Do intelligent tutoring systems benefit K-12 students in the U.S.? A meta-analysis

Walter Leite[1], Huibin Zhang[2], Shibani Rana[1], Yide Hao[1], Amber Hatch[1], Lingchen Kong[1], Huan Kuang[3]

1. University of Florida
2. University of Tennessee, Knoxville
3. Florida State University

Intelligent Tutoring Systems (ITS)

- ITS are software that interact with students as they solve problems on a turn-by-turn basis, providing feedback on the current steps and/or hints
- ITS can have different degrees of granularity of interaction with the student, such as interaction at each answer, step, or sub-step.
- ITS offer a variety of supports for learning, such as on-demand hints, just-in-time hinds, content sequencing, question sequence, feedback, and explanations
- ITS commonly incorporate a model of student knowledge.

Evidence of effectiveness of ITS

- There have been a few meta-analyses of ITS, but none focusing on U.S. K12 student populations.
- VanLehn (2011) found positive effects of step-based (d = 0.76), sub-step-based (d = 0.40) and answer-based (d = 0.31) ITS as compared to no tutoring.
- Steenbergen-Hu and Cooper (2013) found no effect of ITS in K12 education as compared to classroom instruction.
- Ma et al. (2014) found positive effects of ITS use as compared to large-group human instruction (g=0.44) with a sample of both K12 and Higher Ed. Studies.
- Kulik and Fletcher (2016) found an effect size of 0.41 with K12 samples.

Limitations of previous meta-analyses

- Previous meta-analyses have summarized studies performed worldwide, which may not generalize to U.S. K-12 student populations.
- Most previous meta-analyses of ITS did not separate K-12 from higher education studies.
- Most previous meta-analyses did not focus on experimental or quasi-experimental studies.

Objective of the Current Study

- Summarize the treatment effects of ITS in U.S. K-12 student populations.
- Evaluate heterogeneity of treatment effects.
- Evaluate studies that meet the What Works Clearinghouse standards without reservations or with reservations.

Method

- Stages:

1. Literature search
2. Article screening independently by 2 reviewers, with a 3rd resolving conflicts
3. Coding independently by 2 reviewers, with a 3rd resolving conflicts.
4. Model fitting

- Search Databases:
  - Learntechlib
  - ERIC
  - PsycInfo
  - Academic Search Premier
  - IEEE Xplore Digital Library
  - ACM Digital Library
  - Proquest Dissertation and Theses.

Method: Keywords

- [[Abstract: "intelligent tutor*"] OR [Abstract: "artificial tutor*"] OR [Abstract: "computer tutor*"] OR [Abstract: "computer-assisted tutor*"] OR [Abstract: "computer-based tutor*"] OR [Abstract: "intelligent learning environment*"] OR [Abstract: "computer coach*"] OR [Abstract: "online tutor*"] OR [Abstract: "e-tutor*"] OR [Abstract: "electronic tutor*"] OR [Abstract: "web-based tutor*"] OR [Abstract: "intelligent virtual"] OR [Abstract: "intelligent agent"] OR [Abstract: "cognit* tutor*"] OR [Abstract: "adapt* tutor*"] OR [Abstract: "virtual companion"] OR [Abstract: "intelligent coach*"]] AND [Abstract: student*] AND NOT [Abstract: college] AND NOT [Abstract: undergraduate] AND [Publication Date: (01/01/2011 TO 12/31/2021)]

Methods: Inclusion Criteria

1. One of the systems examined in the study meets definition of intelligent tutoring system
2. Experimental study or propensity score analysis (matching, weighting, stratification) study or regression discontinuity design of intelligent tutoring systems.
3. Studies published between January 1st 2011 and December 31st 2021.
4. Studies had to focus on students in grades K–12.
5. Studies had to measure the effectiveness of ITS on student achievement
6. Studies had to have used an independent comparison group that was non-ITS.
7. Studies had to be conducted with a sample from the United States of America.
8. Studies published in academic journals, dissertations/thesis, and conference proceedings.

Methods: Data Extraction

- MUTOS Framework (Becker, 2017) was used to extract data about five dimensions of a study:
  - Methods (M),
  - Units (U)
  - Treatments (T)
  - Observing operations (O)
  - Setting (S)

Methods: Multivariate random effect model

$$T_{ik} = \theta_{ik} + \varepsilon_{ik} = \mu + \eta_k + \phi_{ik} + \varepsilon_{ik}$$

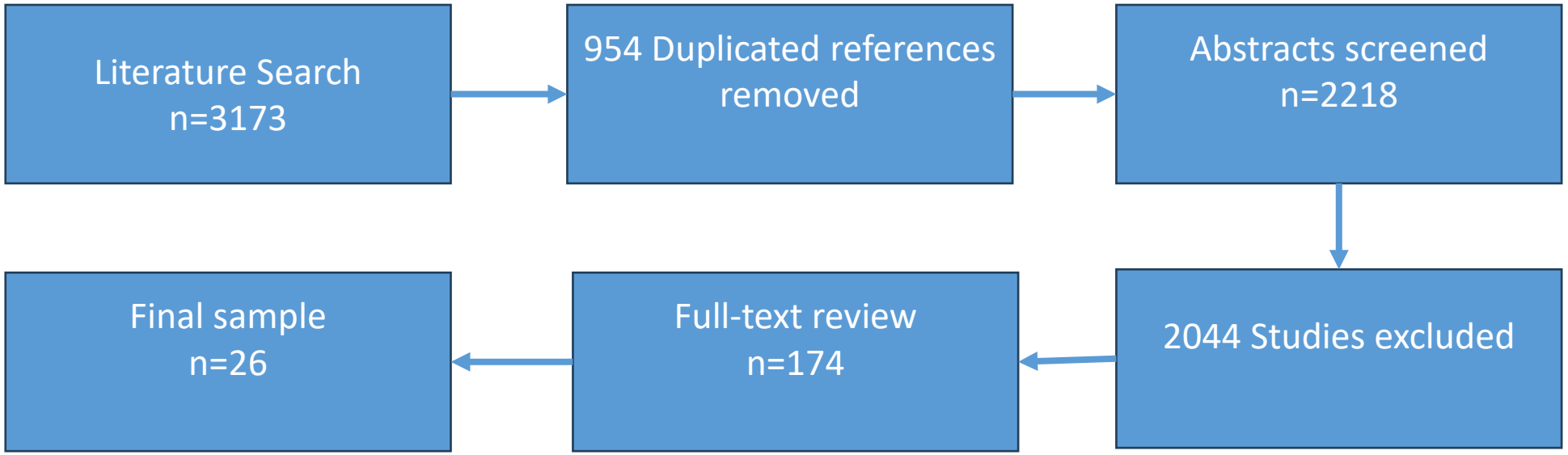
μ is the average effect size across studies, Var $(\eta_k) = \tau^2$ indicates between-study variation of true study-average effect size, Var $(\phi_{ik}) = \omega^2$ is within-study variation of true effect size, and ε_{ik} is the sampling error. Var $(\varepsilon_{ik}) = s_{ik}^2$ is known.

Methods: Moderator Analysis

$$T_{ik} = \sum_{1}^{P} \beta_p X_{pk} + \eta_k + \phi_{ik} + \varepsilon_{ik}$$

- X_{pk} indicates the p^{th} covarite in the study k and β denote the corresponding regression coefficients.

Results: Prisma



Results: Overall Effect Size

- There were 95 effect sizes in 26 articles.
- There was a significant positive effect size of ITS on U.S. K-12 students' learning outcomes (g=0.360, SE=0.046, p<0.001)
- **Publication Bias:** For Rosenthal approach, the fail-safe N is 51008 with a target significance level of 0.05, while the fail-safe N is 299 for the Orwin approach with a target effect size of 0.10.

Meta-regression Analysis for M Dimension

Moderator	Coefficient (β)	SE	t-Statistic	p-value
Type of Publication				
Conference proceeding	<0.01	0.16	0.02	0.99
Dissertation or thesis	**0.44**	**0.16**	**2.78**	**0.04**
Journal article	**0.32**	**0.06**	**5.79**	**0.02**

Meta-regression Analysis for T Dimension

Moderator	Coefficient (β)	SE	t-Statistic	p-value
ITS use in school				
As a separate activity	0.80	0.34	2.34	0.14
As the main instruction method	0.83	0.35	2.39	0.16
As a partial substitute for the regular curriculum	**0.53**	**0.19**	**2.73**	**0.03**
Not applicable	0.54	0.47	1.16	0.29
Other	0.63	0.47	1.35	0.24

Meta-regression Analysis for O Dimension

Moderator	Coefficient (β)	SE	t-Statistic	p-value
Type of learning outcome				
Mathematics	**0.91**	**0.11**	**8.38**	**<0.001**
Reading	**1.21**	**0.14**	**8.88**	**0.01**
Science	0.54	0.18	3.02	0.06
Writing	**1.27**	**0.14**	**8.97**	**<0.001**
Instructor type				
Different instructors for treatment and control groups	**-0.52**	**0.06**	**-8.49**	**0.01**
No instructor	-0.23	0.14	-1.69	0.21
Not specified	-0.47	0.00	-114.60	0.01
Same instructor for both treatment and control groups	-0.52	0.49	-1.06	0.39
Measurement Timing				
End of School Year	0.00	0.04	0.04	0.97
End of Semester	**0.48**	**0.07**	**6.99**	**<0.001**
End of the unit	0.11	0.12	0.89	0.46
Immediately after Intervention	0.27	0.10	2.65	0.07

Results: Moderation by Type of Control Group

Type of Publication	K	g	95% CI	p-value
Classroom Instruction	48	0.414	0.257- 0.571	<0.05
Small Group	1	-0.181	-1.226- 0.865	0.732

Results: Moderation by Type of Publication

Type of Publication	K	g	95% CI	p-value
Dissertation or Thesis	9	0.506	0.088- 0.924	<0.05
Journal	84	0.357	0.238- 0.475	<0.05
Conference	2	0.061	-0.672-0.793	0.870

Results: Moderation by ITS Use in School

ITS Use in School	K	g	95% CI	p-value
As a partial substitute for the regular curriculum	6	0.249	-0.168-0.666	0.242
As a separate activity	34	0.408	0.205-0.612	<0.05
As the main instruction method	47	0.399	0.243-0.555	<0.05
Other	1	0.090	-0.946-1.125	0.864

Results: Moderation by Learning Outcome

Learning Outcome	K	g	95% CI	p-value
Mathematics	26	0.272	0.060-0.483	<0.05
Reading	62	0.358	0.226- 0.490	<0.05
Science	3	0.137	-0.437-0.711	0.637
Writing	4	1.011	0.503- 1.519	<0.05

Results: Moderation by Instructor Type

Instructor Type	K	g	95% CI	p-value
Different instructors for treatment and control groups	73	0.346	0.217-0.476	<0.05
Same instructor for treatment and control groups	4	0.691	0.071-1.311	<0.05
No Instructor	3	0.979	0.086-1.873	<0.05

Results: Moderation by Measurement Timing

Measurement Timing	K	g	95% CI	p-value
End of Semester	1	0.564	-0.700-1.829	0.378
End of School Year	41	0.273	0.109-0.437	<0.05
End of the Unit	6	0.218	-0.211-0.647	0.316
Immediately after Intervention	26	0.620	0.387-0.854	<0.05

Conclusion

- The effect size of ITS we identified with studies of US K-12 samples was moderate (i.e. g=0.360), which corroborates Kulik and Fletcher's (2016) finding of a significant moderate effect (Glass $\Delta = 0.41$.) with K-12 samples.

- Our results contrast with Steembergen-Hu and Cooper's (2013) findings of no effect of K-12 mathematics ITS.

- Studies in dissertations or theses had a higher effect size than studies published in journals.

- School districts could increase the benefits of having computers available to all students by making ITS available to students on these computers.


Thank you!
walter.leite@coe.ufl.edu
https://virtuallearninglab.org/
UF College of Education