Automated Analysis of Physics Lesson Plans

David Buschhüter
Physics Education
Institute for Physics and Astronomy
University of Potsdam

Introduction: lesson plans as learning outcomes of pre-service physics teachers

Session Title:

„Challenges in measuring learning outcomes in STEM fields"

Computer-based classification of lesson phases

Lesson planning

Pre-service physics teachers

Can machines hypothesize about learning processes?

Universität Potsdam

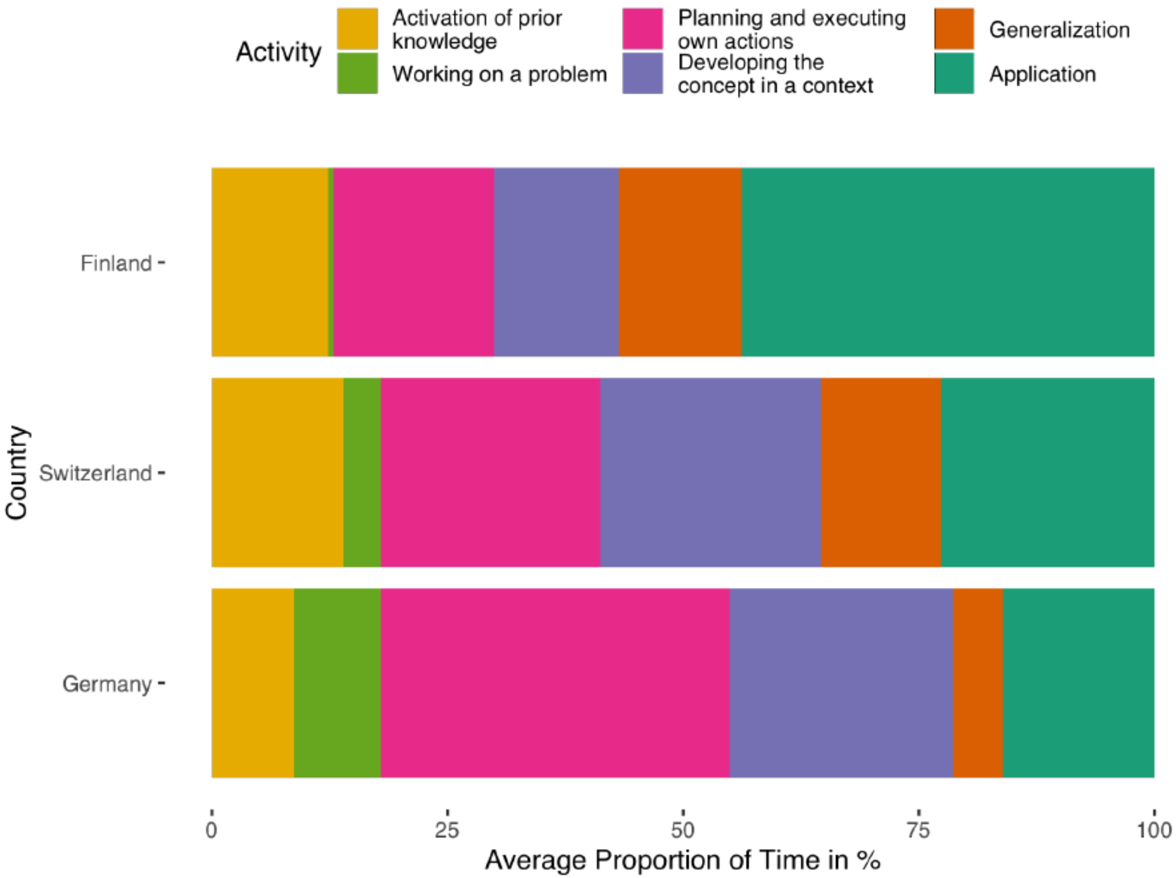
Agenda

» Why automation of classification of phases is beneficial from a practical perspective
  - Problem: Lesson planning in physics
  - Basis models: A theory to solve the problem
  - Machine Learning models as a basis to support automated feedback

» Study: A first step towards feedback

» A perspective for future research

Teachers in Germany tend to use experience to build theory

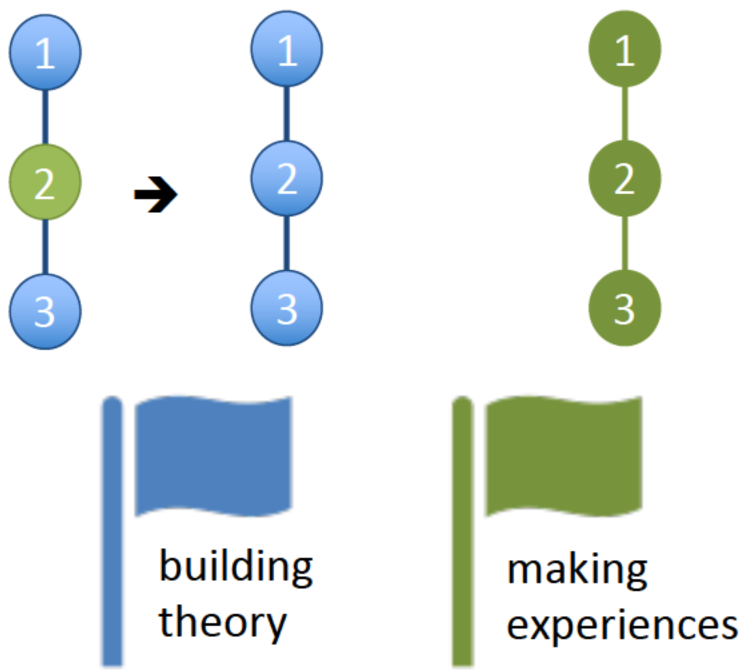


- German teachers provide a mix of both learning by experience and building theory
- German students show less learning gains than Finnish students
- Students can't directly infer a new concept from their experience (Muckenfuß, 1996)

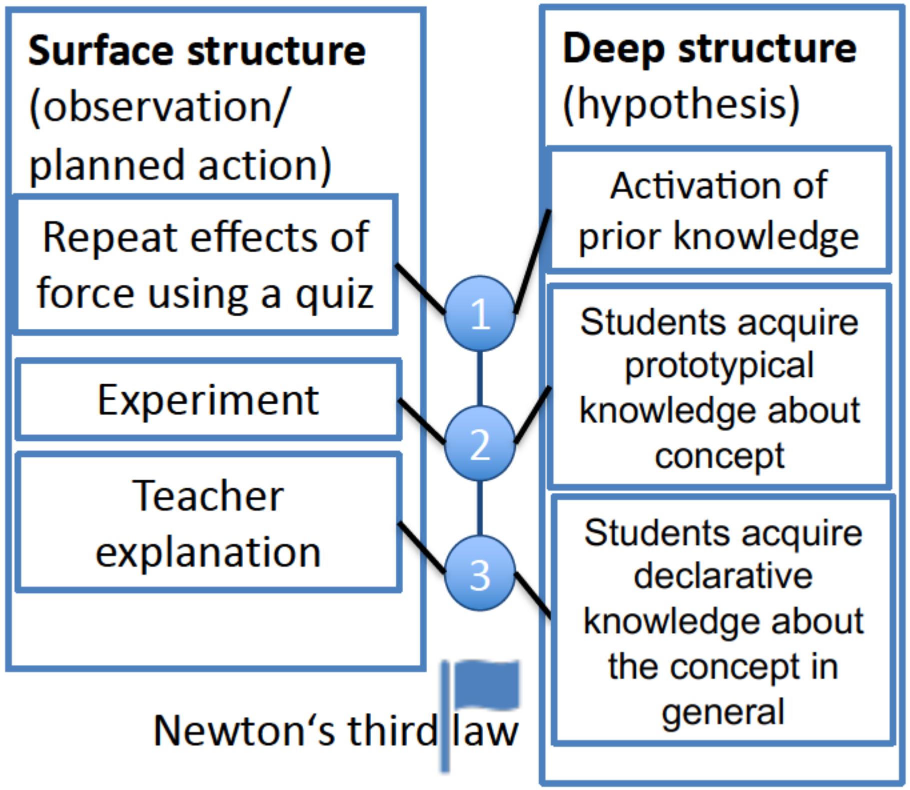
(Geller, 2015, p. 111; graph replotted and translated; values extracted from graph; for exact values see original)

Planning lessons means hypothesizing about learning

Basis Models of Teaching and Learning (e.g. Oser & Baeriswyl, 2001): Different chains of actions (phases) for different teaching goals



+ 12 other basis models



Planning lessons means **hypothesizing about learning**

Providing more feedback would be useful but challenging

» Problem seems insensitive to instruction followed by experience for pre-service physics teachers (Mutschler et al., 2021)

» More frequent feedback would be beneficial (Hattie & Timperley, 2007, Lee et al., 2021) & more research is needed in this field (Mutschler et al., 2021; Geller, 2015)

» Coding related to **deep structures** is resource intensive

- **Experts need extensive experience**
- **Low interrater-agreement**
  - Kappa: 0.39-0.50 (Geller, 2015, p. 85)
  - Kappa: 0.4 median (Reyer, 2004, p. 200)
  - Consensus-validation (Gerber, 2007)
  - Others failed finding agreement and switched research question (Borowski, personal communication)

Universität Potsdam

A machine learning model would enable us to store the expertise

- **For science education/assessment (Zhai, Shi, & Nehm, 2021; Zhai, Yin et al., 2020) many examples for students:**
  - Extracting student ideas of generality of explanations (Rosenberg & Christ, 2020) unsupervised
  - Computer Assisted Scoring, Biology, Subject Matter Knowledge: (Kappa > 0.8) (Nehm & Härtig, 2012; Nehm, Ha, & Mayfield, 2012)
  - Learning progressions on the concept of flux Kappa > 0.6 (Jescovitch et al., 2021)
- **Less for teacher performance**
  - Teacher Reflection (e.g. Wulff et al., 2020)
  - Araya et al. (2011): Teacher activities and mathematical content (agreement mostly high but scatters by characteristic) from lesson transcripts.
- **Can we also classify deep structural characteristics?**
  - High complexity and high diversity (Zhai, Haudek, et al. 2020; Haudek & Zhai, 2021) and many ways how the same deep structure can be expressed (Zhai, Shi, & Nehm, 2021) makes classification challenging
  - Unclear if even worth trying?
- **The study: Just a step in that direction**
  - Can we classify the steps of concept building automatically via machine learning in very restricted lesson planning scenario?

Universität Potsdam

Task: Machine

Activation

Prototype

Concept

Application/ Transfer

Outside of Basis Model

Order	Description of Classroom Action	HKS
1	Teacher greets the students	Outside
2	Students obeserve an experiment	Activation
3	...	

Numeric

Text

Classifier (logistic regression, decision tree, ...)

Design

Data (Schröder et al., 2020):

» Paper and pencil-test

» 126 Lesson plans (1231 segments) of pre-service physics teachers

Task: Plan a lesson on Newton‘s third law of motion

Goal: Students should be able to explain the law after the lesson

Human Analysis

» Paraphrasing

» Segmentation

» Digitalization

» Assignment of chain of phases ➔

order	Description	Phase (label)
1	Start by explaining the general rule of Newton‘s third law	Concept
2	...	...

Machine learning

- Pre-Processing
- Machine learning: model selection & evaluation (cross-validation, hyperparameter tuning, 30% test data)

Universität Potsdam

A first step: Shallow Learning to assign chains of lesson phases

Multinomial Logistic Regression

	precision	recall	f1-score	N
Activation	0.74	0.68	0.71	72
Prototype	0.68	0.82	0.75	142
Concept	0.96	0.61	0.75	44
Application / transfer	0.70	0.75	0.72	64
Outside of the basis model	0.67	0.13	0.22	15
accuracy				**0.72**
macro avg	0.75	0.60	**0.63**	337
weighted avg	0.73	0.72	0.71	337

- Cohen’s Kappa = 0.60 (acceptable, Caspar & Wirtz, 2002 p. 49; Geller, 2015, p. 85)
- This would allow group-feedback for this planning situation
    - Are “all” steps of concept building in the lesson plan?
    - Is there enough application?

Universität Potsdam

The algorithm is interpretable

Higher regression coefficient

Concept	
Feature	English (Example word in German)
merksatz	mnemonic sentence (Merksatz)
tafelbild	blackboard image (Tafelbild)
na	short for Newton's Law (Newtonsches Axiom)
order	-
tafel	blackboard (Tafel)

Generic term

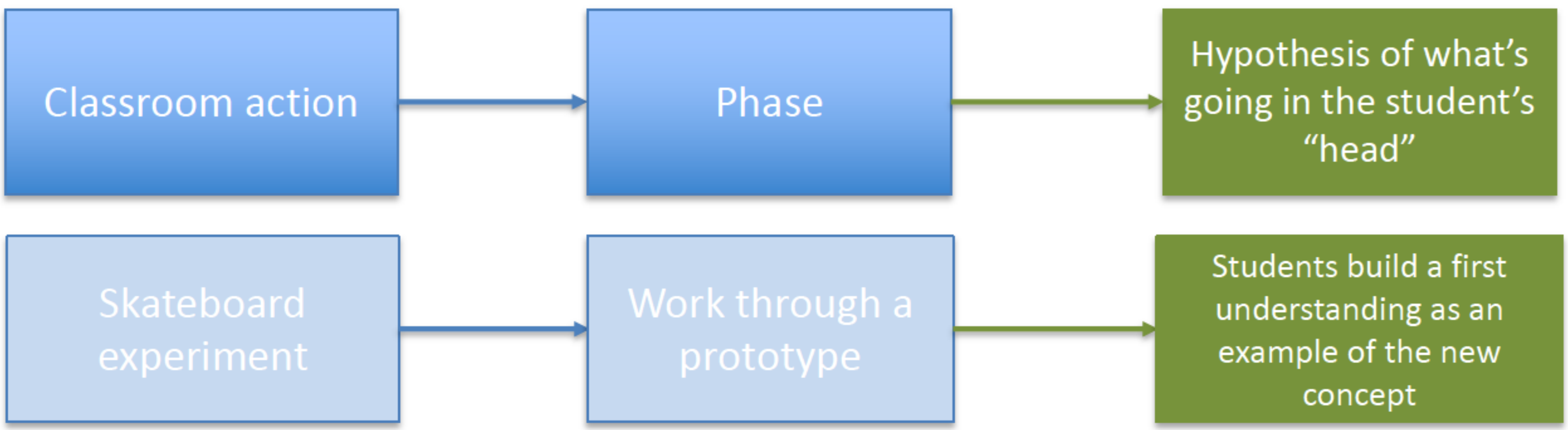
Content specific term

Position in the lesson

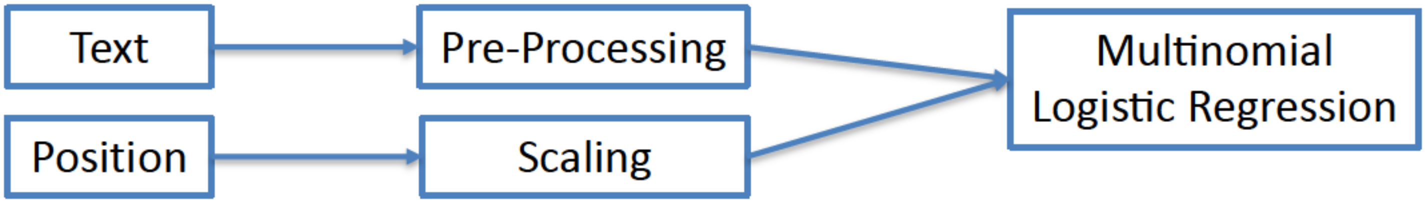
Universität Potsdam

In principle we build a machine which makes hypothesis on learning process

» Does this machine make hypotheses about (intended) learning processes?



But how should it think? Limitations and possible extensions



- Account for context of the classroom action to be classified (Gerber, 2007)

Demonstration of an experiment with skateboards and explaining, why the active person will also be accelerated

Is this a prototype or an application?

- Other basis models

What do we learn for future research?

Common Question: „Will Artificial Intelligence Replace Teachers in the Future?“ (AnalyticsInside, 2021)

What are the possibilities?

- ➢ Science Education and STEM Education have to provide answers related to categories most important to us
  - » Science Assessment (Zhai, Shi, & Nehm, 2021; Zhai, Yin et al., 2020)
  - » For teaching performance: Deep structure characteristics (Fischer et al., 2003) ➔ less evidence
- This study provides first evidence that this might be possible
- Challenges
  - Cohen’s kappa of 0.6 in a very restricted scenario shows that it will be challenging to provide such a classifier in general
  - Also content specific terms (Newton’s third law) are used for prediction
  - Logistic regression seems to be a good classifier: accuracy vs. substantial validity of the reasoning process

What do we learn for future research?

"[...] planning of instructional action according to hypothesized learning processes is usually primarily intuitive." (Gerber, 2008, p. 35 translated)

A machine would add to the teacher

We should try it

Thank you!

Dr. Peter Wulff
Tanja Mutschler
Prof. Dr. Andreas Borowski

Jan Schröder
Prof. Dr. Josef Riese

Dr. Christoph Vogelsang
Prof. Dr. Christoph Kulgemeyer

Get in touch: David.Buschhueter@uni-potsdam.de

References

- Araya, R., Plana, F., Dartnell, P., Soto-Andrade, J., Luci, G., Salinas, E., & Araya, M. (2012). Estimation of teacher practices based on text transcripts of teacher speech using a support vector machine algorithm. *British Journal of Educational Technology*, *43*(6), 837–846. https://doi.org/10.1111/j.1467-8535.2011.01249.x
- Fischer, H. E., Klemm, K., Leutner, D., Sumfleth, E., Thiemann, R., & Wirth, J. (2003). Naturwissenschaftsdidaktische Lehr-Lernforschung : Defizite und Desiderata. *Zeitschrift für Didaktik Der Naturwissenschaften*, *9*, 179–208.
- Geller, C. (2015). *Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb [Learning process-oriented sequencing of physics lessons in connection with knowledge acquisition]*. Berlin: Logos.
- Gerber, B. (2007). *Strukturierung von Lehr-Lern-Sequenzen im Physikunterricht*. Universität Bern.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487
- Haudek, K. C., & Zhai, X. (2021). Exploring the effect of assessment construct complexity on machine learning scoring of argumentation (preprint). In *Virtual Annual International Conference of the National Association for Research in Science Teaching (NARST)* (pp. 3–22).
- Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Merrill, J., Urban-Lurain, M., Doherty, J. H., & Haudek, K. C. (2021). Comparison of machine learning performance using analytic and holistic coding approaches across constructed response assessments aligned to a science learning progression. *Journal of Science Education and Technology*, *30*(2), 150–167. https://doi.org/10.1007/s10956-020-09858-0
- Lee, H. S., Gweon, G. H., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine learning-enabled automated feedback: supporting students’ revision of scientific arguments based on data dawn from simulation. *Journal of Science Education and Technology*, *30*(2), 168–192. https://doi.org/10.1007/s10956-020-09889-7
- Muckenfuß, H. (1996). Grundpositionen Wagenscheins – kritisch hinterfragt [Basic positions of Wagenschein - critically discussed]. *Der Mathematische und Naturwissenschaftliche Unterricht*, *49* (8), 455–462.
- Mutschler, T., Buschhüter, D., Schröder, J., Riese, J., & Borowski, A. (2021). Theoriekonformität von Unterrichtsplanungen im Fach Physik vor und nach dem Praxissemester. In S. Habig (Ed.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch? Gesellschaft für Didaktik der Chemie und Physik online Jahrestagung* (pp. 258–261). Universität Duisburg-Essen.
- Nehm, R. H., & Härtig, H. (2012). Human vs. computer diagnosis of students’ natural selection knowledge: testing the efficacy of text analytic software. *Journal of Science Education and Technology*, *21*(1), 56–73. https://doi.org/10.1007/s10956-011-9282-7
- Nehm, R. H., Ha, M., & Mayfield, E. (2012). Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations. *Journal of Science Education and Technology*, *21*(1), 183–196. https://doi.org/10.1007/s10956-011-9300-9
- Oser, F., & Baeriswyl, F. J. (2001). Choreographies of teaching: bridging instruction to Learning. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed., pp. 1031–1065). Washington: American Educational Research Association.
- Reyer, T. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht. Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*. Berlin: Logos.
- Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., … Schecker, H. (2020). Die Messung der Fäigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten Performanztests [Measuring the skills of lesson planning in the subject of physics using a standardized performance assessment]. *Zeitschrift für Didaktik Der Naturwissenschaften*. https://doi.org/10.1007/s40573-020-00115-w
- Wulff, P., Buschhüter, D., Westphal, A., Nowak, A., Becker, L., Robalino, H., … Borowski, A. (2020). Computer-based classification of preservice physics teachers’ written reflections. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-020-09865-1
- Zhai, X., Shi, L., & Nehm, R. H. (2021). A meta-analysis of machine learning-based science assessments: Factors impacting machine-human score agreements. *Journal of Science Education and Technology*, *30*(3), 361–379. https://doi.org/10.1007/s10956-020-09875-z
- Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, *56*(1), 111–151. https://doi.org/10.1080/03057267.2020.1735757
- Zhai, X., Haudek, K.C., Shi, L., Nehm, R.H., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching*, *57*(9), 1430–1459. https://doi.org/10.1002/tea.21658