# Assessing student learning and improving instruction with transition matrices

Paul J. Walter, and Gary A. Morris

*Mathematics Department, St. Edward's University, 3001 S. Congress Ave., Austin, TX, 78704*

For common multiple-choice assessments, we can investigate progress in student understanding by creating simple transition matrices that identify the percentage of students who select each possible pre-/post-test answer combination on each question of a diagnostic exam such as the Force Concept Inventory. In order to create a transition matrix, we first rank answer choices from worst to best using item response curves. This allows us to determine changes in understanding of concepts and misconceptions even when students do not get the correct answer on the post-test. Our work outlines the utility of transition matrices and the availability of a new tool for adoption by high school and college level physics teachers who use a common assessment. We then construct transition probabilities matrices and find that students who chose an incorrect answer to a question on the pre-test are equally likely to choose the correct answer on the post-test regardless of which wrong answer she or he selected on the pre-test.

## I. INTRODUCTION

This work outlines features of transition matrices and describes the approach of constructing transition matrices for common assessments. We also investigate which students who select incorrect answers are closer to full conceptual understanding. Sec. II describes how transition matrices provide detailed information about the progress students make in the understanding of physics concepts. In Sec. III, we describe a tool for the widely used Force Concept Inventory (FCI) [1]. The tool enables instructors to view the transition matrices based on their own student data and to make changes to improve instruction in an informed way. While our tool and examples relate to the FCI, a similar approach can be applied to any multiple-choice assessment that has functional distractors. Numerous common assessments for physics are listed and available on PhysPort [2]. In Sec. IV we demonstrate for one data set that students who choose better wrong answers are not closer to Newtonian thinking.

## II. TRANSITION MATRICES

For each question on a multiple-choice assessment, a transition matrix displays the percentage of students who selected each combination of pre-test and post-test answers where the sum of the entries is 100%. Figure 1 shows the transition matrix with the second author as the instructor for Question 5 on the FCI using matched pre-/post-test data for 506 students taking a calculus-based pre-medical introductory physics course at Rice University from 2000 – 2003. The answer choices are listed from worst to best where the correct answer is along the bottom row for the pre-test and is along the rightmost column for the post-test. Entries along the diagonal denote students who made no change in the answer they selected. The cells shaded in green show an improvement in student understanding and the cells shaded in red show the percentage of students whose responses have gotten worse.

In this case, Question 5 refers to an object moving in uniform circular motion and students are asked to list the forces present on the object. Consider the cell in the transition matrix where 10.98% of students went from answer choice 5 (E) on the pre-test to answer choice 4 (D) on the post-test. Both answer choices are incorrect since they include a force along the direction of motion that is not present. However, the transition matrix encapsulates that students are improving in this case since answer choice 4 (D) has correctly denoted the direction of the centripetal force being toward the center of the circle instead of outward (as in answer choice 5 (E)). This type of improvement in student understanding and information about what misconceptions still remain is typically missed when instructors only make use of the overall pre-/post-test gains or the Hake normalized gain [3].

Figure 2 shows the overall results of the transition matrix in Fig. 1 for Question 5 on the FCI. While over 55% of students got the question wrong on the post-test, the use of transition matrices permit a more nuanced view of changes in student understanding. For instance, 69% of students either got the question correct or moved to a better wrong answer while only 12% of students moved to a worse answer choice. It can be troubling for the instructor, however, that nearly 19% of students made no change based on instruction.

| Question 5 | | | | | |
|---|---|---|---|---|---|
| | Post | | | | |
| Pre | 1 | 3 | 5 | 4 | 2 |
| 1 | 0.40% | 0.20% | 0.00% | 0.00% | 2.00% |
| 3 | 0.80% | 3.19% | 3.99% | 9.58% | 8.18% |
| 5 | 1.00% | 1.60% | 5.59% | 10.98% | 9.38% |
| 4 | 0.40% | 1.80% | 2.00% | 9.58% | 13.77% |
| 2 | 0.80% | 0.80% | 0.60% | 2.00% | 11.38% |

FIG. 1. A transition matrix for Question 5 on the FCI. Each entry in the matrix denotes the percentage of students who selected a particular combination of answers on the pre-test and post-test. Answer choices are ranked from worst to best, in which correct answer choices are the bottom row on the pre-test and the rightmost column on the post-test.

*2016 PERC Proceedings,* edited by Jones, Ding, and Traxler; Peer-reviewed, doi:10.1119/perc.2016.pr.089

| | | | Raw | Normalized |
|---|---|---|---|---|
| Pre-Right | 15.57% | Right -> Right | 11.38% | 73.08% |
| Pre-Wrong | 84.43% | Wrong -> Right | 33.33% | 39.48% |
| | | Wrong -> Better Wrong | 24.75% | 29.31% |
| Post-Right | 44.71% | Kept Wrong | 18.76% | 22.22% |
| Post-Wrong | 55.29% | Wrong -> Worse Wrong | 7.58% | 8.98% |
| | | Right -> Wrong | 4.19% | 26.92% |

FIG. 2. Summary of the results obtained from the transition matrix from Question 5 on the FCI.

## III. FCI TRANSITION MATRICES TOOL

The FCI Transition Matrices Tool is a macro-enabled Excel spreadsheet that requires entering item-level data that includes every answer choice by each student on both the pre-test and the post-test. If a student did not take one of either the pre-test or the post-test, the data for that student can be left blank. The tool uses matched data where only students who have taken both the pre-test and the post-test contribute to the transition matrices. Along with the information provided by the transition matrices themselves and the summary results as shown in Figs. 1 and 2, the tool also provides feedback on changes in understanding of concepts and remaining misconceptions that are listed in Hestenes et al. (1992) (using the 1995 revised Tables I and II) [1]. The FCI Transition Matrices Tool is openly available for physics instructors on the American Modeling Teachers Association website [4].

### A. Ranking the Answer Choices

Ranking the answer choices and creating a transition matrix allows for instructors to obtain feedback on whether students are improving in their understanding, even if they are still selecting an incorrect answer. Wang and Bao (2010) show that students' level of understanding is highly correlated with their overall scores on the the FCI [5]. Item response curves for each question on the diagnostic assessment are constructed by plotting the percentage of answer choices selected as a function of the overall score of the students, which serves as a proxy for ability level. Our method of ranking answer choices for each question on multiple-choice assessments from worst to best is based on analyzing the item response curves constructed from $> 4500$ student responses for the FCI in Morris et al. (2012) [6].

Fig. 3 shows the item response curves for Question 5 on the FCI. The correct answer choice is $\star = 2$ (B). To determine the next best answer choice, we see that stronger students who score $\sim 18$ on the FCI are most likely to choose answer choice $\triangle = 4$ (D) when they get the question wrong. Thus, we consider answer choice 4 to be the second-best answer. For the third and fourth best answer choices, we choose $\square = 5$ (E) and $\diamond = 3$ (C), respectively. Stronger students who score $\sim 18$ are slightly more likely to select answer choice 5 than 3. Also, students of lower ability level are not drawn to answer choice 5 and are more likely to select answer choice 3. Answer choice $+ = 1$ (A) is hardly selected by students of any ability level and is taken to be the worst answer.

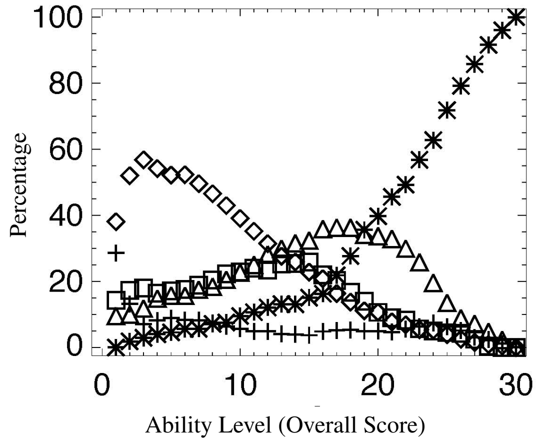


FIG. 3. Item response curves of Question 5 on the FCI. The ordinate is the percentage of students who selected a given answer choice. The abscissa is the overall score of the student, which is taken as a proxy for ability level. Each item response curve corresponds to a particular answer choice: $+ = 1$ (A); $\star = 2$ (B); $\diamond = 3$ (C); $\triangle = 4$ (D); $\square = 5$ (E). Reproduced from Fig. 1 of Morris et al. (2012).

If there is sufficient amount of student data, instructors can generate personalized item response curves and rank the answer choices themselves. While the rankings of answer choices for most questions would likely remain the same as provided in the FCI Transition Matrices Tool, such an approach may lead to some changes for each individual instructor. Making changes to the ranking of answer choices to the FCI Transition Matrices Tool is trivial and only requires reordering the answer choices in the transition matrices.

Item response curves further allow for testing the effectiveness of the assessment instrument [7]. As is often the case, the wrong answer choices contain common misconceptions and are meant to serve as distractors that attract students who lack full understanding. However, some questions have as few as two functioning answer choices that students are likely to select. As shown in Fig. 4 one such example is on Question 4 of the FCI where students almost entirely select one of two answer choices. In Question 4, a car and a truck collide and the question is what object exerts a greater force on the other. Students select either that the truck exerts a larger force or that both objects exert forces of equal magnitude on each other. Other answer choices, such as the car exerting a force of larger magnitude are very rarely chosen by students. In such a case, there is an opportunity to revisit the answer choices that are not functioning as distractors and consider what changes can be made. The three non-functioning answer choices may show how many students are randomly guessing on the test and for our sample show that very few are doing so. Further suggestive that students in this sample are not guessing completely at random, two questions have answer

| Question 4 | | | | | |
|---|---|---|---|---|---|
| Post | | | | | |
| Pre | 4 | 3 | 2 | 1 | 5 |
| 4 | 0.00% | 0.00% | 0.20% | 0.00% | 0.20% |
| 3 | 0.00% | 0.00% | 0.00% | 0.20% | 0.00% |
| 2 | 0.00% | 0.00% | 0.20% | 0.40% | 0.20% |
| 1 | 0.20% | 0.00% | 1.19% | 39.96% | 27.24% |
| 5 | 0.20% | 0.00% | 0.20% | 4.57% | 25.05% |

FIG. 4. Transition matrix for Question 4 on the FCI. From item response curve analysis, answer choices 4, 3, and 2 were indistinguishable.

| Transition Probabilities | | | | | | |
|---|---|---|---|---|---|---|
| Post | | | | | | |
| Pre | 5 | 4 | 3 | 2 | 1 | N |
| 5 | 0.09 | 0.08 | 0.11 | 0.22 | 0.50 | 362 |
| 4 | 0.02 | 0.14 | 0.09 | 0.27 | 0.49 | 958 |
| 3 | 0.02 | 0.04 | 0.21 | 0.25 | 0.48 | 2236 |
| 2 | 0.01 | 0.04 | 0.08 | 0.38 | 0.49 | 5113 |
| 1 | 0.01 | 0.02 | 0.04 | 0.11 | 0.81 | 6512 |

FIG. 5. Overall transition probabilities from the pre-test answer choice to the post-test answer choice averaged over all 30 questions on the FCI. Answer choice 5 generically represents the worst answer choice and answer choice 1 represents the correct answer. Each row sums to 1. The column on the right shows the number of times the particular pre-test answer choice was chosen.

choices that are never selected on the pre-test: answer choice 5 (E) on Question 13 and Question 15.

## IV. TRANSITIONS TOWARD NEWTONIAN THINKING

As students gain understanding of concepts, it stands to reason that they often do so in stages while overcoming a series of misconceptions in succession. Morris et al. (2012) show that the ability level of the students correlates to what misconceptions they have based on evaluating item response curves. [6]. Dedic et al. (2010) employ Latent Markov Chain Modeling to show that students form a hierarchical series of 7 reasoning-groups. The transitions between reasoning-groups typically only occur from a lower (Class 7 being the lowest) to higher (Class 1 representing Netwonian thinking) classes of reasoning-groups [8]. An interesting result from that work was that students in Class 3 were more likely to transition to Newtonian thinking (i.e., Class 1) than students in Class 2 at a rate of $65\%$ for Class 3 students compared a rate of $37\%$ for Class 2 students. Despite being deemed to have a lower level of understanding and the fact that students do not transition from Class 2 to Class 3 (but do transition from Class 3 to Class 2), one can argue that Class 3 students are closer to Newtonian thinking than Class 2 students. In our work, the parallel question we wish to address is whether students who chose better wrong answers on the FCI pre-test are more likely to transition to Newtonian thinking.

### A. Methods

For our sample of 506 students, we convert each transition matrix to a transition probabilities matrix where each matrix row of post-test results is normalized for each each pre-test answer choice—each element in a row is the probability that a given post-test answer choice is chosen for a given pre-test answer choice. For a transition matrix the sum of all of the cells is $100\%$; for a transition probabilities matrix the sum of each row is 1. From the transition probabilities matrices we compare the likelihood of students transitioning to Newtonian thinking (i.e., the correct answer on the post-test) from the best wrong answer (second-best answer choice) versus a worse wrong answer choice. When comparing individual questions, we restrict our study to only those questions that had at least two wrong answer choices that were each selected by $\geq 50$ students on the pre-test. We also average over all 30 questions on the FCI to determine overall transition probabilities.

### B. Results

Of the 30 questions on the FCI, 10 questions did not have at least two wrong answers choices that were selected by 50 or more students on the pre-test. For the other 20 questions, 11 questions had a higher transition probability to the correct answer on the post-test from a worse wrong answer than the best wrong answer (i.e., second-best answer choice); 9 questions had a higher transition probability to the correct answer on the post-test from the best wrong answer than a worse wrong answer.

Averaging the results of all 30 questions on the FCI, Fig. 5 shows the pre-test to post-test transition probabilities from the worst answer choice (i.e., answer choice 5) to the best (correct) answer choice (i.e., answer choice 1). Each row in the transition probabilities matrix sums to 1. While there are fewer students choosing each subsequently worse answer choice, the likelihood that a student that chose a wrong answer on the pre-test will choose the correct answer on the post-test is $\sim 50\%$ regardless of what wrong answer he or she chose on the pre-test.

### C. Discussion

Item response curves in Morris et al. (2012) clearly show that students of different ability levels (based on overall score on the FCI) are likely to chose different wrong answers [6]. Examining those answer choices generally shows students of lower ability level have multiple misconceptions on a given question while students of higher ability level that still got a question wrong have fewer misconceptions. That we find students are equally likely to move to the correct answer regardless of which wrong answer choice was selected on the pre-test is a surprising result. The transition probabilities measure suggests that students who choose better wrong answers are not closer to Newtonian thinking than those who chose worse wrong answers but rather that students who chose incorrect answers on the pre-test are equally distant from Newtonian thinking.

While further analysis is needed, this result suggests that when employing transition matrices there is value in also showing the transition probabilities matrices to garner a more complete picture of the transitions that students are making as a result of instruction. The transition matrices approach differentiates large FCI gains as a result of instruction versus smaller gains (or losses)—in either case the transition matrices would look quite different but the transition probabilities matrices may be approximately the same. In future work we will perform a similar analysis of determining transition probabilities for a larger sample both on the FCI and also on the Science Literacy Concept Inventory [9].

## V. CONCLUSIONS

Transition matrices offer a way of gaining a considerable amount of information about the understanding students have that is often not being utilized. Examining the transitions in understanding that occur as the result of instruction provides further insight into how students learn and opens new avenues for physics education research. The FCI Transition Matrices Tool is freely available for usage by high school and college instructors of introductory physics. We have also found a preliminary and perhaps counterintuitive result that students are no more likely to transition to Newtonian thinking from better wrong answers on the FCI pre-test than worse wrong answers.


[1] D. Hestenes, M. Wells, and G. Swackhamer, “Force Concept Inventory," The Physics Teacher **30**, 141 - 158 (1992).

[2] S. McKagan, PhysPort, (American Association of Physics Teachers, College Park, 2011), <https://www.physport.org>.

[3] R.R. Hake, “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," American Journal of Physics **66** (1), 64-74 (1998).

[4] G.A. Morris, P.J. Walter, S. Skees, and S. Schwartz, FCI Transition Matrices Tool, (American Modeling Teachers Association, 2015), <https://modelinginstruction.org/researchers/evaluation-instruments/fci-transition-matrix-tool/>.

[5] J. Wang and L. Bao, “Analyzing force concept inventory with item response theory," American Journal of Physics **78** (10), 1064-1070 (2010).

[6] G.A. Morris, N. Harshman, L. Branum-Martin, E. Mazur, T. Mzoughi, and S.D. Baker, “An item response curves analysis of the Force Concept Inventory," American Journal of Physics **80** (9), 825-831 (2012).

[7] G.A. Morris, L. Branum-Martin, N. Harshman, S.D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley, “Testing the test: Item response curves and test quality," American Journal of Physics **74** (5), 449-453 (2006).

[8] H. Dedic, S. Rosenfield, and N. Lasry, “Are all wrong FCI answers equivalent?," AIP Conference Proceedings, **1289** (1), 125 (2010).

[9] E.B. Nuhfer, C.B. Cogan, C. Kloock, G.G. Wood, A. Goodman, N.Z. Delgado, C.W. Wheeler. “Using a Concept Inventory to Assess the Reasoning Component of Citizen-Level Science Literacy: Results from a 17,000-Student Study," Journal of Microbiology & Biology Education, **17** (1), 143 (2016).