

# Revising Learner Misconceptions Without Feedback: Prompting for Reflection on Anomalous Facts

Leave Authors Anonymous  
for Submission  
e-mail address

Leave Authors Anonymous  
for Submission  
e-mail address

Leave Authors Anonymous  
for Submission  
e-mail address

## ABSTRACT

A challenge for Massive Open Online Courses (MOOCs) is to promote critical thinking, in thousands of learners when there is limited opportunity for real-time feedback. One aspect of critical thinking is the revision of incorrect beliefs. Could online lessons be designed to encourage an individual to generate their own internal feedback about their misconceptions? This paper investigates the pedagogical value of problems designed to prompt learners to reflect on facts that are *anomalous* or inconsistent with their existing incorrect beliefs. Recent psychological theories offer insight into how the structure of anomalous problems and the wordings of question prompts can influence the degree to which learners revise their beliefs. These are tested in two randomized experiments with online crowd workers. We find that reflective prompts for learners to explain why anomalous facts are true successfully guides them to revise their misconceptions and discover or understand the correct concepts. The studies also provide empirical evidence to inform nuanced instructional design strategies for revising misconceptions without feedback: 1. Prompts to explain "why?" (an anomalous fact is true) are more effective than prompts to articulate thoughts in revising beliefs. 2. However, prompts to explain are not a silver bullet. It was necessary to provide learners with a problem structure that contained a sufficient number of anomalous facts that required explaining. 3. While explaining more anomalous facts was better than explaining fewer, the quantity *per se* was less important than whether the facts being explained contradicted all of a learner's potential misconceptions.

## ACM Classification Keywords

K.3.1 Computing Milieux: Computer Uses in Education; J.4 Computer Applications: Social and Behavioral Sciences

## Author Keywords

Online learning; prompts; MOOCs.

## INTRODUCTION

The proliferation of online courses requires educators to consider new ways to prompt reflection and learning. One of the

advantages of online education is the opportunity for asynchronous interactions that allow learners to interact with the material at their own pace, potentially allowing deeper reflection. However, a potential disadvantage of online learning is that a learner may spend more time only passively consuming information, thus having less opportunity for experiencing the immediate, interactive feedback that is available in a live, small-scale, instructor-led classroom discussion. As a result, online learners may miss out on learning opportunities that are difficult to recreate in an asynchronous online environment.

In a live classroom, a teacher may prompt higher-level learning using "Socratic methods" like asking students to generate explanations for why facts are true or why relationships hold. The teacher employing this method can guide the process of inquiry and student self-reflection along a path that encourages critical thinking and learning through belief revision. This method requires teachers' active moderation and immediate feedback for individual students. However, due to a high student-to-teacher ratio, it does not easily scale to massive online learning settings. As a result, online learners may hold incorrect beliefs persistently, because instructors cannot probe for each student's incorrect beliefs and design interactive activities to correct them. Can we replicate this good process without real-time instructor involvement?

Existing approaches try to leverage peers for discussion [8] or assessment [?], and intelligent tutoring systems for feedback [6]. [?] present an automatic hinting interface, [5] use Natural Language Processing to coach answering of domain-specific questions, and [2] provide feedback on correctness of multiple choice explanations in an intelligent tutoring system. However, it is challenging to duplicate the success of these intelligent tutoring technologies in many new online education contexts, as larger numbers of increasingly diverse students have a wide range of thoughts and cognitive processes.

Less work has explored how to design the wording of questions and prompts to encourage online learners to self-identify gaps in knowledge and self-initiate revision of incorrect beliefs. Discovering effective questions would be a highly scalable and broadly applicable way to enhance asynchronous solitary online learning. This paper considers how instructors can design online prompts for learners to reflect on facts that are *anomalous* or conflicting with learners' beliefs. Our results inform design principles for how to select such anomalous facts and how to word prompts for learners' reflection in a way that encourages *learners* to play an active role in identifying and revising their misconceptions into accurate knowledge.

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced.

Every submission will be assigned their own unique DOI string to be included here.

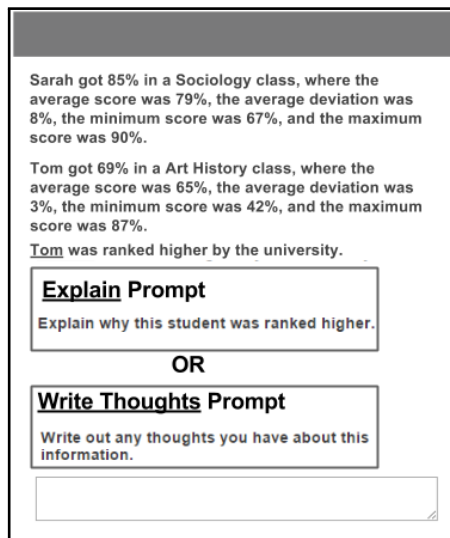


Figure 1. Screenshot of what learners saw for each ranked pair and accompanying Reflective Prompt. Only one prompt was shown, via random assignment to the *Explain* or the *Write Thoughts* Reflective Prompt.

### RELATED WORK

Online environments frequently *provide* explanations of concepts to learners through lecture videos. However, research in education emphasizes the pedagogical value of using technology to prompt learners to generate their own *self*-explanations of what concepts or facts mean in their own words [3, 1]. Of particular relevance to belief revision is the finding from cognitive science theory that proposes that explaining "why?" a fact is true does not merely boost attention or motivation, but drives people to interpret what they are explaining as one instance of a broader pattern [9, ?]. If a fact—like the answer to a math problem—conflicts with a learner’s misconception—like division being the opposite of subtraction—then being prompted to explain why that fact is true might help a learner identify gaps in their knowledge. This leads them to discover new, more accurate principles that account for the anomalous fact.

Research in education, cognitive science, and philosophy of science has explored the role of anomalous facts in learning and conceptual change [4], but found it difficult to get people’s attention to anomalies that can trigger a belief revision. To our knowledge, the research we report in this paper is novel in both human-computer interaction and cognitive science in its investigation of how online learning resources can be enhanced by adding prompts to reflect on facts that are strategically designed to expose common learner misconceptions.

### STUDIES: DESIGNING PROMPTS & ANOMALOUS FACTS

These two randomized A/B comparisons aim to help instructional designers understand how belief revision can be supported by designing prompts for learners to reflect on the facts they are learning. In addition, the studies investigate the best ways to choose which anomalous facts are presented as the target of Reflective Prompts. In this paper we define *anomalous* as being inconsistent with learner misconceptions.

The two studies use similar learning materials, procedure, and participants, which are explained in the next few sections. We

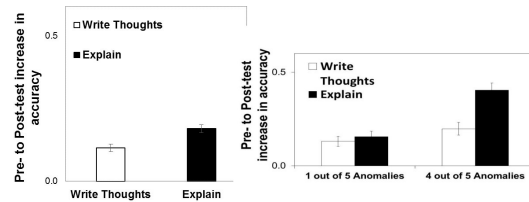


Figure 2. Accuracy Increase from pre-test to post-test on ranking pairs of students in which the correct rule "More deviations above the average" was pitted against the Misconceptions. (a) All learners in Study 1 and Study 2. (b) Study 1. Bars represent +/- 1 standard error of the mean.

Type of Information	Sarah	Tom	Ranking Rule	Use of rule	Higher ranked
Personal Score	85%	69%	Higher score	85 > 69	Sarah
Class Average	79%	65%	Greater distance from average	(85 - 79) > (69 - 65)	Sarah
Class Maximum	90%	87%	Closer to maximum	(90 - 85) < (87 - 69)	Sarah
Class Deviation	8%	3%	More deviations above the average	(85-79)/8 < (69-65)/3	Tom

Figure 3. The misconceptions (Higher Score, Greater Distance from Average, Closer to Maximum) and correct concept (More deviations above the average) that underlie ranking of pairs of students.

then outline the instructional design questions about using Reflective Prompts and targeting Anomalous Facts, and report the results from Study 1 and Study 2 that provide evidence towards answering them.

### Learning about Statistics for Ranking

We chose to conduct the studies using educational materials from statistics problems, because reflecting on facts that are answers to (math) problems has broad relevance. In particular, instructional designers at Khan Academy and university MOOCs on EdX have been sufficiently interested in our proposals to compare question prompts (to reflect on math and biology problems) that they collaborated in launching prompts we proposed to thousands of real students.

The specific statistics task used in the studies is one that prior educational research has identified as particularly challenging, because many students have multiple misconceptions that are hard to uproot even with instructor’s feedback [7]. Figure 3 lists the three common misconceptions and correct concept used in this task. The task therefore provides a fertile setting for us to evaluate strategies for prompting learners to reflect on combinations of anomalous facts with different properties, like which misconceptions they contradict.

Figure 1 shows a screenshot of what learners in our studies observe. Learners see data about the grades of two students who are taking different classes at a university, and are informed that the university ranked one higher than the other. Learners then have to infer from the statistical information about each student (like the student’s score and class average score, see Figure 1, Figure 3) what statistical information is being used to rank students from supposedly non-comparable classes.

### Potential Learner Beliefs: Correct beliefs and misconceptions about how pairs of students are ranked

With respect to IDs’ consideration of how to present anomalous facts, Figure 3 is helpful because it shows four beliefs learners could have about what statistical information is used

to rank pairs of students. Figure 3 also illustrates how holding each belief/misconception would lead a learner to expect Sarah or Tom to be ranked higher, based on the information provided about them and their class scores. If in fact the higher ranked member of the pair is not the student the misconception predicts, then that observation/fact is anomalous with respect to the misconception. For the purposes of this task, the correct belief is that the university will give the higher rank to whichever students is "more deviations above the average"—whichever member of the pair has the score that is more deviations above their class mean. The statistical justification over the misconceptions is based on concepts like standardized normal or z-score. The empirical justification is that it is the only rule that always has perfect accuracy in predicting the higher ranking of all five/six pairs of students. The other three rules are considered misconceptions about how the university might be ranking students, because they are both statistically and empirically problematic in the learning task. They also reflect more general misconceptions in understanding statistical concepts [7].

### Participants

We recruited people who use Amazon Mechanical Turk to engage in these learning tasks, rather than undergraduate laboratory participants or students using online education platforms. This was to have a more representative sample than a laboratory study, while being able to conduct more sophisticated experiments than would be possible with students taking an online course. Participants (659 in Study 1 and 261 in Study 2) were recruited to do a 20-40 minute HIT with compensation around \$3.00–\$6.00/hour.

### Procedure

The previous sections explained most of the details about how learners were introduced to the task of learning a university's system for ranking students from different classes. As a Pre-Test, before seeing the ranked pairs, learners made predictions (no feedback was provided) about which person would be higher ranked, in four *un-ranked* pairs of students. These four pairs pitted the ranking from "more deviations above average" principle against the ranking from all three misconceptions. Learners then learned from five (Study 1) or six (Study 2) of the *ranked* pairs. Each pair remained on the screen for two minutes, with the Reflective Prompt for a text box below. A Post-Test after learning used new items with similar structure to the Pre-Test. Scoring "accuracy" as a learner ranking in accordance with the "more deviations above average" principle, belief revision was operationalized as accuracy increase from before to after study. This is the dependent variable used on the vertical axis of Figures and in statistical tests.

### Instructional design question 1

ID Q1. "To promote belief revision, what kinds of Reflective Prompts should learners receive?"

Study 1 and Study 2 both investigate what kind of Reflective Prompt to present to learners. Belief Revision refers to guiding learners towards the correct but difficult to understand belief about ranking being based on "More deviations above the mean", over the three other more intuitive but incorrect beliefs

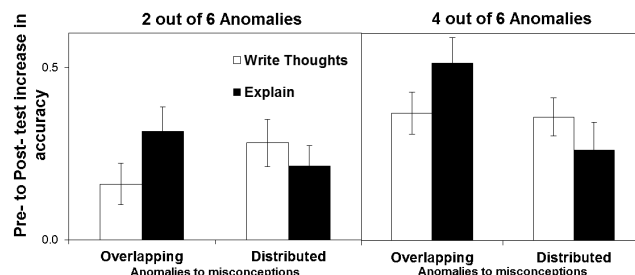


Figure 4. Study 2 Accuracy Increase from pre-test to post-test. Bars represent +/- 1 standard error of the mean.

Experiment 2		2 out of 6 anomalies condition												
		Overlapping condition						Distributed condition						
		1	2	3	4	5	6	1	2	3	4	5	6	
Ranking Rule	Higher score	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	
	Greater distance from average	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	
	Closer to maximum	x	x	✓	✓	✓	✓	x	x	✓	✓	✓	✓	
	More deviations above average	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
			4 out of 6 anomalies condition											
			Overlapping condition						Distributed condition					
			1	2	3	4	5	6	1	2	3	4	5	6
		Higher score	x	x	x	✓	✓	✓	x	x	x	✓	✓	✓
		Greater distance from average	x	x	x	✓	✓	✓	x	x	x	✓	✓	✓
		Closer to maximum	x	x	x	✓	✓	✓	x	x	x	✓	✓	✓
		More deviations above average	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 5. The design of the 6 ranked pairs in Study 2 to have different distributions of anomalous information, based on the 2 (Number of Anomalous Facts per misconception: 2 vs 4) x 2 (Distribution of Anomalous Information: Overlapping vs Distributed). In the Overlapping condition, every ranked pair that is anomalous with respect to one misconception is anomalous with respect to all of them, and every ranked pair consistent with one misconception is consistent with all. In the Distributed Condition, the amount of Anomalous Information is identical, but it is distributed evenly across all of the observations provided for learners to reflect on.

(see Figure 3). The studies experimentally varied whether learners are prompted to *Explain* why a fact is true (why Tom was ranked higher than Sarah) versus *Write Thoughts* about a fact (that Tom was ranked higher than Sarah). Figure 2 (combining studies) shows that the prompt to explain why led to greater belief revision than the prompt to write thoughts ( $p < 0.001$ ).

### Instructional Design Question 2

ID Q2. "Which and what kinds of Anomalous Facts should learners be prompted to reflect on?"

Both studies investigated the instructional design (ID) question of how to choose which anomalous facts learners are prompted to reflect on. What it means for a fact to be *anomalous* with respect to a misconception is that the fact contradicts what a learner with that misconception would predict. All facts (anomalous or not) are consistent with what what is predicted by the correct belief (which is difficult for learners to discover or understand), so the term 'anomalous fact' always refers to a contradiction to one or more misconceptions.

Study 1 investigates how belief revision is influenced by targeting reflective prompts at different numbers of anomalous facts. The number of ranked pairs targeted by Reflective Prompts

is always five, so Study 1 randomly assigns people to have 1 ranked pair that is anomalous (with respect to all three misconceptions) versus having 4 ranked pairs that are anomalous (with respect to all three misconceptions).

### **Results: Is targeting Reflective Prompts at one anomalous fact sufficient for belief revision?**

Figure 2 shows the change in accuracy, as a function of the two independent factors, which was analyzed using a 2 (Reflective Prompt: Explain vs. Write Thoughts) x 2 (Number of Anomalous Facts: 1 vs. 4) ANOVA.

The significant main effects of Reflective Prompt and Number of Anomalies ( $p < 0.01$ ) were superseded by a significant interaction ( $F(1, 659) = 8.20, p < 0.01$ ). Reflective Prompts to Explain promoted belief revision when the targets included 4 anomalous facts, but had no effect when the targets only included 1 anomalous fact.

### **Results: Is it better to target Reflective Prompts at Anomalous Information that is distributed across the facts being explained, or at Anomalous Information that is overlapping in facts that contradict all misconceptions?**

Study 1 assumes that a ranked pair must be anomalous with respect to all of the three misconceptions, or with none of them (consistent with what the misconceptions predict). But IDs might be able to promote belief revision more effectively by distributing anomalous information between the targets of Reflective Prompts. Especially since belief revision did not occur when Reflective Prompts were targeted at one ranked pair that was anomalous with respect to all three misconceptions (and four ranked pairs consistent with the misconceptions). Reflective Prompts might be more effective if they were targeted at observations that had anomalous information distributed across them. For example, the 4 out of 5 anomalies condition from Study 1 might be approximated by having three ranked pairs that contained anomalous information. A first ranked pair could be anomalous with respect to the "higher score" rule, and a second anomalous with respect to the "distance from average" rule, and a third with respect to the "closer to maximum" rule.

Study 2 therefore randomly assigned people to the *Overlapping* or the *Distributed* condition. Figure 4 depicts each of the ranked pairs and which misconceptions they were anomalous with respect to. To allow Overlapping and Distributed to be crossed symmetrically with number of anomalies, Study 2 expanded the ranked pairs to 6, and had a 2/6 Anomalies and 4/6 Anomalies condition.

Figure 4 shows the Accuracy increase as a function of the experimental variables we analyzed in a 2 (Reflective Prompt: Explain vs. Write Thoughts) x 2 (Number of Anomalous Facts: 2/6 vs 4/6) x 2 (Distribution of Anomalous Information: Overlapping vs Distributed) ANOVA. The key result was an interaction between Reflective Prompt and Distribution of Anomalous Information,  $F(1, 261) = 11.23, p < 0.01$ . Reflective Prompts to Explain promoted belief revision when observations of ranked pairs contradicted *all* the misconceptions (Overlapping condition). But the benefit of explaining

was lost when the anomalies were Distributed, and the facts being explained did not contradict all the misconceptions, but could sometimes be accounted for in terms of a misconception. Although Study 1 suggested that the quantity of Anomalous Facts learners were explaining was the key factor in belief revision, Study 2 reveals a more nuanced insight, where the critical design implication is that the Reflective Prompts are most effective when targeted at facts that rule out *all* the misconceptions.

## **DISCUSSION, LIMITATIONS, & FUTURE DIRECTIONS**

The studies presented provide insight into how instructional designers can help online learners revise their erroneous beliefs *without requiring real-time feedback*. By prompting learners to explain why anomalous facts are true, learners can experience in a passive, online environment, the critical thinking process of belief revision, which is traditionally experienced through an instructor-led inquiry. We find that belief revision is not driven by prompts for mere articulation, but only by prompts to explain "why" a fact is true. Our second finding is that explaining is more effective when there are a great number of anomalies. Explaining is not helpful when there is just one anomaly, but is when there are four out of five. However, we note the large amount of anomalous information is *not* sufficient to revise beliefs, as there is no effect of number of anomalies in the *write thoughts* condition.

There are a few limitations of the work that need to be taken into account. First, it could be very hard for instructional designers to come up with the anomalous facts and know what misconceptions are. Designing these facts can be an arduous task, although they can be presented to a large number of students once created. Also, further studies are needed to see if our results generalize to live learning scenarios or other learning tasks. While Turkers may be representative of passive online learners, their motivation differs from voluntary learners. We also plan to test our prompts to ask users to explain anomalous facts on informal learning materials, such as a Wikipedia page.

Our work presents an example of how cognitive science research can be used to inform curriculum design that prompts cognition processes associated with critical thinking, which can occur without live instructor feedback. These prompts can be added to any existing online learning material without significant engineering efforts. We believe that as online consumption of materials rapidly grows to dominate over live-feedback, in-person learning, it is important for instructional designers to explore, test, and apply novel ways to present materials. We present a simple yet effective model for bringing the benefits of an in-person instructional technique to a learning at scale setting. Furthermore, online learning environments allow us an unprecedented opportunity to calibrate learning outcomes as a function of material presentation methods, allowing for learning experiences that are optimized for individual learning effectiveness.

## **REFERENCES**

1. Vincent Alevan, Amy Ogan, Octav Popescu, Cristen Torrey, and Kenneth Koedinger. 2004. Evaluating the

- effectiveness of a tutorial dialogue system for self-explanation. In *Intelligent tutoring systems*. Springer, 443–454.
2. Vincent AWMM Alevan and Kenneth R Koedinger. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive science* 26, 2 (2002), 147–179.
  3. Michelene TH Chi, Nicholas Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. Eliciting self-explanations improves understanding. *Cognitive science* 18, 3 (1994), 439–477.
  4. Clark A Chinn and William F Brewer. 1993. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research* 63, 1 (1993), 1–49.
  5. Cristina Conati and Kurt Vanlehn. 2000. Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education (IJAIED)* 11 (2000), 389–415.
  6. Amy Ogan, Erin Walker, Ryan SJD Baker, Genaro Rebolledo Mendez, Maynor Jimenez Castro, Tania Laurentino, and Adriana de Carvalho. 2012. Collaboration in cognitive tutor use in Latin America: Field study and design recommendations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1381–1390.
  7. Daniel L Schwartz and Taylor Martin. 2004. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction* 22, 2 (2004), 129–184.
  8. Rummel Nikol Koedinger Kenneth R. Walker, Erin. 2011. Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning* 6, 2 (2011), 279–306. DOI : <http://dx.doi.org/10.1007/s11412-011-9111-2>
  9. Joseph J Williams and Tania Lombrozo. 2010. The role of explanation in discovery and generalization: evidence from category learning. *Cognitive Science* 34, 5 (2010), 776–806.