Using a Q Matrix to Assess Students' Latent Skills in an Online Course

Ping-Hung Hsieh
Xiaohui Chang
Andrew Olstad
Oregon State University

Oregon State University

Introduction

Online teaching and learning has become an increasingly important aspect of the educational mission of universities. In person, teachers have time-tested tools for assessing student ability, including a wealth of verbal and nonverbal communication. The online format provides a wealth of data, and promises—but may not yet deliver—useful tools for this sort of just-in-time assessment. Publisher homework websites and quizzes inside a learning management system like Canvas can theoretically provide up-to-the-minute performance data including scores, use of help features, access of resources, and more.

Our setting (teaching introductory online quantitative classes in the College of Business at a large research university) makes these innovations particularly appealing. Publishers have correctly identified our interest in "knowing" our students better via their online performance, but we have not yet seen an off-the-shelf solution that gets at our need: the ability to quickly and effectively react to student data in real time.

In this paper, we discuss a portion of our research conducted in an online quantitative methods class, a 200-level undergraduate course in the College of Business. This research included constructing a Q Matrix as part of a Cognitive Diagnosis Model for our quantitative methods class. A Q Matrix is a mathematical tool that creates a linkage between underlying concept development and students' performance on test items. In order to create assessments of learning which are based on student responses to questions, we must first investigate whether these questions are actually testing the foundational concepts we wish to evaluate. The Q Matrix offers a more holistic view of student achievement, and allows better insight (in terms of specificity regarding particular skills and concepts) into student growth and accomplishment than traditional item response methods. Q Matrix analysis requires serious attention to questions about how students are learning material and what underlying skills are being assessed by test questions. The research is based on two main theoretical foundations: Item Response Theory and Cognitive Diagnosis Models.

Item Response Theory (IRT) vs. Cognitive Diagnosis Model (CDM)

Traditional tests ("item response") measure students' overall course mastery by calculating the number of questions they score correct. IRT assumes a latent ability for each student on each question (item). It focuses more on whether a question (1) is easy or hard, (2) is easier to guess correctly, and (3) is able to discriminate higher or lower ability. Each student will be given an estimated ability but IRT does not tell us what skills have been mastered.

Cognitive Diagnosis Models (CDMs), on the other hand, focus more on whether a student has mastered a certain set of skills, and not so much about each question individually (George & Robitzsch, 2015). CDMs attempt to discern which skills students have mastered by mapping each question onto the underlying skills required to answer that question (Uchiyama & Radin, 2009). The resulting output describes not one overall "mastery" number (like a 75% on a final exam) but instead estimates for each student which skills they have and have not mastered.

Our current project looks at both issues: student's mastery of skills and the alignment of each question to the skills (as opposed to whether a question is easy or hard.) In this research, we have pursued a combination of IRT and CDM.

Explanation of a Q Matrix

The mapping tool which makes this calculation possible is called a "Q Matrix." A Q matrix consists of rows (questions on the exam) and columns (skills assessed on the exam). Each cell is either a "1" (if that row's question requires that column's skill) or a "0" (if that question does not require that skill). At its heart, this is essentially a compact list of which skills are required for which questions. It is still possible to "guess" a question's answer correctly without the required skills, or to "slip" and make a mistake on a question in spite of possessing the required skills. These "guessing" and "slipping" parameters can be estimated using the data.

Although simple in principle, constructing the Q matrix requires careful consideration of each question in a test instrument. It may be especially difficult—and especially valuable—for instructors to carefully assess these groups of questions in order to make sure that the questions really do align with the desired skills. In particular, we found in our own process that small variations in how a question was phrased may lead students to use different approaches in answering the intended question.

Implementation: Application to our current courses

Within the College of Business, we have addressed ongoing challenges for our students in their introductory statistics course. While overall (item-response) scores have been lower than desired, we had no data on students' performance from a cognitive diagnosis model.

In order to better understand our students' successes and challenges in the course, we constructed a "Q Matrix" for portions of the course, based on a subset of their midterm questions. We chose a series of questions similar to those given on previous exams which we felt clearly mapped to required course skills. All three authors separately considered the questions and drafted what skills were required by each question. This process was surprisingly complicated: in particular, we found constructive and illuminating differences in the levels of abstraction to describe skills (for instance "read a z table" vs. two skills, "read a lower-tailed z table" and "find an upper tail based on reading a lower-tailed z table"). Table 1 shows an illustration: this Q matrix shows how 6 questions (1-6) rely on students understanding 5 concepts (A-E). The first row shows that Question 1 hinges on Concepts A & B. Concept D appears in Questions 2, 3, and 6, while E appears in Questions 4-6.

Once we had come to a consensus about a Q matrix, which we all believed described the skills we thought were required for students answering these questions, we used the collected data and the "dina" package in the statistical software, R in order to analyze our results.

The following are examples of six actual questions from our Q Matrix:

Q1. Find $p(Z<0.73)$

Q2. Find $p(Z>0.82)$

Q3. X is normally distributed with $\mu=3.2$ and $\sigma = .9$. Calculate the z-score for $x = 3$.

Q4. X is normally distributed with $\mu=3.2$ and $\sigma = .9$. If x has a corresponding z-score of -1.21, what is x?

Q5. If X is normally distributed, and x is greater than 6.3% of the population, what is its z-score?

Q6. If X is normally distributed, and x is less than 33% of the population, what is its z-score?

Q1 and Q2 only require the single skill: transform a Z score to probability. This is a matter of using a standard normal table, equivalent to an Excel function like *=norm.dist()*. In other words, take a statistic like z =2.33 and transform it into a probability: for a variable following a standard normal distribution, there's a 97% chance that $z<2.33$.

Q3 only requires the single skill: standardization. This is what mathematicians sometimes call "plug and chug": substitute values into the standardization formula ($z = \frac{(x-\mu)}{\sigma}$) and calculate the resulting value. This results in a number describing how many standard deviations our value was from the expected mean. In other words, given x = 2, μ = 3 and σ = 4, calculate $z = \frac{(x-\mu)}{\sigma} = \frac{(2-3)}{4} = -0.25$

Table 1: Illustration of a Q-Matrix

	Concept A	Concept B	Concept C	Concept D	Concept E
Question 1	1	1	0	0	0
Question 2	0	0	1	1	0
Question 3	0	0	0	1	1
Question 4	0	0	1	0	1
Question 5	0	1	1	0	1
Question 6	1	1	1	1	1

Q5 and Q6 requires: Probability to z score. Given a probability, like “there is a 97% chance that z is less than a certain value,” we conclude that z must be 2.33.

Note that there are many choices here: we could potentially define many different “skill sets” based on how much detail we want to use to describe the skills required for these tasks. On a very coarse level, all 6 of these could require the skill “use a z table.” On a finer level of analysis, Q1 and Q2 could be seen as requiring different skills (“read a z table to find the area below a z score” vs “read a z table to find the area above a z score”). Finer gradations of analysis are possible, and there seems to be no firm rule in the literature for how exactly to create these rules. Naïvely, at least some of us thought there might be a simple process for breaking down each item into component pieces: but upon reflection, this is far from trivial: we do not even have a clear vocabulary for what are the “atomic” skills in our discipline. Therefore, we need to make some sort of judgment call: what level of abstraction will we set, and how will we formalize our understanding of the student process? The conclusions we came to were necessarily ad hoc and specific to our own discipline and setting. The process, however, might generalize.

How the final analysis is limited by the construction of the Q matrix

It is important to note that the Q matrix shapes and limits all future data analysis in a CDM approach. While there are no “hard and fast” rules for how to construct the Q matrix, variations in how the model is constructed can lead to profoundly different conclusions. It is therefore imperative that researchers carefully review their assumptions about which skills are required for each question. Much of the focus of our current work has been critically evaluating our own Q matrix to see if (for instance) the same students seem to be correctly answering questions the Q matrix interprets as covering the same concepts.

Possibility of generating a Q matrix from the data

It is possible to treat the skills being developed as latent variables, and attempt to generate a Q matrix from the data itself. This process is described in other papers (Chung, 2014) but we did not attempt such a construction in our case.

Results

Interpreting Output from a Q Matrix

The value of a CDM (and therefore the Q matrix) lies in the potentially much greater explanatory power of its output. After running a simulation, the dina package returns a series of explanatory worksheets. The results of the simulation use responses (in our case, students’ test answers) to generate predictions for required constructs (skills in the case of our particular Q matrix). Each student’s likelihood of possessing a given skill is estimated, so we can find how many students possess each combination of skills.

For instance, a student who scored correctly on the first two items but missed the other four might have a high likelihood of understanding how to use a statistical table to find a p-value from a z score, but a low likelihood of having mastered the other two skills: standardizing a sample statistic and calculating a z score from a p-value.

From a pedagogical perspective, this information is more directly useful than a list of which questions are most often missed, because it instead points directly to which underlying skills (or combinations of skills) are lacking. If large clumps of students are missing particular skills, remediation that goes beyond reviewing the problems lots of people missed becomes possible.

Refinement of a test based on a Q Matrix

An intermediate result for explorations with a Q matrix might (as in our case) be anomalous: in particular, a Q matrix offers a natural way to check assumptions about which questions really are equivalent in their difficulty. Two parameters are generated for each question, a “slipping” parameter, which shows how likely people are to miss the question even if they have the requisite skill, and a “guessing” parameter which shows how likely they are to be able to correctly give an answer in spite of not knowing the requisite skills. This may allow improvement of test instruments. If a question is “easy,” in the sense of easily guessed without

proper understanding, this will be identified by a CDM model.
Our analysis led to some surprising results. In an effort to ensure some level of internal validation, we included pairs of questions which (in our opinion) were essentially similar. In particular, we had three pairs of questions where in each case we expected a question and its partner to require the same skills (as codified in our Q matrix) be similarly difficult to guess, and be similarly easy to miss in spite of having mastered the required skills.

What we found was quite the opposite. In addition to seeing that fairly few of our students had the required skills mastered as of the midterm exam, we saw that for one pair of questions one was much easier to guess, which led us to reconsider the multiple-choice answers. For another pair of questions, one was much easier to "slip" on: we were then prompted to see what additional "hidden" skill might be required to understand one of those two questions.

Discussion of these results quickly led to a focus on which skills students had and were lacking, as well as directing our attention to possible issues with consistency amongst questions we had considered "similar." It became clear that some multiple-choice questions were meaningfully different for the students in terms of difficulty, even though they looked the same to the instructors at first glance. This may mean that multiple-choice format questions are less useful to us than free response questions might be.

These hours of discussion were both valuable and difficult, and we (the authors) found them more informative than previous experience we had mapping course outcomes. In particular, the conversations led to a clear focus on student experience and outcomes, and might therefore be helpful in a wide range of learning communities. This endeavor also highlighted some of the limits of multiple-choice formatting, and served as a reminder that questions which seem similar or even identical, to experts may not seem similar to learners. While we have been consistent in format throughout our work for the sake of comparability, these observations have informed our current development work on upcoming courses.

Conclusion

The Q Matrix is a promising and powerful tool for understanding and interpreting student performance on evaluations in an online course. Interpreting "guessing" and "slipping" parameters allow for a nuanced view of students' responses to exam questions. In the context of our own work, we saw that the tool was valuable both for data analysis in the traditional sense and, perhaps more importantly, as a framework for discussion of our methods and assumptions.

References

Chung, M. T. (2014). *Estimating the q-matrix for cognitive diagnosis models in a bayesian framework* (Doctoral dissertation, Columbia University).

George, A. C., & Robitzsch, A. (2015). Cognitive diagnosis models in R: a didactic. *The Quantitative Methods for Psychology, 11*(3), 189-205.

Uchiyama, K. P. & Radin, J. L. (2009). Curriculum mapping in higher education: A vehicle for collaboration, *Innovative Higher Education*, *33*, 271-280.

About the Research Unit at Oregon State Ecampus

Vision

The Ecampus Research Unit supports Oregon State University's mission and vision by conducting world-class research on online education that develops knowledge, serves our students and contributes to the economic, social, cultural and environmental progress of Oregonians, as well as national and international communities of teachers and learners.

Mission

The Ecampus Research Unit (ECRU) makes research actionable through the creation of evidence-based resources related to effective online teaching, learning and program administration toward the fulfillment of the goals of Oregon State's mission. Specifically, the research unit conducts original research, creates and validates instruments, supports full-cycle assessment loops for internal programs, and provides resources to encourage faculty research and external grant applications related to online teaching and learning.

Contact

Katie Linder, Ph.D.
Director of Research
Oregon State Ecampus
541-737-4629
kathryn.linder@oregonstate.edu

Creative Commons License

Suggested Citation

Hsieh, P.-H., Chang, X., & Olstad, A. (2018). Using a Q Matrix to assess students' latent skills in an online course. *White Paper*. Corvallis, OR: Oregon State University Ecampus Research Unit.