

Software-Supported Active Learning Cycles Allow Combined Learning and Assessment Activities

@Conference for Advancing Evidence-Based Learning (CAEBL) <u>https://learning.northeastern.edu/caebl2019/</u>

Wolfgang Gatterbauer (Khoury College of Computer Sciences, Northeastern) April 30, 2019

Imagine: You are the teacher in a non-STEM class

- You are teaching a business strategy class
 - 30-60 students in class
 - Evolving domain (e.g., Will Netflix survive against Walmart, Blockbuster?)
 - Many case studies (Amazon: books, groceries, health care, what next?)
- You face 2 challenges:
 - how to best help students learn?
 - how to best assess students?
- Thesis: appropriately designed computational systems can enable novel Combined Learning and Assessment Activities

Outline

1.What we do: The approach
 2.Why we do it: Scientific basis
 3.Proof that it works: Preliminary results
 4.Behind the curtain: the "secrete" sauce

Our two challenges lead to two modified challenges

- We face 2 problems:
 - how to best help students learn?
 - how to best assess students?

- We would like learning activities that support
 - Active learning (solving exercises)
 - Constructive learning (creating artifacts)
- We like many of them and scale grading

Basic Solution: Students Create & Answer MCQs



 (1) Students create challenging Multiple-Choice Questions (MCQs)

 (2) Other students answer these MCQs maintaining continuous interaction and receiving feedback

3rd Component: Students Improve & Finalize MCQs



• (3) Students improve MCQs and select among best alternatives

- And the system automatically grades the individual contributions
- resulting in a Virtuous Active
 Learning Cycle where each step
 helps students learn and enables
 the other 2 steps

Additional Benefits



- Constant Signal: System calculates dynamic ranking of students
 - giving instructors a continuous signal ("cockpit") of student participation, and giving students continuous feedback
 - Students getting feedback & constant test statistics; Leaderboard for motivation
- Semi-supervised by instructor: can use existing test banks to "seed" the process
- Scale: ML techniques surface most ambiguous MCQs to verify and focus limited instructor time for optimal use

Current Design of Inquizitiv.com

1. Create

2. Answer & Improve





3. Answer & Select



Students individually create MCQs

Students answer MCQs and make improvements

Students select best improvements

+ Instructor seed interaction



Instructor answers MCQs surfaced by system, chooses improvements, inserts known test questions, etc.

Anon. Leader board + statistics



System dynamically determines best contributors and participation

2 iterations per question



Tool automatically evaluates the quality of questions



The system uses the gathered responses to determine the quality of each MCQ based on Item Response Theory and algorithms from truth discovery Some Screenshots



→ C In (i) inquizitiv.com/courses/ gatterbauer Courses Logout gatterbauer Courses you attend: Course Name Course Number Term Year University Active	•
gatterbauer Courses Logout gatterbauer Courses you attend:	
Courses you attend: Course Name Course Number Term Year University Active	
Courses you attend: Course Name Course Number Term Year University Active	
Course Name Course Number Term Year University Active	
C5 inguizitiv demo 123 Summer 2018 Northeastern University True	
inguizitiv-test 181112 Fall 2018 Free True	
Database design (section 3) 3200 Spring 2019 Northeastern University True	

Add Course



	Password Reset		× Inquizitiv	× +					
_) → C'	۵	inquizitiv.c	om/assignments/44/crea	ate	… ⊠ ☆	<u>↓</u> III\	0	0 👳	-
	gatterbauer	Courses	Logout gatterbauer						
	Create a	new ques	stion						
	Stem text*								
	Type here	the question S	TEM.						
						11.			
	Reason text	*							
	Type here want to en	general informa nphasize.	ation about this question,	e.g., the source of inf	ormation, or the overall l	esson you			
						14.			
	Correct	answer							
	Answer text	t*							
	Type here	the CORRECT	ANSWER.						
						11.			
	Explanation	text*							
	Explain wh	ny this choice is	s correct.						

) → C' û	i inquizitiv.com/assignments/44/create ···· ☑ ☆	111\	0 [) 🔯	-0
	Explanation text*				
ĺ	Explain why this choice is incorrect.				
		11.			
	ncorrect answer				
,	Answer text				
	Type here a DISTRACTOR (incorrect answer).				
		11.			
E	Explanation text				
	Explain why this choice is incorrect.				
l		11.			
ļ	Create question				
	Description				
	Create a challenging MCQ on SQL				
	Create a chanenging MCQ on SQL				

1) Students create and answer questions

Question

Which of the following is not a benefit to Netflix in streaming content?

Textbook Chapter 4. The question seeks to emphasize the large amount of benefits, but also significant disadvantage with the streaming service.

Easier access to customers.

Customers can reach Netflix's content by simply logging on.

It is easy to imitate.

This is the biggest flaw with the streaming service. Competitors can easily replicate a streaming service if they can purchase the content.

High switching costs.

Netflix stores customers data and preferences, that recommends appropriate shows and movies, in a silo that other companies do not have access to.

Lower costs.

Eliminating distribution centers and shipping & handling reduced costs for Netflix.

Design decision: Answer explanations

3) Improvment of assessment items

Version 1	Improved Version 2
Question Stem	Question Stem
Answer text* What is the capital of Germany	Stem text* What is the capital of Germany
Explanation text* Reason	Reason text*
Correct answer	Correct answer
Answer text*	Answer text*
Munich	Berlin
Explanation text*	Explanation text*
sic	sic
Incorrect answer	Incorrect answer
Answer text*	Answer text*
Vienna	Vienna
Explanation text*	Explanation text*
explanation	explanation
Incorrect answer	Incorrect answer
Answer text*	Answer text*
London	London

The system shows MCQs to students and asks them to improve the MCQ. "Improvement" of questions implies making actual changes to the question parts.

4) Selecting best item parts

This is the original version of the question.

Finalize Question

Which of the following is a automated A/B testing?	reason v	vhy Big	Skinny	would	use
Correct: B					

to evaluate alternative sponsored-search campaigns by submitting a variety of ads to a search engine to see which performed best.

Correct.

in order to keep in touch with existing custome

B Incorrect. This is simply incorrect because it the reason Big Skinny used Email Marketing.

> to see whether or not people will stop clicking display ads because they are accustomed to irrelevant ads.

X C Simply incorrect because it is irrelevant/does no make sense. A/B testing is the marketing technique that shows different ads to different users.

> to assess which type of social media is more effectiv to a set of users.

Simply incorrect. It is a distractor.

None of the above.

X

Х

E Simply incorrect

Please choose from the available question parts those that you think improves this question the most. (Please reconstruct this question to create the best version possible, given the following question parts.)

kinny would use	Which of the following is a reason w Why did Big Skinny would use automated A/B testing?	Which of the following is a reason why Big Skinny would use automated A/B testing?	
	Correct: B	Correct B	
arch campaigns earch engine to	to evaluate alternative sponsored- search campaigns by submitting a variety of ads to a search engine to see which performed best compare performances.	to eEvaluate alternative sponsored-search campaigns by submitting a variety of ads to a search engine to see which performed best.	
	Correct.	Correct.	
) customers.	in order to keep in touch To maintain contact with existing customers.	in order to keep in touch with existing customers.	in order to keep in touch contact with existing customers.
ecause it is larketing.	Incorrect. This is simply incorrect because it is the reason Big Skinny used Email Marketing.	Incorrect. This is simply incorrect because it is the reason Big Skinny used Email Marketing.	Incorrect. This is simply incorrect because it is the reason Big Skinny used Email Marketing.
o clicking on comed to	to see whether or not if people will stop clicking on display ads because they are accustomed to irrelevant ads.	to see whether or not people will stop clicking on display ads because they are accustomed to constantly seeing irrelevant ads.	to see whether or not people will stop clicking on display ads because they are accustomed to irrelevant ads.
vant/does not keting to different	Simply incorrect because it is irrelevant/does not make sense. A/B testing is the marketing technique that shows different ads to different users.	Simply incorrect because it is irrelevant/ does not make sense. A/B testing is the marketing technique that shows different ads to different users.	Simply incorrect because it is irrelevant/does not make sense. A/B testing is the a marketing technique that to show a different ads to different users.
is more effective	to assess which type of social media is more effective to a set of users. Simply i Income. It is a distractor		

The system shows multiple versions of a MCQ to student and asks them to select one version for each of the question parts, i.e. for the stem, for the correct answer, and for each incorrect answer.

4) Selecting

Incorrect Answer Explanation 2 (Chosen by 3%)

it allows increased utilization of data centers.

> This is a true as virtualization changed utilization from ~15% to around 80%. This leads to cost savings

it allows increased utilization of data centers.

> Through the use of virtualization, utilization increased from about 15% to around 80%. This leads to large cost savings

It allows increased utilization of data centers.

While VMware is the leading company, it is being challenged by Microsoft, Dell, etc. As this market grows, competition increases. It allows for increased utilization of data centers.

This is true.

Diffs ON

Incorrect Answer Explanation 2 (Chosen by 3%)

it allows increased utilization of data centers.

This is a true as virtualization changed utilization from ~15% to around 80%. This leads to cost savings it allows increased utilization of data centers.

This is a true acrough the use of virtualization changed, utilization from increased from about 15% to around 80%. This leads to large cost savings t allows increased utilization of data centers.

TWhis is a true as virtualization VMware is the leading company, it is being challenged utilization from ~15% to around 80%. This leads to cost savingeby Microsoft, Dell, etc. As this market grows, competition increases, It allows for increased utilization of data centers.

This is a true as virtualization changed utilization from ~15% to around 80%. This leads to cost savings.

Faculty interface for calendar

TAs

Management Information Systems 70451 S 2015 CMU

Add New Assignment

÷				
Assignments				
Assignment	Create Deadline	Improve Deadline	Finalize Deadline	
Strategy & IT-2	Jan. 21, 2015, 9:30 a.m.	Jan. 24, 2015, 11:59 a.m.	Jan. 26, 2015, 9:30 a.m.	View Result
Does IT matter-3	Jan. 26, 2015, 9:30 a.m.	Jan. 28, 2015, 9:30 a.m.	Feb. 2, 2015, 9:30 a.m.	View Result
Developping an App for that (or Strategy and IT)-4	Feb. 2, 2015, 9:30 a.m.	Feb. 4, 2015, 9:30 a.m.	Feb. 9, 2015, 9:30 a.m.	View Result
Moore's law (or Strategy and IT)-5	Feb. 9, 2015, 9:30 a.m.	Feb. 11, 2015, 9:30 a.m.	Feb. 16, 2015, 9:30 a.m.	View Result
Network effects-6	Feb. 16, 2015, 9:30 a.m.	Feb. 18, 2015, 9:30 a.m.	Feb. 23, 2015, 9:30 a.m.	View Result
Disruptive Innovation-7	Feb. 23, 2015, 9:30 a.m.	Feb. 25, 2015, 9:30 a.m.	March 2, 2015, 9:30 a.m.	View Result
Anything goes (any topic that we have seen in class)-8	March 2, 2015, 9:30 a.m.	March 4, 2015, 9:30 a.m.	March 16, 2015, 9:30 a.m.	View Result
Databases and SQL-9	March 23, 2015, 9:30 a.m.	March 25, 2015, 9:30 a.m.	March 30, 2015, 9:30 a.m.	View Result
Databases and SQL-10	March 30, 2015, 9:30 a.m.	April 1, 2015, 9:30 a.m.	April 6, 2015, 9:30 a.m.	View Result
Business Analytics and Cloud Computing-11	April 6, 2015, 9:30 a.m.	April 8, 2015, 9:30 a.m.	April 13, 2015, 9:30 a.m.	View Result
Internet, Google, Information Security (Chapters 14, 15, 16)-12	April 13, 2015, 9:30 a.m.	April 15, 2015, 9:30 a.m.	April 20, 2015, 9:30 a.m.	View Result
Project presentations-13	April 20, 2015, 9:30 a.m.	April 22, 2015, 9:30 a.m.	April 27, 2015, 9:30 a.m.	View Result
Homework for Majd-14	July 21, 2015, 9:30 a.m.	July 22, 2015, 9:30 a.m.	July 23, 2015, 9:30 a.m.	View Result

Assignments

Students

Why it should work

scientific justifications

Scientific Justifications for Learning Activities



Scientific Justifications

- (1) Answer:
 "Active retrieval practice"
- (2) Create:
 "Learning by Explaining"
- (3) Improve/Select:
 Research on Peer Assessment

Studying new material: "Under which study condition do you think you learn better?" → Active retrieval



(1) Active retrieval practice

- "Active retrieval": answering a question = retrieving information
- Active retrieval practice leads to greater learning gains than the same amount of time spent studying
 - Carrier, Pashler. The influence of retrieval on retention. Memory and Cognition, 20:632–642, 1992
 - J. D. Karpicke and J. R. Blunt. Retrieval practice produces more learning than elaborative studying with concept mapping. Science, 331(6018):772–5, Feb 2011.
 - N. Kornell, M. J. Hays, and R. A. Bjork. Unsuccessful retrieval attempts enhance subsequent learning. Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP:LMC), 35:989–998, 2009.
 - M. A. McDaniel and R. P. Fisher. Tests and test feedback as learning sources. Contemporary Educational Psychology, 16:192–201, 1991.
 - M. A. Pyc and K. A. Rawson. Testing the retrieval effort hypothesis. Journal of Memory and Language, 60:437–447, 2009.
 - H. L. Roediger and J. D. Karpicke. Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17:249–255, 2006.

"Use quizzes to re-expose students to information" = strong recommendation in Institute for Education Science Practice Guide

 IES. Organizing instruction and study to improve student learning: a practice guide, 2004. http://educationnorthwest.org/resource/1820.

(1) Active retrieval practice

- While earlier research found short-answer questions produced greater learning than MCQs, more recent, classroom-based work has shown that both question formats enhance test performance
 - K. B. McDermott, P. K. Agarwal, L. D'Antonio, H. L. Roediger, and M. A. McDaniel. Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. Journal of Experimental Psychology: Applied, 20:3–12, 2014.
- Test questions have a greater impact on learning when they are accompanied by feedback
 - W. L. Cull. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. Applied Cognitive Psychology, 14:215–235, 2000.
 - H. Pashler, N. J. Cepeda, J. T. Wixted, and D. Rohrer. When does feedback facilitate learning of words? Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP:LMC), 31:3–8, 2005.
- Well-constructed MCQs have been shown to have the strongest predictors of overall student performance compared to other evaluations (in-class participation, case exams, written assignments)
 - N. Bontis, T. Hardie, and A. Serenko. Techniques for assessing skills and knowledge in a business strategy classroom. Int. J. Teaching and Case Studies, 2(2):162–180, 2009.

(2) Learning by Explaining

- Making sense of new information by explaining it to oneself promotes better learning and performance.
 - M. Chi, M. W. Lewis, P. Reimann, and R. Glaser. Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13:145–182, 1989.
 - M. T. H. Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science, 1:73–105, 2009.

Prompting students to self-explain or providing self-explanation training similarly leads to enhanced learning.

- K. Bielaczyc, P. Pirolli, and A. L. Brown. Training in self-explanation and self-regulation. Cognitive Science, 18:439–477, 1994.
- M. T. H. Chi, N. de Leeuw, M. Chiu, and C. LaVancher. Eliciting self-explanations improves understanding. Cognitive Science, 18:439– 477, 1994.
- T. D. Griffin, J. Wiley, and K. W. Thiede. Individual differences, rereading, and self-explanation. Memory and Cognition, 36:93–103, 2008.
- A. Renkl, R. Stark, H. Gruber, and H. Mandl. Learning from worked-out examples: The effects of example variability and elicited selfexplanations. Contemporary Educational Psychology, 23:90–108, 1998.

"Help students build explanations by asking and answering questions" = strong recommendation in IES Practice Guide

 IES. Organizing instruction and study to improve student learning: a practice guide, 2004. http://educationnorthwest.org/resource/1820.

(2) Learning by Explaining

- Empirical evidence for improvements in comprehension, learning, and memory by "training students to ask good questions"
 - A. C. Graesser and N. K. Person. Question asking during tutoring. American Educational Research Journal, 31(1):pp. 104–137, 1994.
 - B. Davey and S. McBride. Effects of question-generation training on reading comprehension. Journal of Educational Psychology, 78:256–262, 1986.
 - J. R. Gavelek and T. E. Raphael. Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G.
 MacKinnin, and T. G. Waller, editors, Metacognition, cognition, and human performance, vol. 2, pp. 103–136. Academic Press, 1985.
 - A. King. Effects of self-questioning training on college students' comprehension of lectures. Contemporary Educational Psychology, 14:366–381, 1989.
 - A. King. Enhancing peer interaction and learning in the classroom through reciprocal questioning. American Educational Research Journal, 27:664–687, 1990.
 - A. S. Palinscar and A. L. Brown. Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. Cognition and Instruction, 1:117–175, 1984.
 - M. Singer and D. Donlan. Active comprehension: Problem solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly, 17:166–186, 1982.

• Teaching students to generate questions on the text they have read resulted in gains in comprehension.

 B. Rosenshine, C. Meister, and S. Chapman. Teaching students to generate questions: A review of the intervention studies. In Review of Educational Research Summer, vol. 66, pp. 181–221, 1996.

(3) Peer Assessment & Scaffolded Learning Experiences

- Peer assessment = learners provide feedback (and assessment) to each other
- Providing peer assessments have been shown to improve students' writing. Students who provided elaborate forms of feedback (incl. free-form comments) performed significantly better on their own writing than students who provided numerical ratings only
 - R. S. Wooley, C. A. Was, C. D. Schunn, and D. W. Dalton. The effects of feedback elaboration on the giver of feedback. In Proceedings
 of the Annual Meeting of the Cognitive Science Society, pp. 2375-2380, 2008.
- With technical tools that provide scaffolded peer review students' peer assessments become more valid and reliable
 - K. Cho, T. R. Chung, W. R. King, and C. D. Schunn. Peer-based computer-supported knowledge refinement: an empirical investigation. Commun. ACM, 51(3):83–88, 2008.
- One issue: how to incentivize students to grade "correctly"

Preliminary results

Natural questions to ask and preliminary results

- 1. Assessment: How well can it predict ranking in a class?
 - Rank in class vs. rank by tool
 - Details on algorithms later (keyword "truth discovery")
- 2. How much do students actually learn?
 - Hard to verify: A/B test, dividing students into two groups, what is the baseline process; making sure both groups spend same time
- 3. Do students improve their ability to ask good questions
 - Do question scores improve over time?
- 4. Are the questions contributed by students "interesting"?
 - Anecdotal evidence: let's look at some of the questions produced

1. Assessment: can we predict class rankings

- Pearson's r (Correlation Coefficient PCC, product-moment CC):
 - PCC = 1: perfect correlation, PCC = 0: no correlation
- Class at CMU in 2015, n=32 students, 223 questions: PCC between final rankings in class vs. rankings according to our tool*
 - Inquizitiv / Final scores: PCC = 0.535
- Comparison: PCC on midterm exam between parts of the exam
 - BCQ / MCQ: PCC = 0.49
 - BCQ / Essay: PCC = 0.44
 - MCQ / Essay: PCC = 0.57
- *Notice: some methodological detail: Final score included 5% class participation, part of which came from "overall engagement" with the tool

- 3. Quality of students questions: start vs. end of semester
- Tool assigns question scores in [-1,1] for each homework.
 - Question: do questions become better over time?



Average Score = 0.12 n = 34 students 38% (13/34) < 0 AVG score = 0.26 n =32 students 16% (5/32) < 0

3. Anecdotal evidence: Students improving stems, ...

What are all reasons that newcomers to an industry, where there is a firm with dominant network effects, find entering the market hard?

Key reasons as to how newcomers face difficulties in competing against the firms with dominant network effects. What are all reasons that newcomers to an industry, where there is a firmchallenges do newcomers face when entering an industry with a dominant network effects, find entering the market hard?firm.

Key reasons as to how newcomers face difficulties in competing against the firms with dominant network effects. What are all reason<u>of the main</u> challenges that newcomers to an industry, face when entering a market where there is a firm with dominant network effects, find entering the market hard?

Key reasons as to how newcomers face difficulties in competing against the firms with dominant network effects.

... improving explanations, ...

Correct Answer Explanation 1 (Chosen by 57%)

Lower switching costs

Wrong. Not mentioned in the book.

Lower switching costs

Wrong. Not mentioned in the book: Lowering switching costs would allow customers to move freely to other competitors. Lower switching costs for your product.

Wrong. Not mentioned in the booA firm would want its products to have a high switching cost to build a network.

... bringing entirely new perspectives, ...

Incorrect Answer Explanation 3 (Chosen by 15%)

Touch screen on the Microsoft's Office was a real problem for the death of e-readers.

This is true because the official said that "the real problem for his colleagues was the touch screen".

Touch screen on the Microsoft's Office was a real problem for the death of e-reader<u>A predisposition</u> against touch screen interfaces that differed from the desktop interface for Windows.

This is true because the official said that "the real problem for his colleagues was the touch screen"Executives, including Bill Gates, wanted to maintain the look and feel of Windows. They did not see a reason to change their interface when their software was being used on new devices. Touch screen on the Microsoft's Office was a real problem for the death of e-readers.

This is true-because the official said that "the real problem for his colleagues was the touch screen"., as a former senior marketing manager Kurt Massey said that ""In the 40s, 50s, and 60s, Sears had it nailed. It was top-notch."

..., and even coding assignments

Incorrect Answer Explanation 2 (Chosen by 25%)

SELECT name, population FROM states WHERE name = "A%"

Simply incorrect

SELECT name, population FROM states WHERE name = "A%"

Simply incorrectThe '%' wildcard is not defined for

equality '=' operations.

Behind the curtain

Computational "truth discovery" in social data



Students

Questions

Given: • Various interactions of students with questions (create, answer, improve, ...)

- Goal: rank students by mastery of subject
- Existing work in Item Response Theory is not transparent and does not scale ☺
- We: Axiomatic approach combined with iterative updates. Discrete solution with new hardness results. Continuous solution with linear scalability

JRG [WALCOM'17] (invited to Theory of Computer Science'18)

Interesting Challenges from CS point-of-view

• "Truth discovery"

- Developing fast reliable algorithms for ranking
- Understanding some iterative algorithm (understanding convergence, solid comparison to max likelihood solution, place into related work)
- Data set versioning
 - Building a scalable "Question Management System" (QMS)
 - How to store, index, search, compare 1) sets (different units) of 2) versions (evolving) of 3) MCQs (itself with some structure)
- Others:
 - ML-type "active learning": make best use of instructor time
 - Adaptive assignment of exercises

Take-away points &

Take-away questions

Food for thought: virtuous learning cycles

- Example of computational approach that:
- sequences learning activities s.t.
 - (i) each promotes learning, and
 - (ii) they mutually support each other
- aligns the incentives of students to perform well on *each* task with goal for assessment
- creates an "instructor cockpit" to observe students engage in a structured discussion on the class material and the high-level misunderstandings "flagged" for the instructor

Students (1)(2) Create Answer (3) Improve/ Select

Question:

• How else can we design such tools? Thank you 😳