

### Untersuchung der Kompetenzstruktur im Bereich Erkenntnisgewinnung

Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung sind ein Kernelement naturwissenschaftlicher Grundbildung. Die Bildungsstandards definieren im Fach Chemie für den mittleren Schulabschluss im Kompetenzbereich *Erkenntnisgewinnung* acht Standards. Für diese wurde im Projekt ESNaS (Evaluation der Standards in den Naturwissenschaften der Sekundarstufe 1) eine Unterteilung in die Teilbereiche *Naturwissenschaftliche Untersuchungen*, *Naturwissenschaftliche Modellbildung* und *Wissenschaftstheoretische Reflexion* sowie eine weitere Untergliederung in Aspekte vorgenommen (siehe Abbildung 1) (Wellnitz et al., 2012).

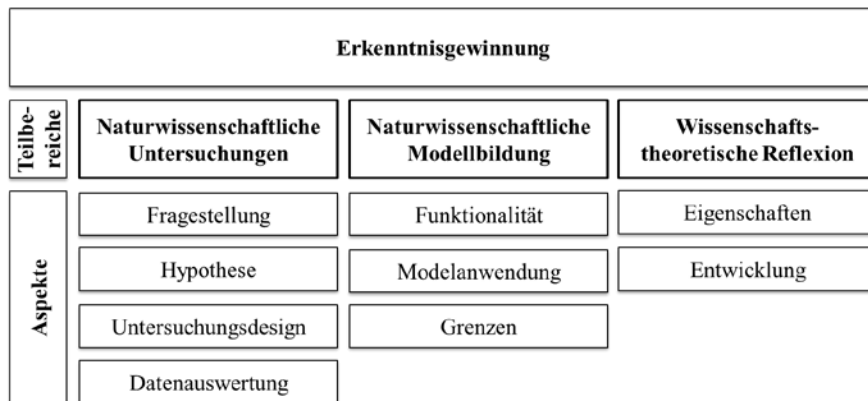


Abbildung 1. Differenzierung des Kompetenzbereichs Erkenntnisgewinnung

#### Ziele der Studie

Ein elementares Ziel der Studie ist die Entwicklung eines Testinstruments zur Erfassung der Schülerfähigkeiten in den drei Kompetenzteilbereichen. Mit Hilfe dieses Tests soll der Kompetenzbereich *Erkenntnisgewinnung* auf empirische Trennbarkeit bei der Erfassung von Schülerfähigkeiten hinsichtlich der Kompetenzteilbereiche geprüft werden. Zusätzlich sollen die Kompetenzteilbereiche *Naturwissenschaftliche Untersuchungen* und *Naturwissenschaftliche Modellbildung* auf ihre empirische Trennbarkeit hinsichtlich der jeweiligen Aspekte untersucht werden.

#### Methodisches Vorgehen

Basierend auf der Sichtung der vorhandenen Aufgaben auf dem Projekt ESNaS wurden gezielt neue Items entwickelt. Ziel war es, einen Itempool von 20 Items pro zu untersuchendem Aspekt für die Hauptstudie zu erhalten. Nach einer Pilotstudie für die ersten beiden Kompetenzteilbereiche mit  $N = 95$  Items und  $N = 603$  Probanden und einer anschließenden zweiten Pilotstudie für den dritten Kompetenzteilbereich mit  $N = 24$  Items und  $N = 148$  Probanden wurden 95 neu entwickelte und 65 ESNaS Aufgaben in der Hauptstudie eingesetzt.

Die 160 Items der Hauptstudie wurden im Multi-Matrix-Design auf 24 verschiedene Testhefte mit unterschiedlichen Schwerpunkten verteilt und an Gymnasien in der 9. Jahrgangsstufe und an Gesamtschulen in der 10. Jahrgangsstufe in NRW eingesetzt.

Zusätzlich wurden zwei Skalen des kognitiven Fähigkeitstests nach Heller und Perleth (2000) und zwei Tests zum allgemeinen Sprachstand (C-Tests) eingesetzt (Robitsch et al., 2008; Wockenfuß et al., 2006). Als weitere Kontrollvariable wurde die letzte Zeugnisnote im Fach Chemie erhoben.

### Ergebnisse

Die Ergebnisse der Hauptstudie basieren auf den Daten von 1043 Schülerinnen und Schülern von 17 Schulen (12 Gymnasien und 5 Gesamtschulen) mit einem Durchschnittsalter von 15.1 Jahren ( $SD = .89$ ). Die Probanden sind zu 48 % weiblich.

Tabelle 1 zeigt sowohl die Reliabilitäten des Gesamttests zur *Erkenntnisgewinnung*, als auch die der einzelnen Kompetenzteilbereiche und der Kontrollvariablen C-Test und KFT. Alle Werte liegen in einem zufriedenstellenden bis sehr guten Bereich.

| Testinstrument                        | EAP/PV Reliabilität | Cronbach's Alpha |
|---------------------------------------|---------------------|------------------|
| Kompetenztest                         |                     |                  |
| <i>Erkenntnisgewinnung</i>            | .75                 | -                |
| Naturwissenschaftliche Untersuchungen | .73                 | -                |
| Naturwissenschaftliche Modellbildung  | .72                 | -                |
| Wissenschaftstheoretische Reflexion   | .63                 | -                |
| Allgemeiner Sprachstand (C-Test)      | .88                 | .92              |
| Kognitiver Fähigkeitstest (KFT)       | -                   | .88/.93          |

Tabelle 1. Reliabilitäten der einzelnen Testinstrumente

Die empirische Trennbarkeit des Kompetenzbereichs *Erkenntnisgewinnung* wurde durch den Vergleich dreier möglicher Modelle untersucht. Dazu wurde ein eindimensionales Modell (1DM), welches keine Trennung der Kompetenzteilbereiche annimmt, das dreidimensionale Modell (3DM), welches die Trennbarkeit aller drei Kompetenzteilbereiche postuliert und ein zweidimensionales Modell (2DM), welches eine Trennbarkeit des dritten Kompetenzteilbereichs (*Wissenschaftstheoretische Reflexion*) von den anderen beiden Kompetenzteilbereichen annimmt, berechnet und verglichen.

Die Analyse erfolgte durch die Berechnung eines Rasch-Modells für jedes der drei postulierten Modelle. Die drei Modelle wurden anhand ihrer Deviance und den drei informationstheoretischen Maßen AIC, BIC und CAIC verglichen (siehe Tabelle 2).

|                 | 1DM   | 3DM   | 2DM   | Z3DM  |
|-----------------|-------|-------|-------|-------|
| <b>Deviance</b> | 22172 | 22156 | 22167 | 22166 |
| <b>AIC</b>      | 22466 | 22460 | 22465 | 22470 |
| <b>BIC</b>      | 23193 | 23212 | 23203 | 23223 |
| <b>CAIC</b>     | 23340 | 23364 | 23352 | 23375 |

Tabelle 2. Vergleich der Modelle anhand Deviance und informationstheoretischen Maßen

Die geringste Deviance, und damit einhergehend die beste Passung der Daten, weist das dreidimensionale Modell auf. Der Vergleich der informationstheoretischen Kriterien zeigt

die beste Modellpassung für das dreidimensionale Modell, wenn die Stichprobengröße bzw. die Stichprobengröße und die Itemanzahl keiner weiteren Gewichtung unterzogen werden (Vergleich AIC). Die stärkere Berücksichtigung dieser Kriterien, welche die komplexeren Modelle bestrafen, führt zu einer etwas besseren Passung des eindimensionalen Modells im Vergleich zum zwei- und dreidimensionalen Modell (Vergleich BIC und CAIC). Bei der Interpretation ist allerdings zu berücksichtigen, dass die Deviancemaße der komplexeren Modelle durch die Anzahl der Modellparameter unter Berücksichtigung der Stichprobengröße (BIC) bzw. Stichprobengröße und Itemanzahl (CAIC) relativiert werden. Eine statistische Absicherung des Unterschieds zwischen zwei Modellen unter Berücksichtigung der Parameterzahl bietet der Likelihoodquotiententest, welcher eine signifikant bessere Passung des dreidimensionalen Modells (3DM) im Vergleich zum eindimensionalen Modell (1DM) zeigt. Die Differenz der in Tabelle 2 dargestellten Deviance für das ein- und dreidimensionale Modell tritt mit einer Wahrscheinlichkeit von  $p \leq .001$  zufällig auf. Dies indiziert, dass das dreidimensionale Modell (3DM) die Daten signifikant besser abbildet als das eindimensionale Modell (1DM). Das zweidimensionale Modell (2DM) hingegen zeigt keine signifikant bessere Passung auf die Daten ( $p = .08$ ) im Vergleich zum eindimensionalen Modell (1DM).

|                   | $\Delta$ Deviance | $\Delta$ df | Signifikanz   |
|-------------------|-------------------|-------------|---------------|
| <b>1DM / 3DM</b>  | 16                | 5           | $p \leq .001$ |
| <b>1DM / 2DM</b>  | 5                 | 2           | $p = .08$     |
| <b>1DM / Z3DM</b> | 6                 | 5           | $p = .31$     |

Tabelle 3. Likelihoodquotiententests

In weiterführenden Analysen konnte die bessere Passung des dreidimensionalen Modells (3DM) durch die Berechnung von einem dreidimensionalen Zufallsmodell (Z3DM) bestätigt werden (Tabelle 2 & 3). Dennoch ist anzumerken, dass die latenten Korrelationen zwischen den Dimensionen mit .89-.96 sehr hoch sind. Dies spricht nach Prenzel, Walter & Frey (2006) allerdings nicht gegen eine bessere Passung des komplexeren Modells.

Basierend auf den Ergebnissen kann die Schlussfolgerung gezogen werden, dass sich der Kompetenzbereich *Erkenntnisgewinnung* durch die drei Kompetenzteilbereiche *Naturwissenschaftliche Untersuchungen*, *Naturwissenschaftliche Modellbildung* und *Wissenschaftstheoretische Reflexion* abbilden lässt, die sich allerdings sehr ähnlich sind.

#### Literatur

- Heller, K.A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. Manual*, Göttingen: Beltz Test GmbH.
- Prenzel, M., Walter, O., & Frey, A. (2007). PISA misst Kompetenzen. Eine Replik auf Rindermann (2006): Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 58(2), 128-136.
- Robitzsch, A., Karius, I. & Neumann, D. (2008). *C-Tests for German Students: Dimensionality, Validity and Psychometric Perspectives.*, Berlin.
- Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., Sumfleth, E., & Walpuski, M. (2012). Evaluation der Bildungsstandards -eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 261-291.
- Wockenfuß, V. & Raatz, U. (2006). Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In R. Grotjahn (Hrsg.), *Der C-Test. Theorie, Empirie, Anwendungen* (S. 211-242). Frankfurt am Main: Lang.