (vgl. Krüger et al., 2018). Diese Position wird in aktuellen epistemologischen Diskussionen aufgegriffen. Mahr (2015) unterscheidet das Modellobjekt deutlich vom zu Modellierenden Sachverhalt und formuliert zwei Perspektiven auf das Modellobjekt, nämlich Modell *von* etwas und Modell *für* etwas zu sein. Die subjektgesteuerte Zweckbestimmung eines Modells wird so von der Frage getrennt, wie real es den jeweiligen Sachverhalt tatsächlich abbildet. Außerdem unterscheidet er auf diese Weise prinzipiell zwischen der Tätigkeit des Modellierens und der Modellnutzung. Einen noch stärkeren Fokus auf das Subjekt (hier: den Modellnutzer) legt Knuuttila (2011), indem Modelle als *epistemische Werkzeuge* (ebd.) keinen unmittelbar repräsentativen (d.h. strukturell abbildenden) Charakter mehr benötigen, um für Erkenntnisgewinnung nützlich zu sein (ebd., S. 266). Für einen Übergang zu empirischen Untersuchungen bedarf diese theoretische Rahmung aber einer Konkretisierung für eine einheitliche Aufgabenkonstruktion (Terzer, Hartig & Upmeyer zu Belzen, 2013). Einerseits sollen die problemhaltigen Aufgaben chemiespezifisch sein, andererseits soll ein generalisierbares Framework verwendet werden, um die Aufgaben vergleichbar zu halten. Zu diesem Zweck wurde eine Beschreibung für Modelle (Lesh & Doerr, 2003) verwendet, die in der chemiedidaktischen Forschung bereits eingesetzt wurde (Schwarz et al., 2009; Bodner & Briggs, 2005), dessen Dimensionalitätsprüfung bisher aber aussteht. Danach können Modellobjekte kontextunabhängig mit vier Komponenten beschrieben werden.

- Die *elements* sind die kleinsten Sinneinheiten in einem Modell (bspw. Elektronen).
- Die *relations* sind Verhältnisse zwischen den Sinneinheiten (bspw. Energiedifferenzen).
- Die *operations* sind Veränderungen der Verhältnisse (bspw. Wechsel eines Energieniveaus).
- Die *rules* sind formallogische Verknüpfungen zur Verknüpfung der anderen drei Komponenten (bspw. Wenn-Dann-Sätze).

Wären diese Komponenten empirisch in problemorientierten, chemiespezifischen Aufgaben wiederzufinden, so könnte zu einem stärker verallgemeinerbaren Begriffsrahmen für hypothetisch-deduktive Modellnutzung beigetragen werden.

Die leitende Forschungsfrage des Projekts lautet: *Können Aufgaben konstruiert werden, die die theoretisch angenommene vierdimensionale Struktur von Modellen in chemiespezifischen Aufgaben empirisch widerspiegeln?*

### **Methode**

Die Konstruktion der Aufgaben wurde in vier Hauptschritten durchgeführt. *A)* Nach einem Expertenrating zur Zuordnung der Modellbestandteile in verschiedenen Kontexten ( $N = 7$ ,  $\kappa = .43-.82$ ) und anschließender Überarbeitung mit Rücksprache, wurden offene Aufgaben konzipiert. *B)* SuS ( $N = 42$ ) der 9. Jahrgangsstufe eines Berliner Gymnasiums wurden aufgefordert bspw. die kleinsten Sinneinheiten („Beschreibe die einzelnen Bestandteile des Modells!“) eines Modells zu benennen. *C)* Mit den so gewonnenen Informationen wurde ein Konstruktionsmanual für Multiple-Choice-Aufgaben (1 aus 4) erstellt, von denen 60 Aufgaben in einer quantitativen Pilotstudie in der 10. Jahrgangsstufe von Berliner Gymnasien und Sekundarschulen bearbeitet wurden ( $N = 254$ ). Es wurde eine Distraktorenanalyse, eine deskriptive Untersuchung von Zusammenhängen mit Kovariablen (Lesegeschwindigkeit- und Verständnis, kognitive Fähigkeiten, Chemie-, Deutsch- und

Mathematiknote, Interesse an den Kontexten, kontextspezifisches Fachwissen) und eine Rasch-Skalierung ( $wMNSQ_{min} = .89$ ,  $wMNSQ_{max} = 1.15$ ,  $EAP/PV = .41$ ) der Aufgaben vorgenommen. Zudem gibt es aus dem Datensatz der Vorstudie bereits empirische Hinweise auf eine Mehrdimensionalität nach der DETECT-Statistik ( $DETECT > 1.0$ ,  $ASSI > .25$ ,  $RATIO > .36$ , vgl.: Jang & Roussos, 2007; Zhang, 2007). D) Parallel wurde das Konstruktionsmanual überarbeitet und die 60 Aufgaben finalisiert, indem bspw. die Anzahl der vorhandenen Fachbegriffe oder Abbildungen so einheitlich wie möglich gestaltet wurden. Mit den finalen Aufgaben wurde die Hauptstudie an Berliner Gymnasien in der 10. Jahrgangsstufe durchgeführt ( $N = 524$ ). Aufgrund der hohen Missing-by-Design-Rate Rechnung zu tragen, wurde eine konfirmatorische Faktorenanalyse unter dem IRT-Paradigma angewendet (Hartig & Frey, 2013; Chalmers, 2012; Hartig & Höhler, 2009). Dabei wurden die in der Vorstudie geschätzten Aufgabenschwierigkeiten ( $M = .62$ ,  $SD = .65$ ) unter Normalverteilungsannahme (Shapiro-Wilk:  $W = .97$ ,  $p = .19$ ) als Prior gesetzt, um die Berechnung zu stabilisieren. Für den Modellvergleich wurde ein G-Faktor als plausibel angenommen. Dieser bildet im Messmodell die übergeordnete Fähigkeit zur Modellnutzung ab.

### Ergebnisse

Beim Modellvergleich nach bayesianischem Ansatz, ist das G-Faktormodell zwar eine gute Erklärung für die Daten, unter dem Aspekt der Modellsparbarkeit wäre aber dem eindimensionalen Modell der Vorrang zu geben (Tabelle 1). Die eindeutige Rückführung der Ergebnisse auf eine vierdimensionale Struktur kann darum nicht aufrechterhalten werden. Weil der Bayes-Faktor aber einen direkt interpretierbaren Zugang auf die Wahrscheinlichkeit für oder gegen ein Modell bietet (Stefan, Gronau, Schönbrodt & Wagenmakers, 2019) und der direkte Modellvergleich allein an Informationskriterien ambivalent diskutiert wird (Fox, 2010), gibt es auch empirische Evidenz für das komplexere Modell. Die EAP-Reliabilitäten für den eindimensionalen Fall ( $EAP/PV = .52$ ), sowie für die Dimensionen des G-Faktormodells ( $EAP/PV = .17-.52$ ) sind niedrig, liegen aber im üblichen Bereich für diese Art Fähigkeitstests (Hartmann, Upmeyer zu Belzen, Krüger & Pant, 2015).

Tabelle 1: Modellvergleich von ein- und vierdimensionaler Skalierung, sowie unter Annahme einer G-Faktorstruktur.

N (Dim)	AIC	BIC	logLik	df	Bayes-Faktor
1	8881	9126	-4382	-	-
4	8936	9219	-4401	9	>100
4 + G	8963	9475	-4360	63	0

### Ausblick

Die Befunde stützen den Versuch, einen über alle verwendeten Problemkontexte generalisierbaren Ansatz zur Modellnutzung für hypothetisch-deduktives Denken zu beschreiben. Es ist die verhältnismäßig geringe Fallzahl zu berücksichtigen, die höherdimensionale Strukturschätzungen schwierig macht. Die strenge Standardisierung zur Operationalisierung der epistemologischen Grundannahmen und das Multiple-Choice-Format lassen wenig Spielraum für einen kreativen Umgang mit problemhaltigen Umgebungen. Der iterative Charakter der Modellnutzung (Schwarz et al., 2009) wird ebenfalls nicht berücksichtigt. Trotzdem können die empirischen Hinweise auf die Mehrdimensionalität und der theoretische Unterbau konsistent aufeinander bezogen werden: Die Modellkomponenten erklären einen Teil der Antwortmuster. So können zukünftig Problemlösestrategien unter expliziter Modellnutzung zielgenauer realisiert werden, als es bisher der Fall ist (Nicolaou & Constantinou, 2014).

## Literatur

- Berliner Senatsverwaltung für Bildung, Jugend und Familie (2016), Rahmenlehrpläne Teil C. Chemie. Jahrgangsstufen 7-10. URL: [https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche\\_Fassung/Teil\\_C\\_Chemie\\_2015\\_11\\_10\\_WEB.pdf](https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/rahmenlehrplaene/Rahmenlehrplanprojekt/amtliche_Fassung/Teil_C_Chemie_2015_11_10_WEB.pdf). Zuletzt abgerufen: 12.10.2019.
- Bodner, G. M. & Briggs, M. W. (2005), A Model of Molecular Visualization. In J. K. Gilbert (Hg.), *Visualization in Science Education*. Dordrecht: Springer, 90-105.
- Chalmers, R. P. (2012), mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29.
- Devitt, M. (2006), Scientific Realism. In P. Greenough & M. P. Lynch (Hg.), *Truth and Realism*. Oxford: Oxford University Press, 100-124.
- Fox, J.-P. (2010). *Bayesian Item Response Modeling. Theory and Applications*. New York, Dordrecht, Heidelberg, London: Springer.
- Gilbert, J. K. & Justi, R. (2016). Models of Modelling. In J. K. Gilbert & R. Justi (Hg.), *Modelling-based Teaching in Science Education. Models and Modeling in Science Education 9*. Switzerland: Springer, 17-40.
- Hartig, J. & Frey, A. (2013), Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? *Zeitschrift für Erziehungswissenschaft*, 16, 47-51.
- Hartig, J., Höhler, J. (2009), Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35, 57-63.
- Hartmann, S.; Upmeyer zu Belzen, A.; Krüger, D. & Pant, H. A. (2015), Scientific Reasoning in Higher Education. Constructing and Evaluating the Criterion-Related Validity of an Assessment of Preservice Science Teachers' Competencies. *Zeitschrift für Psychologie*, 223(1), 47-53.
- Jang, E. E. & Roussos, L. (2007), An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44, 1-21.
- Knuutila, T. (2011), Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42, 262-271.
- Krüger, D.; Kauertz, A. & Upmeyer zu Belzen, A. (2018). Modelle und das Modellieren in den Naturwissenschaften. In D. Krüger, H. Schecker & I. Parchmann (Hg.), *Theorien in der naturwissenschaftsdidaktischen Forschung*. Berlin: Springer, 141-157.
- Lesh, R. & Doerr, H. M. (2003), Foundations of models and modeling perspective on mathematics teaching, learning, and problem solving. In R. Lesh & H. M. Doerr (Hg.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah (NJ): Erlbaum, 3-33.
- Mahr, B. (2015), Modelle und ihre Befragbarkeit. *Grundlagen einer allgemeinen Modelltheorie. Erwägen Wissen Ethik – Streitforum für Erwägungskultur*, 26(3), 329-342.
- Nehring, A.; Nowak, K. H.; Upmeyer zu Belzen, A. & Tiemann, R. (2015), Predicting Students' Skills in the Context of Scientific Inquiry with Cognitive, Motivational, and Sociodemographic Variables. *International Journal of Science Education*, 37(9), 1343-1363.
- Nicolaou, C. T. & Constantinou, C. P. (2014), Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52-73.
- Schwarz, C. V.; Reiser, B. J.; Davis, E. A.; Kenyon, L.; Achér, A.; Fortus, D.; Schwartz, Y.; Hug, B. & Krajcik, J. (2009), Developing a Learning Progression for Scientific Modeling: Making Scientific Modeling Accessible and Meaningful for Learners. *Journal of Research in Science Teaching*, 46(6), 632-654.
- Stefan, A. M.; Gronau, Q. F.; Schönbrodt, F. D. & Wagenmakers, E.-J. (2019). A tutorial on Bayes Factor Design Analysis using an informed prior. *Behavior Research Methods*, 51, 1042-1058.
- Terzer, E.; Hartig, J. & Upmeyer zu Belzen, A. (2013), Systematische Konstruktion eines Tests zu Modellkompetenz im Biologieunterricht unter Berücksichtigung von Gütekriterien. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 51-76.
- Van Fraassen, B. (1980), Arguments Concerning Scientific Realism. In B. Van Fraassen (Hg.), *The Scientific Image*. Oxford: Oxford University Press, 6-40.
- Zhang, J. (2007), Conditional covariance theory and DETECT for polytomous items. *Psychometrika*, 72, 69-91.