

Tobias ROLFES, Weingarten, Boris GIRNAT, Hildesheim,
Christian FAHSE, Anne M. HUPFER, Landau &
Alexander ROBITZSCH, Kiel

Quantitative Ergebnisse zur Kompetenzstruktur des Wahrscheinlichkeitsbegriffs

Der Wahrscheinlichkeitsbegriff ist ein Schlüsselbegriff des Stochastikunterrichts der Sekundarstufe. In den Anfängen der unterrichtlichen Behandlung des Wahrscheinlichkeitsbegriffs wurde fast ausschließlich der klassische Ansatz nach Laplace gelehrt und gelernt. Jedoch erfuhr diese Einführung auf den klassischen Ansatz Kritik. Heutzutage wird daher der frequentistische Ansatz zur Einführung in den Wahrscheinlichkeitsbegriff befürwortet und auch in curricularen Standards festgeschrieben (Common Core State Standards Initiative, 2010; MBWJK, 2007). In welcher Weise Schülerinnen und Schüler mit dem klassischen und dem frequentistischen Zugang zum Wahrscheinlichkeitsbegriff adäquat umgehen können, ist jedoch bisher noch nicht ausreichend untersucht. Das Ziel der empirischen Erhebung war daher, die Kompetenzstruktur zum Wahrscheinlichkeitsbegriff quantitativ-empirisch zu untersuchen.

Methode

Insgesamt $N = 608$ Schülerinnen und Schüler aus insgesamt 23 Klassen der Klassenstufen 8, 9 und 10 nahmen an der Untersuchung teil. Der Papier- und Bleistifttest war im Multi-Matrix-Design mit 4 Booklets gestaltet und enthielt 93 Items in 56 Testlets. Auf theoretischer Grundlage wurden die Items fünf unterschiedlichen Itemkategorien zugeordnet. Die quantitative Auswertung der Itemschwierigkeiten erfolgte mit Hilfe der probabilistischen Testtheorie, indem eine eindimensionales Partial-Credit Modell geschätzt wurde. Dazu wurden teilweise mehrere Items (z.B. Ja/Nein-Items) wegen eines besseren empirischen Modellfits zu einem Item aggregiert. Zur Untersuchung, ob Schwierigkeitsunterschiede zwischen den Itemkategorien vorlagen, wurde ein linear-logistisches Testmodell (LLTM, vgl. Fischer, 1995) berechnet.

Ergebnisse

Die Mehrheit der (aggregierten) Items wiesen einen akzeptablen Modellfit für ein eindimensionales Partial-Credit Modell auf ($r_{it} > .25$, Infit < 1.15). Die Items sind in Abbildung 1 entsprechend ihrer Itemschwierigkeit in einer Wright-Map dargestellt. Eine geringe Itemschwierigkeit (z. B. -3 auf der Logit-Skala) bedeutet, dass es sich um ein sehr leichtes Item handelt. Entsprechend sind Items mit einem großen Schwierigkeitsparameter (z. B.

+3 auf der Logit-Skala) sehr schwierige Items, wohingegen der Wert 0 einen mittleren Schwierigkeitsgrad repräsentiert. Die Itemschwierigkeiten streuten relativ breit auf der latenten Fähigkeitsskala (vgl. Abbildung 1).

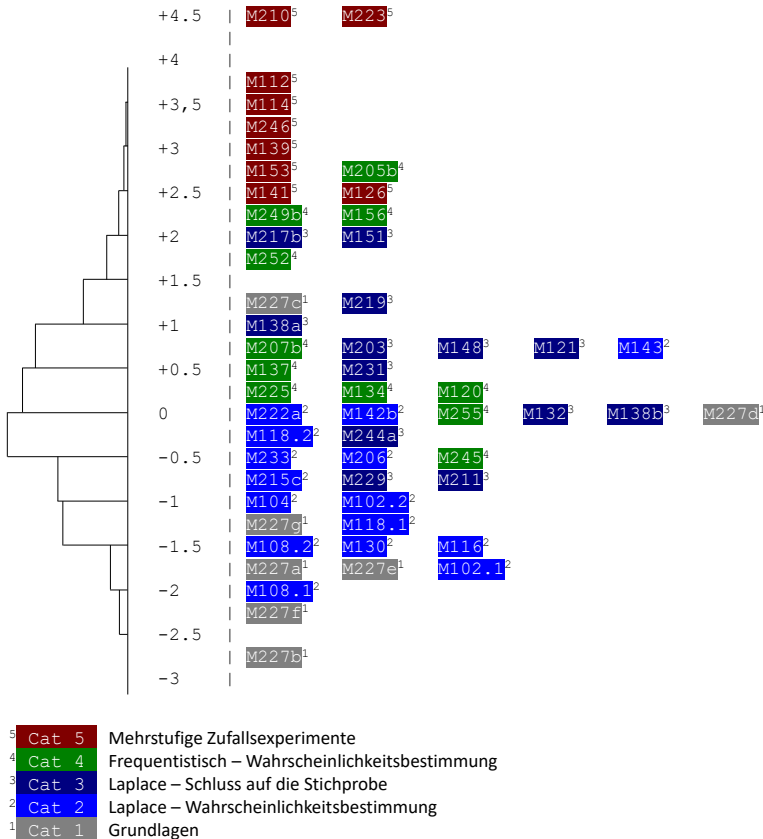


Abb. 1: Wright-Map der Itemschwierigkeiten.

Items zu mehrstufigen Zufallsexperimenten wiesen die höchste Schwierigkeit auf. Die Bestimmung der Wahrscheinlichkeit über den frequentistischen Ansatz war für die Schülerinnen und Schüler schwieriger als die Bestimmung über den klassischen Ansatz. Dabei zeigten sich Unterschiede bei den Items zum frequentistischen Ansatz: Die drei schwierigsten Items erforderte den Umgang mit einer Sequenz von Stichprobenergebnissen, wohingegen bei den leichteren Items nur ein Stichprobenergebnis gegeben war. Items, die nach der Bestimmung der Wahrscheinlichkeit mit Hilfe des klassischen Ansatzes einen Schluss auf eine Stichprobe erforderten, hatte mittlere

Schwierigkeit. Das linear-logistische Testmodell zeigte, dass sich die durchschnittlichen Itemschwierigkeiten signifikant zwischen den Itemkategorien unterschieden (vgl. Tabelle 1).

Kategorienvergleich	Δ (SE)
Cat 1 vs. Cat 2	0.38 (0.02)
Cat 2 vs. Cat 3	1.11 (0.03)
Cat 3 vs. Cat 4	0.21 (0.04)
Cat 4 vs. Cat 5	2.33 (0.09)

Anmerkungen. Δ = durchschnittliche Schwierigkeitsdifferenz der Itemkategorien (in Logit), SE = Standardfehler.

Tabelle 1: Vergleich der Itemkategorien mit Hilfe eines linear-logistischen Testmodells.

Eine Analyse in Abhängigkeit von der Klassenstufe ergab, dass sich die durchschnittliche Schülerfähigkeit im Umgang mit dem Wahrscheinlichkeitsbegriff mit jedem Schuljahr signifikant erhöhte. So hatten die Schülerinnen und Schüler in der Klassenstufe 8 im Mittel ein Fähigkeitsniveau (in Logit) von $M = -0.35$, $SD = 0.95$, in der Klassenstufe 9 betrug der Durchschnitt $M = 0.01$, $SD = 0.95$, $d = 0.35$ und in der Klassenstufe 10 erreichten die Schülerinnen und Schüler im Schnitt ein Fähigkeitsniveau von $M = 0.43$, $SD = 1.11$, $d = 0.38$.

Diskussion

Die empirischen Ergebnisse stellen die in breiten Teilen der Stochastikdidaktik befürwortete Herangehensweise, den Wahrscheinlichkeitsbegriff über den frequentistischen Ansatz einzuführen, in Frage. In den quantitativen Analysen waren die Items basierend auf dem klassischen Ansatz für die Schülerinnen und Schüler deutlich leichter zu lösen. Unter dem einfachen didaktischen Prinzip „Vom Leichten zum Schweren“ scheint es den empirischen Ergebnissen besser zu entsprechen, wenn zunächst mit dem klassischen Ansatz begonnen wird und darauf aufbauend Eigenschaften von Stichproben analysiert werden (z. B. Erwartungswert und Variabilität). Erst danach erscheint ein Übergang zum frequentistischen Ansatz sinnvoll, bevor mehrstufige Zufallsexperimente behandelt werden. Dementsprechend könnte der Erwartungswert möglicherweise als didaktische Verbindung zwischen dem klassischen und dem frequentistischen Ansatz genutzt werden (vgl. Fahse & Rolfes, 2019; Rolfes & Fahse, 2019).

Die Untersuchung ist einigen Einschränkungen unterlegen. So stellt die Stichprobe ein Convenience Sample dar und kann nicht als repräsentativ

angesehen werden. Daher sind beispielsweise die entdeckten Fähigkeitsunterschiede zwischen den Jahrgangsstufen nur eingeschränkt interpretierbar. Außerdem ist unbekannt, in welcher Weise und in welchem Umfang die Schülerinnen und Schüler bereits Unterricht zum Wahrscheinlichkeitsbegriff erhalten hatten. Deshalb ist es möglich, dass die Untersuchungsergebnisse nicht unerheblich ein Resultat des Unterrichts sind. Das Ergebnis, dass der Laplace-Ansatz leichter als der frequentistische Ansatz war, könnte auch daraus resultieren, dass die Lehrkräfte weiterhin im traditionellen Curriculum verhaftet sind und den unterrichtlichen Schwerpunkt auf die Behandlung des klassischen Ansatzes zum Wahrscheinlichkeitsbegriff legen und den frequentistischen Ansatz dagegen kaum thematisieren.

Für die weitere Forschung wäre daher eine experimentelle Untersuchung wünschenswert, in dem unter kontrollierten Bedingungen die Wirkung von Unterricht zum klassischen Ansatz und zum frequentistischen Ansatz verglichen wird. Des Weiteren könnte eine Interventionsstudie zu mehrstufigen Zufallsexperimenten evaluieren, in wie weit sich die relative Schwierigkeit von Items zu mehrstufigen Zufallsexperimenten zu übrigen Items zum Wahrscheinlichkeitsbegriff verändert.

Literatur

- Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association for Best Practices and the Council Chief State School Officers.
- Fahse, C. & Rolfes, T. (2019). Mathe-Welt: Alles Zufall oder was? Ein Crash-Kurs in 14 Aufgaben. *Mathematik lehren*, (213).
- Fischer, G. H. (1995). The Linear Logistic Test Model. In G. H. Fischer & I. W. Molenaar (Hg.), *Rasch models: Foundations, recent developments, and applications* (S. 131–155). New York, NY: Springer.
- Ministerium für Bildung, Wissenschaft, Jugend und Kultur Rheinland-Pfalz. (2007). *Rahmenlehrplan Mathematik (Klassenstufen 5 - 9/10)*. Mainz: Autor.
- Rolfes, T. & Fahse, C. (2019). Zufallsexperimente erfassen. Wahrscheinlichkeit, Erwartungswert und Variabilität. *Mathematik lehren*, (213), 3–8.