

**Abstract PAEPSY 2021**

**Title:** Stability of Value-Added Models: Comparing Classical and Machine Learning Approaches

**Authors:** Emslander, V.<sup>1</sup>; Levy, J.<sup>1</sup>; Scherer, R.<sup>2</sup>; Brunner, M.<sup>3</sup>; Fischbach A.<sup>1</sup>

<sup>1</sup> Luxembourg Centre for Educational Testing (LUCET) at the University of Luxembourg, Faculty of Humanities, Education and Social Sciences, University of Luxembourg, Luxembourg

<sup>2</sup> Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo, Norway

<sup>3</sup> Department of Education, University of Potsdam, Potsdam, Germany

**Abstract**

**Background:** What is the value that teachers or schools add to the evolution of students' performance? Value-added (VA) modeling aims to answer this question by quantifying the effect of pedagogical actions on students' achievement, independent of students' backgrounds (e.g., Braun, 2005). A plethora of VA models exist, and several outcome measures are in use to estimate VA scores, yet without consensus on the model specification (Everson, 2017; Levy et al., 2019). Furthermore, it is unclear whether the most frequently used VA models (i.e., multi-level, linear regression, and random forest models) and outcome measures (i.e., language and mathematics achievement) indicate a similar stability of VA scores over time.

**Objectives:** Drawing from the data of a highly diverse and multilingual school setting, where leveling out the influence of students' backgrounds is of special interest, we aim to (a) clarify the stability of school VA scores over time; (b) shed light on the sensitivity toward different statistical models and outcome variables; and (c) evaluate the practical implications of (in)stable VA scores for individual schools.

**Method:** Utilizing the representative, longitudinal data from the Luxembourg School Monitoring Programme (LUCET, 2021), we examined the stability of school VA scores. We drew on two longitudinal data sets of students who participated in the standardized achievement tests in Grade 1 in 2014 or 2016 and then again in Grade 3 two years later (i.e., 2016 and 2018, respectively), with a total of 5875 students in 146 schools. School VA scores were calculated using classical approaches (i.e., linear regression and multilevel models) and one of the most commonly used machine learning approaches in educational research (i.e., random forests).

**Results and Discussion:** The overall stability over time across the VA models was moderate, with multilevel models showing greater stability than linear regression models and random forests. Stability differed across outcome measures and was higher for VA models with language achievement as an outcome variable as compared to those with mathematics achievement. Practical implications for schools and teachers will be discussed.

**Keywords:** Value-Added Modeling, School Effectiveness, Machine Learning

## References

- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Educational Testing Service.
- Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research*, 87(1), 35–70.  
<https://doi.org/10.3102/0034654316637199>
- Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2019). Methodological issues in value-added modeling: An international review from 26 countries. *Educational Assessment, Evaluation and Accountability*, 31(3), 257–287. <https://doi.org/10.1007/s11092-019-09303-w>
- LUCET. (2021). *Épreuves Standardisées (ÉpStan)*. <https://epstan.lu>