

Chapter 47

Error Discovery Learning



Christopher Lee

What Is Error Discovery Learning?

The case for active learning has for decades highlighted three basic flaws in conventional lecture courses: poor student engagement; poor conceptual understanding (e.g., as measured by concept inventories); and poor transfer (ability to apply learning to real-world problems) (Posner et al. 1982; Hestenes et al. 1992; Halpern 1998). The evidence for all three of these criticisms has become overwhelming, and the benefits of active learning in remedying them have also been amply demonstrated (Crouch and Mazur 2001; Walczyk and Ramsey 2003; Knight and Wood 2005; Michael 2006; Deslauriers et al. 2011; Haak et al. 2011; Gasiewski et al. 2012; Watkins and Mazur 2013; Freeman et al. 2014). Yet after decades of such results, active learning remains the exception rather than the rule (Kober 2015).

This suggests that active learning is caught in a squeeze between two opposing forces. On the one hand, asking faculty to go all in to “transform” their courses is too high a barrier: too much work for too little institutional reward for the instructor (even though *students* clearly benefit) (Dionisio and Dahlquist 2008; Fairweather 2008; Austin 2011; Henderson et al. 2011). On the other hand, common steps such as adding some clicker questions to lecture, or including some existing concept inventory questions on the final exam, seem to fall short of an active learning “transformation” in one crucial way: instruction is still essentially a *one-way broadcast from instructor to students*. This issue goes to the heart of all three core criticisms: a persistent theme in active learning research has been that students cannot effectively engage, understand, or transfer the concepts they are “learning” without a thoroughly *two-way communications process* that exercises all three of those meta-cognitive “muscles” (Mazur 1997; Smith et al. 2009; Freeman et al. 2014).

C. Lee (✉)

Department of Computer Science, UCLA, Los Angeles, CA, USA

e-mail: cjlee@ucla.edu

© Springer Nature Switzerland AG 2020

J. J. Mintzes, E. M. Walter (eds.), *Active Learning in College Science*,
https://doi.org/10.1007/978-3-030-33600-4_47

757

Hence it is interesting to ask: Is there an *easy* way to start turning large-course instruction into a two-way communications process that exercises all three of these muscles? This question has been the point of departure for a data-driven approach to active learning that we have dubbed *error discovery learning* (Lee et al. 2018). It is based on several premises:

- *Instructor blindspots*: instructors often cannot see what each student in a large course actually thinks about a given concept and specifically how they misunderstand it (Hestenes et al. 1992).
- *Student blindspots*: students often do not immediately see important implications of a concept and furthermore can only become aware of their own blindspots when they attempt to apply the concept to a real-world example problem that convinces them they're missing a vital implication (Mazur 1997).
- *Data-driven, open-response concept testing*: the only way to overcome these blindspots is to collect and analyze enough student solutions to such “challenge problems” to convincingly identify the underlying causes of all student errors. A crucial and obvious aspect of this is that it must be open-response (not multiple choice; Griffard and Wandersee 2001), encouraging students to explain their own thinking in their own words (Camfield and Land 2017), sufficient to diagnose their inner thought processes directly from their answers.
- *Specific error models and frequencies*: each of these underlying causes must be cataloged as an “error model” that identifies exactly where students’ thinking is going astray (Klymkowsky and Garvin-Doxas 2008; Andrews et al. 2012; Leonard et al. 2014) and its frequency measured across the student population.
- *Resolutions and validations*: once an instructor has identified a specific error model as blocking a large fraction of students, she can provide *resolution lessons* that explain the misconception, why it is wrong and how to fix it (van Gelder 2005). Students who make this error need to be redirected to these resolution lessons and then re-evaluated with a validation step to assess whether this actually helped them overcome that misconception.

This process changes instruction to a two-way communication cycle (Fig. 47.1; an example session is shown in Fig. 47.2) that focuses almost entirely on the explicit analysis and resolution of *blindspots*, an educational element almost wholly missing in action from conventional textbooks and course materials. Beginning in 2011, this error discovery learning process has been developed as an open-source software platform (Courselets.org) and tested in a number of UCLA life and computer science courses, comparing several stages of instruction. For example in a bioinformatics theory course of approximately 80 students per class (Lee et al. 2018), we began with a conventional lecture (2003–2008); converted to a “Socratic method” of posing challenge problems (Prince and Felder 2006) for students to answer verbally (2009) or on their laptops (2011–2013) during class; combined in-class web-based exercises with follow-up stages outside class (2015–2017), and added entirely online exercises and projects (2016–2017).

This chapter seeks to review the empirical findings from these tests on student learning outcomes and to summarize our practical experience with this method as

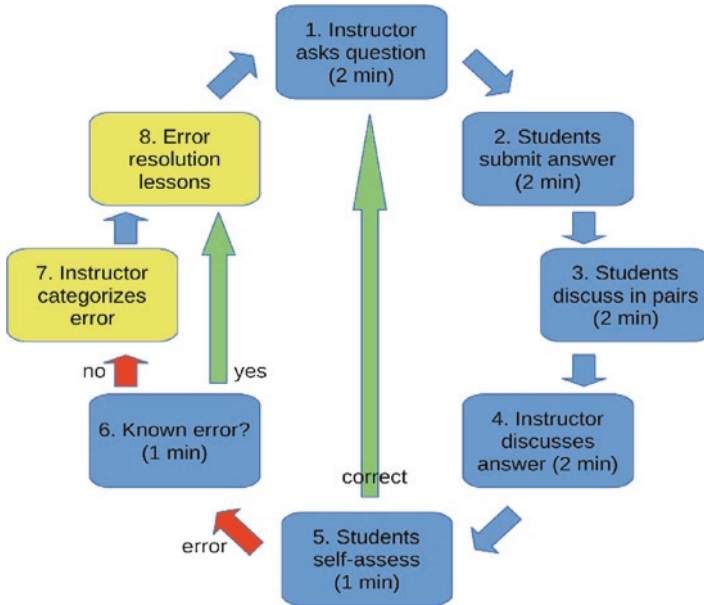


Fig. 47.1 Stages of the error discovery learning process

Students answer challenge problems by writing text on their laptop or smartphone and then briefly discuss their answers in pairs, before each assessing their own answer against the correct answer and against known conceptual errors that have been previously observed on that question (see text for more details). Optional stages that can be performed outside of class (online) are highlighted in yellow

proposed “best practices” that seem to make it easiest and most effective for other instructors to adopt in their own teaching. All of the tools we describe are freely available as an open-source software platform (Courselets.org), so interested readers can immediately inspect or try out any aspect of this.

Evidence from Classroom Studies

We examined EDL’s effects on student engagement, exam scores, and course completion rates, focusing especially on measuring disparities across all students in a class. Several main conclusions emerged consistently from all 5 years of EDL instruction in the study (Lee et al. 2018):

- In the presence of blindspots, *adding more exercises can actually increase disparities* between the least- vs. most-engaged students. For example, switching from lecture to a verbal Socratic method in 2009 succeeded in boosting the total number of times students answered questions in class, from 0.2 per class (2008) to 21.3 per class (2009). However, this boost was conspicuously limited to a

Fig. 47.2 Example student session on [Courselets.org](https://courselets.org). The student participates in a “chat session” where they answer questions posed by the instructor, self-assess, identify, and resolve misconceptions

A biotech company has developed a new test for AIDS, and has published validation data for a large, random population sample. The validation data report each person's result from the new test (positive test result T^+ vs. negative test result T^-) as well as whether that person really had AIDS or not by the "gold standard" clinical criteria (actually had the disease D^+ vs. did not have the disease D^-). Say you're a physician advising a patient who has just received his test result T . The patient asks "How reliable is this result?" Using the symbols defined above, **define exactly what probability you would provide such a patient as the measure of the reliability of his test result.**

The reliability is $p(T^+|D^+)$

Select the level of your confidence?

Not quite sure

The key thing to realize is that what matters for a patient is the conditional probability $p(D|T)$. I.e. given *what they know* (the test result T), what is the reliability vs. uncertainty in forecasting *what they want to know* (whether the patient has disease, D)? Note that the opposite conditional probabilities $p(T|D)$ are **not relevant to a patient** or doctor because they do not go from "what you know" (T) to "what you want to know" (D).

Different

Below are some common misconceptions. Select one or more that is similar to your reasoning.

I didn't consider what was hidden (unknown) vs. observable (known)

small number of students. Since this method focused in the usual way on “what is the right answer?”, its main effect on students who didn’t “get the right answer” was probably to discourage them – many times per class.

- By contrast, posing the same questions through *the EDL software platform dramatically boosted the number that the average student answered* during the course, from a mean of 14 per student in 2009, to 30 per student in 2011, to 60 in 2015. Most importantly, the biggest shift was for the least-engaged students. For example, in 2015, 90% of students answered *at least 30* challenge problems or more each (whereas in 2009, most students answered few if any). As a second example of engagement disparities, the large difference in the number of questions answered by undergraduate vs. graduate students in 2009 shrank in 2011 and virtually disappeared in the subsequent years of EDL.
- *Exam scores displayed a similar boost, with the biggest shifts for the least-engaged students:* whereas no significant increase in exam scores was observed in 2009, they increased in each EDL year (2011–2015), to the extent that the lowest 50% of student exam scores were shifted into the range of the top 50% of student exam scores from 2008 to 2009. Independent assessment of exam cognitive rigor found the 2011–2015 exams equal or more challenging than 2008–2009.
- *The EDL cycle also appeared to boost course completion rates, especially among women:* this bioinformatics theory course (Computer Science 121, covering probabilistic modeling of genomics data) had always had a high attrition rate (48%) each year from 2003 to 2009, but during each EDL year 2011–2015, it experienced much lower attrition (11%). This effect was especially strong for women.
- The EDL data showed that *each student had around 20 distinct misconceptions* that they needed to identify and address in the course. This strongly suggests that each student must complete at least 20 challenge problem EDL cycles, and likely many more, to identify and overcome each of these. This in turn suggests an important threshold effect for student engagement; for example, although 2009 used much the same active learning materials (challenge problems) as the subsequent EDL period, most students in 2009 answered far fewer than this threshold. This may explain why no attrition or exam score improvements were observed in 2009.
- The EDL data cataloged *over 220 distinct misconceptions* that students shared (Fig. 47.3), many of which were unanticipated by the instructor and many of which were surprisingly common (e.g., shared by 30–40% of students). Moreover, when we searched the textbook for statements addressing these specific misconceptions, we found they were not addressed. While it is unsurprising that textbooks focus on “right ways” of thinking rather than “wrong ways,” the EDL data indicated that on each concept, around 50% of students were blocked by such an unaddressed blindspot.

Students	Misconception
47%	Mixing up <i>disjoint</i> vs. <i>statistical independence</i> of random events, e.g. presented with a Venn diagram showing two disjoint events, students assert "They are independent since they have no overlap".
35%	Failure to distinguish the two directions of a conditional probability, i.e. a conditional probability vs. its converse.
33%	Failure to realize that we can always eliminate a variable from a joint probability by summing over all its possible values, and that this is a crucial tool for manipulating probability models.
32%	Failure to consider which direction of conditional probability measures real-world prediction performance, for a hidden variable H vs. observable variable O , i.e. $p(H O)$ vs. $p(O H)$.
31%	Confused about the difference between the concept of a <i>random state</i> vs. a <i>random variable</i> .
28%	Believe that $\sum_{Cp}(S C)=1$, i.e. that a conditional probability is normalized over the <i>condition variable</i> .
23%	Failure to realize that if a hidden state is rare, the false discovery rate can be high even if the false positive rate is low.
17%	Failure to realize that $p(X_1, X_2, \dots, X_n)$ for I.I.D. X_1, X_2, \dots, X_n goes down exponentially with increasing n .
14%	Failure to realize that a likelihood $p(O \text{model})$ is only meaningful in <i>comparing</i> one model vs. another, not an absolute measure of our confidence in the model.
9%	Some people thought two independent experiments should agree simply because of the large sample size. No, the key unknowns are the False Positive and False Negative rates.

Fig. 47.3 Examples of statistical misconceptions discovered among students in a computer science course (that required completion of a statistics and probability course as a course prerequisite)

- The data showed that student misconceptions display a strong 80–20 rule: for any given concept, a small number of misconceptions are very common and explain the vast majority of student errors. Overall, the top four error models for each question addressed 70–80% of all student errors on that question (Lee et al. 2018). This makes EDL potentially both scalable and efficient: identifying a misconception in just one student can automatically help *all* subsequent students who make that error; and an initial sample of ten students is enough to address 70–80% of all student errors.

For any instructor who has not previously used an EDL platform such as Courselets, these data may seem only to sketch a goal that is both far away and lacking many essential details about how to actually do it, how much work it takes, and where to begin. Therefore in the remainder of this review, we “fast-forward” to our practical conclusions for *EDL best practices* that furnish precisely these details.

EDL in Practice: Can Instructors Use This Now?

The EDL work described above was done using [Courselets.org](https://www.courselets.org), a free, open-source web platform developed specifically for EDL, that any instructor or student can use. It has been used predominantly in life science and computer science courses, with class sizes ranging from 20 to 700 students each. Students using [Courselets.org](https://www.courselets.org) have completed over 120,000 EDL exercise cycles (challenge problems) in approximately the last 2 years. [Courselets.org](https://www.courselets.org) exercises are designed to integrate directly into university course management systems (CMS), so that instructors can assign a Courselets exercise as easily as copying and pasting a URL into their campus CMS. Student simply click these links within the campus CMS to immediately perform the exercise, without any additional steps such as creating a Courselets account or having to sign in to Courselets. Instructors can also simply email a clickable enroll code link to any student, independent of the campus CMS.

Where Is the Best Place to Start Using EDL? *Best Practice: Safety + Urgency*

One quick way to gain insight into how EDL works is to examine carefully a practical question: if you wanted to pick one set of your *existing* class materials to run as a Courselets exercise, what would be the best choice: an assigned reading; graded homework; practice exam problems; or a graded project? EDL depends first and foremost on getting students to actively expose their misconceptions. For the students that means hard work, both intellectually and psychologically (van Gelder 2005; Smith et al. 2009; Hochanadel and Finamore 2015). We have found that this in turn depends on two factors that may seem opposed: *safety + urgency*.

- *Urgency*: in the high pressure and pace of modern university education, students are driven to invest such hard work only in assignments that directly determine their grade. Hence a reading assignment (even assuming it contains good challenge problems) would not be a top choice, because it lacks this immediate urgency.
- *Safety*: however, urgency contains a paradox; it punishes students for exposing their misconceptions. The standard grading method of “points off for errors” switches students to *tell the instructor what (you think) she wants to hear*, rather than naively (i.e., honestly) *thinking out loud*. In other words, the prime directive is to *get the points* (Hughes et al. 2014; Schinske and Tanner 2014), rather than to self-reveal. While this may be unavoidable for conventional graded homeworks and projects, it intrinsically destroys the value of EDL. Just how fundamental this fork in the road is for students is nicely captured by the very first thing students said when Eric Mazur tried out the Force Concept Inventory on his class: “Professor, are we supposed to answer these questions according to the

way you taught us, or the way we really think?” (Even though it was given to them as an ungraded exercise, the students apparently did not forget the prime directive! Mazur 2009). This raises the question, how can we create a “safety zone for errors” in which students are positively motivated to expose all their misconceptions?

In our experience, the best practice for creating both safety + urgency is to *pair a Courselets exercise with a subsequent high-stakes assessment*, for example, by giving students *practice exam problems* as a Courselets exercise preceding the actual exam. The Courselets exercise should be assigned on a credit-for-completion basis (rather than “points off for errors”) – so that students will do it – with a strong message that “this is your chance to boost your exam score, by showing us how you think through the kinds of problems that will be on the exam, so we can help you identify where specifically you’re going wrong and how to fix it in time.” The student experiences this unusual combination of safety + urgency roughly as follows: “OK, here’s a typical exam problem... uh-oh, looks like I got it quite wrong! Oh wait, here’s some help that explains where my thinking is going wrong. OK that makes a lot more sense now. I better try the other practice problems ASAP!”

In addition to these key motivational advantages, using existing practice exam problems as your first Courselets exercise is also a great place to start because it is almost no work. That is, most instructors already have practice exam problems at hand; they are obligated to provide those to students anyway (so there is little additional work on platform B instead of platform A); and this involves *no change* to your existing class assignments, lecture materials, or any other aspect of your course.

How Should I Run an EDL Exercise for Maximum Learning? ***Best Practice: The 10% Rule for Immediate Resolution***

Here again, the obvious practical questions for a new EDL instructor are how do you actually do EDL, to generate error models, and resolutions that fix them. To make this concrete, let’s say we want to use practice exam problems as a Courselets exercise prior to a Friday midterm exam. When should we make it due? What are the time-critical events for making it succeed?

Safety + Urgency is the obvious make-or-break principle here, whose logical implications must drive all our detailed decisions. For example, the urgency comes solely from Friday’s midterm. Within the crowded schedules of modern university students, the number of hours required to study for a midterm cannot (reasonably) much exceed the number of spare hours a student has in 1 day (8 h, say). This produces an unfortunate corollary: if students *can* study for a midterm in 1 day, when *will* they begin studying? For too many the answer is the day before. But if their studying reveals misconceptions, that will be too late to address them.

Intelligently scheduling the Courselet assignment can help solve this. If we assign it to start on Monday and be due Wednesday, this makes students start the practice exam by Tuesday. And once they begin, an interesting captive-audience

effect kicks in: since they are obligated to spend this time on the Courselet no matter what, they might as well treat this as their midterm study time. In effect this “back-ports” the midterm’s urgency from Thursday to Tuesday while giving the Courselet the special combination of safety + urgency that is essential for EDL. However, this captive-audience effect breaks down if we try to schedule the Courselet too far in advance of the midterm; that is why urgency is a crucial ingredient.

This opportunity will translate into learning gains only if it gives each student immediate help in identifying and overcoming the specific blindspots that are blocking her. This suggests another time-critical step in the EDL process and an associated best practice: if the instructor quickly identifies misconceptions in the first ten student answers, our data show that that will be sufficient to address 70–80% of errors in the rest of the class (Lee et al. 2018, Supp. Fig. 1). More to the point, those students will have their errors addressed *as soon as they answer the question*. Courselets.org automatically emails the instructor when the first ten students have answered the exercise, with a link the instructor can click to immediately view their wrong answers and diagnose which misconceptions are present. Simply following this policy will ensure that the remaining 90% of students will find their misconceptions addressed the instant they answer each Courselet question. Applying this rule to our example timeline would typically run as follows:

- *Monday*: the practice exam Courselet link is made available to students.
- *Later Monday*: instructor notified that ten students have answered and writes error models for their mistakes; the first ten students receive back that individualized help.
- *Tuesday*: most students are completing the practice exam and immediately receiving help.
- *Wednesday*: all students must complete the Courselet. The instructor now has a statistical prioritization of the frequencies of all misconceptions in the class, and for each one, how many students (if any) are still confused and want further material to resolve it. The instructor writes resolution lessons for the highest priority confusions, and those students are automatically notified.
- *Thursday*: as students work through each resolution lesson, they re-rate whether it resolved their questions. As major blindspots are gradually overcome, the changing statistics show the instructor what the remaining priorities are, and the instructor continues to add resolution lessons to address them.

Where Should I “Individualize” Student Learning with EDL?

Best Practice: Run a Prerequisites Inventory

One major advantage of Courselets is that it can individualize what each student puts effort into, depending on their distinct needs. This can be a very helpful complement to standard course materials, which assume a *uniform set of demands* across all students. This suggests an obvious practical question: Where is the best place in a course to first take advantage of this individualization?

In many courses, one of the best places is to provide a “*prerequisites inventory*” *courselet* that identifies where each student needs help with a crucial prerequisite skill and helps them address it. Many students’ understanding of prerequisite concepts and skills falls far below 100% (Hestenes et al. 1992; Crouch and Mazur 2001; Smith et al. 2008; Shi et al. 2010), causing big disparities before the course even starts. On Courselets instructors can remedy these individual needs in an individualized way, easily and without impacting the existing course structure.

- An *online courselet of quick prerequisite exercises* will take little time for a student who has those skills. But any prerequisite that a student lacks will redirect her to additional steps to help her with that specific skill.
- The gap between what is *uniform* in a class (its assignments) and what is *not* (the students’ differing backgrounds) is arguably the most valuable place for individualized exercise. *Within* the class, instructors provide a uniform experience, but everything that comes *before* the class is both out of their control and *not uniform* across the students.
- Adding such a prerequisites courselet as an online exercise has *little effect on the existing assignments and syllabus*. The instructor doesn’t have to change anything in the class to take full advantage of this individualized approach.

How can we best give the prerequisites inventory *safety + urgency*? We give it urgency by telling students “This will help you get solid on the key skills you’ll need to use over and over in the course, starting with this week’s homework.” The details depend on how students enter the course:

- In certain cases such as freshman chemistry or calculus, a course may be widely regarded as a key gateway, with “intake infrastructure” already in place for involving students well in advance of the first day of class (e.g., placement exams). In that case, mandating that students complete the prerequisite inventory during the existing process (e.g., freshman orientation) helps shift the exercise from one of “deficit identity” to one of “grit” (where students find that they immediately get help on their specific difficulties) (Rittmayer and Beier 2008; Trujillo and Tanner 2014; Hochanadel and Finamore 2015).
- Otherwise, the prerequisite inventory should be the key “initiation” for entering the course, the first thing students do, backed by credit-for-completion, a deadline, and follow-up to get laggards to complete it. If absolutely necessary, urgency can be added by pairing the Courselet with a follow-up graded quiz.

How Can I Apply EDL to Graded Homework and Projects?

Best Practice: The Weekly Mission Training Cycle

To make this concrete, let us look at an example of how the single-nucleotide polymorphism (SNP) scoring project (from my course on introduction to bioinformatics theory) was transformed into an EDL design. This project sought to give computer

science students an opportunity to apply the modeling concepts taught in class in order to write their own program for scoring genomic sequencing data. In the original lecture course, students had a 1-week deadline to submit their program and received their grades one week later.

In practice, the 2-week turnaround time for feedback meant that *any* blindspot during the first week blocked students from being able to do the project. Their questions and concerns in turn forced me to make the project instructions ever more encyclopedic in supplying students every possible detail in both theory and programming, contrary to the learning value of an independent project. This problem was resolved by a weekly “Mission Training” cycle that created safety + urgency, eliminating blindspots (with immediate resolution).

Each 1-week topic begins on Monday and presents the “mission” (project) the students have to complete by the following Monday, with a three-part timetable: “today and Wednesday’s classes will help you learn the basic principles you need; by Friday you each need to solve all the theory for the project; over the weekend you have to finish the mission and write the program.”

The theory part, called “Mission Training,” is run first as a set of challenge problems on Courselets made available on Monday and due by Friday’s class. The graded project relying on the theory is due 3 days later providing a strong element of urgency. In effect, all the materials previously given to the students as “project instructions” were simply turned into a carefully sequenced series of questions for the students to solve *themselves*, with the EDL process guiding them every step of the way, identifying their blindspots, and providing immediate resolutions (readers can view the Mission Training at <https://www.courselets.org/chat/enrollcode/74f4690fd169491c8ee084b5fa8ddc44/>). Second, the graded project was moved to an online testing platform that gives students instant feedback on whether their program is successful (readers can try the online project at <https://stepik.org/lesson/32858/step/1>).

Our experiences with the Mission Training cycle suggest several basic conclusions: (1) a fundamental principle for graded assignments (“summative assessment”) is that it is only fair to give them *after* students have had enough *practice* on those skills (“formative assessment”) to learn them solidly; (2) an exercise cannot be considered “formative” for most students unless it achieves “zero blindspots”; (3) an exercise cannot be *formative* unless it provides immediate resolution of student blindspots; (4) conventional graded assignments fail these zero blindspots + immediate resolution requirements, greatly reducing their learning value; and (5) adding Mission Training using EDL resolves this paradox.

The Mission Training cycle provides another useful bridge, between strictly conceptual learning and real-world application. To state the matter simply, students can only work on one major thing at a time, e.g., in Mission Training they work on concepts and then in the project they work on application. By contrast, in the original project design, there was only a single stage, so students struggled (unsuccessfully) with both the concepts and applications (e.g., writing code) simultaneously. Without the bridge of a Mission Training cycle, instructors are stuck; they must either give up on seriously focusing on applications or “spoon-feed” students the conceptual implications. It is unrealistic to expect them to figure out *both* simultane-

ously. Concepts and mechanics are again two things that conflict when “overloaded” into a single stage but when split into two linked stages become powerfully supportive of each other.

How Do I Write the Most Effective Challenge Problems for EDL? Best Practice: *Translation Problems*

The best practices started from simply reusing *existing* problems (e.g., practice exam questions) but ended with suggestions that requires writing *new* challenge problems (e.g., Mission Training questions). EDL changes the focus for what makes a “good problem”: whereas a conventional homework problem mainly has to be quickly *gradable*, an open-response concept test must meet the more challenging requirement of *zero blindspots*. That is, if a student has a misconception, it *will be exposed* and second that it should be *diagnosable* directly from the student’s answer. This is more challenging.

Our experience with EDL suggests that there is a straightforward best practice for doing this, which we call *translation problems*. The question should require students to answer using a *different representation* than the question was asked in. For example, we might pose a question about a probability problem in *words*, using a specific vocabulary taught in class. Then we could ask students to show how to solve it by drawing *pictures* using a specific kind of diagram taught in class (e.g., Venn diagrams). Such translation tasks have been directly shown to improve retention (Fernandes et al. 2018). Alternatively, we could pose a problem by drawing a diagram and ask students to show how to solve it using *equations*.

For example, “Draw a Venn diagram representing two independent variables, X and Y . Restrict each variable to just three discrete states and draw the diagram so area represents (i.e., is proportional to) probability.” Since it is no longer possible to answer the question “in the same language,” students must perform *translation*. That is, they must first interpret the conceptual meaning of the question representation (e.g., diagram), then reformulate that meaning in the second representation (e.g., equations), and finally reason correctly in that second representation to solve the problem. This translation process will immediately expose any flaw in their conceptual understanding at any point in this extended chain. Moreover, because of this *exposure guarantee*, translation problems need not be very difficult, that is, translation problems that are *easy* for a student with good understanding of the basic concept can still give *high probability of exposure* for any student who has a misconception. This in turn maximizes *diagnosability*. For such simple translation problems, it is generally straightforward to see *precisely where* a student went wrong conceptually.

This best practice can actually be an easy step for most instructors because their materials *already* teach students multiple representations of a concept, typically in *words* (a specific vocabulary that defines the key relationships in the concept), *pictures* (a specific way of drawing those key relationships that makes it easy to see and manipulate them), and *symbols* (or equations). Students have difficulty understand-

ing a complex idea if it is presented in only a *single* form, such as a lecture that is 100% equations or a textbook without a single picture or diagram. So the use of such multiple representations is just standard practice for good teaching.

The instructor's task is made easier because translation problems don't have to be *hard* to be *effective* at exposing misconceptions. In our experience, just about any basic question "rerouted" to use translation will expose copious misconceptions: e.g., students who "know" the right words but not some of their basic meanings; students who misunderstood the original representations that were taught; and students who can't use the representations to see their basic implications. For instance, in the example above, the instructor's task is relatively simple (like asking what is the Spanish word for "hello"), yet it exposes basic misconceptions in about 50% of students who have *completed* a statistics and probability course.

Conclusion

In closing, I wish to distill the underlying logic of what EDL *is* and what it *is not*. Although elements of *automation* are obviously crucial to EDL's scalability (it can work as well in a class of 700 students as in a class of 20), it is important to understand that it contains no aspect of "artificial intelligence." It is *not* based on trying to use machine learning to "classify" student answers, "auto-grade" them, or "predict" what information will help them most. Instead, each step of EDL, such as an instructor identifying a new error model in student responses or a student performing self-assessment, is 100% *human intelligence* rendered "self-efficacious" by ensuring first that there are no blindspots (the instructor can easily see what every student in the class is *actually* thinking on each concept) and second that there is always an immediate next step.

Zero blindspots, immediate resolution. These are the crucial characteristics of effective *two-way* communication that large-class instruction has largely lacked. Indeed, if we reflect on the best of "two-way" teaching as epitomized by the one-on-one tutorial, the EDL "best practices" will be seen as nothing more than what good tutors have always been doing: starting with a prerequisites inventory; providing immediate resolution; finding and creating opportunities for safety + urgency; running a Mission Training cycle; and posing translation problems.

The difference is that EDL makes this two-way communication scalable, moving it from the domain of a one-on-one tutorial to large classrooms and even online. EDL turns a one-on-one tutorial into a *social network*. The individual student experiences EDL as simply a conversation like a traditional tutorial (a chat user interface), but behind the scenes, the system is connecting her to the right people and insights that will give her the specific help she needs now. Once a misconception is identified in a *single* student's answer, that insight will be made available to *all* students who make that error in the future, and their remediation experiences will in turn help others. The design goal of the whole system is to make those human learning connections.

The idea of EDL as a social network that brings learners (and instructors) together in flexible ways whenever they can help each other is a powerful idea. Its implementation so far (on [Courselets.org](https://www.courselets.org)) has already proved itself highly useful for boosting student engagement, learning outcomes, and persistence, but this clearly has only scratched the surface of what is possible. It is now easy for any instructor to try out an EDL in their own class by giving their students practice exam problems and immediate resolution on Courselets. Beyond that, many things are possible. It only awaits a community of teachers and learners to take it wherever they want it to go.

References

- Andrews, T. M., Price, R. M., Mead, L. S., McElhinny, T. L., Thanukos, A., Perez, K. E., Herreid, C. F., Terry, D. R., & Lemons, P. P. (2012). Biology undergraduates' misconceptions about genetic drift. *CBE Life Sciences Education, 11*, 248–259.
- Austin, A. E. (2011). *Promoting evidence-based change in undergraduate science education. Commissioned Paper*. Washington, DC: National Academies National Research Council Board on Science Education.
- Camfield, E. K., & Land, K. M. (2017). The evolution of student engagement: Writing improves teaching in introductory biology courses. *Bioscene: Journal of College Biology Teaching, 43*(1), 20–26.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970–977.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science, 332*(6031), 862–864.
- Dionisio, J. D. N., & Dahlquist, K. D. (2008). Improving the computer science in bioinformatics through open source pedagogy. *Inroads – SIGCSE Bulletin, 40*(2), 115–119.
- Fairweather, J. S. (2008). *Linking evidence and promising practices in Science, Technology, Engineering, and Mathematics (STEM) Education: A Status Report*. Washington, DC: Commissioned paper presented at NRC workshop on Evidence on Selected Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education.
- Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science, 27*, 302–308.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.
- Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education, 53*(2), 229–261.
- Griffard, B., & Wandersee, J. H. (2001). The two-tier instrument on photosynthesis: What does it diagnose? *International Journal of Science Education, 23*, 1039–1052.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science, 332*(6034), 1213–1216.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53*(4), 449–455.
- Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching, 48*(8), 952–984.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141–158.

- Hochanadel, A., & Finamore, D. (2015). Fixed and growth mindset in education and how grit helps students persist in the face of adversity. *Journal of International Education Research, 11*, 47–50.
- Hughes, B. E., Hurtado, S., & Eagan, M. E. (2014). *Driving up or dialing down competition in introductory STEM courses: Individual and classroom level factors*. Washington, DC: A paper presented at the Association of the Study of Higher Education.
- Klymkowsky, M. W., & Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed's tools and the biology concept inventory. *PLoS Biology, 6*(1), e3. <https://doi.org/10.1371/journal.pbio.0060003>.
- Knight, J. K., & Wood, E. B. (2005). Teaching more by lecturing less. *Cell Biology Education, 4*(4), 298–310.
- Kober, N. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, DC: National Academies Press.
- Lee, C. J., Toven-Lindsey, B., Shapiro, C., Soh, M., Mazrouee, S., Levis-Fitzgerald, M., & Sanders, E. R. (2018). Error-discovery learning boosts student engagement and performance, while reducing student attrition in a bioinformatics course. *CBE Life Sciences Education, 17*(ar40), 1–13.
- Leonard, M. J., Kalinowski, S. T., & Andrews, T. C. (2014). Misconceptions yesterday, today, and tomorrow. *CBE Life Sciences Education, 13*(2), 179–186.
- Mazur, E. (1997). *Peer instruction: A user's manual series in educational innovation*. Upper Saddle River: Prentice Hall.
- Mazur, E. (2009). *Confessions of a converted lecturer* (Lecture at UMBC). Retrieved November 20, 2018, from <https://www.youtube.com/watch?v=WwslBPj8GgI>
- Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education, 30*, 159–167.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conceptions: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227.
- Prince, M. J., & Felder, R. M. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education, 95*(2), 123–138.
- Rittmayer, A. D., & Beier, M. E. (2008). Self-efficacy in STEM. In B. Bogue & E. Cady (Eds.), *Applying research to practice (ARP) Resources*. http://www.engr.psu.edu/awe/misc/arps/arp_selfefficacy_overview_122208.pdf. Accessed 13 July 2016.
- Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE Life Sciences Education, 13*, 159–166.
- Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE Life Sciences Education, 9*, 453–461.
- Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The genetics concept assessment: A new concept inventory for gauging student understand of genetics. *CBE Life Sciences Education, 7*, 422–430.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves performance on in-class concept questions. *Science, 323*, 122–124.
- Trujillo, G., & Tanner, K. D. (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE Life Sciences Education, 13*, 6–15.
- van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 53*(1), 41–46.
- Walczyk, J. J., & Ramsey, L. L. (2003). Use of learner-centered instruction in college science and mathematics classrooms. *Journal of Research in Science Teaching, 40*(6), 566–584.
- Watkins, J., & Mazur, E. (2013). Retaining students in Science, Technology, Engineering, and Mathematics (STEM) majors. *Journal of College Science Teaching, 42*(5), 36–41.