FREIE UNIVERSITÄT BERLIN

Development of a Test to Measure the Data Literacy of Pre-Service Teachers

Dominik Endemann & Holger Gärtner

THEORY

- Data-driven decision-making as a promising strategy for enhancing the quality of education and instruction (Lai & Schildkamp, 2013)
- In Germany: Mandatory standardized student assessments (Vergleichsarbeiten / VERA) in grades 3 (German and mathematics) and grade 8 (German, mathematics, first foreign language) aimed at supporting instruction development
- The use of data from standardized assessments to improve teaching often fails to meet expectations (Gärtner, 2025)

- Limited focus on data literacy in teacher education; many teachers struggle to interpret data and act on it (Kippers et al., 2018)
- At present, no assessment fully measures teachers' ability to understand, interpret, and make decisions based on student achievement data

→ **GOAL: Development of a test to measure teachers' specific data literacy, intended for use in pre-service teacher education and in-service professional development**

CONSTRUCT DEFINITION AND TEST DEVELOPMENT

Understanding: Identifying single datapoints in the report, merging two or more datapoints, comparing them, or positioning them in relation to each other; multi-step analysis of the results and searching for patterns and trends that are not explicitly recognizable

Interpreting: Identifying relevant contextual information that provides a possible explanation and linking it to the information in the report; drawing conclusions and explanations

Deciding: Setting objectives and prioritizing measures while considering available resources; planning the measure

(Endemann et al., in press)

Stimuli

- Original VERA-3-Feedback-Reports (primary school)
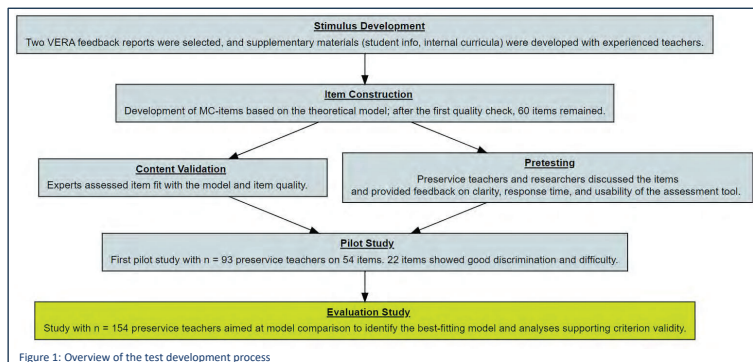- Context information on curriculum

Figure 1: Overview of the test development process

EVALUATION STUDY

Study: n = 154 pre-service teachers (76% female); n = 21 Items; online-test (surveydown), average completion time: approx. 30 minutes

1. Item selection

- Item analysis based on difficulty parameter, item-fit (wMNSQ) and discrimination (CTT)
- → Final Test: Understanding (D1): 7 Items; Interpreting (D2): 5 Items; Deciding (D3): 5 Items
- → EAP-Reliability: EAP_{D1} = 0.61; EAP_{D2} = 0.62; EAP_{D3} = 0.57
- → Item discrimination between .10 and .44
- → Infit (wMNSQ) ranged from 0.87 to 1.10
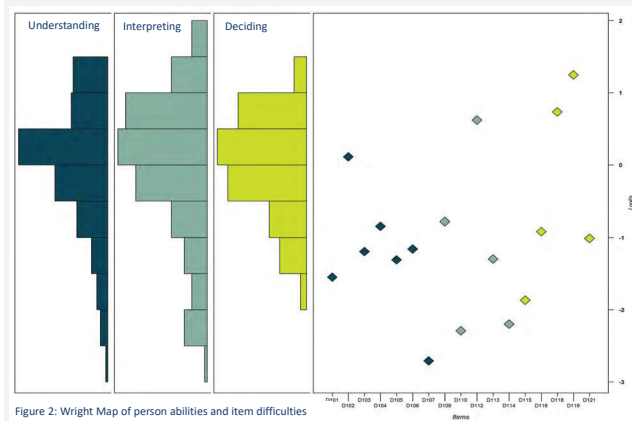- → Rasch Item difficulties between -2.71 and 1.25 (Figure 2)

2. Model comparison

- Model comparison using likelihood-ratio test and information based indices (AIC, BIC)
- → 3-dimensional Rasch model fits data significantly better than 1-dimensional model

Table 1: Information criteria and results of the model comparison with the unidimensional model

Model	Deviance	N_{par}	AIC	BIC	χ^2	df	p
3-Dim	2529.63	23	2575.63	2645.48	-	-	-
1-Dim	2546.24	18	2582.24	2636.91	16.61	5	< .05

3. Criterion validity

- Correlation analyses and group comparisons (t-tests) to examine criterion validity:

H1: Data literacy is positively associated with perceived clarity of the feedback report.
→ r = .27 – .38 ($p < .05$)

H2: Data literacy is positively associated with the acceptance of standardized assessments in schools.
→ r = .22 – .30 ($p < .05$)

H3: Preservice teachers for primary education outperform those for secondary education, likely due to their greater expertise in diagnostic assessment and didactics within core subjects such as German and Mathematics (Figure 3).
→ $d_{understanding}$ = 0.35*, $d_{interpreting}$ = 0.39*, $d_{deciding}$ = 0.08 (n.s.)
*significant at $p < .05$

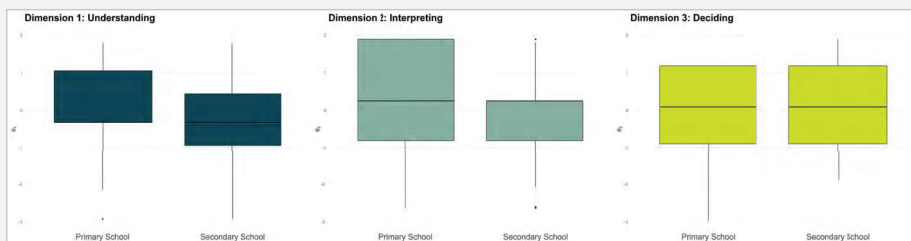Figure 2: Wright Map of person abilities and item difficulties

Figure 3: Comparison of data literacy scores between pre-service primary and secondary school teachers

CONCLUSIONS AND IMPLICATIONS

Conclusions

- Three-dimensional structure with 17 items showing good discrimination
- The three-dimensional model enables targeted feedback on areas where teachers need further support in handling data
- Reliabilities are on the lower end of the acceptable range, but suitable for the intended use

Implications / Outlook

- Study with in-service teachers
- Revision of items with low discrimination and to improve reliability
- Intervention study designed to improve data literacy
- Assess test–retest reliability

References
Endemann, D., Holder, K. & Gärtner, H. (in press). Teachers' data-competent use of results from standardized achievement tests – construct definition and operationalization. In C. Schuhmacher & D. Ifenthaler (Hrsg.), International Perspectives on Educational Data Literacy: Frameworks, Contexts, and Practices. Routledge.
Gärtner, H. (2025). Nutzung von Lernstandserhebungen zur Förderung adaptiven Unterrichts. In A. Albers & N. Jude (Hrsg.), Blickpunkt Bildungsmonitoring – Bilanzen und Perspektiven (pp. 142–157). Weinheim, Basel: Beltz Juventa.
Kippers, W. B., Poortman, C. L., Schildkamp, K. & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? Studies in Educational Evaluation, 56, 21–31. https://doi.org/10.1016/j.stueduc.2017.11.001
Lai, M. K. & Schildkamp, K. (2013). Data-based Decision Making: An Overview. In K. Schildkamp, M. K. Lai & L. Earl (Eds.), Data-based Decision Making in Education. Challenges and Opportunities (Studies in Educational Leadership, vol. 17, pp. 9–22). Dordrecht: Springer.

Demo version of the data literacy test

View sample materials and feedback reports

Dominik Endemann | d.endemann@fu-berlin.de
Department of Education and Psychology
Division for the Evaluation of School and Instruction Quality
Freie Universität Berlin
Habelschwerdter Allee 45, 14195 Berlin