Trustworthy Machine Learning for Automated Writing Feedback

David Adamson
High School Teacher → CMU → Lightside Labs → **Turnitin**

Applied Machine Learning Days @ EPFL 2019

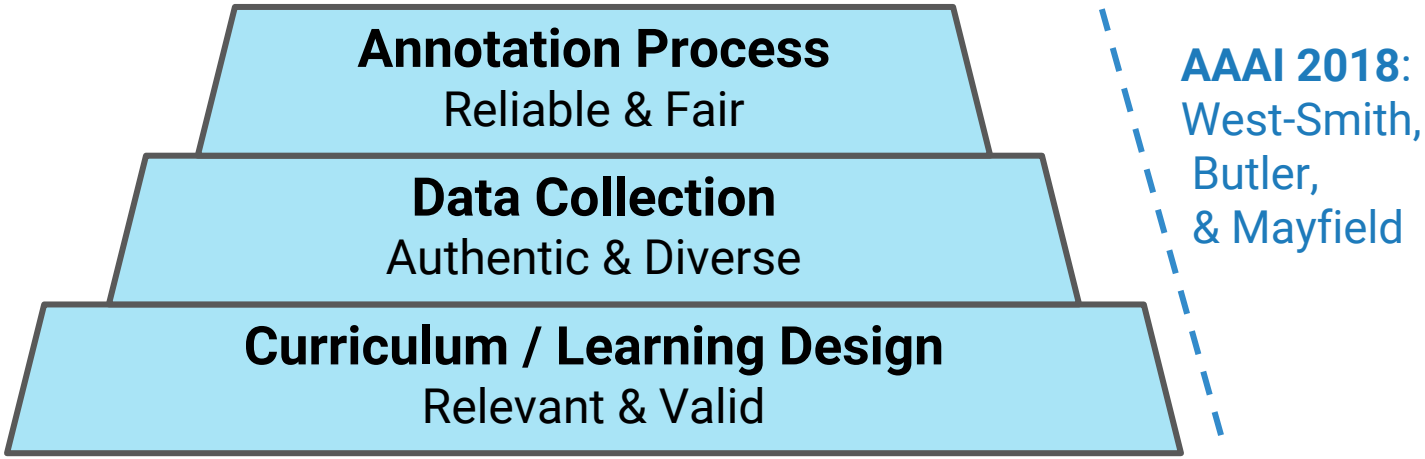
Outline

- A Framework for Trustworthy ML, for Education
  - Because we want teachers and administrators to trust it enough to buy it
  - Because we want to make a difference in the classroom
- Turnitin's *Revision Assistant*
- Efficacy Evaluation in the Wild
- A cautionary tale of curricula and datasets
  (in RA, and everywhere!)

Trustworthy Machine Learning

for Education
(and Writing Feedback!)

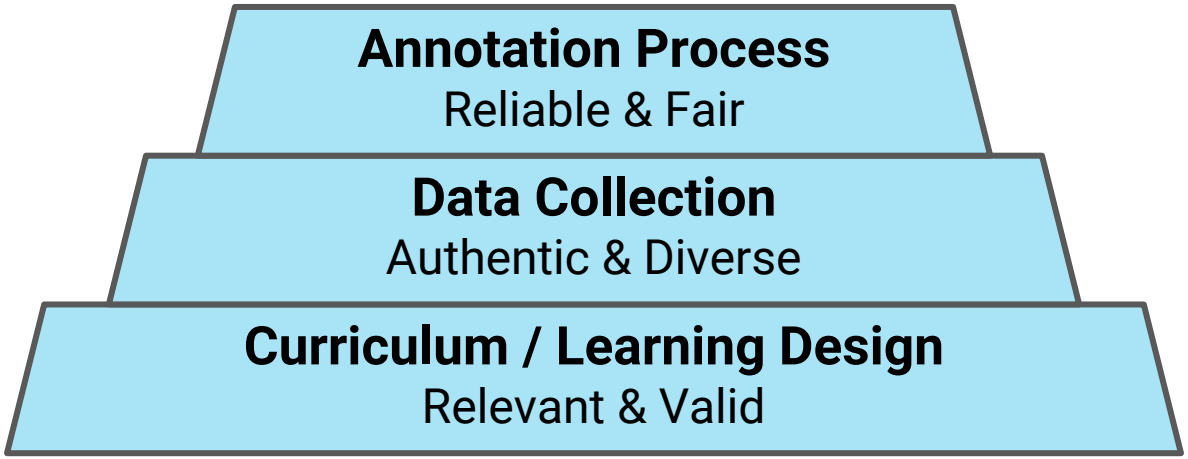
Trustworthy Machine Learning for Education



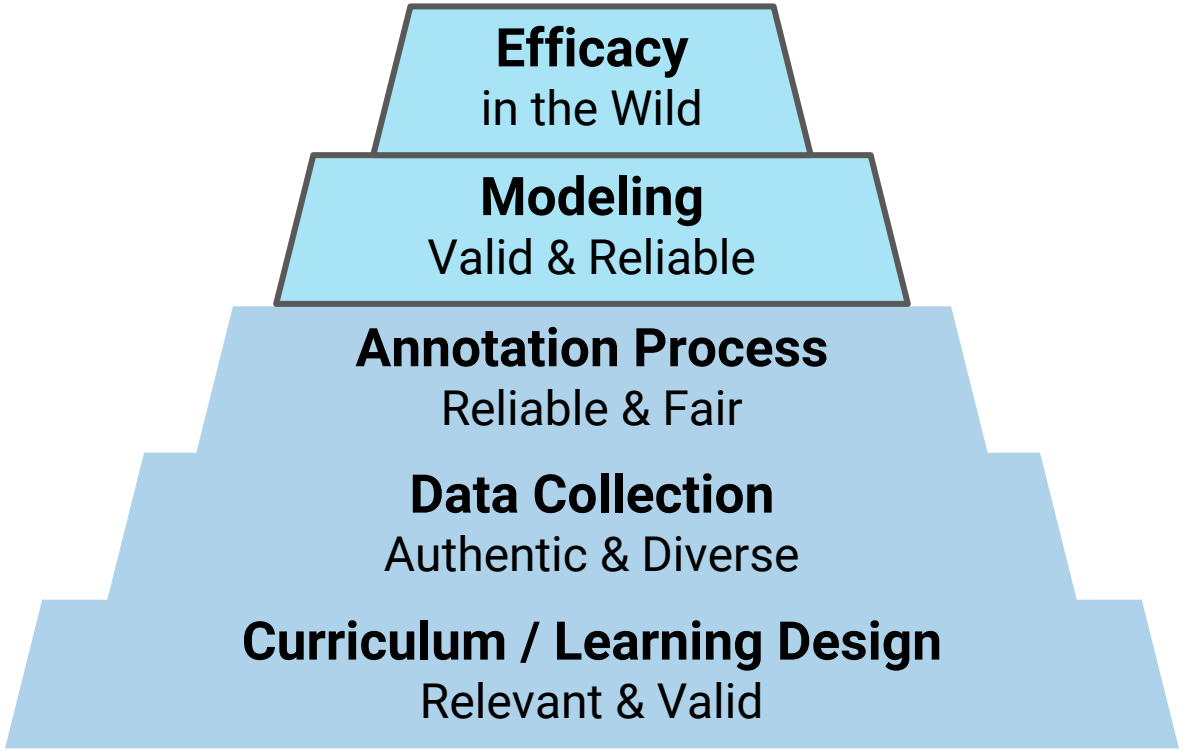
Trustworthy Machine Learning for Education

- **Curriculum & Learning Design**
  - Must be relevant and authentic – so teachers actually use it
  - Difficult to anticipate how/when modules will be applied (Ball & Cohen 1996, Nguyen, Huptych, & Rientes, 2018)
- **Training Data**
  - Must be authentic, representative, and diverse
- **Annotation Process**
  - Rubric must be clearly articulated and authentic for the task
  - Inter-rater reliability is not enough (Hovy and Lavid, 2010)

Trustworthy Machine Learning for Education



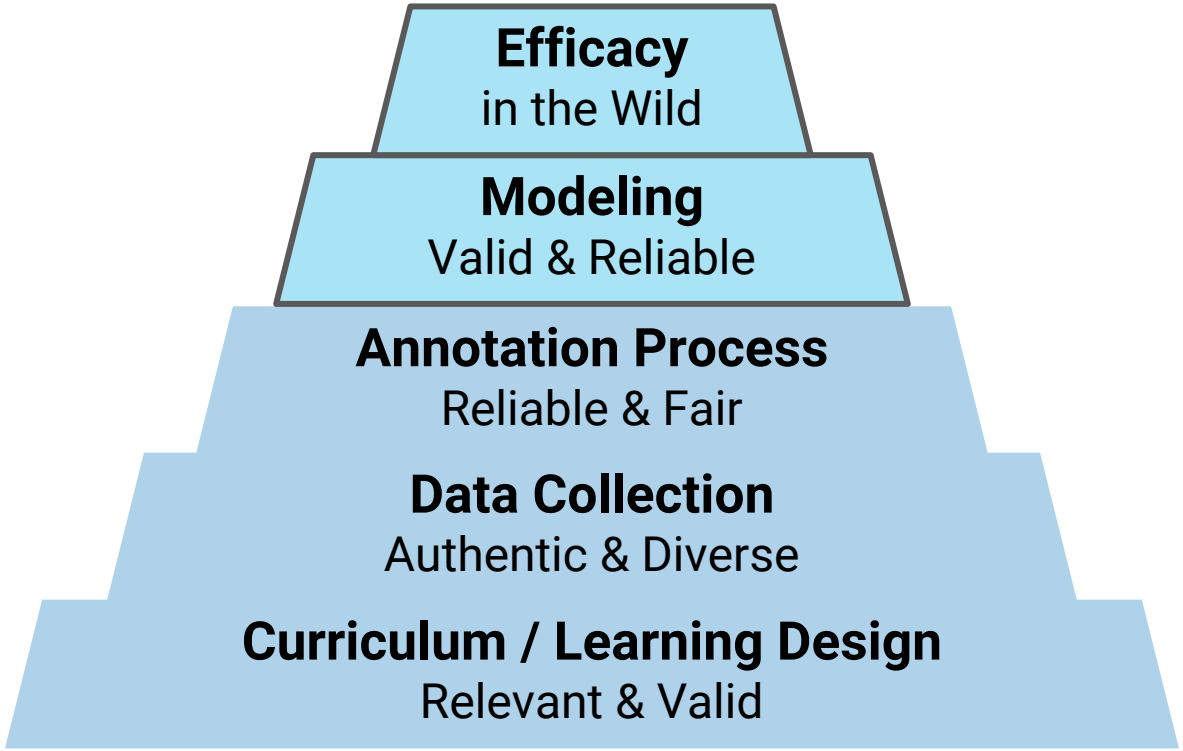
Trustworthy Machine Learning for Education



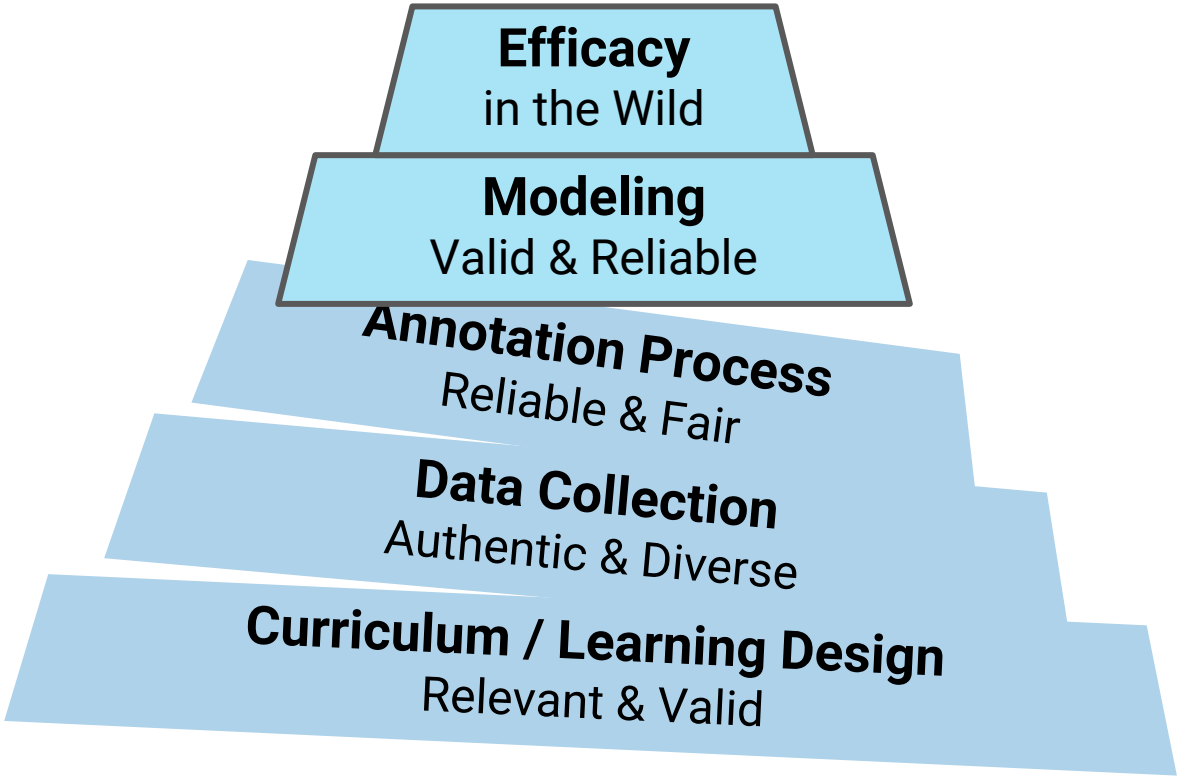
Trustworthy Machine Learning for Education

- **Model** reliability
  - Agreement with raters
  - Robustness to actual student input
  - More important to user trust than fancy features (West-Smith et al, 2018)
- Does a tool based on your model(s) **make a difference** to students and teachers?
  - Teachers and students integrate the tool into their practice
  - Predict (and improve!) student outcomes
  - Evaluate at scale (Grimes & Warschauer 2010, Wilson & Czik 2016)

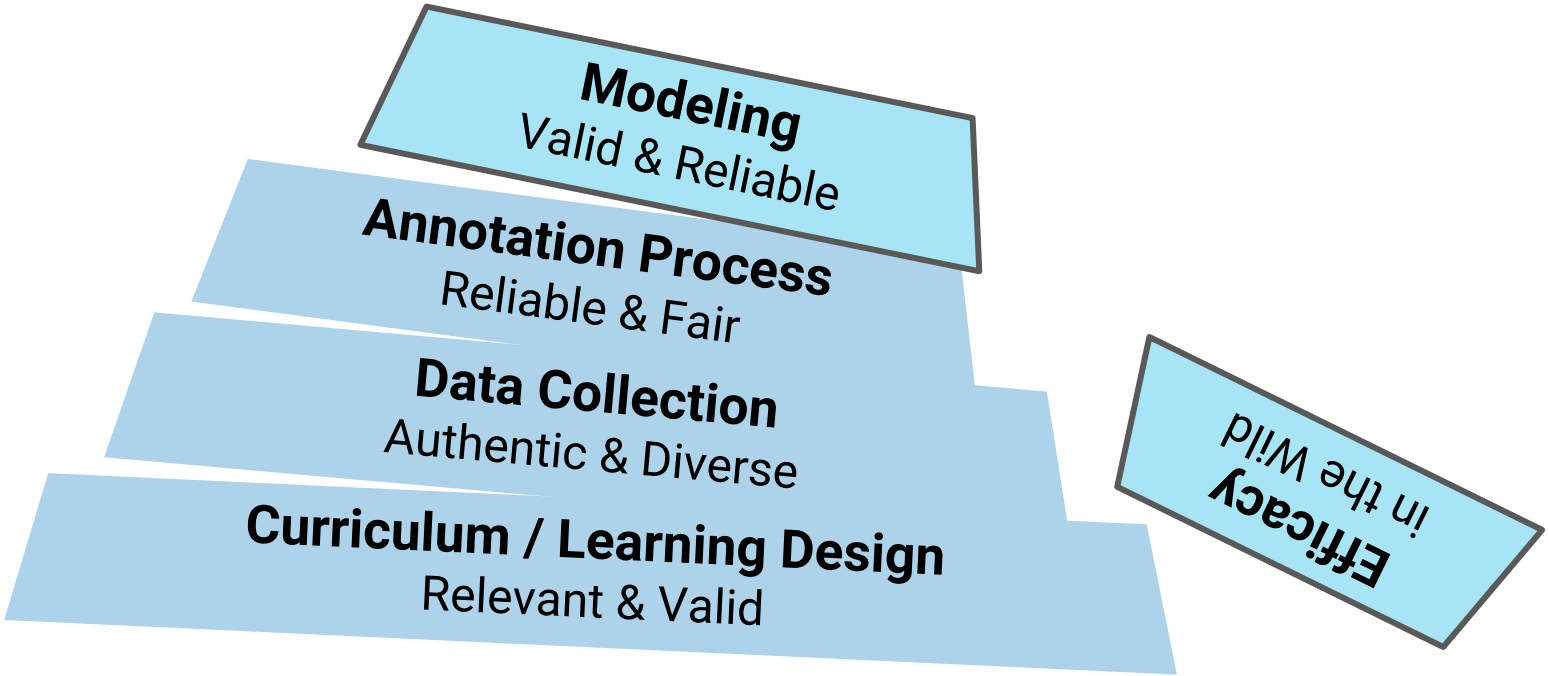
Trustworthy Machine Learning for Education



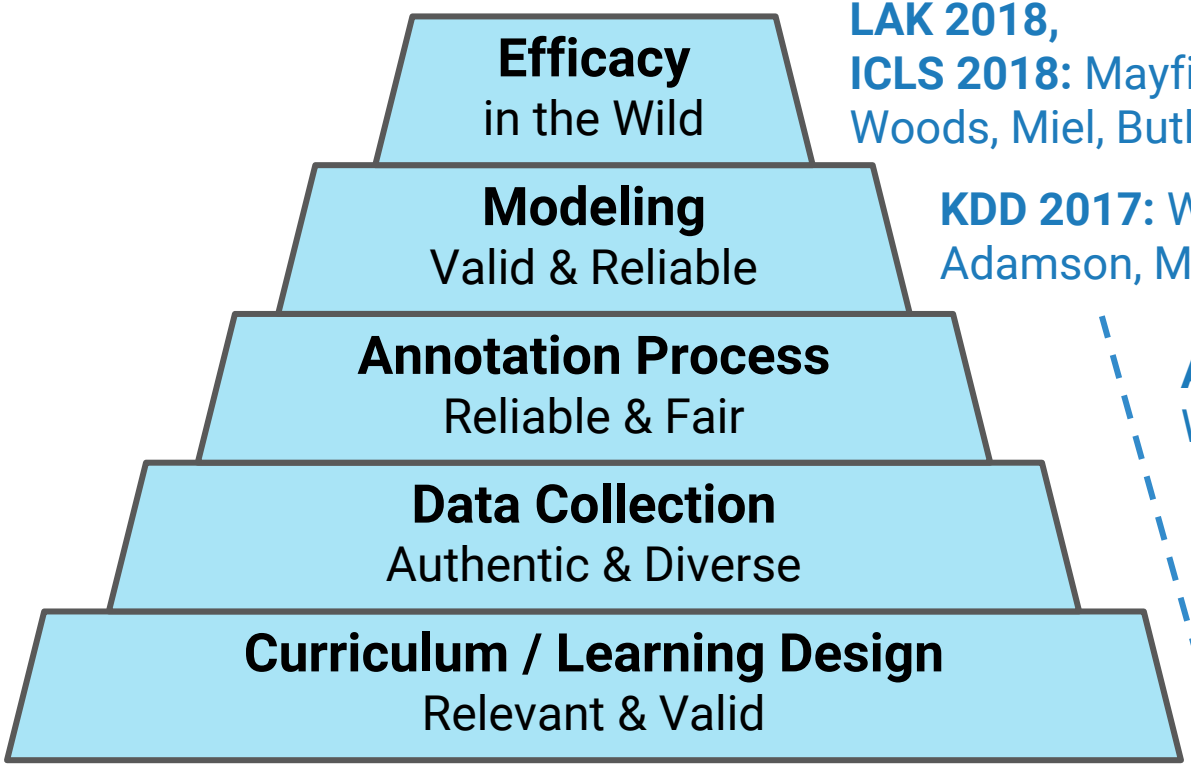
Trustworthy Machine Learning for Education



Trustworthy Machine Learning for Education



Trustworthy Machine Learning for Education



LAK 2018,
ICLS 2018: Mayfield, Adamson, Woods, Miel, Butler & Crivelli

KDD 2017: Woods, Adamson, Miel, & Mayfield

AAAI 2018: West-Smith, Butler, & Mayfield

Revision Assistant

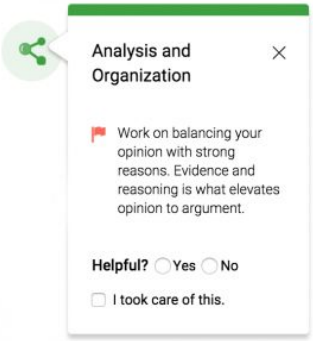
Automated Writing Feedback

People Are Great

Curriculum, User Experience, Product, and Machine Intelligence teams

Revision Assistant

- On-demand feedback for a growing library of writing **prompts** (173+)
- Prompts, scores and feedback are grounded in genre-specific, standards-aligned **rubrics** (West-Smith et al., 2018)
- **“Signal Check”** assignments offer targeted, actionable, model-driven formative feedback – see *Formative Essay Feedback Using Predictive Scoring Models* (Woods et al., 2017)
- **“Spot Check”** assignments can be used as formative assessments, allowing teachers to collect student writing with score predictions, without Signal Check feedback.



Organization

How well do I organize my ideas so that readers can follow my analysis?

State your claim in the introduction, group related ideas into body paragraphs, and restate your claim in the conclusion. Use topic sentences and clear transitions to show the relationships between and among your ideas.

The Railroad

Harriet Tubman was born in Maryland in 1820 and escaped slavery in 1849. Harriet went to the South to rescue family members and other people living in slavery through the Underground Railroad. In the narrative, ""Harriet had found it hard to leave the warmth and friendliness, too. But she urged them on"" (3). This shows that Harriet wasn't willing to give up and she was motivating everyone to move forward. Another quote stated, ""she had promised her passengers food and rest and warmth, and instead of that, there would be hunger and cold and more walking over the frozen ground"" (2). This quote from the narrative shows that Harriet cares for the slaves and that she is willing to risk anything for their safety.

Harriet Tubman was an inspiration to many people. Unlike some slaves, Harriet wasn't planning on giving up. One of the slaves wanted to give up, but Harriet pointed a gun at him then said, ""Go on with us or die."" Harriet was obviously more motivated than the passengers. The

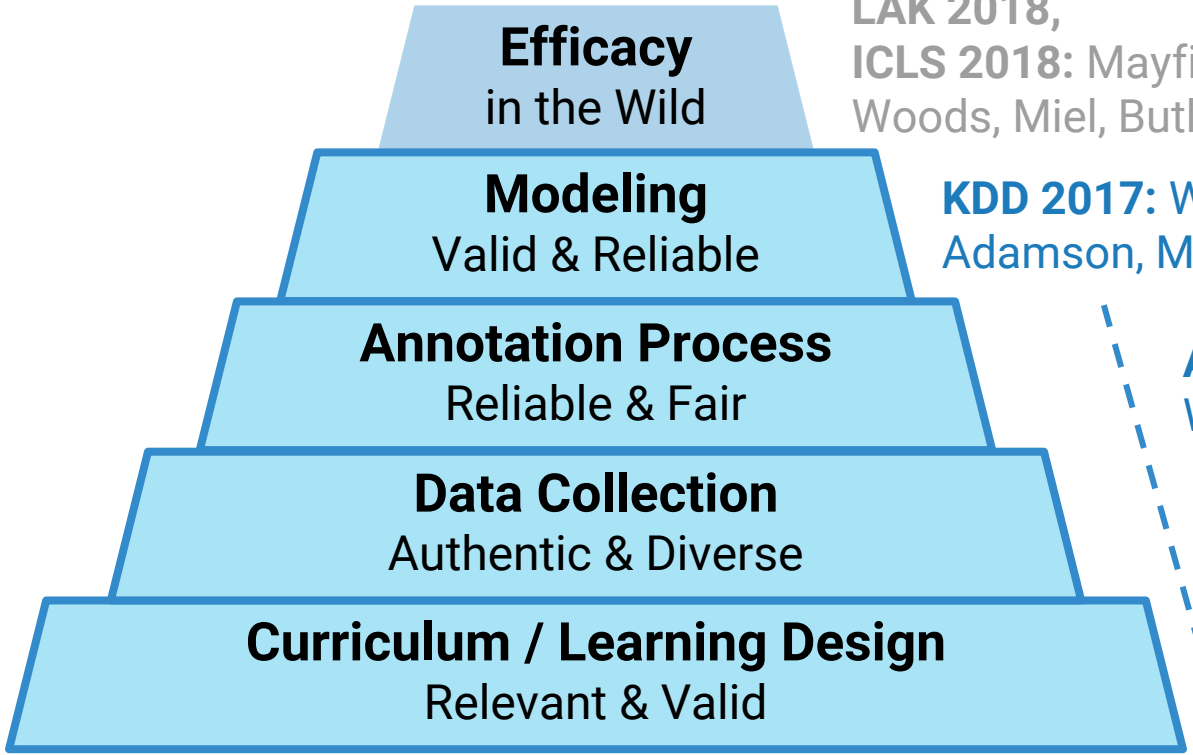
Organization

You are beginning to present your ideas in a purposeful way. For each paragraph, include a clear topic sentence, supporting evidence, and an explanation of how that evidence supports your claim.

Helpful? Yes No

I took care of this.

Building Trust in Revision Assistant



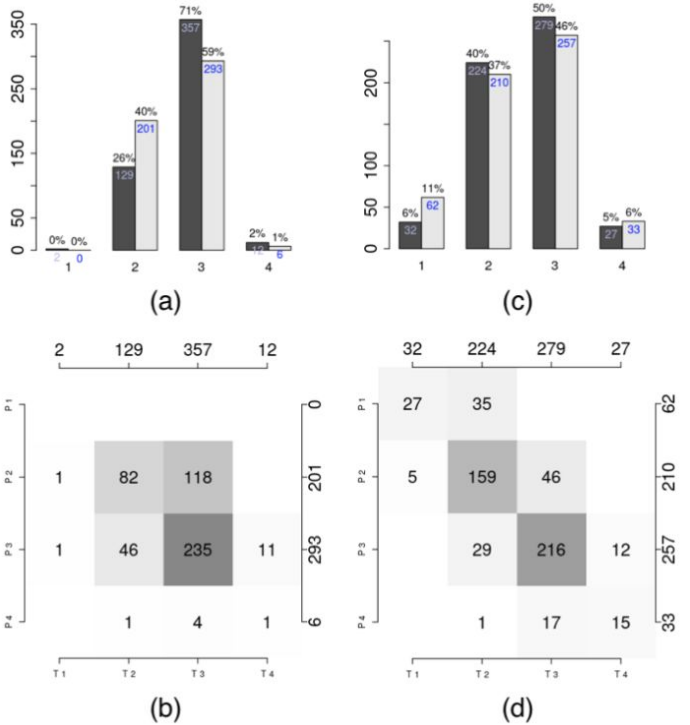
LAK 2018,
ICLS 2018: Mayfield, Adamson, Woods, Miel, Butler & Crivelli

KDD 2017: Woods, Adamson, Miel, & Mayfield

AAAI 2018: West-Smith, Butler, & Mayfield

Foundational Practices

- **Curriculum:** Prompt development and rubric design led by in-house curriculum specialists and educators, in partnership with school districts
- **Data and Annotation:** High-touch involvement in dataset collection and scoring practices. Iterate on rubric design with annotators.
- **Models:** Strict standards for model acceptance (gatekeepers for production deployment)



Before

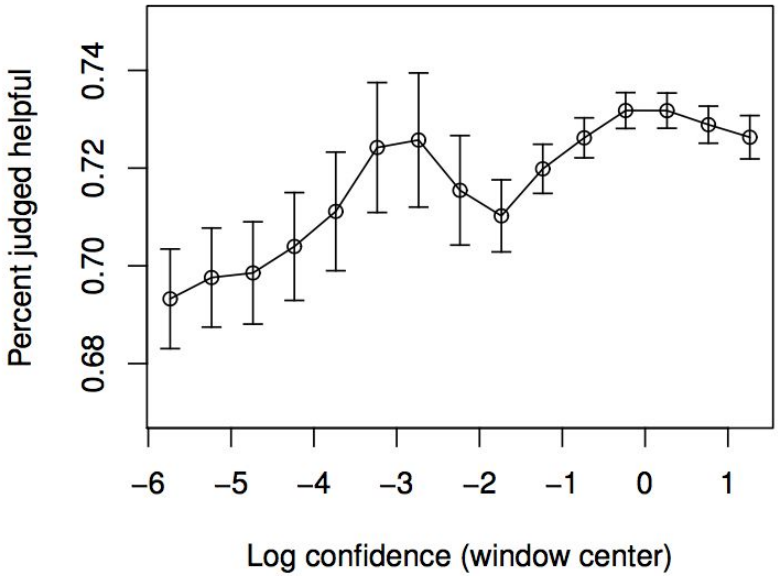
After

Internal Efficacy Measures

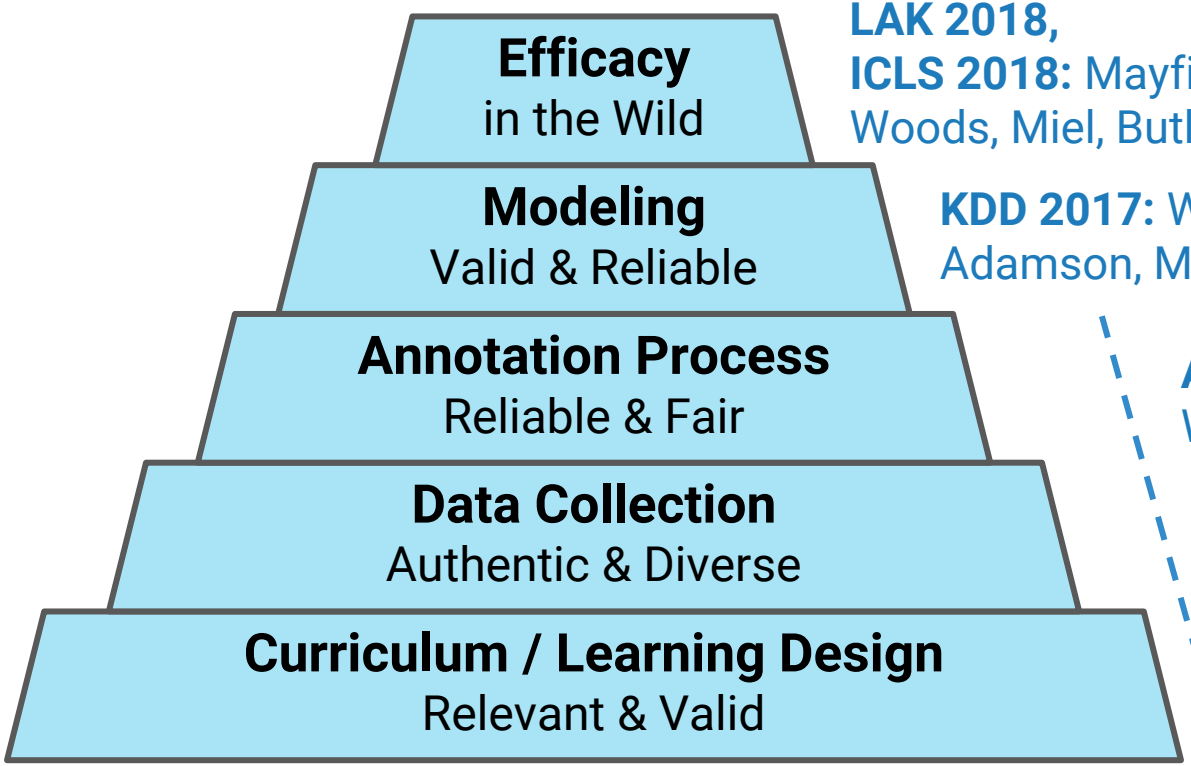
- This school year to date (Aug 2018 through Jan 2018): About 86,000 active students, working with 17,000 active teachers, requested feedback about 750,000 times, in over 300 schools (150% growth vs. last year)

- Student rating of comment helpfulness is significantly correlated with RA's predictions of sentence impact. (Woods et al., 2017)

- Predicted essay scores increase with continued feedback and drafting (Woods et al., 2017)



Building Trust in Revision Assistant

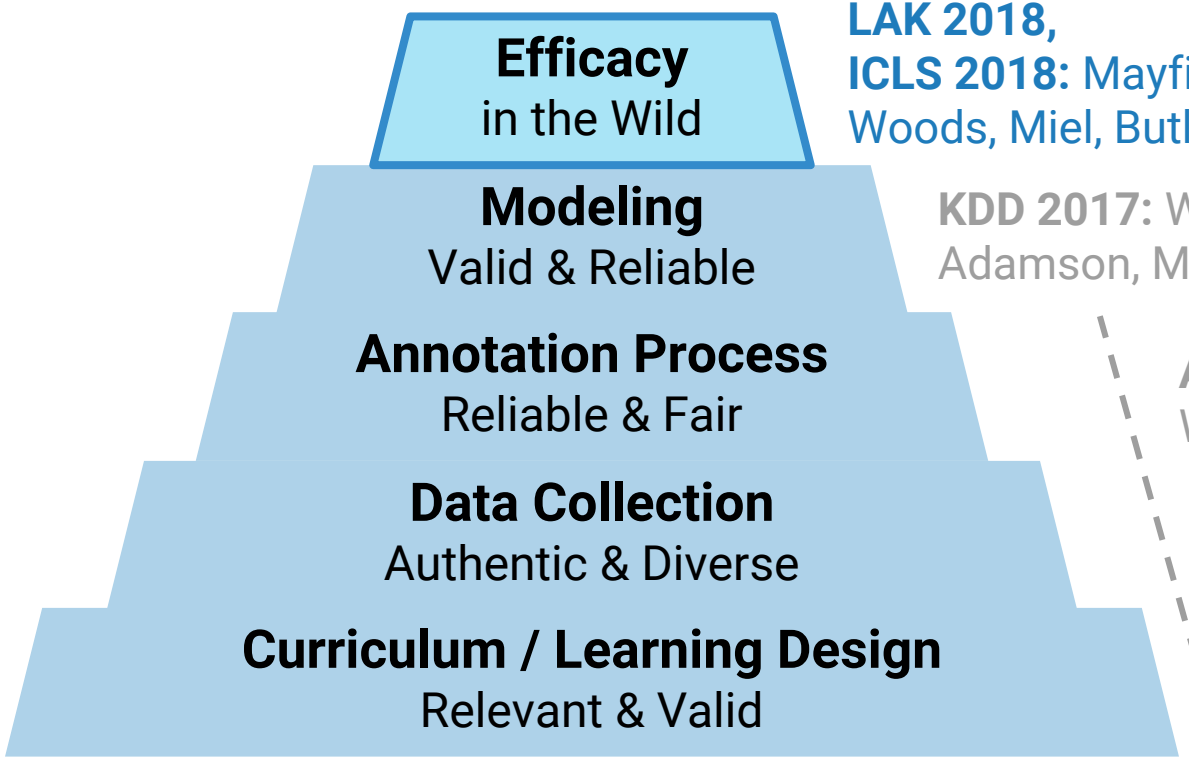


LAK 2018,
ICLS 2018: Mayfield, Adamson, Woods, Miel, Butler & Crivelli

KDD 2017: Woods, Adamson, Miel, & Mayfield

AAAI 2018: West-Smith, Butler, & Mayfield

Building Trust in Revision Assistant



LAK 2018,
ICLS 2018: Mayfield, Adamson, Woods, Miel, Butler & Crivelli

KDD 2017: Woods, Adamson, Miel, & Mayfield

AAAI 2018: West-Smith, Butler, & Mayfield

Efficacy in the Wild

LAK 2018
ICLS 2018

Automated Writing Evaluation in the Wild

Criterion, WriteToLearn, MyAccess

- Feedback systems based on classical AES approaches have demonstrated mixed results around revision, classroom management, increased motivation, but haven't demonstrated changes in school outcomes.
  (Scharber et al., 2008) **(Grimes and Warschauer, 2010)**

WriteLab, PEG, **Revision Assistant**

- More recent focus on feedback-oriented automated writing evaluation – actionable next steps, that students perceive as informative and valuable.
  (Riedel et al., 2006) (Roscoe et al., 2013)
- Modest-at-best improvements on transfer tasks. (Wilson and Czik, 2016)
- Little work on the longitudinal effect of AWE in the classroom, **or at a larger scale.**

Questions

What can prolonged Revision Assistant usage tell us about individual growth and school outcomes?

- Can *RA* be used to help forecast end-of-year student outcomes?
- How does *RA* fit in to classroom practice and district-wide initiatives?

Forecasting ELA Outcomes

In Texas

Forecasting Outcomes

In a mid-sized Texan high school (~2600 students),

Four participating teachers used Revision Assistant in their 9th-grade classes (85 students)

Bookended by a school-administered benchmark and the statewide ELA exam (MC + Written Composition)

Can Revision Assistant be used to help forecast exam outcomes?

Benchmark Assessment
Fall 2016

Revision Assistant
Spot Check Pretest
Feb 2017

Revision Assistant
Signal Check Practice
Feb-April

Revision Assistant
Spot Check Posttest
April 2017

STAAR English I Exam
May 2017

Results

- The benchmark test is a reasonable predictor of end-of-year raw score, but doesn't tell us much about the Written Composition component.
- Full model with benchmark and Revision Assistant data significantly outperforms ($p < 0.01$) benchmark alone.

Model	Written Composition		ELA Multiple Choice Score	
	r	RMSE	r	RMSE
Benchmark Only	0.20	0.99	0.58	6.56
Spot Check Predictions	0.43	0.90	0.28	7.36
Benchmark + Spot + RA	**0.51**	**0.87**	**0.67**	**5.99**

Discussion

- RA predictions and usage data could supplement benchmark tests in forecasting end-of-year student outcomes.

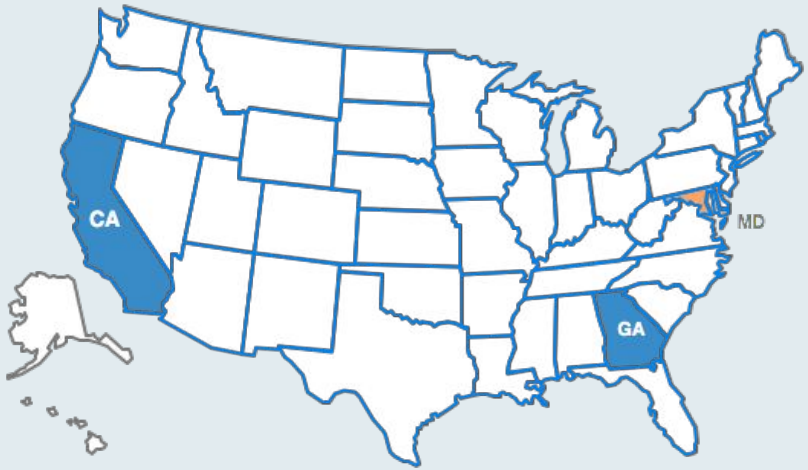
- Teacher pacing and usage varied - we'd prefer to repeat this study with more consistent classroom implementation.
- Sample is small
- No evidence that RA feedback improved student outcomes

Supporting District Initiatives

RA + district improvement

Supporting District Initiatives

In Maryland



Supporting District Initiatives

At a medium-large school district in Maryland, five middle schools integrated Revision Assistant into their curriculum as part of a 2016-2017 district-wide emphasis on writing across all subject areas.

This initiative included intensive district-led professional development, in addition to hands-on training with Revision Assistant.

"We were in the process of working with our English, social studies, and science teachers to develop writing prompts that could work across content areas and trying to push writing in those three areas. Revision Assistant really fit into the plans we had already put in place." *- Content Specialist*

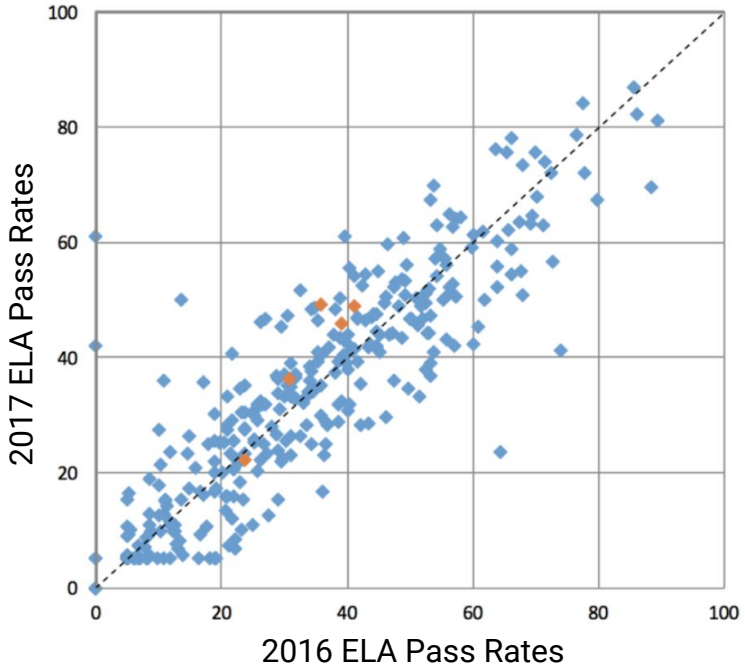
Supporting District Initiatives

PARCC 8th grade English/Language Arts exam (aligned to the Common Core state standards) administered in Spring 2016 and Spring 2017.

How did the target schools' improvement compare to other Maryland schools?

Growth by School

School	2016	2017	Change
1	35.8	49.1	+13.3%
2	41.1	49.0	+7.9%
3	39.0	45.8	+6.8%
4	30.7	36.2	+5.5%
5	23.7	22.2	-1.5%
Treatment (n=5)	**34.1**	**40.3**	**+6.4%**
Control (n=3)	40.4	38.7	-1.9%
Maryland (n=352)	31.8	32.0	+0.2%



Schools using Revision Assistant as part of a district initiative (with lots of support and involvement from Turnitin) showed an encouraging growth in pass rates on MD's ELA exam.

Discussion

- The five Revision Assistant schools in this district showed greater growth than 94% of randomly-selected subsets of Maryland schools.

- Response to Revision Assistant adoption was positive:
  - ***“I like to know that my students have worked on a writing piece several times before I see the finished product”***

- This was part of a larger district initiative, and we can’t yet tease out RA’s role in the improvement.

Replication!

ICLS 2018

Replication in Georgia

- Maryland findings were replicated in another school district (9% growth, vs about 4% statewide), in another state, against a different end-of-year ELA exam, at both 9th and 11th grades.

- The district implemented RA without any special support – RA's effect on student outcomes, at the school level, is not necessarily dependent on Turnitin staff's close involvement in implementation.

- Sample is twice as large as in Maryland study, but still only 10 schools, one school year, and one timed standardized test as an outcome variable.

Finding Correlated Factors

ICLS 2018

Finding Correlated Factors

in California

When Are RA Schools Successful?

Setting

- 33 high schools in California in 2016-2017, in all regions of the state.
- About 170,000 essay drafts, about 15% of all RA usage nationwide
- Students tested on CASPP, part of Smarter Balanced (Common Core)

Research Questions

- Describe demographic variables that characterize RA schools - who buys the product and is there a bias?
- Identify variables (in demographics or in usage) that predict growth.

When Are RA Schools Successful?

What characterizes an RA-implementing high school?

- Large student body (about 2000 students, double state average)
- Higher pre-existing ELA scores (66% pass rate vs. 51% state average)
- Higher overall graduation rate (95% vs. 82% state average)

How pervasive is RA usage in adopting districts?

- 15 schools in 4 districtwide implementations (all schools participate)
- 9 schools in partial district adoptions (2+ schools in a district)
- 9 schools implemented RA standalone (no district support)

Measuring Impact

Outcome Variables

- Overall passing (3+) as % of test-takers
- “Exceeds expectations” (4/4) as % of test-takers
- Passing rates (3+) in each of 4 subscores

Method

- Test pre-existing school variables and *RA* implementation variables
- For each possible predictor, compute *t*-test measuring impact on growth from 2016 to 2017 for each outcome variable

When Are RA Schools Successful?

Do pre-existing demographics predict success?

- Most school demographics had no significant effect on outcomes
- Lower total enrollment does correlate with growth ($r = -0.34$, $p < 0.05$)

Does *RA* usage predict success?

- No significant effect from whether a school uses *RA*
- Quantitative metrics from *RA* usage (# drafts, etc.) are not predictive.

When Are RA Schools Successful?

Does *RA* ***district-wide adoption*** predict success?

- 3.3% growth in districtwide *RA* passing rates vs. 0.7% state average
- Larger growth in three of four subscores (including +9.3% in Reading).
- No differences at all among standalone schools or partial adoptions.

When Are RA Districts Successful?

What characterizes *RA*'s districtwide implementations?

- **Larger** districts
  ~= capacity for administration support and coordination?
- Higher graduation rate than state average

- In at least two of these districts, RA use was explicitly **coordinated** between schools as part of a district-wide initiative. (as in MD)

Discussion

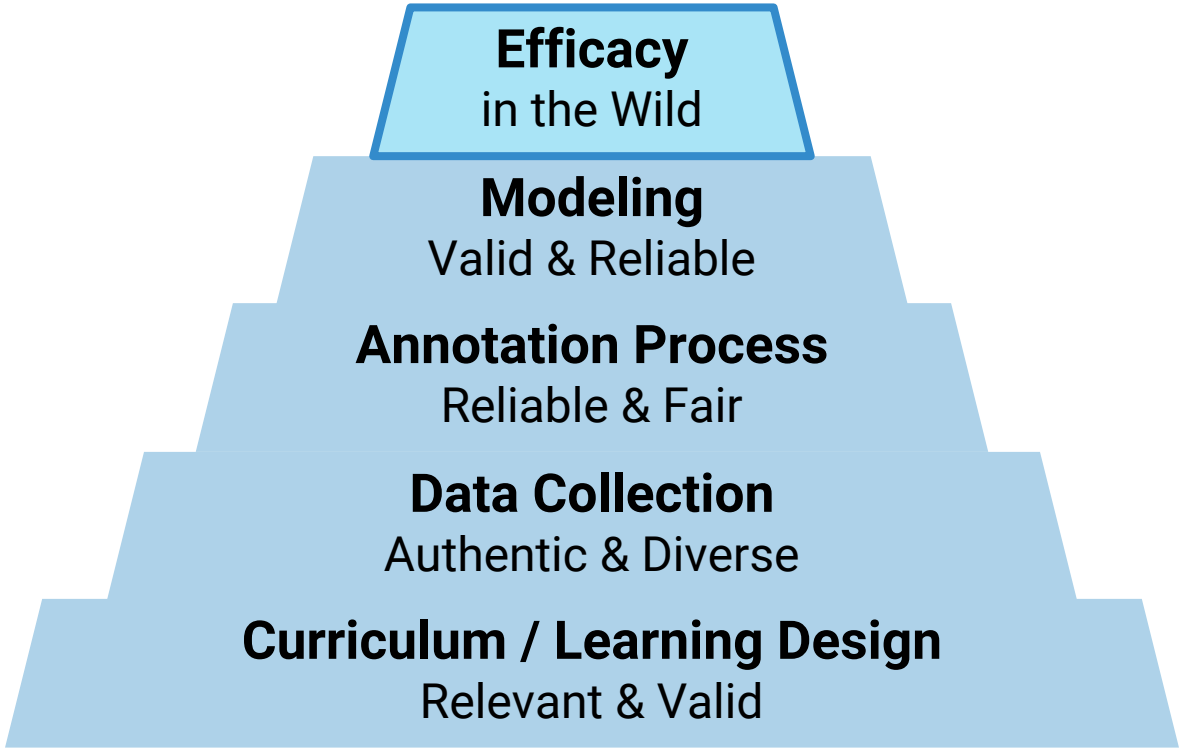
Primary Findings

- District-wide implementations of *RA* in California replicated the outsized growth seen in studies of districts in Maryland and Georgia.
- Schools using *RA* in California without district-wide participation saw no impact on test scores.

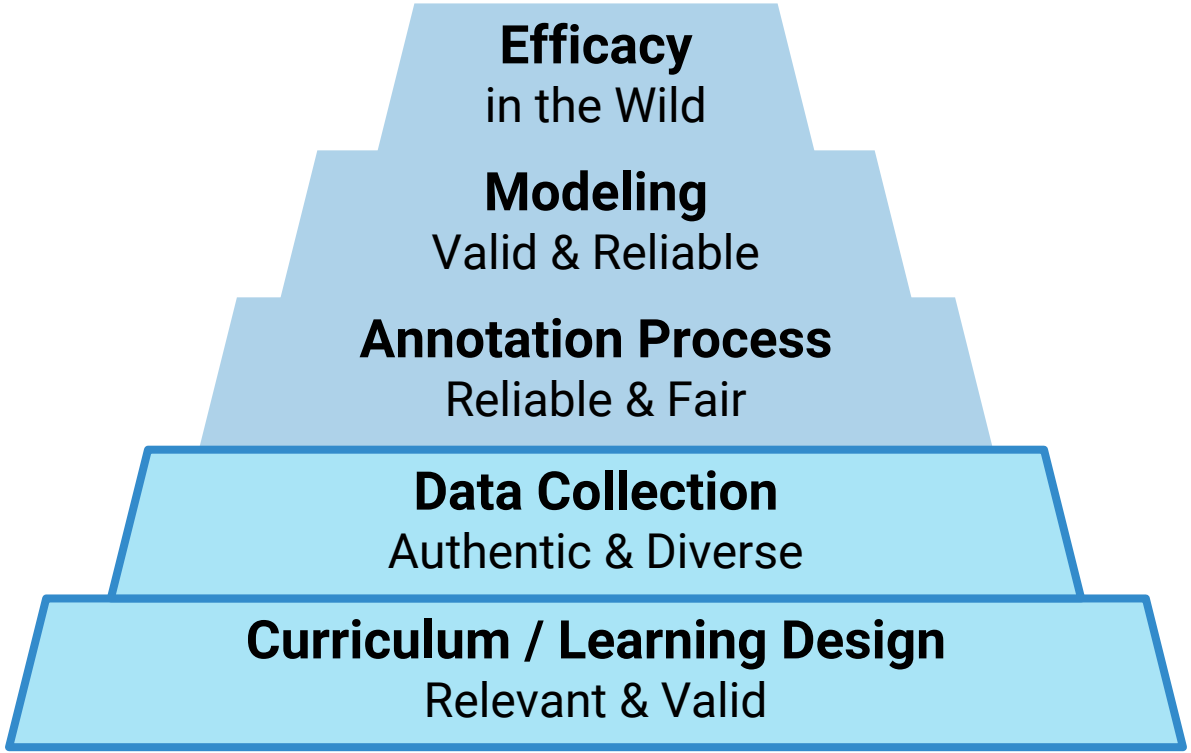
Limitations

- No way of identifying pre-existing factors that led to *RA* purchase (subgroups are quasi-experimental)
- No separating the *RA* technical intervention from alternate hypotheses:
  - Does a forward-thinking district office, which purchases new products, produce better results?
  - Does the purchase of *RA* produce better cross-district communication between buildings?
  - Does using *RA* improve conversations between teachers during and after training?

Trustworthy Machine Learning for Education



Trustworthy Machine Learning for Education



Curriculum & Data

AAAI 2018
and something new

Curriculum & Data

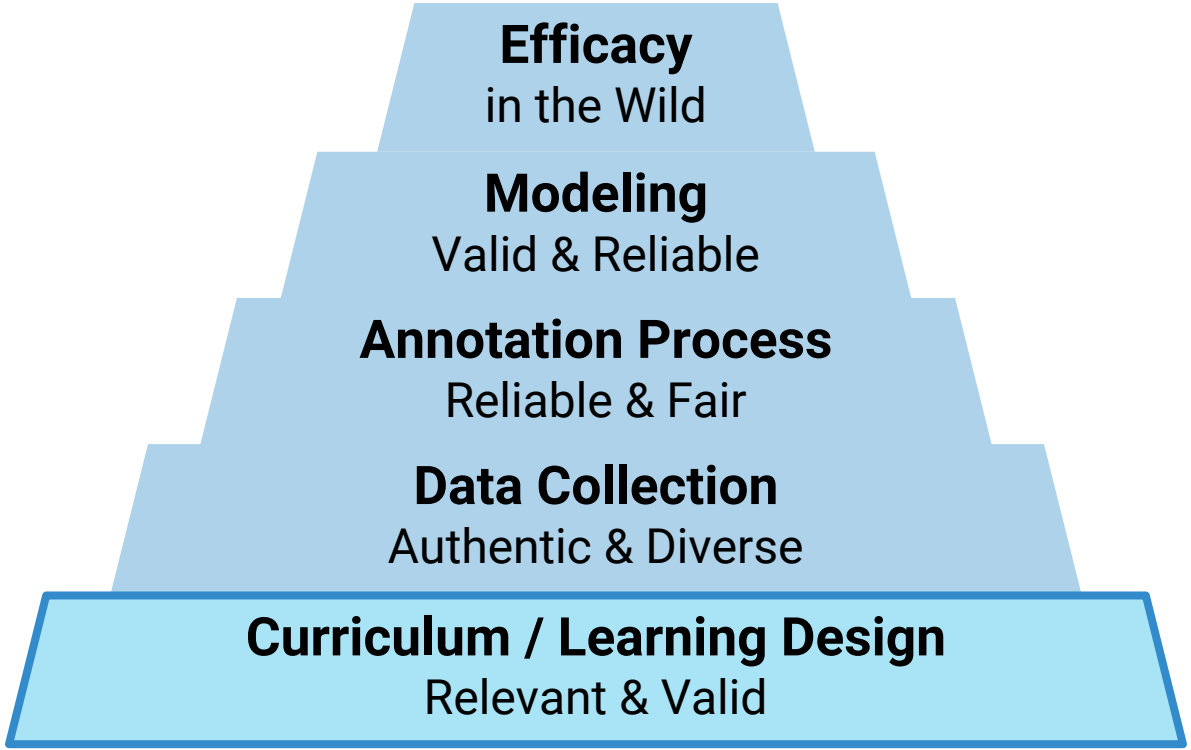
AAAI 2018
and something new

Curriculum Validity & Trustworthy Training Data

- Before you even consider modeling approaches, make sure you're aiming for a well-defined and curriculum-relevant task
- Source training data from a diverse set of students, in authentic settings, with trustworthy raters
- These matter more to educators than "construct validity" or rater agreement metrics

Let the Teachers Lead

Trustworthy Machine Learning for Education



Gender Bias in Secondary Curricula

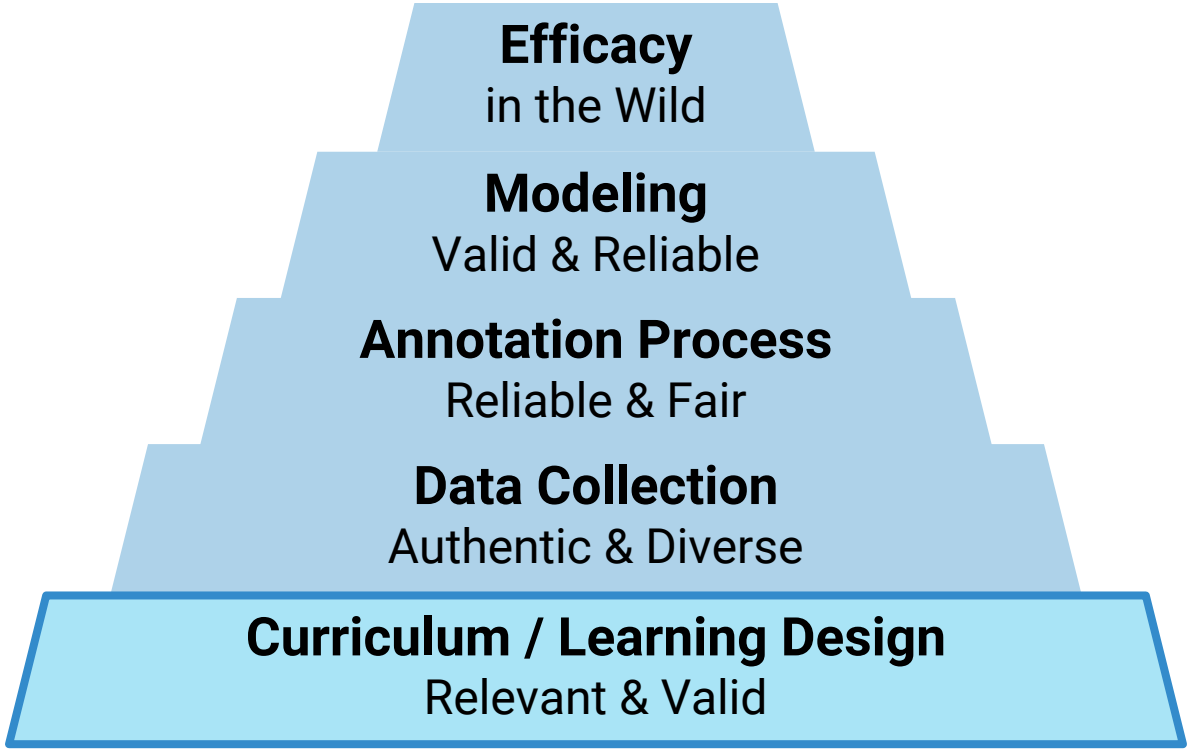
- In a sample of US History textbooks published in 2005, over 85% of named individuals are male – fewer than 15% are women (Chick, 2006)

- Only one of the ten most common book-length works taught in high school English Courses (To Kill A Mockingbird) is written by a woman (Applebee, 2009)

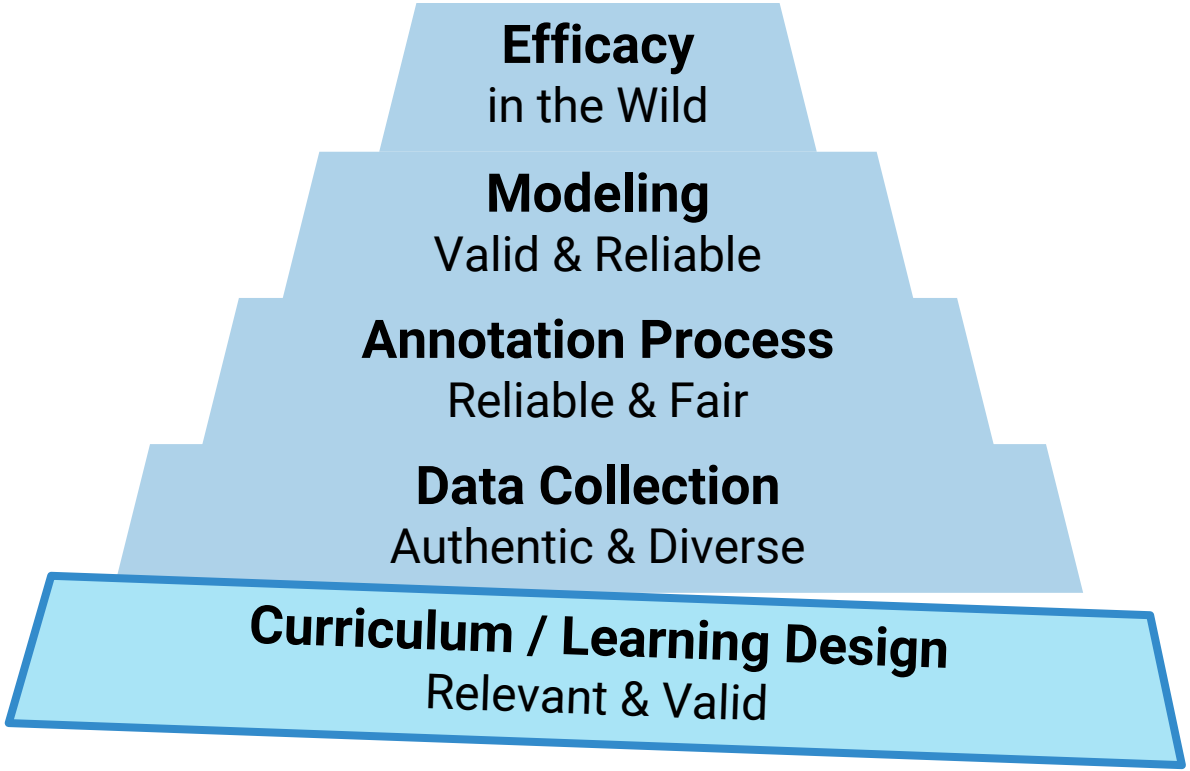
Gender Imbalance in Revision Assistant?

Prompt Genre	Prompts with sources	Prompts with any sources by women	% Prompts with any sources by women	% Total sources by women
Analysis	28	11	39%	31%
Argumentative	33	15	45%	18%
Historical Analysis	27	8	30%	8%
Informative	35	14	40%	22%
Narrative	11	5	45%	42%
All Genres	**134**	**53**	**40%**	**18%**

Trustworthy Machine Learning for Education



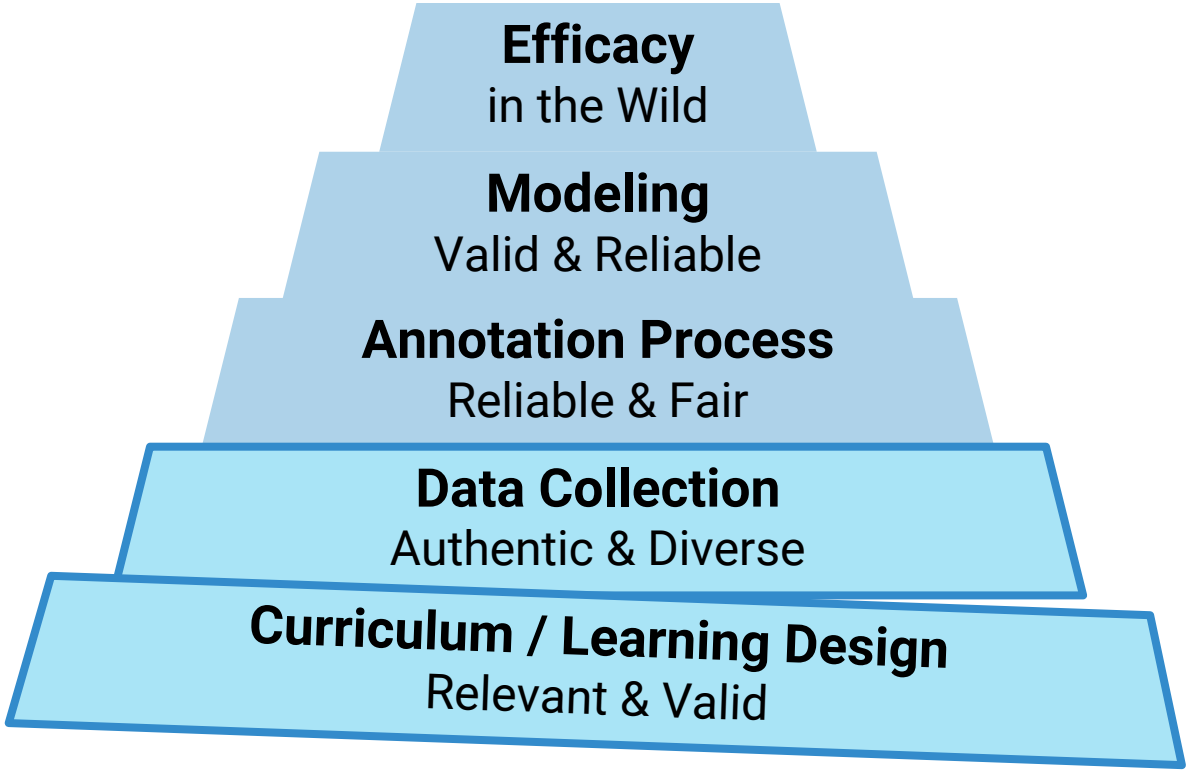
Trustworthy Machine Learning for Education



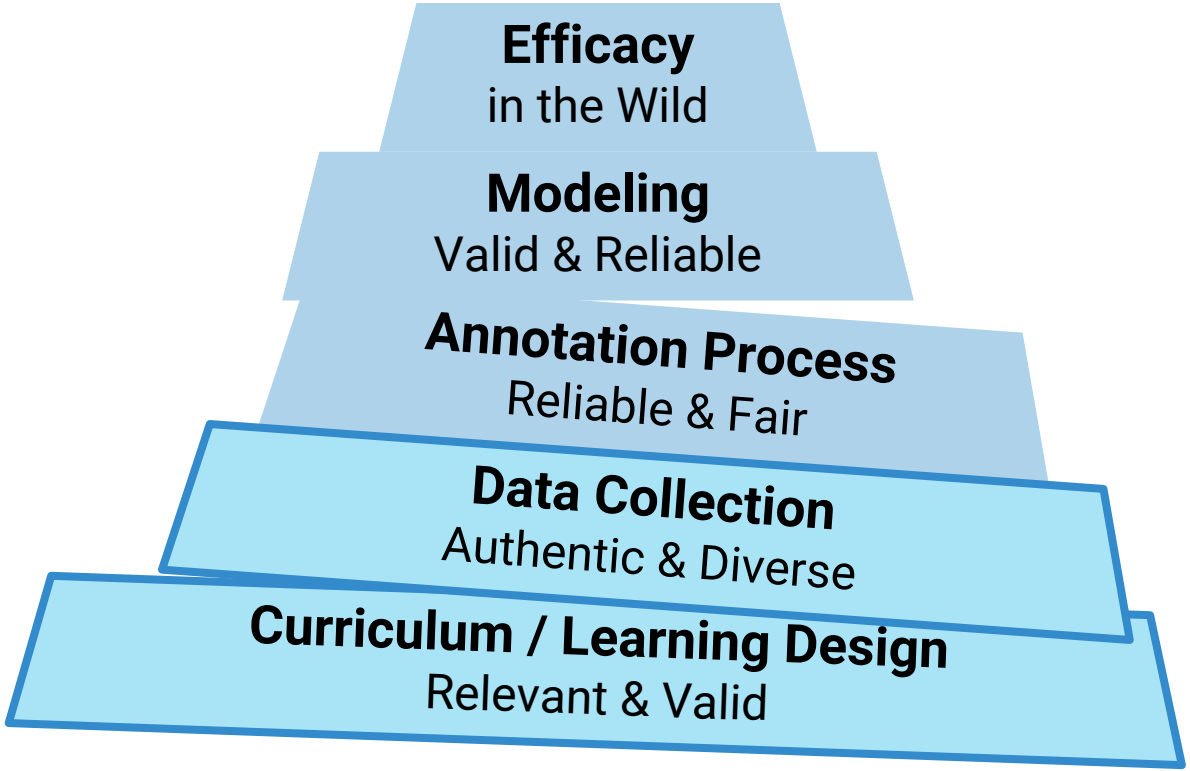
Rubric-Scoped Models & Training Data

- Collecting and scoring prompt-specific training data is a bottleneck!
- We can use our multi-prompt corpus to train models that are suitable for *any possible prompt* on a given evaluation rubric!
- Use leave-one-prompt-out cross validation to estimate performance on brand-new prompts – from here, everything looks fine.

Trustworthy Machine Learning for Education



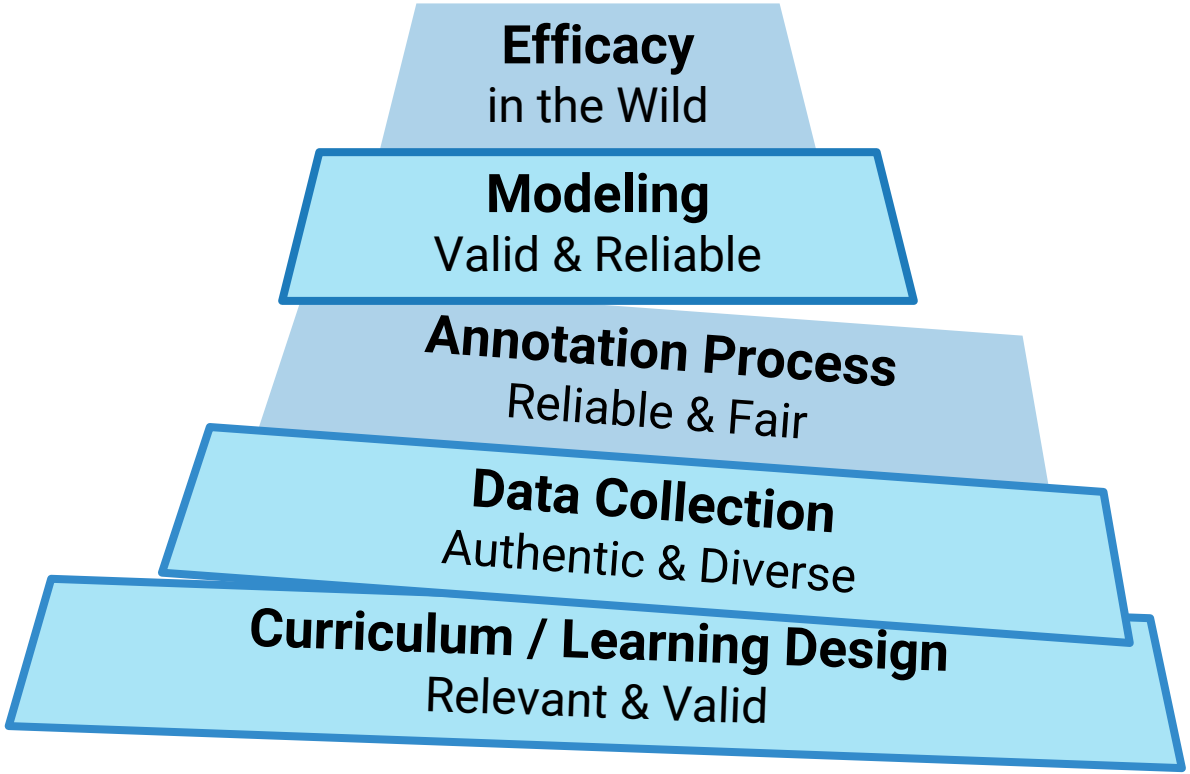
Trustworthy Machine Learning for Education



“Algorithms are still made by human beings, and those algorithms are still pegged to basic human assumptions.
They’re just automated assumptions.
And if you don’t fix the bias, then you are just automating the bias.”

-Representative Alexandria Ocasio-Cortez, 21 January 2019

Trustworthy Machine Learning for Education



Rubric-Scoped Models & Gender Bias

- **Features:**
  - N-grams which include female personal pronouns (she/her/hers) are roughly 30% as frequent as their male counterparts
  - 1.2% of model “feature importance” comes from n-grams with male pronouns, vs **0.0%** from n-grams with female pronouns.

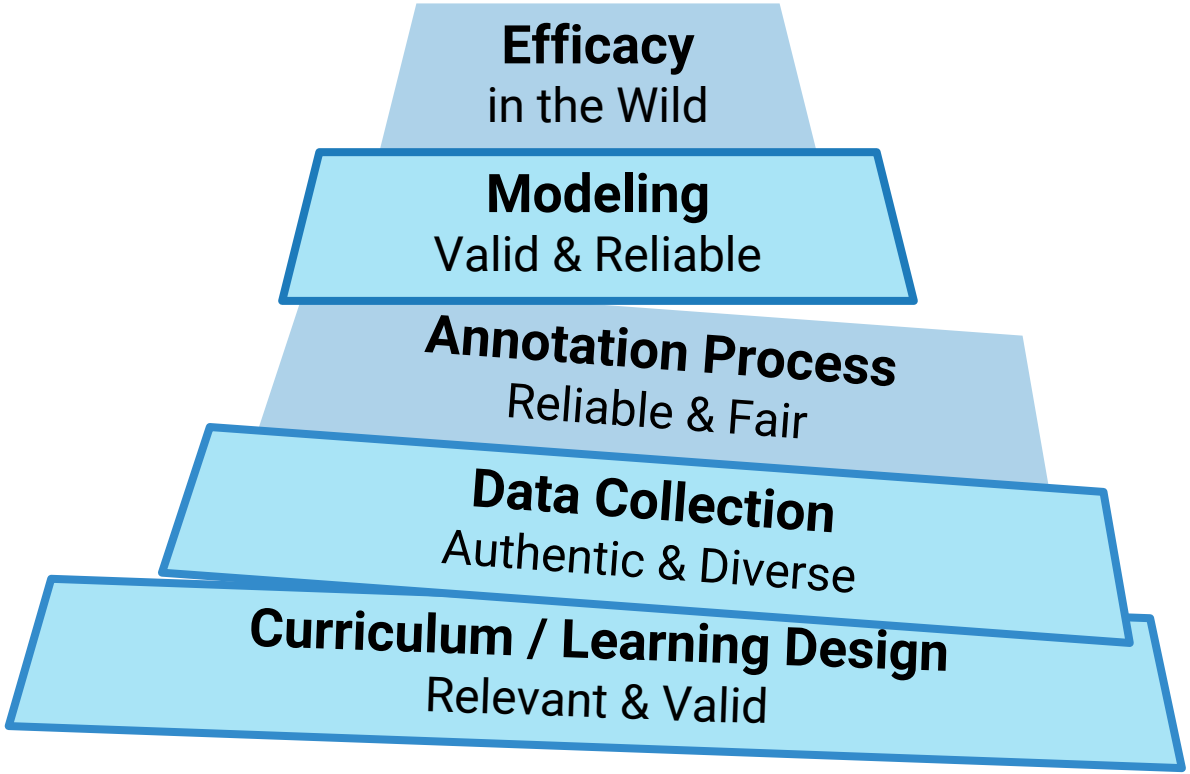
- **Model Stability:**
  Change the all the personal pronouns in an essay text to female, as if the essays were written about women source authors.
  Do the predicted essay scores change?

Rubric-Scoped Models & Gender Bias

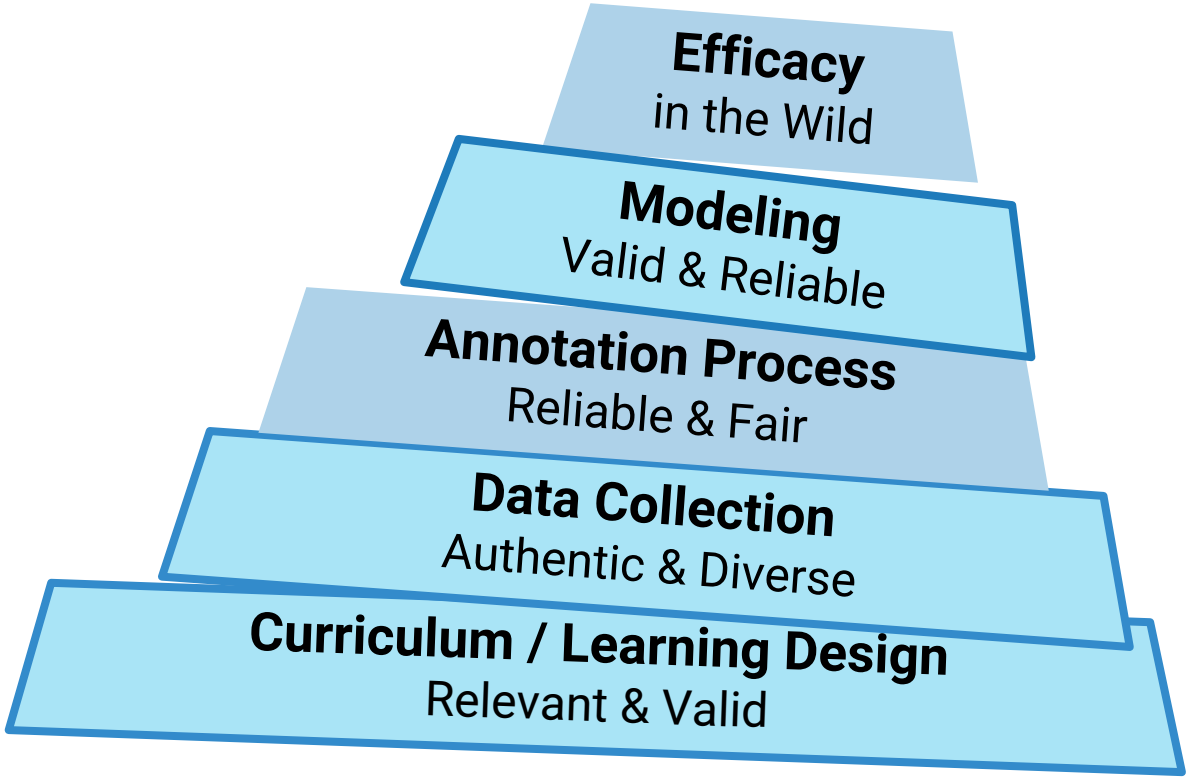
- **Features:**
  - N-grams which include female personal pronouns (she/her/hers) are roughly 30% as frequent as their male counterparts
  - 1.2% of model “feature importance” comes from n-grams with male pronouns, vs **0.0%** from n-grams with female pronouns.

- **Model Stability:**
  Change the all the personal pronouns in an essay text to female, as if the essays were written about women source authors.
  Do the predicted essay scores change? **16%** of essays change!

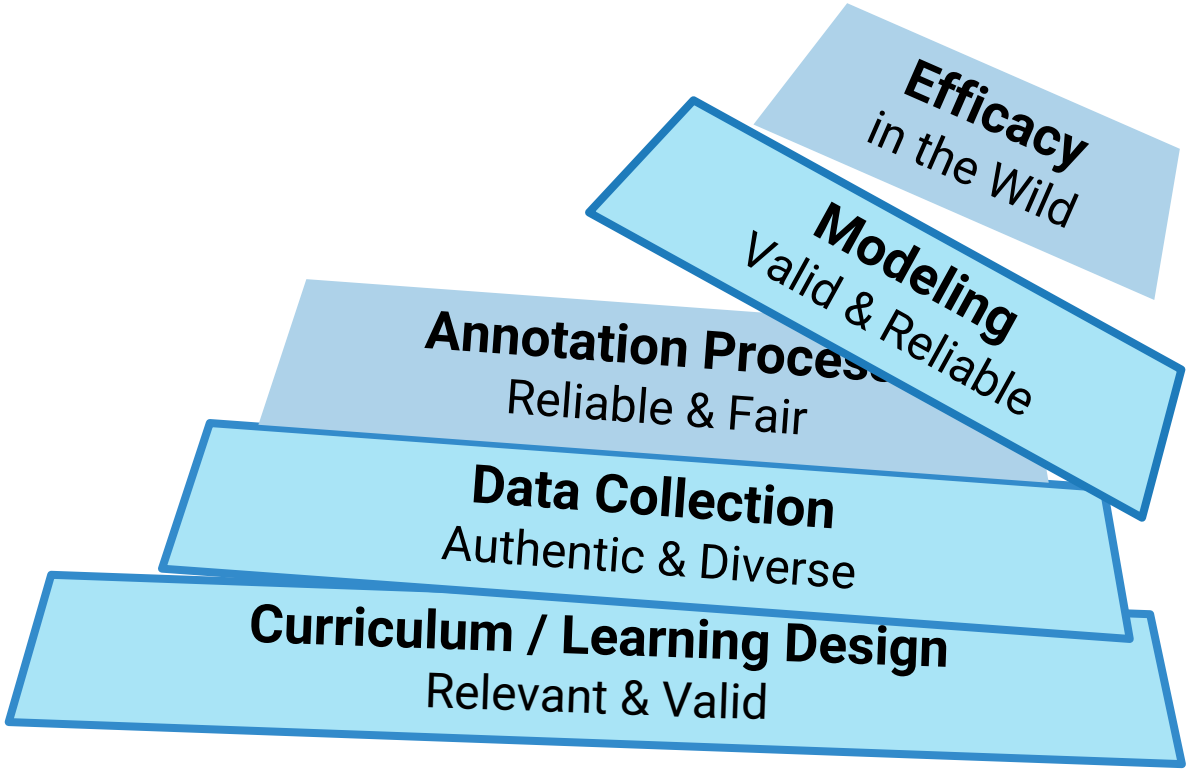
Trustworthy Machine Learning for Education



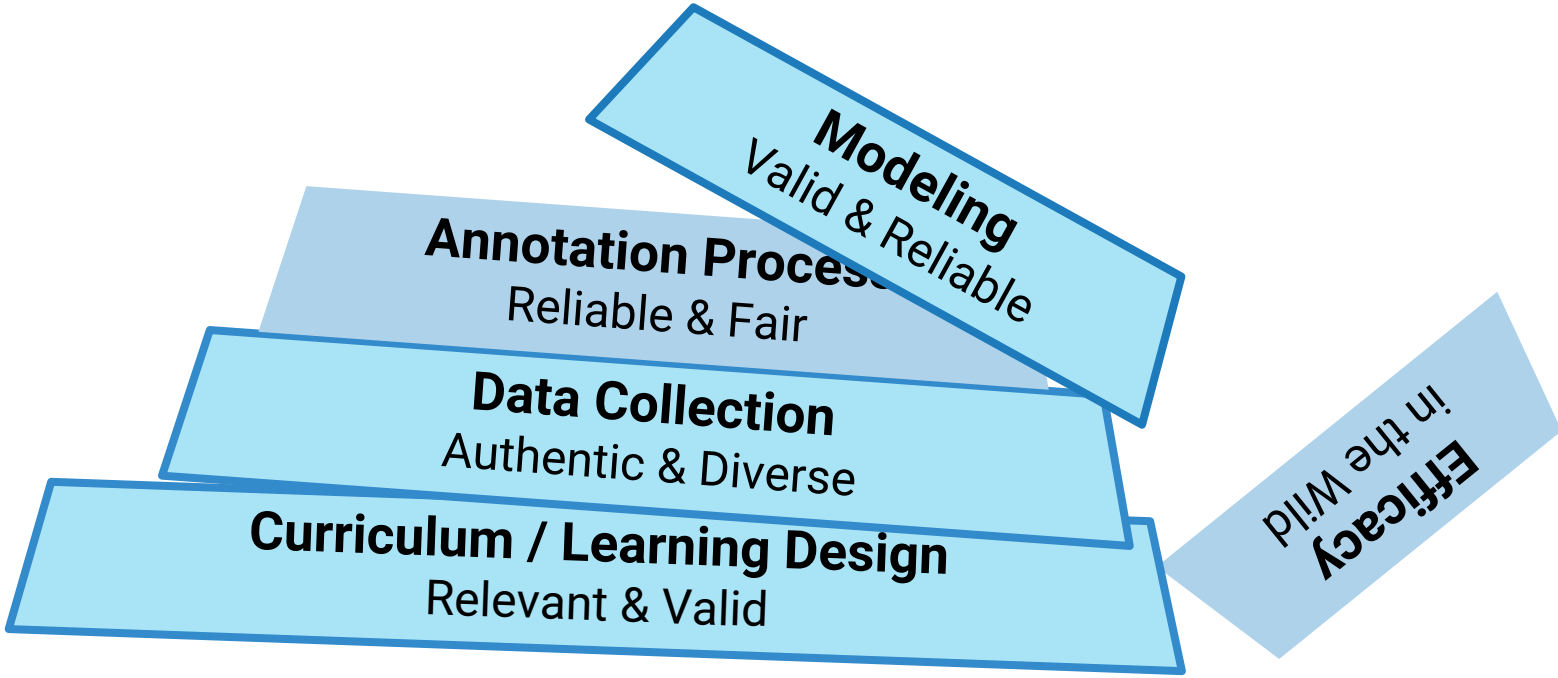
Trustworthy Machine Learning for Education



Trustworthy Machine Learning for Education



Trustworthy Machine Learning for Education



Fix The Data

- During preprocessing (before extracting lexical features),
  replace all gendered personal pronouns with placeholders:
  “In **her** ruling, **Ruth Bader Ginsburg** said that **he** had broken the law.”
  “In **[HER]** ruling, **[AUTHOR]** said that **[SHE]** had broken the law.”

- So every essay has been “blinded” to the gender of its subject…
  can we trust the generality of the rubric-scoped models now?

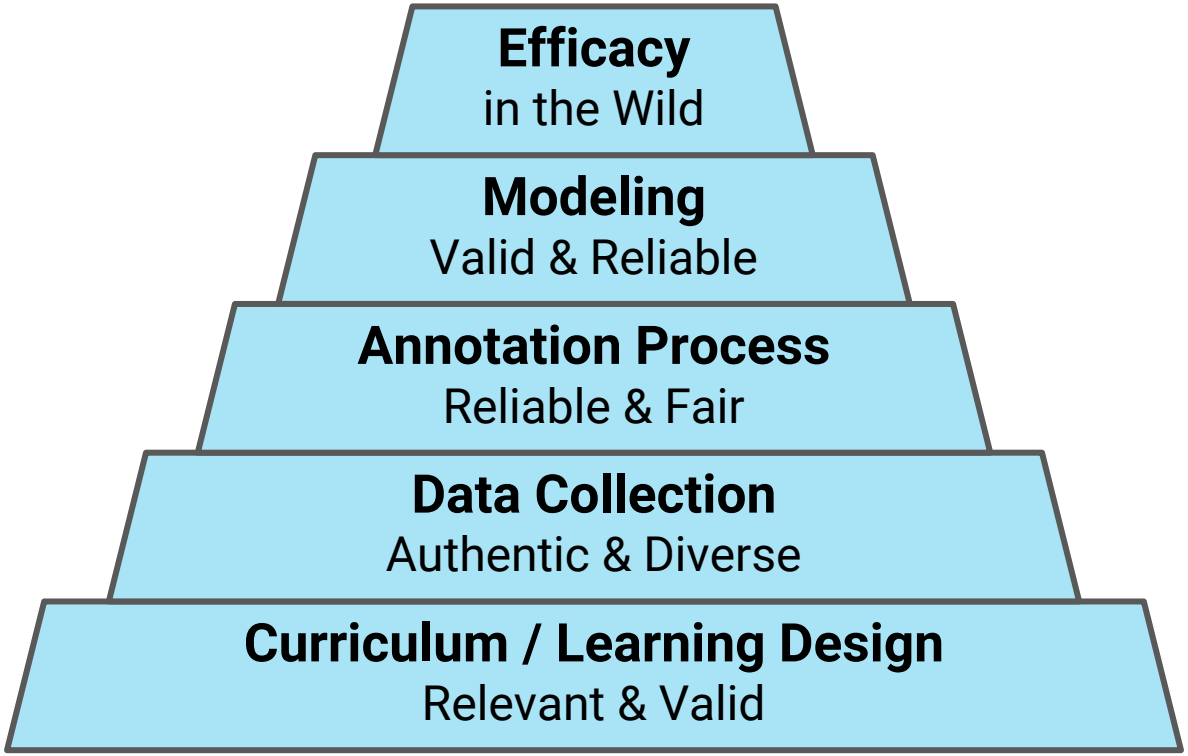
Fix The Data

- During preprocessing (before extracting lexical features), replace all gendered personal pronouns with placeholders:
  “In **her** ruling, **Ruth Bader Ginsburg** said that **he** had broken the law.”
  “In **[HER]** ruling, **[AUTHOR]** said that **[SHE]** had broken the law.”

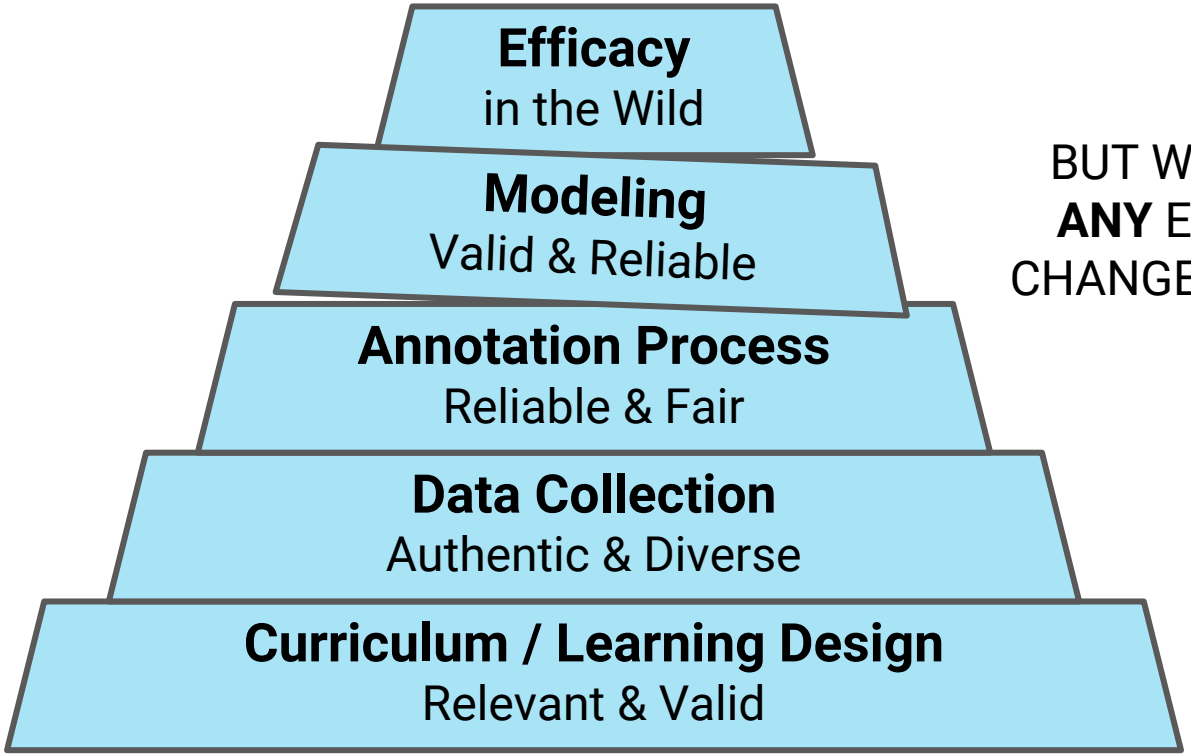
- So every essay has been “blinded” to the gender of its subject… can we trust the generality of the rubric-scoped models now?

- On the same sample of essays, ***almost* zero essays** change score when the gender of pronouns is changed.

Trustworthy Machine Learning for Education

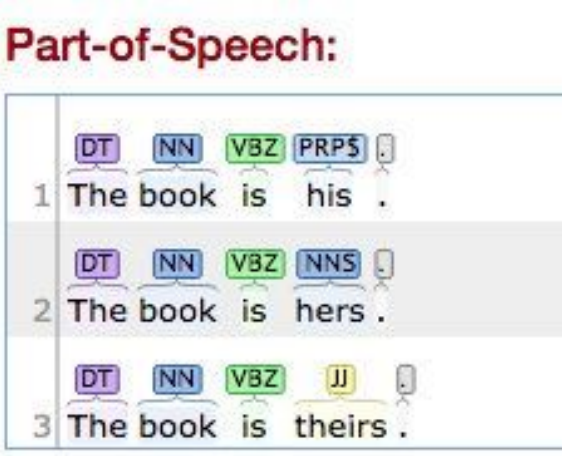


Trustworthy Machine Learning for Education



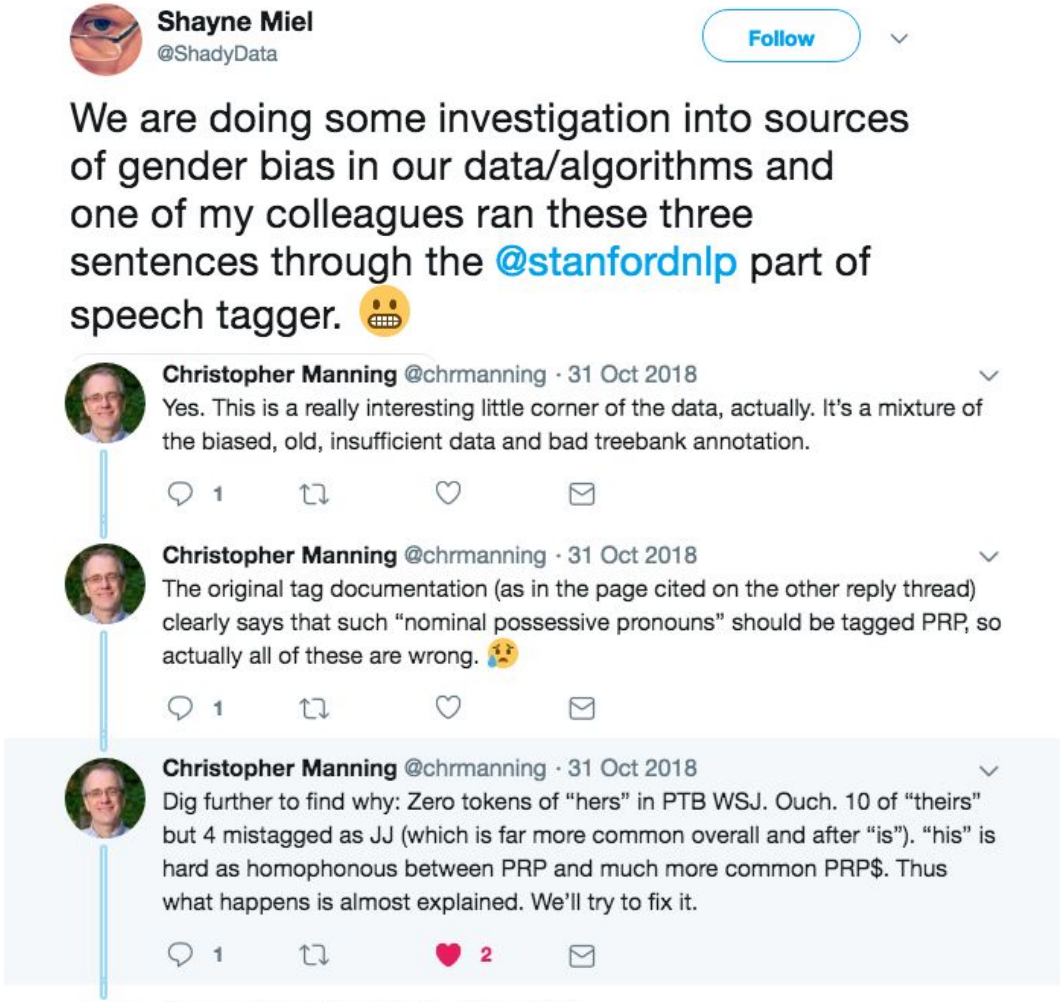
BUT WHY DID
ANY ESSAYS
CHANGE SCORE?

All The Way Down



Zero tokens of “hers” in Penn Treebank.

(there are 1.2 million tokens in the PTB corpus)



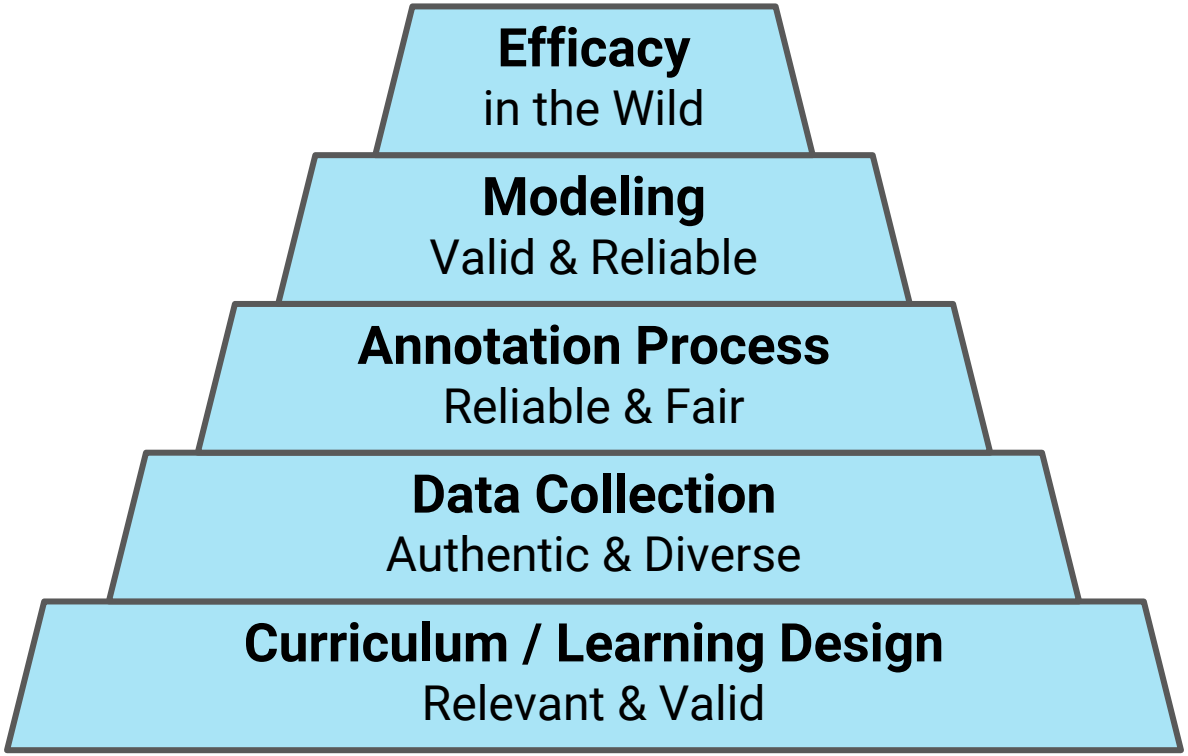
Discussion

- Curriculum decisions and learning design should influence the data available for modeling, and must inform dataset collection practices.

- Small changes to the feature extraction process can introduce (and also correct for) unexpected bias.

- Check for gender/racial/etc bias as part of your model evaluation practices!
  (and in the tools you rely on)

Trustworthy Machine Learning for Education



Punchlines

- Revision Assistant's predictions can forecast students' end-of-year writing outcomes.

- Districtwide implementations of Revision Assistant are correlated with school growth on end-of-year ELA pass rates.

- Check your data, your tools, and your models for bias!

Future Work

Bias Correction

- Uncover and address more subtle demographic issues

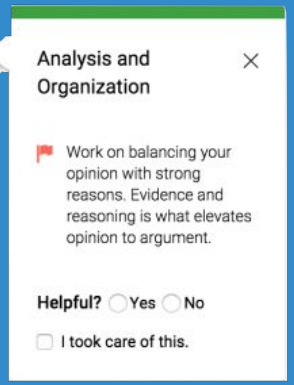
Causal Findings

- Randomized controlled trials with school cooperation and consent
- A/B testing of features at the individual user level, tracked to outcomes *(not done here due to privacy constraints on individual student data)*

Rigorous Correlative Findings

- Cohort studies of paired districts, rather than statewide comparisons
- Follow-on analysis of these districts after the 2017-2018 school year.

Questions + Feedback



David Adamson
dadamson@turnitin.com

Efficacy in the Wild

- Large-scale deployments of lab-tested, theory-backed technologies have presented mixed results. (Aleven & Koedinger 2002) vs (Cabalo et al., 2007)
- Non-academic factors can drive differences in usage (Warschauer et al,. 2004)
- Sometimes teachers and students use the tools in unexpected ways (Ogan et al., 2012)

Forecasting Outcomes

Texas' 9th grade English I "STAAR" assessment:

Raw ELA Score
Multiple-choice questions including reading comprehension, etc
(60 point scale)

Written Composition
Expository writing task, evaluated on a rubric that includes
Language, Organization, and Development of Ideas
(8 point scale)

Can Revision Assistant be used to help forecast exam outcomes?

Models (TX)

- **Benchmark Only**
  Just the school-administered fall benchmark score

- **Spot Check**
  Spot Check predictions from the Pretest only

- **Full Model**
  Linear model with *Benchmark* + *Pretest* + *Posttest* + *Invalid Draft Count*
  (other RA usage factors discarded by SLR)

Supporting District Initiatives

Use of Revision Assistant at the five participating middle schools.

Signal Check feedback is associated with increases in predicted essay scores.

School	Total Signal Checks	Signal Checks per Submission	Mean Increase in Summed Score
1	2011	14.1	5.5
2	3187	9.9	2.7
3	596	5.6	1.9
4	6155	11.3	3.0
5	4744	11.0	2.1
Treatment Schools	**16693**	**11.1**	**2.8**
RA 2016 to 2017*	937000	7.7	2.6

* Woods et al., 2017

Growth in Target District

2016-2017 change in 8th grade ELA pass rate

School	2016 pass rate	2017 pass rate	Change
1	35.8	49.1	+13.3%
2	41.1	49.0	+7.9%
3	39.0	45.8	+6.8%
4	30.7	36.2	+5.5%
5	23.7	22.2	-1.5%
Treatment Avg (n=5)	34.1	40.3	**+6.4%**
Non-Treatment Avg (n=3)	40.4	38.7	-1.9%
Maryland Avg (n=352)	31.8	32.0	+0.2%

Growth by School

2016-2017 change in 8th grade ELA pass rate

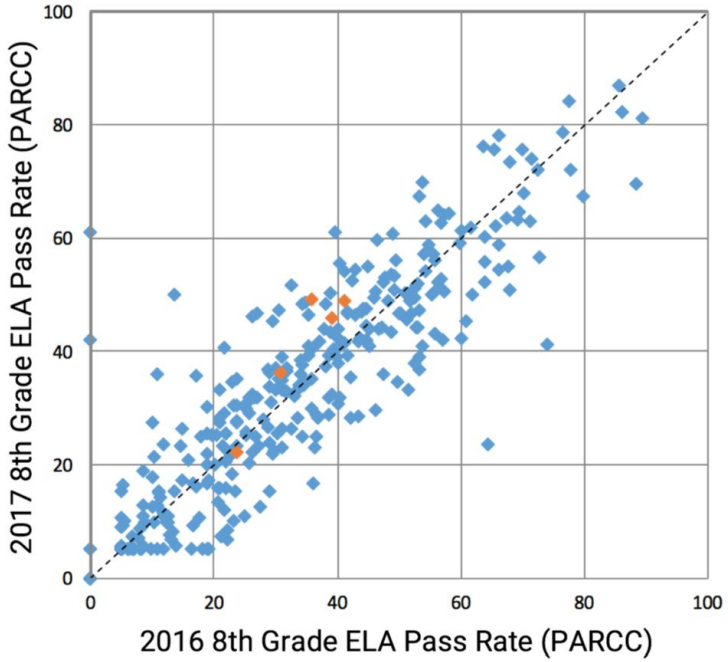


Figure 3: Performance growth of 5 treatment schools (orange) against all other MD schools (blue). The diagonal dashed line represents no year-over-year change.

Georgia Replication Study

School	2016	2017	Change
1	22	34	+12
2	29	36	+6
3	22	21	-1
4	43	56	+13
5	15	24	+9
6	10	26	+16
7	2	14	+12
8	7	19	+12
9	76	88	+13
10	93	96	+3
Treatment	**25**	**34**	**+9**
Georgia	70	74	+4

9th Grade

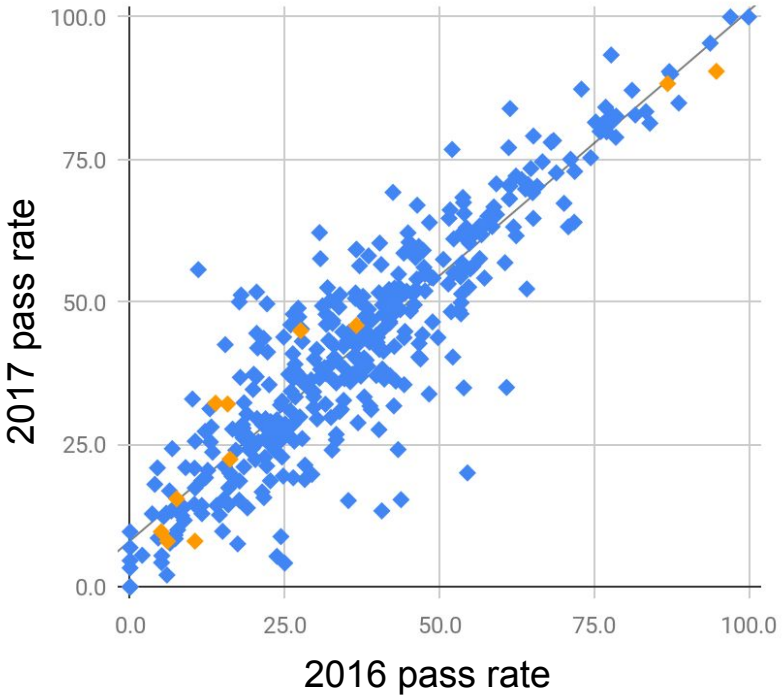
School	2016	2017	Change
1	14	32	+18
2	28	45	+17
3	16	32	+16
4	37	46	+9
5	8	16	+8
6	16	22	+6
7	5	10	+5
8	6	8	+2
9	87	88	+1
10	95	91	-4
Treatment	**22**	**31**	**+9**
Georgia	65	70	+5

11th Grade

Pass rates grew by an average of +9% in both 9th grade (left) and 11th grade (right)

Smallest effects seen in high-performing magnet schools (blue)

Georgia Replication Study



Pass rates grew by an average of +9% in 11th grade, compared to +5% statewide

About CASPP

- California statewide standardized test, consistent since 2015
- English Language Arts (ELA) is tested in each of grades 3-8, and 11
- Overall score and four subscores
  - Reading, Writing, Speaking/Listening, Research
  - All scores have a range from 1-4 (“passing” is 3+)

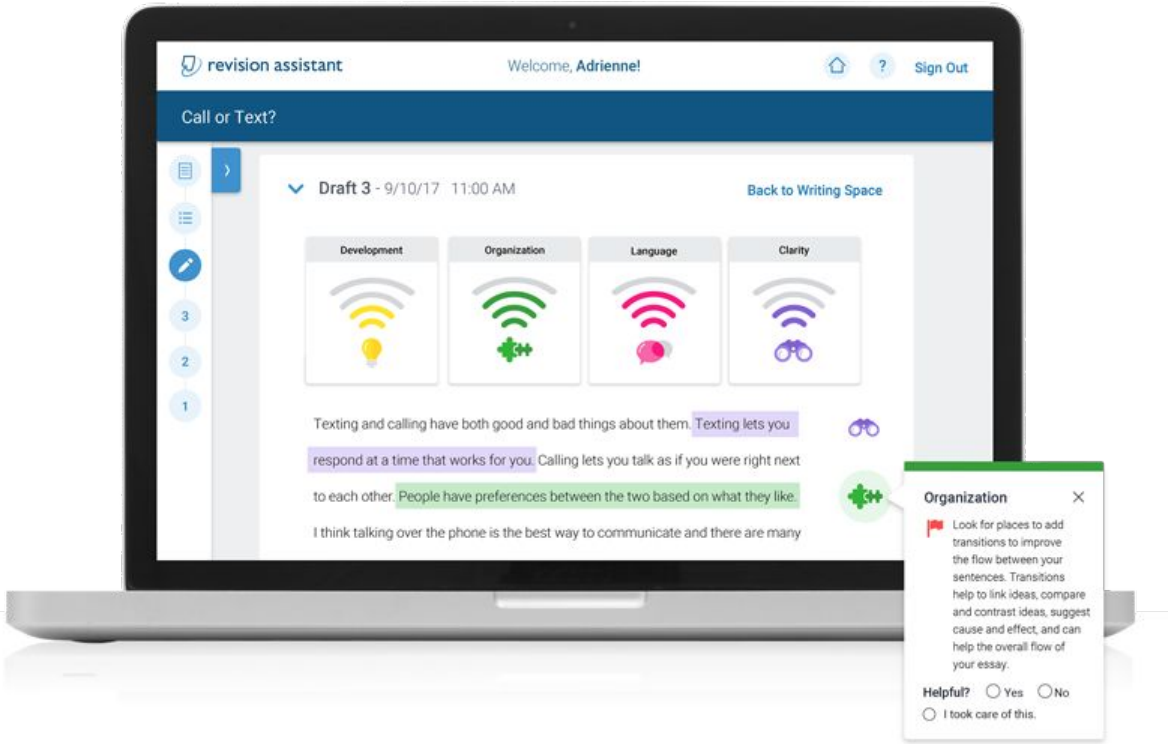
5/6 students have turned in work

View Class Reports | Download Assignment Report

Download Current Draft

	Last Name	First Name	Clarity	Development	Organization	Language	Turned In
☑	Alexander	Zachary	4	4	4	4	✓
☐	Brown	Ashley	2	1	2	2	✓
☐	Gardner	Sharon	1	1	2	1	✓
☐	Morris	Darlene	4	4	3	3	✓
☐	White	Destiny	3	3	2	3	✓

Revision Assistant



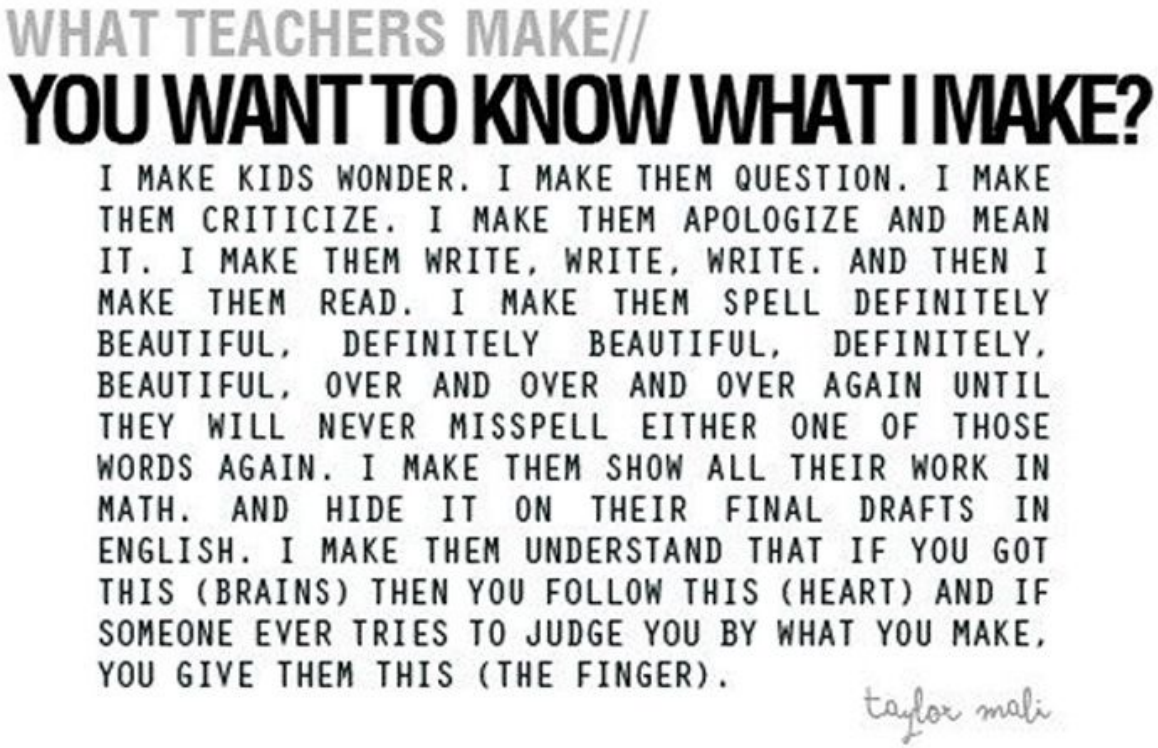
	Advanced	Proficient	Developing	Emerging
 **Clarity and Focus** Present a clear central idea early in the essay and focus on proving it.	The essay contains a clear, focused, and effective central idea that thoroughly addresses the demands of the prompt and fulfills the writing purpose.	The essay contains a mostly clear and focused central idea. The writing is effective in addressing the demands of the prompt and fulfilling the writing purpose.	The essay contains a central idea that may not be completely clear or focused. The writing does not completely address the demands of the prompt, nor fulfill the writing purpose.	The essay does not have a clear, focused, and effective central idea that addresses the demands of the prompt and fulfills the writing purpose.
 **Development** Use facts, definitions, and information from other sources to support and develop your central idea about the issue or topic.	The essay develops the central idea with well-chosen, relevant, and sufficient facts, extended definitions, concrete examples, quotations, etc. that address the audience's understanding of the topic.	The essay develops the central idea with well-chosen, relevant facts, definitions, concrete examples, quotations, etc. that purposefully address the audience's understanding of the topic.	The essay develops a central idea with some facts, definitions, examples, quotations, etc.; however, they may not be the most appropriate or effective supports and/or may be used inconsistently throughout the essay.	The essay does not develop a central idea with facts, definitions, examples, quotations, etc. Some details to develop the ideas may be present, but may not be used effectively to develop the central idea.
 **Organization** Include an engaging introduction and strong conclusion. Use transitions throughout the essay to make connections clear.	The essay uses an organizational structure with appropriate and varied transitions that show relationships between and among complex ideas. The structure creates a sense of cohesion throughout the essay, and includes both an introductory paragraph, as well as a concluding statement/paragraph, that clearly follows from and supports the ideas given.	The essay uses an organizational structure with appropriate transitions that show relationships between and among ideas throughout the essay. An introductory paragraph is present and the concluding statement/paragraph follows from the information given.	The organizational structure of the essay is inconsistent and/or ineffective. Some transitions may not completely or effectively show relationships between and among ideas throughout the essay. Either an introductory paragraph or a concluding statement/paragraph may be incomplete or unclear.	The lack of structure and effective transitions make the essay difficult to understand. The essay is missing entire structural elements, such as an introductory paragraph and/or concluding statement/paragraph.
 **Language and Style** Use specific, interesting language and clear sentence structure to communicate ideas.	The essay has an established, formal style and objective tone that is maintained throughout. The essay uses mostly correct, varied sentence structure and uses precise language and domain-specific vocabulary in a way that addresses the complexity of the topic. Few errors are present, and they do not interfere with meaning.	The essay has an established, formal style that is maintained throughout. The essay uses mostly correct, varied sentence structure and generally uses precise language and domain-specific vocabulary in way that generally addresses the complexity of the topic. The essay may have some errors, but they do not interfere with meaning.	The essay attempts to establish a formal style that may not be maintained throughout. The essay attempts to vary sentence structure and uses some precise language that may be domain-specific at times in a way that may address the complexity of the topic inconsistently. The essay contains some errors that may, at times, interfere with meaning.	The essay does not establish and/or maintain a formal style. The essay uses little variety in sentence structure, and the language is general and not domain-specific. The essay contains errors that interfere with meaning.

California Study

What characterizes *RA*'s districtwide implementations?

District	Enrollment	Graduation %	Free/Reduced Lunch %	2016 Pass %	2017 Pass %	2016 Exceed %	2017 Exceed %
1	20,000-25,000	93	45	63	64 (+1)	25	30 (+5)
2	10,000-15,000	93	46	66	70 (+4)	33	39 (+6)
3	10,000-15,000	96	69	60	63 (+3)	24	25 (+1)
4	30,000-35,000	93	17	78	81 (+3)	49	55 (+6)
California Statewide		83	59	59	60 (+1)	26	28 (+2)

- Large enrollment ~ district administration support and coordination
- Higher graduation rate than state average, but no difference in F/RL %
- Relatively greater impact on *Exceeds Expectations* scores (4/4)
- In at least two of these districts, RA use was explicitly **coordinated** between schools as part of a district-wide initiative. (as in MD)



Teachers make a difference.

taylormali.com/poems/what-teachers-make/