Measuring Experimental Skills in Large Scale Assessments: Developing a simulation-based test instrument

Martin Dickmann[1], Bodo Eickhorst[2] , Heike Theyßen[1], Knut Neumann[3], Horst Schecker[2]
([1] University Duisburg-Essen, [2] University Bremen, [3] IPN Kiel)

Motivation

- Experimental skills are key objectives of science education.
  (c.f. edcuational standards; e.g. KMK, 2005, NRC, 2012)
- Hence, adequate assessment tools for experimental skills are required.
- There are low correlations between students' achievements in written tests and in hands-on tests.
  (Shavelson, Ruiz-Primo, & Wiley, 1999; Stecher et al., 2000)
- Computer-simulations might be valid substitutes for hands-on tests.
  (Shavelson et al., 1999; Schreiber, 2012)

Aims

Development of an instrument that…

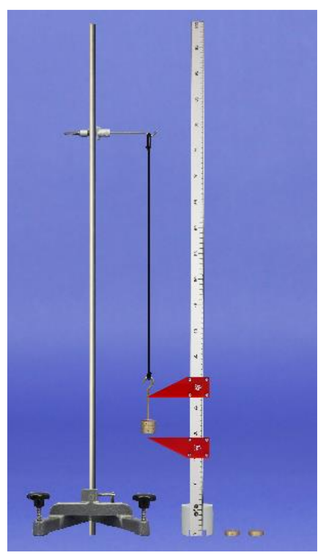
- … measures important experimental skills
  - ➢ reliably and validly
  - ➢ in the three phases of experimenting: preparation, performance, evaluation
- … is applicable in large scale assessments.

Target group: students aged 14 to 16 (secondary school)

Test instrument

- 12 tasks with typical school experiments
- Online simulation environments

  Example: *Test this assumption: "The expansion of a rubber band is proportional to the attached weight".*
- 6 items per task
- Each item starts with a sample solution of the preceding item
- Required physics content knowledge is given



Task structure

PREPARATION	PERFORMANCE	EVALUATION
describe basic idea	assemble and test an experimental setup	plan the evaluation of data
specify procedure	perform and document measurements	process data
prepare measurement report		draw conclusions

Validation studies

Aspects	Research questions	Studies
Content	Do the tasks represent experiments students are likely to know?	Analyses of syllabi and schoolbooks - Expert ratings
Content	Are the tasks consistent with typical classroom experimenting?	
Individual strategies	Do experimental considerations dominate in students' thinking?	Think aloud (intro- and retro-spective)
Individual strategies	Do the tasks offer adequate support to overcome deficits in physics content knowledge?	
Relationship with external variables	Is the test performance a sufficient predictor for performance in hands-on tests?	CBT vs. hands-on
Relationship with external variables	Do test results differ sufficiently from physics content knowledge and cognitive abilities?	Large Scale
Internal test structure	Are the three phases of experimentation empirically separable in students performances?	

Content Analysis

Syllabi and schoolbooks:

- high consistency of the obligatory experiment-oriented contents across the 16 federal states of Germany
  - ➔ identification of 22 typical experimental tasks

(cf. Dickmann & Theyßen, 2013)

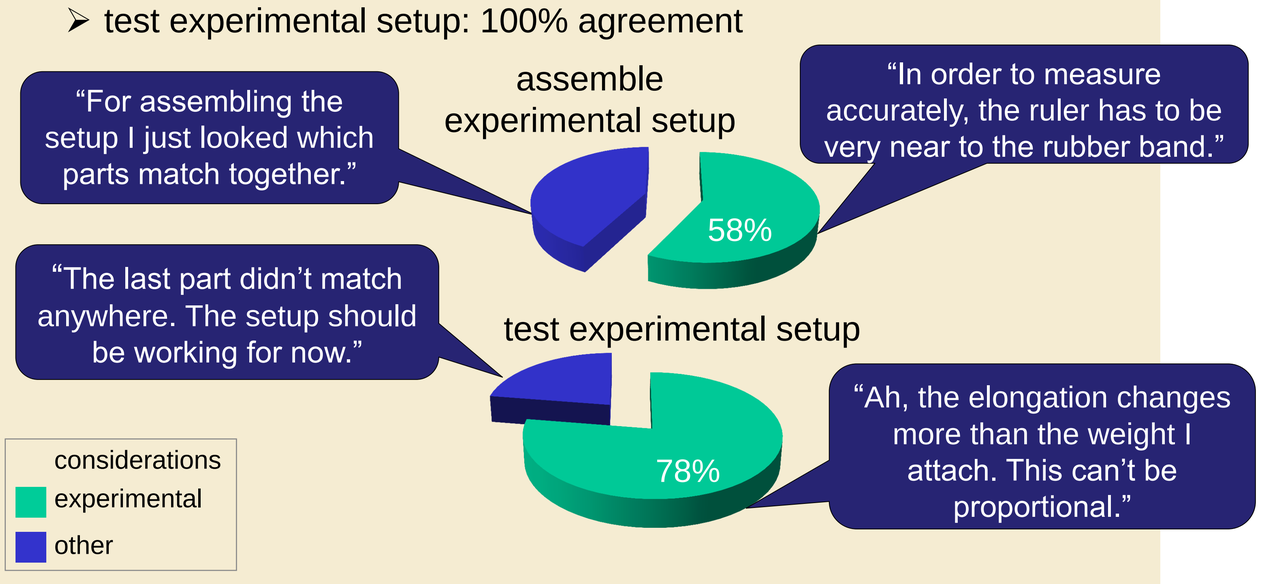
Expert rating (53 teachers):

- for most of the tasks the experts see compatibility with actual classroom practice
  - ➔ choice of 12 tasks with high content validity

(cf. Eickhorst et al., 2013)

Individual strategies

- Sample item: *assemble and test the experimental setup*
- Think-aloud data: 40 students working on the rubber band task
- Do experimental considerations dominate in students' thinking?
- Indicators: students considerations are based on experimental arguments (e.g. safety issues, measurement accuracy, …)
- Interrater reliability:
  - ➢ assemble experimental setup: κ = .714 (88 % agreement)
  - ➢ test experimental setup: 100% agreement



Contacts:
Martin Dickmann | martin.dickmann@uni-due.de
Bodo Eickhorst | bodoe@uni-bremen.de
Heike Theyßen | heike.theyssen@uni-due.de
Knut Neumann | neumann@ipn.uni-kiel.de
Horst Schecker | schecker@physik.uni-bremen.de

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung