

Abstract Title Page

Title: Incorporating Learning into the Cognitive Assessment Framework

Authors and Affiliations: Cassandra Studer¹, Brian Junker¹, Helen Chan²

¹ Carnegie Mellon University Department of Statistics

² Carnegie Mellon University Department of Modern Languages

This work was supported in part by the Program for Interdisciplinary Education Research, Carnegie Mellon University, under Institute for Education Sciences, Department of Education, Grant R305B040063.

Abstract Body

Background / Context:

Haertel et al. (2012) recently reviewed the importance of knowing learning trajectories as a means for educators to make better “instructional decisions regarding the sequencing of ... skills.” They point out that knowing intermediate proficiency levels and how students are likely to pass through them would be useful to educators and yet few of these learning trajectories have been defined. This is largely due to the fact that there are few dynamic models that education researchers can use to track student trajectories.

We introduce a model to incorporate learning over multiple time points into the cognitive assessment framework. In order to extend this framework to be dynamic, we expand the static likelihood function to account for time. We then introduce the Parameter Driven Process for Change + Cognitive Diagnosis Model (PDPC + CDM; Studer, 2012) which can be used as part of this framework. We implement the model on both simulated and real data sets. In particular, we apply PDPC + CDM to a data set with a pre and post test. The intervention exposes native Chinese speakers who are learning English to a cognitive tutor aimed to teach English article rules (Chan, 2012). In this scenario, the model fits well and we are able to detect high rates of learning where 92% of students transitioned to a higher proficiency state.

Purpose / Objective / Research Question / Focus of Study:

Our goal is to employ dynamic cognitive diagnosis models (CDM; Junker and Sijtsma, 2001; Rupp and Templin, 2008) to account for learning. Currently, repeated measures data in the education field is typically modeled using statistical techniques like a paired t-test, repeated measures ANOVA or ANCOVA, or hierarchical linear models. However, these techniques suffer two drawbacks. First, the dependent variable in each of these techniques is a sum score. Therefore, they do not consider item properties meaning that we lose information if students do not answer the same items as one another and at each time point. The second drawback is that these techniques tell us about learning on average but not for an individual student. These concerns have been addressed in the static case using CDM. Previous attempts to create dynamic cognitive assessment models are either highly specific to the situation for which they were created (e.g. Anderson et al., 1995) or continue to model learning at the group as opposed to the individual level (e.g. Cen et al., 2006; Rijmen et al., 2005; 2008; Cho et al., 2010). We introduce PDPC + CDM, a model that extends static CDM to account for multiple time points in a way that is generalizable and tracks individual student learning.

Setting:

The English Article experiment was conducted in March of 2011 at a university in Beijing, China that specializes in foreign language education and research.

Population / Participants / Subjects:

The subjects were 64 native Chinese speakers in their first or second years as English majors. Students had studied English an average of 8.5 years and were 18.5 years old on average.

Intervention / Program / Practice:

The cognitive tutor is a web-based program with 650 cloze task items where students choose the appropriate article from choices 'a'/'an', 'the', or \emptyset , the zero article. The rules for choosing the

most appropriate article were split into 23 skills. Skills 1 – 10 correspond to \emptyset , skills 11 and 12 correspond to 'a'/'an', and skills 13 – 23 correspond to 'the'. Each skill was covered in twenty-five items in the tutor except for one skill corresponding to 'the' which appeared in 100 items. Students could not advance until they correctly answered each item. Immediate feedback was given regardless of whether the input answer was correct. No hints were available. Students worked in the tutor over two 60 minute sessions with two days between sessions.

Research Design:

To assess performance, Chan (2012) administered a pretest immediately before the first tutor session and a post test immediately following the second. Both assessments were administered online over thirty minutes. Each item was dependent on a single skill and each skill was covered by sixteen items - eight at both pre and post test for a total of 184 items at each test. Because of the time constraint, students did not see all items. At pretest, they saw an average of 142.4 items, with standard deviation, $sd = 10.2$ items. They correctly answered an average of 67.2% ($sd = 7.2\%$). At post test, students saw an average of 120.7 items ($sd = 11.1$ items). They correctly answered 85.6% ($sd = 7.8\%$) on average. We used a paired t-test to find that the percentage correct is significantly higher on the post test with an average gain of 18.5% ($t = 21.9, p < 0.001$). Therefore, we expect to find learning gains when we apply PDPC + CDM.

Statistical, Measurement, or Econometric Model:

Our goal is to measure learning of these 23 article skills. To do this, we let $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$ be the complete response pattern for student i where $X_{it} = (X_{it1}, X_{it2}, \dots, X_{itJ})$ is the response vector at time t of the J items. Items are graded dichotomously so that X_{itj} equals one if student i correctly answers item j at time t and zero otherwise. We assume that $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iT})$ is a vector of latent student features and $z_i = (z_{i1}, z_{i2}, \dots, z_{iT})$ is a vector of unobserved states to describe each student's status at each time point. In general, this latent state can be identical to θ or, as in PDPC + CDM, an indicator of membership in latent states that describe the θ distribution.

(please insert figure 1 here)

Regardless of the definition, we assume that the relationships between student responses and latent states can be described by the Attributes Assessment Model (Junker, 1999), a directed acyclic graph (Wasserman, 2004) depicted in Figure 1. The conditional independences inherent in a DAG allow us to assume that observations at one time point are independent of the next given a student's latent state, i.e. $X_{it} \perp X_{it-1} / z_{it}$, that observations are independent of latent states given the student parameter, i.e. $X_{it} \perp z_{it-1} / \theta_{it}$, and the Markov property, i.e. $z_{it+1} \perp (z_{it-1}, \dots, z_{i1}) / z_{it}$. Using these assumptions, we define a general marginal distribution which can be used as a base to all dynamic models including PDPC + CDM. The likelihood is as follows:

$$P(X_i, z_i) = P(z_{i1})P(X_{i1}|z_{i1}, \beta) \prod_{t=2}^T P(z_{it}|z_{it-1})P(X_{it}|z_{it}, \beta) \quad (1)$$

In PDPC + CDM, we assume students are in one of S latent states at each time point. These latent states describe groups of students with similar response and/or skill patterns. We then allow students to transition between states according to a time homogenous hidden Markov model. Then, $P(z_{i1})$ is the initial state probability and $P(z_{it}|z_{it-1})$ is the transition probability in a Markov chain. To define $P(X_{i1}|z_{i1}, \beta)$, we must first introduce CDM in more depth.

In CDM, we assume that $\theta_{it} = (\theta_{it1}, \theta_{it2}, \dots, \theta_{itK})$ is a skill vector where θ_{itk} equals one if student i possesses skill k at time t and zero otherwise. The distribution of θ_{it} is dependent on z_{it} ,

student i 's latent state at time t . Specifically, we assume $\theta_{it}|z_{it} = z \sim \prod_{k=1}^K \text{Bernoulli}(p_{kz})$ where p_{kz} is the probability of knowing skill k in latent state z . These probabilities define what it means to be in each latent state and transition between them.

In Equation 1, we can define $P(X_{it}|z_{it}, \beta) = \sum_{\theta} \prod_{j=1}^J P(X_{itj}|\theta_{it}, \beta_j) \prod_{k=1}^K P(\theta_{itk}|z_{it})$. We derive this equation using the following facts. First, $P(X_{it}|z_{it}, \beta) = \sum_{\theta} P(X_{it}|\theta_{it}, \beta)P(\theta_{it}|z_{it})$ where we use the law of total probability to expand the responses to be dependent on student skills, θ . By the local independence assumption, we can expand $P(X_{it}|\theta_{it}, \beta) = \prod_{j=1}^J P(X_{itj}|\theta_{it}, \beta_j)$. Finally, we assume that the K skills are independent of one another meaning that $P(\theta_{it}|z_{it}) = \prod_{k=1}^K P(\theta_{itk}|z_{it})$.

The probability of a successful response $P(X_{itj} = 1|\theta_{it}, \beta_j)$ can then be defined as any static CDM (e.g. the NIDA, DINA, or RUM models described in Junker and Sijtsma, 2001). For the article data set, we assume the NIDA model (Maris, 1999; Junker and Sijtsma, 2001) where $P(X_{itj} = 1|\theta_{it}, \beta_k) = \prod_{k=1}^K ((1 - s_k)^{\theta_{itk}} g_k^{1-\theta_{itk}})^{q_{jk}}$. In this equation, q_{jk} is expert defined to equal one if item j depends on skill k and zero otherwise. Then the item feature, $\beta_k = (s_k, g_k)$, is a skill, as opposed to an item, parameter and is therefore subscripted by k . The slip probability, $s_k = P(X_{ij} = 0 | \theta_{ik} = 1)$, is the probability that a student incorrectly applies skill k even if he knows it. The guess probability, $g_k = P(X_{ij} = 1 | \theta_{ik} = 0)$, is the probability that a student correctly applies skill k even though he does not know it.

With PDPC + CDM, we can estimate each student's trajectory through the latent state space, z_i , and the probability of knowing each skill in each latent state, p_{kz} ; these probabilities define what it means to be in each latent state. In particular, an education researcher could use this information to see which skills each student is likely to know at each time point.

Usefulness / Applicability of Method:

Adding a dynamic component to CDM has the potential to benefit many in education research. In particular, it will allow researchers to assess whether an intervention is effective and compare experimental conditions to find which most promote learning (Feng et al., 2009). In addition, researchers could use this model to define learning trajectories which could better inform curriculum development (Haertel, 2012). It will allow educators to better focus their teaching by making explicit topics which have been learned by the majority and those upon which the teacher should dedicate more time (Anozie and Junker, 2007). Finally, it will allow schools to make better predictions about performance on end of year accountability exams which in turn will allow them to better prepare for the high stakes assessments (Ayers and Junker, 2008)

Data Collection and Analysis:

Using WinBUGS (Lunn et al., 2000), we simulated 500 values for each parameter which we use to calculate the median and 95% credible intervals.

(please insert figure 2 here)

Figure 2 is a summary of the p_{kz} estimates, the probability of knowing skill k in latent state z . We see that all skills except 7, 16, and 20 have significantly higher probabilities in latent state 2. In fact, 17 of the state 2 probabilities are estimated to be exactly one. Because so many of the probabilities in state 2 are significantly higher than state 1, $p_{k2} > p_{k1}$, we can think of students in latent state 2 as being more proficient than those in latent state 1. In a simulation built to mirror the article data set, we correctly estimated 43 of the 46 (93.5%) p_{kz} parameters within measurement error; this is similar to the 95% we would expect.

We then consider the trajectory of student i through the latent state space, z_i . In the article data set, we found that all 64 students were estimated to be in latent state 1 at the pretest and 59 (92.2%) transitioned to latent state 2 at the post test. This indicates that the tutor was effective at teaching Chinese speakers to use English articles. In the mirrored simulation, we were able to correctly estimate 62 of the 64 (96.9%) trajectories.

As a measure of confidence about the latent state assignments, we looked at the posterior probability of the estimated state. We found that 121/128 (94.5%) of the student time points had posterior probability equal to one. Five of the seven (71.4%) cases where the probability was less than one were still greater than 0.95. In the simulated data set, no incorrectly classified cases had posterior probability this high. One student had a particularly low posterior probability less than 0.6. He was found to fit a profile different from the two latent states we estimated for the other students. In particular, at the post test, this student answered the items depending on half of the skills perfectly and the other half at only 50%. This type of information would likely have been missed using traditional analyses.

Findings / Results:

After applying PDPC + CDM to the article data set, we find that students make large proficiency gains from pre to post test. Therefore, we conclude that the cognitive tutor is a highly effective tool for teaching English articles to Chinese speakers. In addition, we can specifically say that 59 students transition to a more proficient state where 20 of 23 skills have a significantly higher probability of being known ($p_{k2} > p_{k1}$) and 17 are perfectly known ($p_{k2} = 1$). Knowing the probabilities and latent state memberships allow us to say which skills a student is likely to know. This type of analysis could be particularly interesting for researchers as it is more informative than traditional methods which do not tell us about individual learning or the particular skills a student is likely to know. With this type of knowledge, the researcher could assess whether the 6 skills that are not estimated to be perfectly known at the post test could be improved upon in the tutor. A teacher could use this information to realize that the majority of students have learned proper usage of English articles and move to other topics while possibly providing extra support to the 5 students still struggling.

Conclusions:

We aimed to incorporate learning into the cognitive assessment framework that exists for static assessment data. In order to accomplish this, we derive a common likelihood function for dynamic models and introduce PDPC + CDM, a dynamic model which tracks learning indirectly through student membership in latent states which drive the distributions of the student parameter in the static portion of the model. We described this model both theoretically and empirically through application to the article data set (Chan, 2012). One limitation of this data set is that the items are single skill. In order to truly test PDPC + CDM, we need to find data that have items with multiple skills.

In general, by adding a dynamic component to the cognitive assessment framework, we provide education researchers with a method to track individual student learning while taking item and skill features into consideration. In addition, one could use a model such as this to define learning trajectories which could lead to better instructional methods and sequences (Haertel, 2012). Teachers could also use this information to better focus their lessons. One goal for the future would be to make these models accessible to researchers and teachers who can use the results to further student learning and the field of education research.

Appendices

Appendix A. References

- Anderson, J., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167–207.
- Anozie, N. & Junker, B. (2007). Investigating the utility of a conjunctive model in Q matrix assessment using monthly student records in an online tutoring system. *Proposal submitted to the National Council on Measurement in Education 2007 meeting*.
- Ayers, E. & Junker, B. (2008). IRT Modeling of Tutor Performance To Predict End-of year Exam Scores. *Educational and Psychological Measurement*, 68, 972–987.
- Cen, H., Koedinger, K., & Junker, B. (2006). *Lecture Notes in Computer Science: Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement*. Berlin: Springer.
- Chan, H. (2012). Explicitness, Cue Competition, and Knowledge Tracing: A Tutorial System for Second Language Learning of English Article Usage. *Department of Modern Languages, Carnegie Mellon University, Unpublished dissertation*.
- Cho, S., Cohen, A., Kim, S., & Bottge, B. (2010). Latent Transition Analysis with a Mixture Item Response Theory Measurement Model. *Applied Psychological Measurement*, 34(7), 483–504.
- Feng, M., Heffernan, N., & Koedinger, K. (2006). Addressing the Testing Challenge with a Web-Based E-Assessment System that Tutors as it Assesses. *Association for Computing Machinery*, 34(7), 307–316.
- Haertel, E., et al. (2012). NAEP: Looking Ahead—Leading Assessment into the Future. A *whitepaper commissioned by the National Center for Education Statistics*, US Department of Education, Washington DC.
- Junker, B. (1999). Some statistical models and computational methods that may be useful for cognitively relevant assessment. *Prepared for the Committee on the Foundations of Assessment, National Research Council*.
- Junker, B. & Sijtsma, K. (2001). Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory. *Applied Psychological Measurement*, 25(3), 258–272.
- Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 197–212.
- Rijmen, F., DeBoeck, P., & VanderMaas, H. (2005). An IRT Model with a Parameter Driven Process for Change. *Psychometrika*, 70(4), 651–669.
- Rijmen, F., Vansteelandt, K., & DeBoeck, P. (2008). Latent Class Models for Diary Method Data: Parameter Estimation by Local Computations. *Psychometrika*, 73(2), 167–182.
- Rupp, A. & Templin, J. (2008). Unique characteristics of diagnostic models: a review of the current state-of-the-art. *Measurement*, 6, 219–262.
- Studer, C. (2012). Incorporating Learning into the Cognitive Assessment Framework. *Department of Statistics, Carnegie Mellon University, Unpublished dissertation*.
- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer Texts in Statistics.

Appendix B. Tables and Figures

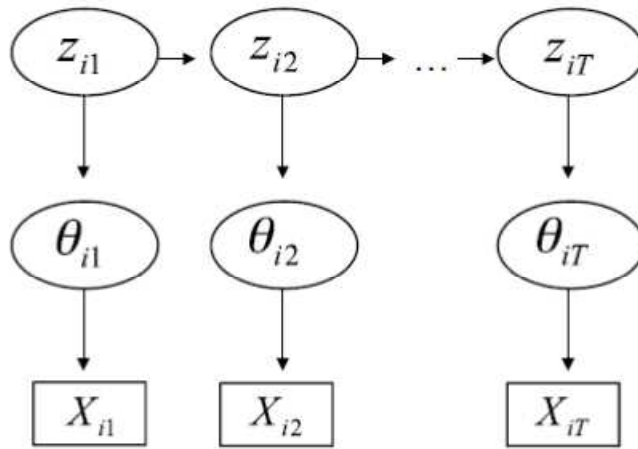


Figure 1

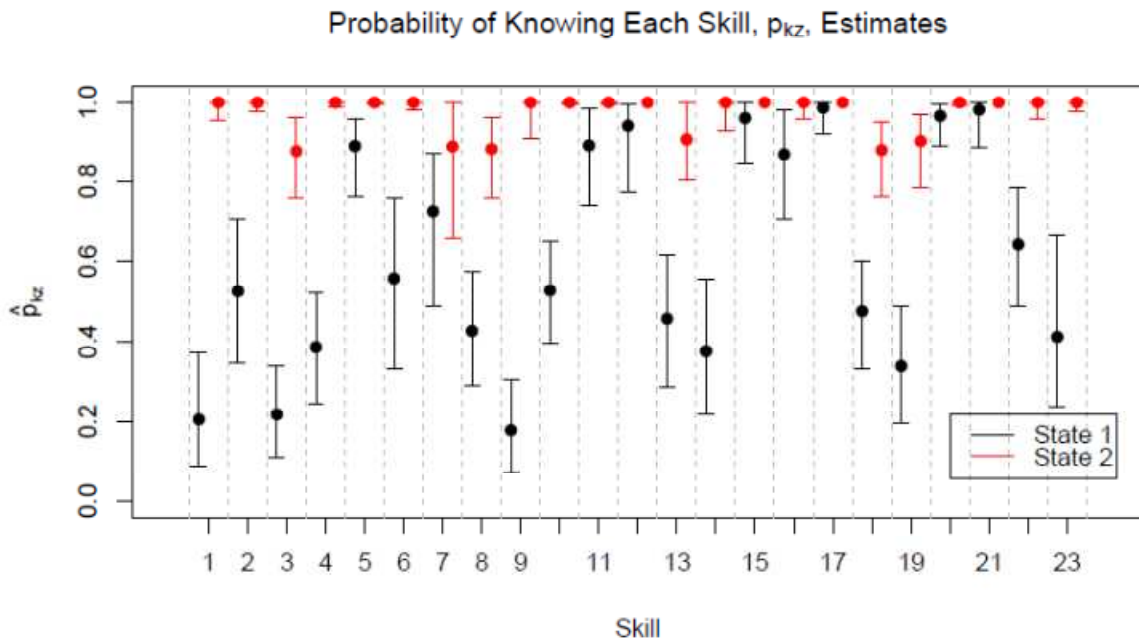


Figure 2