Registry of Efficacy and Effectiveness Studies

Study Title:
AI2Teach: Effectiveness of a teacher training on technology-supported teaching and learning

Registry ID: 20365.1v1

Version History

The first version of this entry was published on September 12, 2024 11:06:42 AM EDT
Currently viewing this version.

Section I: General Study Information

PI name: Florian Berens
PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Co-PI name: Katharina Wendebourg
Co-PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Co-PI name: Mareike Kholin
Co-PI affiliation: German Institute for Adult Education - Leibniz Centre for Lifelong Learning (DIE), Bonn

Co-PI name: Leona Colling
Co-PI affiliation: Leibniz-Institut für Wissensmedien (IWM), Tübingen

Co-PI name: Julia Schmidt-Peterson
Co-PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Co-PI name: Manuel Hopp
Co-PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Co-PI name: Tanja Heck
Co-PI affiliation: Leibniz-Institut für Wissensmedien (IWM), Tübingen

Co-PI name: Stephen Bodnar
Co-PI affiliation: Tübingen Center for Digital Education, University of Tübingen

Co-PI name: Walid El Hefny
Co-PI affiliation: Leibniz-Institut für Wissensmedien (IWM), Tübingen

Co-PI name: Florian Nuxoll
Co-PI affiliation: Department of Computational Linguistics, University of Tübingen

Co-PI name: Christoph Deeg
Co-PI affiliation: Center for School Quality and Teacher Training in Baden-Württemberg

Co-PI name: Katja Krey
Co-PI affiliation: Center for School Quality and Teacher Training in Baden-Württemberg

Co-PI name: Josef Schrader
Co-PI affiliation: German Institute for Adult Education - Leibniz Centre for Lifelong Learning (DIE), Bonn

Co-PI name: Hannes Schröter
Co-PI affiliation: German Institute for Adult Education - Leibniz Centre for Lifelong Learning (DIE), Bonn

Co-PI name: Benjamin Nagengast
Co-PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Co-PI name: Detmar Meurers
Co-PI affiliation: Leibniz-Institut für Wissensmedien (IWM), Tübingen

Co-PI name: Ulrich Trautwein
Co-PI affiliation: Hector Research Institute of Education Sciences and Psychology, University of Tübingen

Primary Funding Source(s):
Akademie für Innovative Bildung und Management Heilbronn-Franken gemeinnützige GmbH

Award Number(s):
-

IRB Name:
Faculty of Economics and Social Sciences Ethics Committee, University of Tübingen

IRB Approval Date:
2023-07-24

IRB Approval Number:
A2.5.4.305_vb

Other Registration Name:
also approved by: Ministerium für Kultus, Jugend und Sport Baden-Württemberg (State Ministry of Education)

Other Registration Date:
2023-08-28

Other Registration Number:
KM31-6499-3/139/3

Study Start Date:
2023-09-01

Study End Date:
2024-09-18

Intervention Start Date:
2024-02-01

Timing of entry:
Prior to analysis of outcome data

Brief Abstract:

Intelligent Tutoring Systems (ITS) can provide direct feedback, respond to the learning process adaptively, and motivate learners through gamification, to name only a few features, depending on their purpose and design (Mousavinasab et al., 2021). In previous work, we developed an ITS that is designed for use in English language teaching in German public schools, with students in their third year of learning English as a second language. The focus of this ITS is on supporting students directly through feedback on the one hand (Rudzewitz et al., 2017) and on adaptive task allocation on the other. In a study on its use in the field, we found that students achieve higher learning success with feedback of the ITS than without (Meurers et al., 2019). In more recent work, we have further developed this ITS and have, for example, added dashboards for students and teachers (Colling et al., 2024).

To fully exploit the potential of this ITS, teachers need further training to orchestrate teaching with the ITS appropriately. Using a design-based research approach (Anderson & Shattuck, 2012), we have developed a teacher training program in a co-constructive process with educational researchers, psychologists, second language acquisition experts, teacher trainers and active teachers over several rounds of evaluation that meets criteria for effective teacher training (Desimone & Garet, 2015; Lipowski & Rzejak, 2021). In particular, our training program extends over a long period of time (five meetings spaced over 4.5 months), combines input, testing and reflection phases, interleaves scientific content with practical experience, strengthens collegial cooperation and facilitates coaching. The content is based on the current state of educational research, refers to central teaching requirements and goes into selected content in greater depth (Lipowski & Rzejak, 2021).

In our RCT, we test the hypothesis that students working with an ITS show better learning performance if their teacher participated in such a teacher training. To this end, 45 teachers introduced our ITS in their 7th grade English classes at the beginning of February 2024 and used it till mid July. A randomly selected 23 of these teachers also took part in our teacher training, whereas the other 22 teachers only received technical instruction.

Teachers and students took part in extensive surveys and tests before and after the study. In surveys, both were asked about their perception of teaching (Jaekel et al., 2021), their motivation regarding the course (e.g. Gaspard et al., 2017), their personality (e.g. Soto et al., 2017) and their socio-demographics. In addition, the students were tested on their general English skills using C-tests and on the grammar topics covered during the study period. Teachers were tested regarding their Technological Pedagogical Content Knowledge (TPACK; Lachner et al., 2021).

References:

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research?. Educational Researcher, 41(1), 16–25. https://doi.org/10.3102/0013189X11428813

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). Computers & Education, 52(1), 154-168. https://doi.org/10.1016/j.compedu.2008.07.006

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. Contemporary Educational Psychology, 48, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Colling, L., Kholin, M., & Meurers, D. (2024). A Learning Analytics Dashboard for K-12 English Teachers-Bridging the Gap Between Student Process Data and Teacher Needs. In Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (pp. 538-548). https://doi.org/10.1145/3631700.3665228

Desimone, L. M. & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. Psychology, Society & Education, 7(3), 252–263.
https://doi.org/10.25115/psye.v7i3.515

Jaekel, A.-K., Göllner, R., & Trautwein, U. (2021). How students' perceptions of teaching quality in one subject are impacted by the grades they receive in another subject: Dimensional comparisons in student evaluations of teaching quality. Journal of Educational Psychology, 113(4), 770–783. https://doi.org/10.1037/edu0000488

Lachner, A., Fabian, A., Franke, U., Preiß, J., Jacob, L., Führer, C., Küchler, U., Paravicini, W., Randler, T., &

Thomas, P. (2021). Fostering preservice teachers' technological pedagogical content knowledge (TPACK): A quasi-experimental field study. Computers & Education, 174, 104304. https://doi.org/10.1016/j.compedu.2021.104304.

Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. Annual Review of Applied Linguistics, 39, 161–188. https://doi.org/10.1017/S0267190519000126

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments, 29(1), 142-163. https://doi.org/10.1080/10494820.2018.1558257

Rudzewitz, B., Ziai, R., De Kuthy, K., & Meurers, D. (2017). Developing a web-based workbook for English supporting the interaction of students and teachers. Proceedings of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa 2017, 134, 36–46.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. Journal of Research in Personality, 68, 69-81. https://doi.org/10.1016/j.jrp.2017.02.004

Keywords:

intelligent tutoring system, language learning, RCT, effectiveness, teacher training, artificial intelligence

Comments:

Data collection for the primary outcome ended in July 2024. Data collection for the secondary outcomes is not fully finished yet.

Section II starts on the next page.

Section II: Description of Study

Type of Intervention:
Curriculum/Product, Practice, Professional Development

Topic Area of Intervention:
Education Technology, English Language Learners

Number of intervention arms:
1

Target school level:
7

Target school type:
Rural, Suburban, Urban

Location of Implementation:
International : Europe

Further description of location:
German Federal State of Baden-Wuerttemberg

Brief Description of Intervention Condition:
Teachers in the intervention took part in a professional development course at the same time as they introduced the intelligent tutoring system (like the teachers in the control group did). This training lasted 2.5 days in presence plus two online sessions of 90 minutes. These five meetings were spaced over a period of about 4.5 months to allow the training content to be tested and to enable further joint work based on the experiences. We attached a detailed description of our teacher training as a separate document.

Brief Description of Comparison Condition:
Waiting control condition: teachers and their classes had access to the digital tutoring system, but did not participate in the accompanying teacher training. They will be invited to do so in the next school year.

Comparison condition:
Other

Comments:
All participating teachers introduced the ITS in their classrooms. Therefore, there is no condition with business as usual.

Section III: Research Questions

Confirmatory research questions:

Question 1:
Students who are taught by teachers participating in the teacher training show higher learning performance (measured as the average across three tests on the grammar topics covered during the study period) than those whose teachers did not receive training.

Question 2:

Students who are taught by teachers participating in the teacher training show higher learning performance (measured using a C-test) than those whose teachers did not receive training.

Question 3:
Students who are taught by teachers participating in the teacher training report higher motivation (SEVT mean score) than those whose teachers did not receive training.

Question 4:
Students who are taught by teachers participating in the teacher training report higher levels of perceived internal differentiation (Jaekel et al., 2021) than those whose teachers did not receive training.

Question 5:
Students who are taught by teachers participating in the teacher training show higher numbers of actions in the ITS (click stream) than those whose teachers did not receive training.

Question 6:
Teachers who participate in the teacher training show larger TPACK gains than those without teacher training.

Exploratory research questions:
Question 1:
Do students who are taught by teachers participating in the teacher training differ regarding the perceived teaching quality (as measured with the other constructs from Jaeckel et al., 2021) from those teachers that did not receive training?

Question 2:
Do students who are taught by teachers participating in the teacher training differ regarding the perceived freedom of choice from those whose teachers did not receive training?

Question 3:
Do teachers who participate in the teacher training show larger gains in a TPK test than those without teacher training?

Question 4:
Do teachers who participate in the teacher training differ in perceived TPACK, PK, TK, and TPK from those without teacher training?

Question 5:
Do teachers who participate in the teacher training evaluate the ITS FeedBook more positively than those without teacher training?

Comments:
Although we have 6 research questions and corresponding outcomes, we do not consider them to be equally important. We only consider the variable addressed in question 1, average performance in three grammar tests covering the most important content of the study period, as the primary outcome variable of interest. Further information can be found in the statistical analysis plan.

Section IV-A: Study Design (Selection)

Study Design:
Randomized Trial (RT)

Comments:
-

Section IV-B: Study Design (Input)

Study Design: Input

Unit of random assignment of intervention:
School

Assignment within sites or blocks:
No

Probability of assignment to treatment:
0.5

Unit outcome data measured:
Student

Intermediate clusters between unit of random assignment and unit of measurement:
Yes

Description of intermediate clusters:
teachers, classes

Comments:
As we have almost the same number of schools, teachers and classes in our data. Therefore, specifying a model with 4 levels would not work. This is why we base our power analysis on a two-level model. As the treatement is given to the teachers, we use the teachers as the upper level. We have 23 teachers in the treatement condition (teaching 24 classes) and 22 teachers in the control group (teaching 23 classes).

Design Classification

Based on the responses above, this study has been classified as:
RT: 3-level Cluster Randomized Trial

Section V: Sample Characteristics

Approximate number of students per intermediate cluster (teachers, classes): 24

Approximate number of intermediate clusters (teachers, classes) per school: 1

Number of schools in the comparision condition: 18

Number of schools in the intervention condition: 20

Were there certain students that were targeted for the study?
No

Were there certain students that were excluded from the study?
No

Were there certain intermediate clusters (teachers, classes) that were targeted for the study?
No

Were there certain intermediate clusters (teachers, classes) that were excluded from the study?
No

Were there certain schools that were targeted for the study?
Yes - only Gymnasiums (academic track schools) were eligible to participate

Were there certain schools that were excluded from the study?
No

Comments:
-

Section VI: Outcomes (Input)

Confirmatory question 1: Outcome Measure 1

Outcome domain: Student Achievement- English language competence (grammar)
Minimum detectable effect size: 0.275
Outcome measure: Average across three English language grammar competence tests for grade 7
Scale of outcome measure: Continuous
Normed or state test: No
Test-retest reliability: N/A
Internal consistency: N/A
Inter-rater reliability: N/A
Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 1: Outcome Measure 2

Outcome domain: Student Achievement- English language competence (grammar)
Minimum detectable effect size: 0.275
Outcome measure:
Scale of outcome measure:
Normed or state test:
Same outcome measure in treatment and comparison groups:

Confirmatory question 2: Outcome Measure 1

Outcome domain: Student Achievement- Student Achievement- English language competence
Minimum detectable effect size: 0.275
Outcome measure: C-Test
Scale of outcome measure: Continuous
Normed or state test: Yes
Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 3: Outcome Measure 1

Outcome domain: Student Social, Emotional, & Behavior - Attitudes
Minimum detectable effect size: 0.275
Outcome measure: Motivation SEVT (measured following Gaspard et al., 2017)
Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability: N/A

Internal consistency: N/A

Inter-rater reliability: N/A

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 4: Outcome Measure 1

Outcome domain: Student Outcome Domain - perceived internal differentiation

Minimum detectable effect size: 0.275

Outcome measure: perceived internal differentiation (Jaekel et al., 2021)

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability: N/A

Internal consistency: N/A

Inter-rater reliability: N/A

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 5: Outcome Measure 1

Outcome domain: Student Social, Emotional, & Behavior - Behavior

Minimum detectable effect size: 0.298

Outcome measure: number of actions in the ITS that are related to tasks (clickstream data)

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability: N/A

Internal consistency: N/A

Inter-rater reliability: N/A

Same outcome measure in treatment and comparison groups: Yes

Confirmatory question 6: Outcome Measure 1

Outcome domain: Student Outcome Domain - Teacher Outcome Domain - Technological Pedagogical Content Knowledge (TPACK)

Minimum detectable effect size: 0.534

Outcome measure: teacher achievement in a Technological Pedagogical Content Knowledge (TPACK) test (Lachner et al., 2021)

Scale of outcome measure: Continuous

Normed or state test: No

Test-retest reliability: N/A

Internal consistency: N/A

Inter-rater reliability: N/A

Same outcome measure in treatment and comparison groups: Yes

Comments:

As we have almost the same number of schools, teachers and classes in our data, specifying a model with 4 levels would not work. We therefore base our power analysis on a two-level model. As the treatement is given to the teachers, we use the teachers as the upper level. We have 23 teachers in the treatment condition (teaching 24 classes) and 22 teachers in the control group (teaching 23 classes).

Section VII: Analysis Plan

Baseline data collected prior to start of intervention:
Yes

Description of baseline data:
We have baseline data for all for all aspects mentioned in the research questions (except for clickstreams) and in addition on big five personality and some socio demographics. However, for baseline equivalence we only consider the six primary and secondary outcomes.

Covariates you plan to include in the model:
Student Pretest

Covariates you plan to include in the model:

Covariates you plan to include in the model:

Analytic model:

We describe our analysis plan and the analytic model in additional material.

Plan to handle cases with missing outcome data:
full information maximum likelihood estimation

Planned multiple comparisons adjustment, confirmatory question 1 (Student Achievement):
No

Comments:
Regarding multiple comparisons, see statistical analysis plan

---

Section VIII: Additional Information

Links:
No links have been added yet.

Files:
File Name: AI2Teach_Description-ITS FeedBook.pdf
Description: AI2Teach: Description of the ITS in use

File Name: AI2Teach_Preregistration_Power-analysis.html
Description: AI2Teach: Power Analysis (HTML version)

File Name: AI2Teach_Preregistration_Power analysis.Rmd
Description: AI2Teach: Power Analysis (Markdown version)

File Name: AI2Teach_Preregistration_Statistical Analysis Plan.pdf
Description:

Comments:
-
