

Marvin Rost<sup>1</sup>  
 Ines Sonnenschein<sup>2</sup>  
 Stephanie Möller<sup>2</sup>  
 Anja Lembens<sup>1</sup>

<sup>1</sup>AECC Chemie (Universität Wien)  
<sup>2</sup>Fachhochschule Münster

## **Übersetzung und Pilotierung des SUMS-Inventary für Chemiestudierende Eine Replikationsstudie**

### **Einführung**

Mit dem steigenden Bedarf an Energiespeichertechnologien, bspw. „Power-to-Gas“, steigt auch der Bedarf an Fachkräften mit chemiebezogenen Kompetenzen. Im Chemieingenieursstudium der FH Münster werden, wie in zahlreichen MINT-Studiengängen, allerdings hohe Abbruchquoten registriert. In Modellierungen für die Vorhersage solcher Studienabbrüche, können zahlreiche Einflussfaktoren identifiziert werden (Fleischer et al., 2019). Dabei kommt dem Vorwissen aus der Schule als kognitive Facette eine besondere Bedeutung zu, weil es direkt auf die Abbruchintention wirkt und nicht, wie bspw. Interesse oder Engagement, über die Zufriedenheit mit den Studieninhalten mediiert wird. Um die Einführungsvorlesung in die Chemie bei Chemieingenieur\*innen, Physiker\*innen und Ökotropholog\*innen anschlussfähiger zu gestalten, wurde darum beschlossen, das Vorwissen der Studierenden stärker in die Vorlesungsinhalte einzubeziehen und die Präsentation von Fachinhalten konsequenter an Vorstellungen über naturwissenschaftliche Modellierungsprozesse anzubinden. Zu diesem Zweck bedarf es eines verlässlichen Instruments, das das modellorientierte Verstehen ökonomisch erfasst, um es anschließend auf die jeweiligen Prüfungserfolge, bzw. die entsprechende Abbruchsintention beziehen zu können.

### **Theorie**

Replikationsstudien sind in den Fachdidaktiken verhältnismäßig wenig dokumentiert (Makel & Plucker, 2014). Amrhein et al. (2019) diskutieren das Gleichgewicht zwischen sozialwissenschaftlichem Erkenntnisinteresse, Theorie und statistischer Abbildung am Beispiel des  $p$ -Werts und geben zu bedenken, dass die Beforschung menschlichen Verhaltens ohne epistemologische, wissenschaftstheoretische oder soziologische Perspektiven unzulässig verengt ist. Allerdings wird dieses Gleichgewicht auch dadurch beeinflusst, dass das systematische Berichten von standardisierten Kennwerten für quantitative Erhebungsinstrumente in den Naturwissenschaftsdidaktiken noch vor der Anwendung eines klassischen Hypothesentest substantiell vernachlässigt wird (Taylor et al., 2016, S. 1221). Gleichzeitig ist die Entwicklung von validen und reliablen Testinstrumenten aber eine notwendige Bedingung, um dem Bedarf nach generalisierbaren Aussagen über größere Populationen von Lernenden nachkommen zu können. Mit dem SUMS-Inventary (Treagust et al., 2002) liegt im englischsprachigen Raum ein häufig eingesetztes Instrument (Mathesius & Krell, 2019) zur Verfügung. Für diesen Multiple-Choice-Fragebogen liegen inkonsistente Befunde bezüglich der Validität der vorgeschlagenen 5-faktoriellen Struktur vor (Lazenby & Becker, 2021; Villablanca et al., 2020). Die fünf Faktoren sind Modelle als multiple Repräsentationen (*MR*), Modelle als exakte Replikat (*ER*), Modelle als Erklärungswerkzeuge

(ET), Nutzen von Modellen (USM) und Revidierbarkeit von Modellen (CNM). Darüber hinaus sind auch Befunde publiziert, die bezüglich des Zusammenhangs zwischen SUMS und selbst erstellter Modelle von Lernenden und Lehrkräften theoriekonform sind (Cheng & Lin, 2015), aber im Hinblick auf den Zusammenhang mit Fachwissenszuwachs widersprüchlich bleiben (Gobert et al., 2011).

Vor diesem Hintergrund sollte SUMS als Basis für ein valides und reliables Erhebungsinstrument verwendet werden und es ergaben sich zwei wesentliche Forschungsfragen:

1. Kann die ursprünglich berichtete Faktorstruktur in einer übersetzten Variante des Fragebogens wiedergefunden werden?
2. Sind die Items des Fragebogens ausreichend reliabel für Zusammenhangsanalysen, bspw. mit der Abschlussnote aus der Einführungsvorlesung Chemie?

### Design und Methode

Abbildung 1 zeigt den Ablaufplan im Überblick. Die Übersetzung ins Deutsche wurde durch zwei Personen vorgenommen und zur Absicherung von einer Person mit Englisch als Erstsprache rückübersetzt. Dabei traten nur wenige Verzerrungen durch den Übersetzungsprozess auf, die wiederum durch eine gemeinsame Überarbeitung minimiert werden konnten. Die Aufgaben wurden bzgl. ihrer Zuordnung zu den theoretischen Faktoren einer Korrelationsanalyse unterzogen. Anschließend wurden mit konfirmatorischen Faktorenanalysen die Messmodelle auf Eindimensionalität geprüft und ein finales Strukturgleichungsmodell mit einem übergeordneten Faktor „Modellverständnis“ getestet (Bühner, 2021; Ziegler & Hagemann, 2015). Die Reliabilitäten wurden, inklusive des 95%-Konfidenzintervalls, als McDonald's  $\omega$  (Hayes & Coutts, 2020) geschätzt.

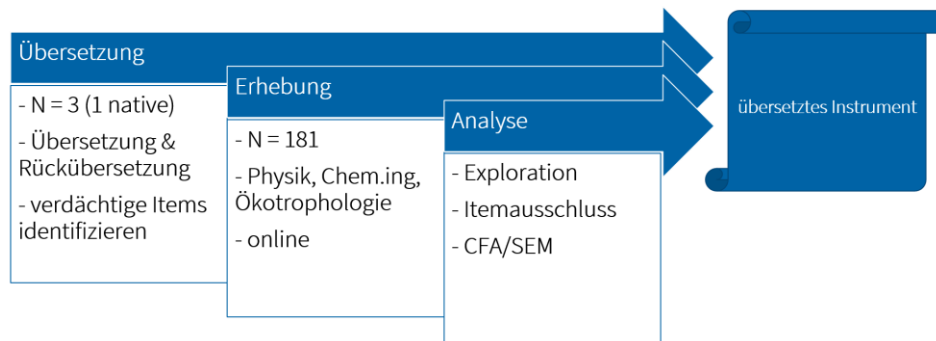


Abb. 1: Ablaufplan der Übersetzung, Datenerhebung und -analyse.

### Ergebnisse

Sowohl über die deskriptiven Kennwerte, als auch nach der Prüfung der konfirmatorischen Faktorenanalysen, wurde der Faktor *Modelle als exakte Replikat* als nicht zu den Daten passend bewertet. Darüber hinaus wurden aus allen Skalen Items entfernt, weil sie sich als nicht zusammenhängend mit den anderen Items zeigten. Das verbliebene Strukturgleichungsmodell zeigte zufriedenstellende Kennwerte (Abbildung 2), die Reliabilitäten mussten aber als nicht ausreichend für Zusammenhangsanalysen eingeordnet werden (Abbildung 3).

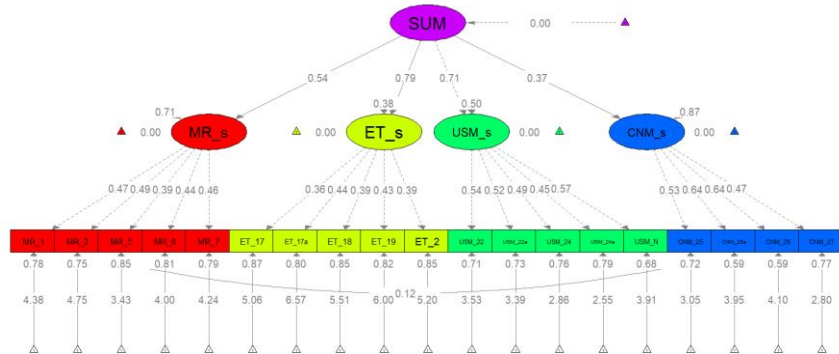


Abb. 2: Strukturgleichungsmodell der verbliebenen Items in den Faktoren MR, ET, USM & CNM, jeweils als essentiell tau-äquivalente Messmodelle mit  $\chi = 169.3(161)$ ,  $p = .31$ ; CFI = .97; RMSEA = .02; 95% CI [0; .04]; SRMR = .07.

**Diskussion**

Sowohl die Ablehnung der ursprünglichen Faktorstruktur, als auch die relativ niedrigen Reliabilitäten haben Implikationen für die Überarbeitung des Instruments in der deutschsprachigen Version. In einem nächsten Schritt werden die Items nach dem Question Appraisal System (QAS-99) (Willis & Lessler, 1999) systematisch überarbeitet. Die Falsifikation der Anwendbarkeit des Instruments an der getesteten Population trägt dazu bei, die Qualität zukünftiger Fragebögen zu modellbezogenen Lehr-/Lernumgebungen zu erhöhen (Haefel, 2022). Eine Erhebung mit der überarbeiteten Version, in der inhaltliche Erwägungen zur Neuordnung und Reformulierung der Items ebenfalls einfließen (Ziegler, 2014), wird mit einer Teilnehmendenzahl  $N = 200$  antizipiert. Darüber hinaus soll eine überarbeitete Version mit dem Potential zur Überführung in eine Kurzskala zur freien Nutzung in chemiebezogenen Lehrveranstaltungen entstehen.

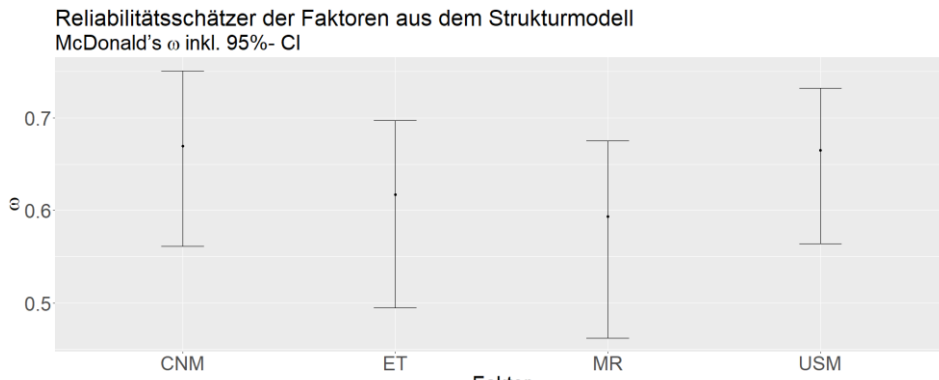


Abb. 3: Reliabilitäten der Faktoren aus dem Strukturgleichungsmodell.

### Lietratur

- Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication. *The American Statistician*, 73(1), 262–270.
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4., korrigierte und erweiterte Auflage). Pearson.
- Cheng, M.-F., & Lin, J.-L. (2015). Investigating the Relationship Between Students' Views of Scientific Models and Their Development of Models. *International Journal of Science Education*, 37(15), 2453–2475. <https://doi.org/10.1080/09500693.2015.1082671>
- Fleischer, J., Leutner, D., Brand, M., Fischer, H., Lang, M., Schmiemann, P., & Sumfleth, E. (2019). Vorhersage des Studienabbruchs in naturwissenschaftlich-technischen Studiengängen. *Zeitschrift für Erziehungswissenschaft*, 22(5), 1077–1097. <https://doi.org/10.1007/s11618-019-00909-w>
- Gobert, J. D., O'Dwyer, L., Horwitz, P., Buckley, B. C., Levy, S. T., & Wilensky, U. (2011). Examining the Relationship Between Students' Understanding of the Nature of Models and Conceptual Learning in Biology, Physics, and Chemistry. *International Journal of Science Education*, 33(5), 653–684. <https://doi.org/10.1080/09500691003720671>
- Haefffel, G. J. (2022). Psychology needs to get tired of winning. *Royal Society Open Science*, 9(6), 220099. <https://doi.org/10.1098/rsos.220099>
- Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But... *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Lazenby, K., & Becker, N. M. (2021). Evaluation of the students' understanding of models in science (SUMS) for use in undergraduate chemistry. *Chemistry Education Research and Practice*, 22(1), 62–76. <https://doi.org/10.1039/D0RP00084A>
- Makel, M. C., & Plucker, J. A. (2014). Facts Are More Important Than Novelty: Replication in the Education Sciences. *Educational Researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189X14545513>
- Mathesius, S., & Krell, M. (2019). Assessing Modeling Competence with Questionnaires. In A. Upmeyer zu Belzen, D. Krüger, & J. van Driel (Hrsg.), *Towards a Competence-Based View on Models and Modeling in Science Education* (Bd. 12, S. 117–131). Springer International Publishing. [https://doi.org/10.1007/978-3-030-30255-9\\_7](https://doi.org/10.1007/978-3-030-30255-9_7)
- Taylor, J., Furtak, E., Kowalski, S., Martinez, A., Slavin, R., Stuhlsatz, M., & Wilson, C. (2016). Emergent Themes from Recent Research Syntheses in Science Education and Their Implications for Research Design, Replication, and Reporting Practices. *Journal of Research in Science Teaching*, 53(8), 1216–1231. <https://doi.org/10.1002/tea.21327>
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' 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>
- Villablanca, S., Montenegro, M., & Ramos-Moore, E. (2020). Analysis of student perceptions of scientific models: Validation of a Spanish-adapted version of the Students' Understanding of Models in Science instrument. *International Journal of Science Education*, 42(17), 2945–2958. <https://doi.org/10.1080/09500693.2020.1843735>
- Willis, G. B., & Lessler, J. T. (1999). *Question Appraisal System QAS-99*.
- Ziegler, M. (2014). Stop and State Your Intentions: Let's Not Forget the ABC of Test Construction. *European Journal of Psychological Assessment*, 30(4), 239–242. <https://doi.org/10.1027/1015-5759/a000228>
- Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of Items: Pitfalls and Loopholes. *European Journal of Psychological Assessment*, 31(4), 231–237. <https://doi.org/10.1027/1015-5759/a000309>