An Empirical Research on Identifiability and Q-matrix Design for DINA model

Peng Xu
Polytechnique Montreal
peng.xu@polymtl.ca

Michel C. Desmarais
Polytechnique Montreal
michel.desmarais@polymtl.ca

ABSTRACT
In most contexts of student skills assessment, whether the test material is administered by the teacher or within a learning environment, there is a strong incentive to minimize the number of questions or exercises administered in order to get an accurate assessment. This minimization objective can be framed as a Q-matrix design problem: given a set of skills to assess and a fixed number of question items, determine the optimal set of items, out of a potentially large pool, that will yield the most accurate assessment. In recent years, the Q-matrix identifiability under DINA/DINO models has been proposed as a guiding principle for that purpose. We empirically investigate the extent to which identifiability can serve that purpose. Identifiability of Q-matrices is studied throughout a range of conditions in an effort to measure and understand its relation to student skills assessment. We compare identifiability to other Q-matrix design principles through simulation studies of skills assessment with synthetic data. Results show that identifiability is an important factor that determines the capacity of a Q-matrix to lead to accurate skills assessment with the least number of questions.

1. INTRODUCTION
This paper addresses the question of how to design a test that will bring the most accurate assessment of a student's skill mastery state with the least number of questions items. The investigation is framed within the DINA model, where que question items can involve one or more skills, and where all skills are required to succeed the question, but a success can still occur through a guessing factor, and failure can also occur through a slip factor. However, the specific factor we focus on is the *identifiability* of the matrix that corresponds to the mapping of test items to skills.

This mapping of items to skills is generally referred to as a Q-matrix, whrere items are mapped to latent skills whose mastery is deemed necessary order for the student to succeed at the items. An item can represent a question, an exercise, or any task that can have a positive or negative outcome. In the DINA model, the conjunctive version of the Q-matrix is adopted and all of the matrix 1-cells corresponding to a line item are required in order to lead to a success.

Q-matrix based models are widely researched during recent years in educational data mining. Among all of them, DINA model is the most used and discussed. Most of these research are focused on one of the 2 following problems. The first one is the Q-matrix validation problem, that is, to improve or validate an expert-given Q-matrix (de la Torre & Chiu, 2015; Chiu, 2013; Desmarais & Naceur, 2013). The other one is the Q-matrix derivation problem, that is, to directly derive a Q-matrix out of the test result matrix (Barnes, 2010; Liu, Xu, & Ying, 2012; Desmarais, Xu, & Beheshti, 2015; P. Xu & Desmarais, 2016). During the investigation of these two problems, a fundamental question has been proposed, the identifiability of model parameters, especially the Q-matrix. Detailed statistical analysis has been made under the DINA/DINO situation, which was first discussed under the situation that slip and guess is zero(Chiu, Douglas, & Li, 2009), and then the case that slip and guess exist but is known(Liu, Xu, & Ying, 2013), and finally the case that slip and guess is unknown(Chen, Liu, Xu, & Ying, 2015).

However, how do we use the conditions offered by the discussion to guide our educational test? First, the discussion is centered around the identifiability on Q-matrix, not on the identifiability of students, which is also a critical problem (Beck & Chang, 2007). Fortunately, for the case slip and guess are known, the identifiability of the parameters p are also given(Chen et al., 2015).

2. IDENTIFIABILITY AND DINA MODEL
Identifiability is a general concept for statistical models. Its definition is,

Definition (Casella & Berger, 2002) A parameter θ for a family of distribution $f(x|\theta : \theta \in \Theta$ is identifiable if distinct values of θ correspond to distinct pdfs or pmfs. That is, if $\theta \neq \theta'$, then $f(x|\theta)$ is not the same function of x as $f(x|\theta')$.

The general idea behind identifiability is that two or more configurations of model parameters can be considered as equivalent. Sets of parameters will be considered equivalent if, for example, their likelihood are equal given a data set. Or, conversely, if the parameters are part of a generative model, two sets of equivalent parameters would generate data having the same characteristics of interest, in particular equal joint probability distributions (see Doroudi & Brunskill, 2017, for more details).

The issue of identifiability for student skills assessment was first raised for the Bayesian Knowledge Tracing (BKT) model by Beck & Chang, 2007 and later discussed by Doroudi & Brunskill, 2017, and van De Sande, 2013. In this paper, we consider the identifiability of the Q-matrix related model, particularly DINA model, which was studied by G. Xu & Zhang, 2015 Qin et al., 2015.

For DINA model, its model parameters $\theta = \{Q, p, s, g\}$, in which Q is the Q-matrix, of which each row is a q-vector, indicating what skills are needed to answer the corresponding item. p is a the categorical distribution parameter for the whole population. That is, it indicates the probability of a student belongs to each profile category. For example, in a 3-skill case, there are 8 categories for students to belong to, and the 8-component probability vector of student belongs to each of these categories is the model parameter p. At last, s and g are both vectors denoting the slip and guess of each item.

The identifiability of all parameters in DINA model have been carefully researched and several theorems are given(G. Xu & Zhang, 2015). But for Q-matrix design problem researched in this paper, we mainly need to ensure that the model parameter p is identifiable, meaning that we can distinguish different profile categories. Fortunately, for the case when s and g are known, the requirement is easy to be satisfied, since it only requires the Q-matrix to be *complete*.

Definition (G. Xu & Zhang, 2015) The matrix Q is complete meaning that $\{e_i : i = 1, ...k\} \subset R_Q$, where R_Q is the set of row vectors of Q and e_i is a row vector such that the i-th element is one and the rest are zero.

And our whole research is based on the proposition below,

Proposition (G. Xu & Zhang, 2015) Under the DINA and DINO models, with Q, s and g being known, the population proportional parameter p is identifiable if and only if Q is complete.

3. EXPERIMENT

The Q-matrix design problem essentially is an optimization problem. Basically, we have a pool of Q-matrices, and each of them is formed by a selection with replacement from a pool of q-vectors. Each Q-matrix will offer some capacity to diagnose students, and measured by a loss function. What we want to is choose a Q-matrix that minimize the loss function.

In our experiments, to use Q-matrix to diagnose students under DINA model, we follow a Bayesian framework. First, we use one-hot encoding to denote all profile categories. For example, in 3-skill case, all the 8 profile patterns in its original form is,

$$\begin{array}{c c} & \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{array}{c} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

and their one-hot encodings are $p_1 = (1, 0, 0, 0, 0, 0, 0, 0)$, $p_2 = (0, 1, 0, 0, 0, 0, 0, 0)$,...,$p_8 = (0, 0, 0, 0, 0, 0, 0, 1)$. Then, we set the prior of each student profile to be

$p_0 = (1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8)$,

and posterior will be calculated after answering each question. The posterior profile of the student is

$\hat{p} = (m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8)$,

and then we calculate the loss between this posterior and the true profile p_{true}, which is one of the one-hot encoding vector.

In short, for any Q-matrix configuration, the loss function is defined as $loss(Q) = \sum_{students} ||\hat{p} - p_{true}||^2$

In our experiments, we consider 3-skill case and 4-skill case respectively. For 3-skill case, we test with $N = 200$ students, with 25 students falling into each of 8 categories. For 4-skill case, we test with $N = 400$ students, with 25 students falling into each of 16 categories. We use $J4$ to denote the number of questions and k to denote the number of involved skills.

3.1 Experiment 1: Comparison of three strategies

In the first experiment, we compare three different Q-matrix design strategies. They are all based on repetition of a specific pool of q-vectors.

Strategy 1: Using the identifiability condition. Only repeatedly using the vectors $\{e_i : i = 1, ...k\}$.Q-matrix used in this strategy is denoted as Q-matrix 1.

Strategy 2: Using the vectors $\{e_i : i = 1, ...k\}$ plus a all-one vector $(1, 1, 1)$ or $(1, 1, 1, 1)$. This is inspired by orthogonal array design, which is a commonly seen design of experiments(Montgomery, 2017). Q-matrix used in this strategy is denoted as Q-matrix 2.

Strategy 3: Repeatedly using all q-vectors. Q-matrix used in this strategy is denoted as Q-matrix 3.

For the 3-skill case, all these three Q-matrices have been shown below.

Q-matrix 1

$$\begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \\ \cdots \\ q_{19} \\ q_{20} \\ q_{21} \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \cdots & \cdots & \cdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Q-matrix 2

$$\begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ \cdots \\ q_{17} \\ q_{18} \\ q_{19} \\ q_{20} \\ q_{21} \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ \cdots & \cdots & \cdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

Q-matrix 3
(all combinations)

$$\begin{array}{c} \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \\ q_7 \\ \cdots \\ q_{15} \\ q_{16} \\ q_{17} \\ q_{18} \\ q_{19} \\ q_{20} \\ q_{21} \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ \cdots & \cdots & \cdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

The 4-skill case is similar, which is omitted here. Results of these two cases are shown in Figure 1 and Figure 2.

3.2 Experiment 2: Find best configuration

The second experiment takes the brutal force way. We directly examine all possible Q-matrix configurations. First, for a given pool of q-vectors to choose from and an integer indicating the number of questions, we need to know the number of possible configurations of Q-matrices we have. This is equivalent to a classical combinatorial problem, that is, to allocate distinguished balls(q-vectors) to indistinguished cells(questions). It can be easily computed by combinatorial coefficients and interpreted by using stars and bars methods. For example, in 3-skill case, we have 7 q-vectors, and if we have 4 questions to allocate them, then we have $\binom{4+7-1}{7-1} = 210$ possible configurations. This number grows up sharply as number of questions increases or number of patterns increases. To compare, in 4-skill case, and if we have 5 questions to allocate them, then we have $\binom{5+15-1}{15-1} = 11628$ possible configurations.

For each configuration, we will calculate the MAP estimation for all categories of each student, and compare with the one-hot encoding for their true categories. The total loss is reported as the performance index.

We show 6 results graph to offer an overview of this experiment.

3-skill case, 4 questions: Figure 3, Figure 4

3-skill case, 8 questions: Figure 5, Figure 6

4-skill case, 5 questions: Figure 7, Figure 8

4. RESULT

All results are given in form of figures.

5. DISCUSSION

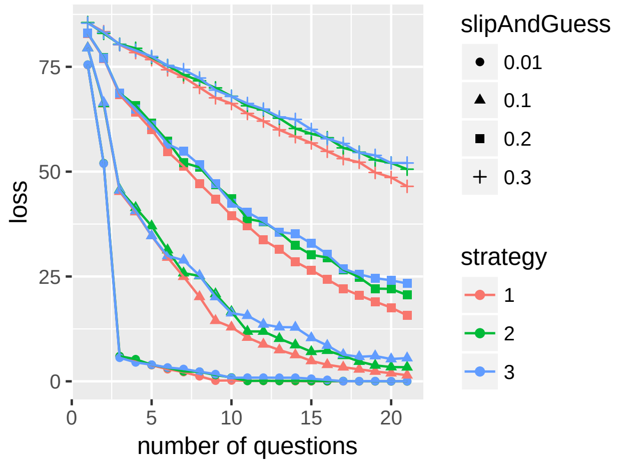

Figure 1: Three Strategy Comparison on 3-skill case

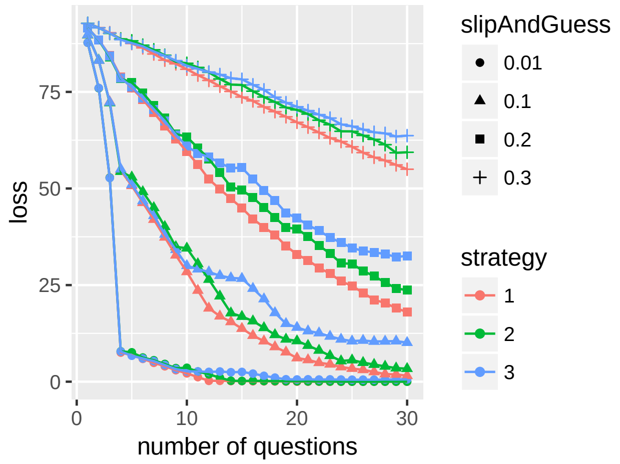

Figure 2: Three Strategy Comparison on 4-skill case

From the result in experiment 1 we can see that strategy 1 always works better than the other two strategies.

From the result in experiment 2, when slip and guess parameters are as low as 0.01, we can see obvious graded patterns among different configurations. This can be explained by the distinguishability of Q-matrix. For example, in Figure 3, we can see there are 7 layers. In fact, the first layer consisted of Q-matrix that can only cluster students into 2 categories. One example of such kind Q-matrix is

$$\begin{array}{c} \\ q_1 \\ q_1 \\ q_1 \\ q_1 \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

This Q-matrix can only tell whether a student mastered skill 1 or not, thus separated students into 2 categories. We know that there are in fact 8 categories of students, the 7 layers in Figure 3 from top to bottom are the Q-matrix that can

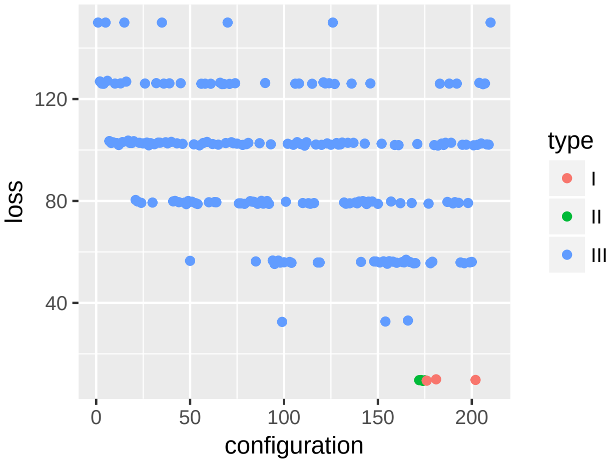

Figure 3: 3-skill case, slip=guess=0.01, J=4

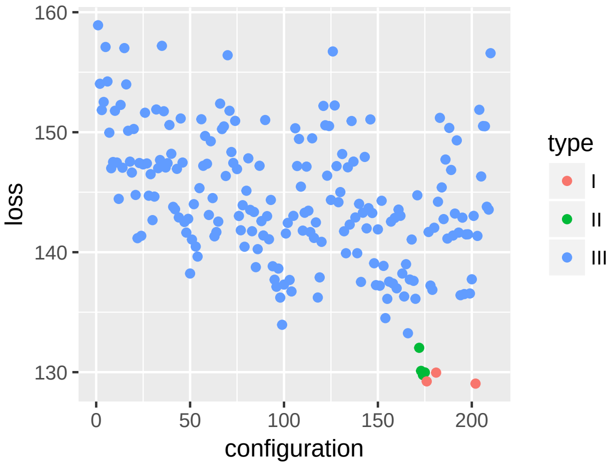

Figure 4: 3-skill case, slip=guess=0.2, J=4

separate students into 2 to 8 categories. We can see that complete Q-matrix always fall in the bottom layer, which agrees with the proposition that we showed in the Section 2. The 4-skill case is similar in Figure 7.

When slip and guess parameter increases, it basically makes the points becomes more divergent which can be easily seen by comparison of Figure 3 and Figure 4. In order to see some more details, we distinguish three types of Q-matrices.

Type I: Complete and confined, meaning it is only consisted of vectors $\{e_i : i = 1, ...k\}$.

Type II: Complete but not confined, meaning it not only contains all vectors $\{e_i : i = 1, ...k\}$, but also contains at least one other q-vector.

Type III: Incomplete Q-matrix.

Type I and Type II Q-matrices performs the same when slip

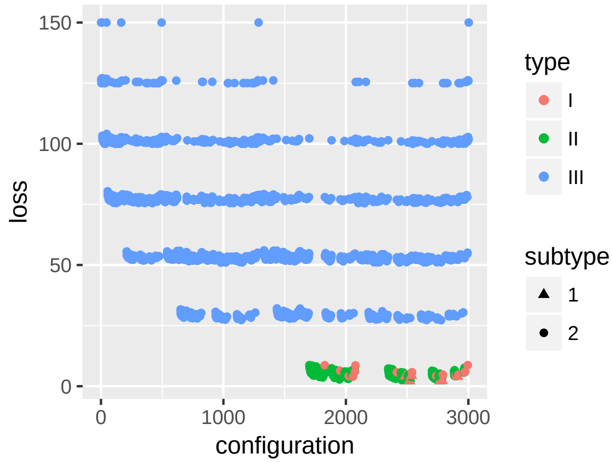

Figure 5: 3-skill case, slip=guess=0.01, J=8

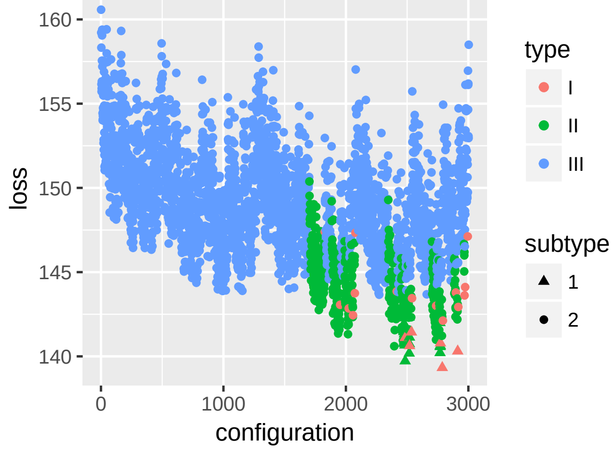

Figure 6: 3-skill case, slip=guess=0.3, J=8

and guess are low(Figure3, Figure 7), but when they get higher, Type II Q-matrices show a better performance(Figure 4, Figure8).

However, when more questions are involved in high slip and guess, the performance becomes more tricky. Therefore, we again consider more subtypes. In 3-skill case for 8 questions, we divide Type I into two subtypes.

Subtype 1: Q-matrix contains each component of $\{e_i : i = 1, ...k\}$ at least twice.

Subtype 2: Other situation(e.g A complete Q-matrix but all the other vectors are just repeated e_1) .

From Figure 6 we can see that the subtype 1(denoted by triangle) shows better performance than subtype 2, meaning that repeating the whole complete Q-matrix is a better strategy just like the strategy 1 we used in experiment 1.

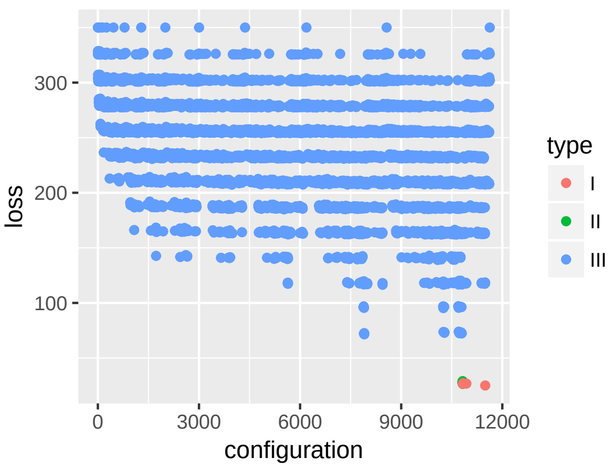

Figure 7: 4-skill case, slip=guess=0.01, J=5

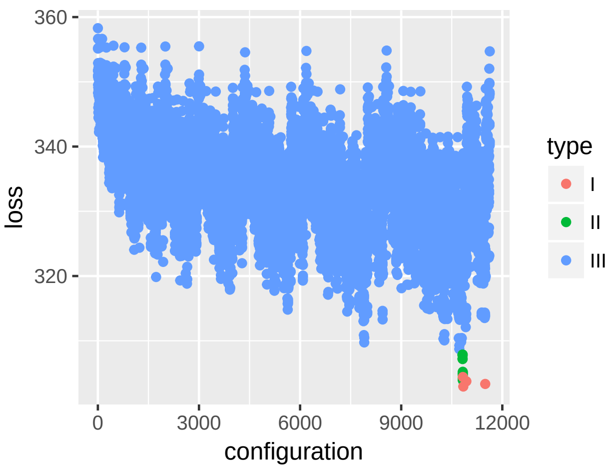

Figure 8: 4-skill case, slip=guess=0.2, J=5

Therefore, we argue that the best Q-matrix design is to use only the vectors $\{e_i : i = 1, ...k\}$ since it offers quicker convergence speed (as shown in experiment 1) and better robustness against slip and guess(as shown both in experiment 1 and 2).

6. FUTURE WORK

More different experiments settings can be considered with different choice on student profiles distribution, and different number of skills involved. Besides, the case that slip and guess are unknown should also be considered, which involves a different identifiability requirement(G. Xu & Zhang, 2015). Moreover, besides the empirical exploration, rigorous mathematical discussion can be conducted for the best Q-matrix design under DINA model.

References

Barnes, T. (2010). Novel derivation and application of skill matrices: The Q-matrix method. *Handbook on educational data mining*, 159–172.

Beck, J. E., & Chang, K.-m. (2007). Identifiability: A fundamental problem of student modeling. In *International conference on user modeling* (pp. 137–146).

Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2). Duxbury Pacific Grove, CA.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866.

Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, *37*(8), 598–618.

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*(4), 633.

de la Torre, J., & Chiu, C.-Y. (2015). A general method of empirical Q-matrix validation. *Psychometrika*, 1–21.

Desmarais, M. C., & Naceur, R. (2013). A matrix factorization method for mapping items to skills and for enhancing expert-based q-matrices. In *Artificial intelligence in education* (pp. 441–450).

Desmarais, M. C., Xu, P., & Beheshti, B. (2015). Combining techniques to refine item to skills q-matrices with a partition tree. In *Educational data mining 2015.*

Doroudi, S., & Brunskill, E. (2017). *The misidentified identifiability problem of bayesian knowledge tracing.* International Conference on Educational Data Mining, EDM2017.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied psychological measurement*, *36*(7), 548–564.

Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, *19*(5A), 1790.

Montgomery, D. C. (2017). *Design and analysis of experiments.* John wiley & sons.

Qin, C., Zhang, L., Qiu, D., Huang, L., Geng, T., Jiang, H., ... Zhou, J. (2015). Model identification and Q-matrix incremental inference in cognitive diagnosis. *Knowledge-Based Systems*, *86*, 66–76.

van De Sande, B. (2013). Properties of the bayesian knowledge tracing model. *Journal of Educational Data Mining*, *5*(2), 1–10.

Xu, G., & Zhang, S. (2015). Identifiability of diagnostic classification models. *Psychometrika*, 1–25.

Xu, P., & Desmarais, M. (2016). Boosted decision tree for Q-matrix refinement. In *Edm* (pp. 551–555).