# Automated Performance Prediction for Personalised Learning

**Samara Asam Banno**

Supervisor: Dr Linda Yang

School of Energy and Electronic Engineering
University of Portsmouth
United Kingdom
December 2018

This thesis is submitted in partial fulfilment of the requirements for the reward of the degree of Doctor of Philosophy of the University of Portsmouth

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

# ACKNOWLEDGEMENTS

This dissertation would have never been completed without the guidance of my advisors and committee members, help from friends, and support from my family.

I would like to express my deepest gratitude to my study director, Dr Linda Yang, for her excellent guidance, encouragement, and providing me with an excellent atmosphere for doing research. I would like to thank Dr David Ndzi and Dr Shikun Zhou for their services as my annual reviewer panels and helping me to develop the scope of my research. Also fully thanks to Dr Micheal Yudelson at Carnegie Mellon University in California-USA for providing the valuable data to be used in this study.

I am indebted to my friends and the members of the School of Engineering, especially from Room A2.09 .

This dissertation is made possible with the undivided support and encouragement from my wonderful mother and my beloved husband, Rafid Denha who both have helped me and stood by my side through the highs and lows.

I would finally like to place on record my sense of gratitude, to all those who directly or indirectly, have lent their helping hand in this venture.

# Abstract

Designing an Intelligent Tutoring System (ITS) that simulates human learning with regard to different knowledge levels is a challenge. Most developed ITSs typically focus on either the building or the testing phase without paying appropriate consideration to the design phase. The result is that these systems offer specific choices in isolation, and can thus prove difficult to apply in situations where multiple factors interact. Researchers have therefore developed tutoring models that did include the design phase in their implementation. However these models do not consider the cognitive factors of students. Such ITS models lack the ability to provide accurate estimations as they do not analyse the students' individual skills against the item skills, particularly when the learning items require multiple skills, which thus reduce student's learning efficiency due to an incomplete representation of the student's knowledge.

This thesis proposes a novel tutoring model, called 'Cognitive Factor Analysis' (CFA), which adapts student cognitive factors, such as guessing/slipping parameters and student proficiency levels, together with each item's parameters to produce a better estimation of student latent performance.. CFA also introduces the concept of the Q-matrix from psychometrics and connects this to the students' prior scores. The model does not only predict the students' performance, but helps students to target the strengths and weaknesses in their knowledge levels. Therefore, CFA has an adverse impact on the student's learning curve and reduce the student's learning time by controlling the amount of time spent practicing the skill several times. It assumes the role of modelling the student's learning by making inferences about their latent performance with multiple skills assessments.

CFA also contributes to cognitive learning psychology by exploring how computational models can be used to understand human behaviour. It shows

how data generated from tutoring systems can be analysed and modelled to create and improve a unified computational theory of human learning. Furthermore, it encapsulates psychological findings in a format that can be used by instructional designers and educational scientists to support the development of tutoring systems.

# Table of Contents

# List of Tables:

# List of Figures:

# Chapter One: Introduction

## 1.1 Motivation

Intelligent tutoring systems (ITSs) are computer-based instructional systems with models of instructional content that specify what to teach, in addition to teaching strategies that specify how to teach. ITSs assess a student's knowledge systematically and present curricula compiled according to their individual skill levels while simultaneously generating appropriate feedback for both student and teacher. The systems make certain inferences related to a student's mastery of specific topics or tasks in order to dynamically adapt the content or style of instruction. Overall, these intelligent tutoring systems have been shown to improve student learning across multiple domains (Koedinger & Anderson, 1997; Ritter et al., 2007).

ITSs aim to engage the students in sustained reasoning activity and to interact with the student on the basis of possessing a deep understanding of the student's cognitive behaviour. Content models (or knowledge bases, expert systems or simulations) give ITSs certain depth such that students can 'learn by doing' in realistic and meaningful contexts (MacLellan, Liu & Koedinger, 2015). However, an incorrect determination of the domain skills may lead to inappropriate learning recommendations and subsequently negatively affect the student's motivation and thus waste their limited learning time (Vanlehn, 2011). With this, building and designing tutoring systems that are pedagogically effective is both difficult and expensive (Murray, 2005).

# Introduction

In an ideal world, the implementation of an intelligent tutor would be based on three phases: design, building, and testing. Each phase requires time, expertise and resources for its appropriate execution, which, in general, makes tutor development a cost prohibitive endeavour (Murray, 1999). As a result, many tools have been created by researchers to support the tutor development process. Unfortunately, these tools provide little insight for the design phase, focusing primarily on supporting the building and testing phases. In other words, while they provide certain capabilities for the interfaces and user profiles, they give little support in terms of what factors need to be considered to estimate the students' performances, or whether the authored content matches their performances and is effective for learning (i.e. the design phase). Furthermore, they do not provide an accurate estimation for the learner's knowledge level due to a lack in the design (Aleven, McLaren, Sewall, & Koedinger, 2006; Sottilare & Holden, 2013; MacLellan, Liu & Koedinger, 2015).

In terms of developing technology to support the design phase, designers must first address two high-level questions: how to have an accurate estimation of the students' performance; and how to present the learning material that matches the students' skill levels. Thus, various models have been used to support this phase, including cognitive task analysis techniques (Clark et al., 2006), learning factor analysis (LFA), and the deterministic inputs noisy 'and' gate model (DINA) to analyse the skills per item and match them with the students' knowledge levels. These models have managed to estimate the students' performances based on the prior scores achieved through answering items while doing assessments. However, there are pitfalls to using intuition to guide tutor design. For example, Clark et al. (2008) argued that much of an expert's knowledge is tacit since although as individuals gain expertise their performance improves, it is difficult to estimate their current performance accurately, especially when there are certain other factors affecting performance estimation (e.g. in the case of multiple choice (MC) assessment, did the student achieve the expertise level through mastering his/her own skills or by guessing the answers or using hints?). Such factors are defined in intelligent tutoring systems as 'cognitive factors'.

Cognitive factors relate to the mechanisms that control human thought, including, for example, language processing, analytical reasoning, learning, and memory. Therefore, an important class of intelligent tutors are those that include cognitive models. Such tutors, often known as 'model tracing tutors' (Anderson & Pelletier, 1991), observe student behaviour and build a fine-grained cognitive model of the student's knowledge, which can be compared with the expert model. Such cognitive models analyse the students' strengths and weaknesses in addition to providing individualised instructions. These models have a relatively deep level of expertise, and thus the student, when stuck, can ask the tutor for a hint to identify the solution to the entire problem. This process will enable the tutor to understand the student's knowledge level and estimate the performance accurately. Furthermore, this will allow teachers to devote more one-on-one time to each student and to work with students of varying abilities simultaneously in addition to designing assignments that match an individual student's needs, thereby enhancing their learning development.

The process of evaluating the student knowledge level during the assessment that contains any form of help from the tutor, such as hints, is similar to a type of educational assessment that is known as 'dynamic assessment' (Kalyuga and Sweller, 2005).

Luckily, design is an iterative process, and poor design decisions can be identified and corrected through testing and redesign. The current best practice for testing a tutor design and improving its pedagogical effectiveness is to conduct a close-the-loop study (Koedinger et al., 2013). This involves creating an initial tutor, deploying it in a classroom, analysing the data from this deployment and using the findings to redesign the tutor before finally deploying it again to test the effectiveness of the redesign. The close-the-loop approach is essentially a data-driven cognitive task analysis technique that guides the redesign of both the material being taught (e.g. the underlying model of skills necessary for domain expertise) and how the material is being presented (e.g. emphasising certain steps in the interface or providing more practice in certain skills).

For their part, Koedinger et al. (2013) applied this approach to a geometry tutor and showed that students achieved mastery more rapidly with the redesigned tutor. While the approach, much like other similar approaches, such as A/B testing (Lomas et al., 2013), is effective for testing and improving initial tutor designs, it is expensive. Classroom studies take time to arrange and run, and often, multiple classroom study iterations are required to achieve a good tutor design.

Given these current practices, the existing tutoring models generally provide less accurate assessments of a student's performance and reduce their learning efficiency due to an incomplete representation of their knowledge.

With this in mind, designing an effective and accurate ITS model can be achieved by implementing the following points:

- Improving the performance estimation accuracy. This can be achieved by including the cognitive factors in terms of analysing the students' skills and comparing them to the given item skills. This process ensures determining the guessing parameters if the student answers correctly without having acquired the necessary skills, or the slipping parameters in cases where the student attempts the item incorrectly even though he or she *has* mastered the required skills.

Therefore, the performance estimation will not only depend on the correct/incorrect scores but will also determine the skills for both the students and the items.

- Determining whether the learning materials are effective for the students. The intelligent model can determine whether the learning materials are effective for the learning process by matching the student's skills with the item skills. With this, the number of practices can be reduced or increased depending on the student's skill levels until he or she reaches the mastery level.

Based on what has been discussed above, the hypothesis of this research is as follows:

Using the cognitive factors (the student's current skills, the guessing/slipping parameters) will improve the prediction of the student's performance, and optimise their learning experience.

## 1.2 Objectives

This work proposes a novel, artificial intelligence model we have termed 'cognitive factor analysis' (CFA). The model targets students in terms of estimating their latent performance. It aims to produce an accurate prediction of the student's ability to answer the next item correctly. CFA analyses their performance level effectively by using multiple skills assessments and adapting the cognitive factors in both problem-solving and learning. In addition, it supports both the design phase and the testing phase of tutor development because it can simulate student interaction with initial tutor prototypes. Finally, it can also produce an accurate estimation of each student's current skills and determine the student's ability to answer the next item correctly.

CFA aims to support the development of any intelligent tutoring model as the data generated from these simulations has an identical format to the data generated from real classroom studies, meaning instructional designers can analyse it using existing tools and techniques. For example, designers might apply learning curve analysis techniques, such as 'additive factors modeling' (Cen, 2009), to gain insights into which skills will be more difficult for each students to learn, such that they can create additional hints that exercise challenging skills (through the given information on whether each student has slipped/guessed the answer). Similarly, a designer might analyse simulated data to test the overall effectiveness of a specific tutor design, or to compare alternative designs to determine which are the most effective for learning. Thus, the developed model can act as a tool to leverage prior learning theory to cost-effectively test and improve any initial tutor designs prior to actual classroom deployment.

Overall, CFA has three main objectives:

1. Precisely identifying the students' weaknesses and strengths to ensure saving their study time.

2. Improving the students' performance estimations by including the cognitive factors.

3. Supporting the building and testing phases of the intelligent tutoring systems by focusing on producing a proper design phase without the need for classroom study.

## 1.3 Thesis Contributions

First of all, the CFA model contributes to the design phase of intelligent tutoring systems by introducing a novel approach that looks at the problem from a very different perspective. It doesn't aim to only predict the student's performance, but also to help the students target their strengths and weaknesses within their knowledge level. By including the cognitive factors (the student's current skills, the slipping and guessing parameters) along with any prior student behaviours, the accuracy of the student's performance estimation will be increased, and, in detecting the knowledge level accurately, the student will both save their learning time and have a better learning curve representation. In essence, it is a tool that aims to take on the role of modelling the student's learning by making inferences about his or her latent performance within multiple skills assessments. Here, CFA doesn't consider that a correct answer from the student is positive evidence of their mastering all the required skills since the student may have simply guessed the answer. Similarly, an incorrect answer is not taken as proof of failure as the student may have slipped up on the answer despite the fact that they had mastered all the required skills. This will have an adverse impact on the student's learning curve and will not waste the student's time in terms of practicing the skill numerous times.

Finally, CFA contributes to the field of cognitive learning psychology by exploring how computational models can be used to understand human behaviour. This is because CFA focuses on modelling the students learning behaviour and understands their needs in improving their knowledge level. This modelling can be achieved first by determining whether the learning material matches the students' skills and second, by estimating the future performance for each student. With this, CFA attempts to show how data generated from the tutoring system can be analysed and modelled to create (and improve) a unified computational theory of human learning and to encapsulate psychological findings in a format that can be used by instructional designers and learning scientists to support tutoring system development.

## 1.4 Overview of thesis contents

**Chapter Two:**

This chapter establishes the technical background and models related to this thesis. It begins by explaining the concept of self-regulated learning and how it can be applied to educational assessment before discussing the theory of dynamic assessment in the field of cognitive psychology (psychometrics). The way in which the assessment item should be presented is shown in the form of multiple choice (MC) assessment. The subsequent sections then discuss the cognitive diagnosis models, which have been used in the area of education data mining area. This includes highlighting the scoring procedure that involves acquiring essential cognitive factors, such as the item difficulty, the student skills and the guessing/slipping parameters and describing the approximation methods used to estimate the required probability of knowledge level and learning.

**Chapter Three:**

This chapter demonstrates the research setup. The proposed model structure is described, and each building block is explained. It also addresses the challenges in improving the student's learning skills by looking at two essential concepts: the probability of identifying the correct answer through a reliance on the prior used hints; and adapting the student's item-skills with the guessing/slipping parameters. The chapter also presents two novel models for estimating student performance, which are derived from two cognitive approaches: the PFA and the DINA models. The first model extends the previous work of PFA by splitting the success rates into ‘prior correct answer’ and ‘prior used hints’. Based on the first model and the consideration of multiple prior student factors, the chapter goes on to present a second model to estimate the probability of identifying the correct answer by adding the slipping/guessing parameters in the form of a logistic regression.

**Chapter Four:**

The first part of this chapter focuses on implementing and evaluating the version of PFA that includes the hints factor. In fact, several experiments have been previously conducted to show the comparisons between the original PFA and our modified version. The subsequent sections then demonstrate the DINA model that includes the same dataset used in PFA in addition to providing a full explanation of the findings related to the obtained results. Finally, our developed model is applied to the dataset and all the factors are extracted and analysed using a novel modification of logistic regression that allows for taking into account any situations resulting in false negative student actions (e.g. slipping/guessing with known/unknown skills). Later in this chapter, CFA defines the student’s learning curve type and determines any factors affecting learning performance. Such factors can be related to the quality of learning, such as learning materials and hints associated with each item/skills, and the student’s expertise level (whether he or she maintains the skill). Meanwhile, the system can potentially save the student’s time as well as recommend certain learning resources that could help improve their current skills by estimating whether they are more likely to make another mistake or answer the next item involving the same skills correctly.

Finally here, how the findings show the significance of the model in relation to the other models is discussed.

**Chapter Five:**

Finally, this chapter presents the conclusions, including the main research contributions, and provides suggestions for future work.

# Chapter Two: Background and Literature Review

## 2.1 Introduction

This chapter establishes the technical background and models related to this thesis. It begins by discussing certain prior work which has been conducted to support ITS. It then explains the concept of cognitive models and how this theory can be applied to educational assessment. The theory of dynamic assessment is discussed in the field of cognitive psychology (psychometrics). Furthermore, a highlight of the process of the scoring procedure is presented which includes essential cognitive factors such as the item difficulty, student's skills and the guessing/slipping parameter and also how the Q matrix is applied. Finally, it describes the approximation methods used to estimate the required probability of knowledge level and learning.

## 2.2 Prior work

Following the brief explanation offered in chapter one regarding the ways in which the design of ITS models could be enhanced by including cognitive factors and improving student performance estimation, this section turns to the literature and review of several prior models that conform to this notion. While there is a vast number of modelling efforts that fit at least some aspect of the current problem, reviewing all of these works is beyond the scope of this thesis. Instead, the literature focuses on highlighting certain models which are directly related to the impact of adopting cognitive models within the design of ITS.

The predominant view within the literature is that estimating the future performance of students consists of heuristic cognitive attributes through a problem-solving strategy (Newell & Simon, 1972), which is typically defined by:

- Tracking student knowledge, i.e. analysing the student's current skills compared to the given item skills to determine student mastery during the practice of the problem.
- Predicting student behaviours within the ITS, such as student performance on the next practice opportunity.

Given this view, prior work characterises ITS as the translation of the latent attributes of students and feedback regarding this with regard to students' behaviour and how to improve this. Although most work can be unified under this characterisation, prior work differs on the cognitive factors and the types of assessment used in the evaluation of students. For example, Feng et al. (2009) employed an item response theory within ITS to estimate student performance in an end-of-year test. This model estimates the student's latent skill based on the number of practices that student might need to reach the correct answer, while ICARUS (Li et al., 2009) is capable of learning by filling in the gaps using means-ends problem-solving. Further, other problem-solving approaches (Anderson, 1993; Fikes et al., 1972; Laird et al., 1987) enable systems to generate their own problem-solving traces. These systems have succeeded in estimating student knowledge, however there is one issue which restricts any system from being the perfect model, namely the way of presenting the items' assessments and the amount of time spent in solving these items. This is because, in ITS, timing is considered to be the most essential factor in assessment which can be affected by the type of assessment and the instructions that might be given to the students while taking the assessment. This subsequently influences the estimation process of knowledge. In particular, if students are well instructed and make good choices regarding the answers, then they will quickly achieve their goals. However, when students are less well informed and fail to make appropriate choices, they will spend more time exploring unyielding areas of the problem. To combat this problem, prior work explores how to leverage correct and incorrect answers by applying the concept of multiple-choice (MC) assessments.

The MC-type item format is easy to administer and inexpensive to score (whether manually or through the use of automated computer systems). However, with MC items, examinees can gain scores by chance through successful guessing when they do not have the required skills, or lose scores by slipping the correct answer even though they possess the appropriate skills to answer correctly. This could seriously negatively affect test validity and reliability because it would introduce another source of measurement error (Han, 2012). Therefore, competent work has adopted MC and attempted to overcome the likelihood of slipping and guessing. For example, Junker and Sijtsma (2001) introduced the concept of guess and slip into the DINA model to improve the accuracy of the model. Pardos et al. (2010) integrated individualisation into knowledge tracing and showed a reliable improvement in the prediction of real-world data (Pardos & Heffernan, 2010). Pavlik et al. (2009) presented a new alternative student model – Performance Factor Analysis (PFA) – which considers the success and failure scores in order to improve the learning performance of students, and found this to be somewhat superior to knowledge tracing. Moreover, there is also a great deal of work which focuses much effort on examining and improving student performance estimations with MC ( Gong, Beck & Heffernan, 2010;

Pardos & Heffernan, 2010). However, without considering the slipping/guessing parameters, they instead focus on prior student scores.

Together, these systems each represent major progress towards predicting human behaviours in the contexts of ITS. However, it is difficult to achieve fair and reasonable evaluations across various models and approaches for a number of reasons. Different studies employ different data sets and report different measurement metrics which makes it difficult to compare the models. Even if the above differences were to be avoided, it is still likely that the studies would use slightly different procedures despite adopting the same model. Additionally, the generality of these previous systems remains unclear. In particular, how much prior knowledge needs to be authored to employ these systems in new domains? For example, CASCADE and STEPS focus mainly on physics, but do these models also explain behaviour in the domains of language or engineering? If so, how much additional prior knowledge must a researcher (or perhaps a teacher) develop in order to apply them? SIMSTUDENT and DINA models have been applied to multiple tutoring domains and, consequently, might have a better claim to generality. However, each SIMSTUDENT and DINA model possesses a specialised set of domain-specific prior knowledge, suggesting that a user would probably need to author additional knowledge. Given this limitation, this research aims to develop a model that will be feasible, generic and practical for application by researchers and teachers by considering multiple cognitive factors with MC assessment, extracting students' skills and matching these with the given item skills to enable better student performance estimations for all types of learning.

The next sections will present more detailed explanations regarding certain significant models which have been used in ITS, analyse their effectiveness and provide a comparative study with respect to their applications.

## 2.3 Cognitive factors

The importance of cognitive factors in education was introduced, developed and thrived during the 1980s and 1990s. The great assumption involved in introducing cognitive factors to student learning is that the way in which humans learn can be modelled as a computational process (Nkambou, Bourdeau et al. 2010). In particular, educational technologies, such as Intelligent Tutoring Systems, can use cognitive factors to assign each student with practice problems that target their specific weaknesses/strengths so that they do not waste time practicing skills they already know. Studies have shown that students can double their learning in the same time (Pane et al., 2014) or learn more in approximately half the time (Lovett et al., 2008) when cognitive factors are considered within the ITS.

Therefore, adopting cognitive factors in tutoring models has a significant impact on student performance estimation. There are several models that have been developed utilising student cognitive factors such as Model Tracing (MT) which is grounded in cognitive psychology based on the ACT-R (Adaptive Control of Thought-Rational) cognitive theory. The belief is that human learning processes can be modelled by some form of structure which describes how a task is procedurally accomplished. The technique is closely related to domain modelling and expert systems. In the Model Tracing framework, student actions are monitored and the solving rules are formed to the path through the problem space (Anderson and Reiser 1985). Although this model has produced effective steps towards a student's learning strategy, nevertheless it still lacks the ability to predict the latent performance of the learner.

Therefore, this model was improved upon and is now known as Knowledge Tracing (KT) which was introduced in (Corbett and Anderson 1994). The model takes the form of the Hidden Markov Model whereby student knowledge is a hidden variable and student performance is an observed variable. The model assumes a causal relationship between student knowledge and student performance; i.e. the correctness of a question is (probabilistically) determined by student knowledge. There are four parameters estimated by the model, 1) prior knowledge, which is the probability that a particular skill was known by the student before interacting with the tutoring system; 2) the learning rate, which is the probability that a student's knowledge transits from an unlearned to a learned state after each learning opportunity; 3) guess, which is the probability that a student can answer correctly even if he/she does not possess the skill required to solve the problem; 4) slip, which is the probability that a student responds to a question incorrectly even if he/she possesses the required skills. Classic KT has been used broadly and successfully across a range of academic domains. However, this model lacks the ability to handle multiple skill problems. The KT model is designed per skill; if to solve a problem requires multiple skills, this raises difficulties in deciding to which specific skill this particular observation should belong.

Furthermore, the Additive Factors Models (AFM) is a generalised linear mixed model which applies a logistic regression to fit a learning curve to student performance data . The central idea of AFM was originally proposed by Draney, Pirolli et al. (1995) and was introduced into the field of ITS by Cen, Koedinger et al. (2006). The model was renamed as the Learning Factors Analysis (LFA) model by these authors.

The model has a logit value representing the accumulated learning for a student using single or multiple skills. The model captures the ability of the student and the easiness of the required skills. It also considers the benefit gained from prior practice by estimating the amount of learning on the skills for each practice opportunity.

Learning Decomposition (LD), a variant of a learning curve, estimates the relative worth of different types of learning opportunities. The approach is a generalisation of learning curve analysis and uses non-linear regression to determine how to weight different types of practice opportunities relative to each other (Beck 2006). Unlike AFM, LD selects the form of exponential curves over the power curves. The model has a free parameter to represent how well students perform on their first trial performing the skill, and a set of free parameters representing how quickly students learn the skill by performing a particular type of practice, such as reading the same story repeatedly or reading a new story. However, neither LD nor LFA consider the prior unsuccessful scores or the prior used hints. Instead, they focus only on the positive responses.

The Performance Factors Analysis model (PFA), which is a reconfiguration of LFA, was presented by Pavlik et al. (2009). PFA disregards the student ability parameter of LFA and thus allows PFA to generalise across different subjects. Depending on the reconfiguration of the difficulty parameter, PFA is able to capture the problem difficulty. Aside from the learning rate parameter, similar to LFA, the model also estimates an additional parameter for each skill reflecting the effect of prior unsuccessful practices.

The Instructional Factors Analysis model (IFA) was presented by Chi et al. (2011) who tailored PFA for their specific needs. This model, other than tracking the effect of prior successful practices, i.e. learning rate, and the effect of prior unsuccessful practices, estimates the effect of an additional variable, what is referred to as ‘tells’, a form of instruction without yielding a correct or incorrect answer.

Finally, there are the DINA and NIDA models which were introduced by Junker and Sijtsma (2001) and Maris (1999), respectively. These are probabilistic models which consider the slipping/guessing parameters in order to produce an estimation of the student responding to the item correctly. These models work on multiple skills assessments, however they are conjugative models that require all the necessary skills to be mastered in order for the examinee to have a high probability of responding correctly, regardless of the student's proficiency and learning rates.

Based on the above literature, developing an accurate performance estimation model is challenging for the following reasons:

- Each learning subject may involve multiple skills which are not explicitly stated in textbooks and students are expected to acquire such skills through problem-solving. Therefore, mastering the skill can only be achieved by the student's performance of tasks that require such skills.
- Most of the cognitive models have been developed by experts. Many previous studies in cognitive psychology have shown that experts often make false predictions about what causes difficulty for students due to 'expert blind spots' regarding the skills that novices need in order to reach aptitude in a particular domain (Pane et al., 2014). This will be reflected in an incorrect estimation of the student's knowledge level.

Previous studies have shown that, due to an incomplete representation of a student's knowledge, the human engineering of these models often ignores distinctions in student skills and learning contents (Koedinger and Nathan 2004; Koedinger and Mclaughlin 2010). This absence will have a negative impact on the performance estimation and will reduce the efficiency of the ITS. Therefore, improving the existing cognitive models has an immediate and significant impact on a student's learning and has a long-term impact on ITS design. The challenge is how to achieve a better intelligent tutoring model taking into account cognitive factors.

**There are several ways to achieve this:**

- Choose the type of assessment items which combine the cognitive skills of the student with the item skills and help to develop learning styles.
- The weaknesses and strengths of a student need to be carefully diagnosed towards a mastery level of a subject by that student (unlike many standard ITS models that aim to produce a prediction at the end).
- The assessment items need to determine the item skill labels in order to be able to author targeted learning items and hint messages. Therefore, there was a necessity to use a specific assessment model designed to be integrated with the student cognitive load. One such type of assessment is the dynamic assessment (Kalyuga, & Sweller, 2005).

The next section will explain the concept of dynamic assessment, how it works and its impact on ITS as well as analysing certain work which has adopted dynamic assessment within ITS.

## 2.4 Dynamic Assessment

The development of dynamic assessment was greatly influenced by L.S. Vygotsky who established the role of social context in children's learning and development and the ways by which to improve their performance level using the assistance of an adult. By proposing the theory of the Zone of Proximal Development (ZPD), which describes the difference between the performance of a learner achieved with and without adult guidance, a greater number of opportunities have been provided to interact with more competent peers and adults such as teachers (Vygotsky, 1980).

According to Sternberg (2002), there are commonly two formats of dynamic assessment: the sandwich format and the cake format. Both formats are presented in the form of 'test–teach–retest'. Sandwich format dynamic assessment means that teaching is held between the pre-test and the post-test, thus constituting a sandwich-like process. In the cake format dynamic assessment, the teaching is a response to the examinee's answers to each item. The main difference between the two dynamic assessment formats is that instruction and assessment are separate in the sandwich format dynamic assessment whereas they are combined in cake format dynamic assessment.

An example of combining dynamic assessment to web-based learning is provided in a study by Wang (2010) which combines the idea of cake format dynamic assessment and the 'graduated prompt approach' (Campione and Brown, 1987) to develop a multiple choice web-based dynamic assessment system. This research treats assessment as a teaching and learning strategy and the main feature is the design of a successive and appropriate series of hints provided with each incorrect answer. It starts with 'general hints' with little specific information regarding the solution, gradually becoming 'specific hints' that offer 'detailed explanations from which learners can generate the correct answer'. This study was developed by GPAM-WATA and the hints (referred to as instructional prompts (IPs)) are provided by its dynamic assessment items. Although this model managed to improve the learning progress of students through learning from instructional hints/assistance to achieve the correct answer, it still did not provide enough information about the examination item and how many skills were delivered to each student. Also, there is an absence in determining the current skills of students and whether these match the given item skills since this may lead to wasting time by students responding to skills which had already been previously mastered.

This issue has been addressed by another study entitled 'The Computerized Ecology Observation Competence Assessment (CEOCA)' which was developed to assess the ecology observation competence of students and adapted the Concept-Map Integrated Dynamic assessment system. The inclusion of concept mapping in learning activities summarises the students' understanding after studying instructional material. Therefore, it provides appropriate ways to improve students' knowledge growth by considering the cognitive load for each individual. The test items are presented with real pictures, films or concept maps. The CEOCA consisted of three facets: knowledge, observation and conceptual relationships. The concept map was integrated into this system to scaffold the structure of students' knowledge. The learning tasks were transferred by wireless communication to students in the field using handheld PDAs. The students' responses to tasks and their observation records were also transmitted back to the learning system. The PDAs functioned as portable notebooks and walking encyclopaedias within the learning system. The learning system also provided guided tasks, immediate feedback, e-library search functions for mangrove wetland ecological systems and an e-diary editor (Pi-Hsia, Hwang, I-Hsiang, 2012). This technology successfully managed to determine the students' skills and matched them with the test items. It did, however, ignore certain other cognitive factors. For example, it did not consider whether the correct answer provided by the students was based on a hint or was derived from their current skills and this may lead the student to depend on guessing the answer rather than using their own skills to successfully answer the item. Therefore, the determination of the student's performance is not as accurate as it should be.

Although these new technologies seem to show promise, researchers have pointed out that students' learning achievements could be disappointing in the absence of effective learning strategies or tools to engage them in improving their knowledge structure (Chu, Hwang, Tsai, & Tseng, 2010), which is considered as an important component of understanding students' ability and their knowledge level (Novak, 1990). In order to identify the advanced scientific ability of various individuals, students need to be able to restructure their prior knowledge which is based on everyday experience and lay culture.

Therefore, our approach aims to extend the 'Graduated Prompting Assessment Module' of the WATA system (GPAM-WATA) by integrating this with an obtained personalised student performance estimation. According to Poehner (2008), the 'graduated prompt approach' (presented in GPAM-WATA) emphasises that, when examinees have difficulty in solving problems, examiners would help them through mediation (hints). Therefore, the interaction between examiners and examinees can help examinees to discover or to apply certain principles to independently solve problems and to learn more. Hence, it is expected that instructional prompts may play the role of a teacher in educating and guiding learners, depending on their previous knowledge, towards an expert performance level.

Furthermore, the developed model determines the current skills of students and matches these with the item skills. This allows the system to present only the assessment items that the student needs in order to reach the mastery level; this approach reduces learning time and facilitates an effective learning progress. Moreover, the system estimates the future performance of the students and the slipping and guessing parameters based on prior scores (these parameters will be explained in the following sections).

The teaching process will be performed by proposing the appropriate learning material depending on the student knowledge level at each stage. The post dynamic test will be performed once again to determine their new level; hence, the student profile will continue to dynamically update the student's new performance state.

To implement dynamic assessment with cognitive factors and ITS, certain factors need to be considered. These points are explained in the following section:

### 2.4.1 Psychometrics in Assessments

Psychometrics represents a branch of statistics that is dedicated to psychological assessment which identifies models with cognitive factors. The aim of a psychological assessment is 'to educate and improve student performance and not merely to audit it'' (Wiggins, 1998, p. xi) and it should be used not only to ascertain the status of learning but also to further learning (Stiggins, 2002). This type of assessment is similar in its principles to dynamic assessment. However, most of the effective work outlined in the previous section regarding dynamic assessment lacks the adoption of cognitive factors and hence does not determine an accurate estimation. As Shepard (2001) noted, assessments typically applied to schools do not provide diagnostic information about individual students in order to develop their learning skills (i.e. these assessments ignore the concept of cognitive factors). These kinds of assessments are based on a measurement model called the item response theory model (or IRT model) which is used for building tests from items and scoring procedures regardless of the skills required for each item for each item. Although scores from these assessments have proved useful in determining the single latent proficiency for each student, these assessments/scores do not permit the evaluation of a student's specific strengths and weaknesses necessary for targeted instruction. They therefore cannot be utilised as a feedback mechanism which would allow teachers to identify effective learning materials that could help to improve a student's learning.

In contrast to the traditional models (i.e. IRT), attempts to provide more informative scores have been made, such as 'Knowing What Students Know' (NRC, 2001), which discuss certain measurement models that integrate the advances in cognitive and psychometric theories and facilitate inferences more relevant to learning. These psychometric models are known as Cognitive Diagnosis Models (CDMs) and can be used to adapt the skills, cognitive processes and problem-solving strategies involved in an assessment. These strategies are usually referred to as Q matrices.

The CDMs are discrete latent variable models developed specifically for diagnosing the presence or absence of multiple skills or the steps required for solving problems in a test. With these models, specific skills or processes that the students have and which they have not mastered can be registered in the student profiles. Such information can be used to direct resources and tailor instruction to optimise student learning.

To illustrate the fundamental difference between IRT and CDM (psychometric) frameworks, consider the fraction-subtraction task: $2\frac{4}{7} - \frac{7}{12}$. By applying the IRT model (which will be explained later in this chapter), the performance of the task might be described as a function of global-subtraction proficiency and students with higher proficiencies are expected to have higher probabilities of answering the item correctly. On the contrary, a CDM might describe the performance as a function of the skills/attributes listed in Table 2.1 which are based on those identified by Mislevy (1995) and Tatsuoka (1990) using cognitive theory and analysis, as the way a student population of interest solves this type of problem. Effective performance of the task requires a series of successful implementations of the relevant attributes. The model might also describe the implications of the absence of one or more of the required attributes. Thus, by adopting cognitive structures in the psychometric model, assessments developed and analysed using CDMs provide information that is more prescriptive, richer and more relevant to the improvement of learning processes.

**Table2. 1**: Attributes in Fraction-Subtraction

| Attributes | Distribution |
|---|---|
| 1 | Borrow one from the whole number to a fraction |
| 2 | Basic fraction subtraction |
| 3 | Reduce/Simplify |
| 4 | Separate whole number from a fraction |
| 5 | Convert the whole number to a fraction |

Based on the above, there are two points that need to be considered:

- Firstly, since the concept of psychometrics or CDM aims to identify the cognitive factors that will be used to later predict students' performance, integrating these into the design phase of ITS has a significant impact. This integration will offer more prescriptive information regarding the student to be used towards improving their learning process. Such information includes determining a student's skills and comparing these to the item skills, so the model is able to detect the missing skills that the student needs in order to improve.
- The presentation of assessment items needs to be identified. There are several types of item presentations, such as full questions, filling in blanks and multiple choice (MC). However, this research is based on MC assessments. The next section explains the importance of and the reasons for including MC assessment items in the design of the proposed model.

### 2.4.2 Multiple Choices in Dynamic Assessment

Although not without its share of criticism, the multiple choice (MC) format has been widely used in educational assessments due to its ability to sample and accommodate diverse contents (Nitko, 2001; Osterlind, 1998). The field of educational testing has developed a successful implementation of many test item formats including short answer and multiple choice questions. For several decades, the multiple choice (MC) type item format has been easy to administer and inexpensive to score (whether manually or using automated computer systems). Unlike other item formats, the MC item scoring process does not involve raters, so there is therefore no rater effect and that might lead to one less source of measurement error. The most straightforward and common way of analysing MC responses with cognitive diagnosis modelling is to treat them as dichotomous data (i.e. the scores are either 0 or 1). For instance, this approach has been adapted by de la Torre (2006) to analyse the National Assessment of Educational Progress (NAEP; 2003) data. Moreover, Birenbaum, Tatsuoka, and Xin (2005) and Tatsuoka, Corter, and Tatsuoka (2004) employed MC to analyse Trends in International Mathematics and Science Study (TIMSS; 2003).

However, such an approach is suboptimal because it does not take into consideration the diagnostic insights regarding student difficulties and alternative conceptions that can be found in the distractors (Haertel & Wiley, 1993; Nitko, 2001; Sadler, 1998). These difficulties are defined by guessing/slipping parameters. With MC items, examinees can gain scores by chance through successful guessing without acquiring the necessary knowledge or, alternatively, they may slip the correct choice despite having acquired the skills to answer it. This could seriously negatively affect test validity and reliability because it would introduce another source of measurement error (Han, 2012).

The developed model of this research will adapt the MC assessments within its design. This adaption is based on the following points:

- Due to the significant impact of using MC in educational assessment (as explained earlier), most ITS models prefer to use this type of assessment, not only because of its ability to stimulate students' active and self-managed learning, but also because of its ability to provide students with rapid feedback regarding their learning. This will contribute towards a reduction in students' study time and help to develop rapid self-learning development.
- From the teaching aspect, using MC assessments allows scoring to be carried out by anyone, or even automatically, thereby increasing efficiency, particularly when teaching large cohorts. Further, with the ITS, the tutor is helped in the design assessment items through the use of quiz tool software, either within or independently of Learning Management Systems (e.g. Moodle). With such software, the tutor facilitates quiz administration, scoring and feedback provision.
- MC assessments help to measure the student's cognitive factors effectively. Each option provided with an item includes certain skills; these skills differ in their complexity and only one option contains the appropriate skills required to answer the item. Therefore, the student will try to choose the option which may match his/her skills regardless of whether it is the right or the wrong answer. As a result, the system will detect the student's strengths and weaknesses by comparing the student's current skills with the correct option skills.

Since the assessment items used to develop the model of this thesis are based on the MC type (due to its significant impact in cognitive assessments as explained earlier), there is a necessity to manage the problem of guessing/slipping parameters and fix the model towards a more accurate estimation of the student's performance. The next section explains the guessing parameters and the effect on student performance.

### 2.4.3 Guessing Parameter with Multiple Choices

In educational multiple-choice testing, guessing is supposed to occur when a test taker does not absolutely know the correct response but still tries to achieve a correct answer. Several ways can be found to address the process for problem-solving and guessing by considering the question, whether the *guessing process* (GP) comes before or after the *problem-solving process* (PSP) (San Martin et al., 2006). There are two kinds of guesses: random and logical. A random guess is based completely at random and not on any other information whereas a logical guess is built on several processes of problem-solving, none of which alone or together are sufficient to lead directly to a correct response. In consideration of both random and logical guesses as outcomes of the guessing process, several works have tried to parameterise and interpret the guessing process in IRT models (Han, 2012).

Therefore, test developers have attempted to eliminate the guessing of an answer by imposing special testing policies such as giving partial points to omitted items, assigning penalties to unsuccessful guesses and/or increasing incorrect item options attractive to low-proficiency examinees. However, it is relatively difficult to completely prevent examinees from obtaining points through successful chance guesses on MC items.

Further, when the guessing parameter is discussed, the slipping parameters should also be considered, as the student might slip or attempt the wrong answer/option even if they have acquired the necessary skills. This might affect the overall performance estimation.

Therefore, the developed model of this thesis has adopted the guessing parameters in its implementation; the model is able to work out whether the student has guessed the answer. The model compares the student's current skills against the given item skills. Therefore, if the student attempts to answer correctly without having mastered the required skills, the parameter of guessing will be included. Likewise, if the student has mastered the skills and incorrectly attempts the answer, the slipping parameter will be considered. Including such parameters within the developed model will increase the accuracy probability of a student's performance estimation and thus improve their learning process.

## 2.5 Q Matrix with multiple choices items

The Q-matrix is a Boolean matrix describing the relationship between items and skills (Tatsuoka, 1983). A cell value of 1 at row i, column j means that the item i requires the use of skill j. A cell value of 0 means otherwise. Table 2.2 shows such a relationship between two testing items and four associated skills. Notice the first item requires only one skill and the second item requires two skills simultaneously.

**Table2. 2**: Sample Q Matrix

| Item/Skill | Add | Sub | Mul | Div |
|---|---|---|---|---|
| 2*8 | 0 | 0 | 1 | 0 |
| 2*8-3 | 0 | 1 | 1 | 0 |

To formulate the Q-Matrix for MC assessment with different skills items, an example of four coded options is given in Figure 1; the skill specifications for these options are given in Table 2.3. By defining the skills and determining the appropriate tasks, options can be developed to be more relevant and informative. For this example, the different options for this item were constructed in consultation with an experienced mathematics educator according to the MC-DINA model (Junker, & Sijtsma, 2001). The option with the largest number of required attributes (i.e. D) is the correct option/key. As can be seen from this example, in addition to the key, some options are also coded under the framework of the Q matrix.

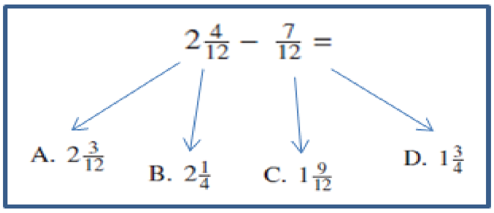



**Figure 2. 1**: A Fraction-Subtraction Item

**Table2. 3**: Q matrix of single item

| Option | Skill 1 | Skill 2 | Skill 3 |
|---|---|---|---|
| A | | ✓ | |
| B | | ✓ | ✓ |
| C | ✓ | ✓ | |
| D | ✓ | ✓ | ✓ |

The options/choices reflect the knowledge scale to obtain the proficiency level of an examinee. The Bear Assessment System (Wilson & Sloane, 2000) utilises construct maps to create the options where the mapping to proficiency levels is determined a priori. That is, the knowledge states represented by the options should be in the subset of the knowledge state that corresponds to the correct choice. In the example above, option C was written to reflect mastery of the skills of borrowing from the whole and basic fraction subtraction but not that of reduction/simplification. By presenting the options in this manner, examinees with the same given skills have a greater ability to choose specific options. Q-matrices for the different coded options are combined in the modified Q-matrix given in Table 2.4. The entry in each cell indicates the number of times the skill is specified in the options. For example, the modified Q-vector for item 10, [1 1 0], indicates that the correct option requires skills 1 and 2.

**Table2. 4**: Q Matrix for items-skills

| Item | Skill 1 | Skill 2 | Skill 3 |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 |
| 4 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 |
| 6 | 1 | 0 | 0 |
| 7 | 0 | 1 | 0 |
| 8 | 0 | 0 | 1 |
| 9 | 0 | 0 | 0 |
| 10 | 1 | 1 | 0 |

### 2.5.1 Scoring Procedure

Multiple choice testing with binary scoring is the most commonly used approach for educational assessments. A score of one refers to the correct response and zero to any incorrectly answered item. However, as several authors have noted (Tatsuoka, 1983), the score of one does not always reflect the knowledge level of examinees. Many erroneous rules occur to produce responses that coincide with the correct answer. Therefore, the performance of examinees cannot be precisely determined from their proficiency level. According to the literature, stochastic models allow for the possibility of 'slips' and 'guesses' in answering any item. A slip could occur when an examinee who has acquired the required attributes fails to correctly answer a subtask or fails to perform the item correctly. A guess refers to correctly answering an item or completing a subtask without possessing one or more of the required attributes. The models we have considered are largely defined by whether slips and guesses are allowed to take place at the subtask level or at the item level. These models are briefly introduced in the following section using the nomenclature for these models in Junker and Sijtsma (2001), with more mathematical descriptions given in the next section.

## 2.6 Students' Cognitive Models

### 2.6.1 The DINA model

The deterministic inputs, "noisy" and "gate" (DINA) model has been broadly used where the probabilities of both slipping and guessing are determined at the item level. It is a popular conjunctive CDM which assumes that a student must have mastered all the required attributes in order to correctly respond to an item in an assessment (Junker & Sijtsma, 2001).

To estimate students' knowledge of attributes, we require information regarding which attributes are required for each item. For this, the Q-matrix is used, which is a J×K matrix where $\boldsymbol{q_{jk}}$=1 if item j requires attribute k, and 0 if not. J is the number of items and K is the number of attributes in the assessment.

A binary latent variable $\boldsymbol{\alpha_{ik}}$ indicates respondent i's knowledge of attribute k, where $\boldsymbol{\alpha_{ik}}$=1 if student i has mastered attribute k, and 0 if they have not. Then, an underlying attribute profile of student i, $\boldsymbol{\alpha_{ik}}$, is a binary vector of length K that indicates whether or not the student has mastered each of the K attributes.

The deterministic element of the DINA model is a latent variable $\boldsymbol{\eta_{ij}}$ that indicates whether or not student i has mastered all attributes required for item j:

$\boldsymbol{\eta_{ij}}=\prod_{\boldsymbol{k=1}}^{\boldsymbol{k}} \boldsymbol{\alpha_{ik}^{q_{jk}}}$ …………. (2.1)

If student i has mastered all the attributes required for item j, then $\boldsymbol{\eta_{ij}}$=1; if the student has not mastered all of the attributes, then $\boldsymbol{\eta_{ij}}$=0.

The model allows for slipping and guessing defined in terms of conditional probabilities of answering items correctly (Yij=1) and incorrectly (Yij=0):

$\boldsymbol{s_j}$=Pr(Yij=0|$\boldsymbol{\eta_{ij}}$=1)

$\boldsymbol{g_j}$=Pr(Yij=1|$\boldsymbol{\eta_{ij}}$=0)

The slip parameter $\boldsymbol{s_j}$ is the probability that student i responds incorrectly to item j although all the required attributes have been mastered. The guess parameter $\boldsymbol{g_j}$ is the probability that student i responds correctly to item j although all the required attributes have not been mastered.

It follows that the probability of a correct response of student i to item j is:

$$P\ (\boldsymbol{Y}_{ij} = \mathbf{1}|\boldsymbol{\alpha}_{ik}) = (\mathbf{1} - \boldsymbol{s}_j)^{\eta_{ij}} \boldsymbol{g}_j^{\ 1-\eta_{ij}} \ldots\ldots\ldots (2.2)$$

Applications of the DINA model, along with multiple choice test algorithms for estimation, are given in Junker and Sijtsma (2001), Tatsuoka (2002) and de la Torre and Douglas (2004). The DINA model is also discussed in Haertel (1989) and Doignon and Falmagne (1999). The representation of the DINA model strategy is shown in Figure 2.2.

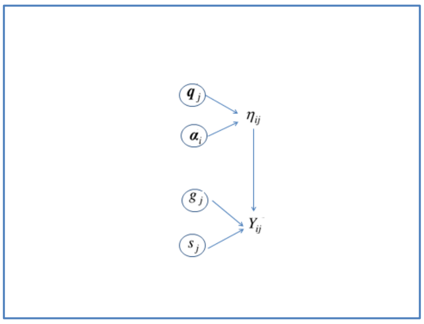


**Figure 2. 2**: DINA model Strategy

### 2.6.2 The NIDA Model:

This model was presented by Maris (1999). It also considers slips and guesses but at the subtask level rather than the whole item as shown:

Let $\boldsymbol{\eta}_{ijk}$ indicate whether the *i*th subject correctly applied the *k*th attribute in completing the *j*th item.

However, this model differs from the DINA model in which the slip $\boldsymbol{s_k}$ and guessing $\boldsymbol{g_k}$ parameters are determined within the attribute rather than by item and are defined by the following:

$$\boldsymbol{s_k} = P\,(\boldsymbol{\eta_{ijk}} = 0 \mid \boldsymbol{\alpha_{ik}} = 1, \boldsymbol{q_{ik}} = 1)$$

$$\boldsymbol{g_k} = P(\boldsymbol{\eta_{ijk}} = 1 \mid \boldsymbol{\alpha_{ik}} = 0, \boldsymbol{q_{ik}} = 1)$$

By assuming the $\boldsymbol{\eta_{ijk}}$s are independent conditional on $\boldsymbol{\alpha_{ik}}$, the item response function which relates the probability of a correct response to the latent attribute pattern has the form shown in equation (2.3):

$$\mathrm{P}(\boldsymbol{Y_{ij}} = \mathbf{1}|\boldsymbol{\alpha_{ik}}, \boldsymbol{s_k}, \boldsymbol{g_k}) = \prod_{k=1}^{k} \boldsymbol{p}\,(\boldsymbol{\eta_{ijk}} = \mathbf{1}|\boldsymbol{\alpha_{ik}}, \boldsymbol{s_k}, \boldsymbol{g_k}\boldsymbol{s_j}) = \prod_{k=1}^{k}[(\mathbf{1} - \boldsymbol{s_k})^{\alpha_{ik}}\ \boldsymbol{g_k}^{\mathbf{1}-\alpha_{ik}}]^{q_{ik}} \quad \ldots\ldots\ldots\ldots\ldots (2.3)$$

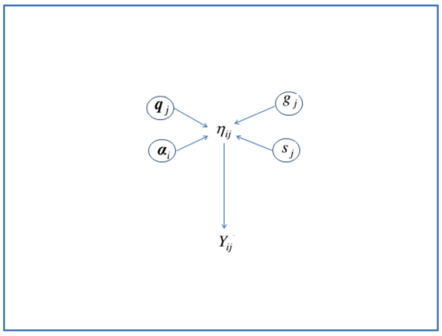



**Figure 2. 3**: NIDA Model Strategy

Neither the DINA model nor the NIDA model consider the possibility that examinees may solve a problem in different ways. According to Mislevy (1996) in his analysis of fraction subtraction data, he considered the notion of multiple strategies in which a strategy refers to the set of required attributes. Moreover, these two strategies do not consider the proficiency level for each examinee and, therefore, they do not provide an accurate estimation for the guess and slip parameters for each individual. Our work extends the DINA model strategy by adding the proficiency level of each student with regard to their correct and incorrect responses for each item.

### 2.6.3 Item Response Theory IRT

In psychometrics, item response theory (IRT) is a paradigm for the design, analysis and scoring of tests and questionnaires. It is a theory of testing based on the relationship between an individual's performances on a test item and the test taker's levels of performance on an overall measure of the ability that item was designed to measure. It is a normal logical form developed by Rasch, (1960). However, Birnbaum (1968) created a logistic version of the IRT model that included three parameters:

$$\boldsymbol{P}_i(\Theta)=\boldsymbol{c}_i+(1-\boldsymbol{c}_i)\frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{(-\boldsymbol{D}\boldsymbol{a}_i(\Theta-\boldsymbol{b}_i))}}\text{……………………}(2.4)$$

Where $Pi(\theta)$ is the probability of a randomly chosen examinee at proficiency level $\theta$ answering item $i$ correctly, and the three item parameters $a$, $b$ and $c$ are often called by their practical interpretations: discrimination, difficulty and guessing, respectively. This formula implements Birnbaum's idea by including a $c$-parameter in the model to allow for the statistical adjustment of IRF for the non-zero performance of low-proficiency examinees on multiple-choice (MC) items.

### 2.6.4 Knowledge tracing (KT)

The knowledge tracing (KT) procedure was discovered by Atkinson (1972) and developed by Corbett & Anderson (1994). It provides a powerful ability to track individual differences with each knowledge component (KC)/skill and this can then be used to make personalised decisions regarding the KCs a student has learned and which KCs need more practice. In KT (Corbett's version), there are four parameters fit to each KC: initial learning, learning rate, guess and slip parameters. These four parameters are interpretable, so it is easy to understand their effects on performance in the model as they use a student's prior history of performance with items for a KC to feed into the model equation by updating the current estimate of student learning based on the student's performance. The KT model has been used extensively by commercial tutors in addition to many experimental studies. While KT does feature desirable properties, it does not support the problem of multiple KC/skills. This prevents the tutor designers from creating practice steps where the student's response requires multiple KCs.

### 2.6.5 Learning Factors Analysis (LFA)

Multiple knowledge skills have been tackled by the Learning Factors Analysis (LFA) model (Cen, Koedinger & Junker, 2006). This model enables the evaluation of cognitive models and the analysis of student-tutor log data. It combines a statistical model and human expertise to measure the difficulty and learning rates of the given skill per item. It supports multiple skills as it captures the Q matrix of each KC, predicts student performance in each KC practice and identifies over-practised or under-practised KCs. The statistical model is shown below:

$$\mathbf{m}\ (\boldsymbol{i},\ \boldsymbol{j} \boldsymbol{\epsilon}\ \mathbf{KCs},\ \boldsymbol{n}) = \boldsymbol{\alpha}_i + \sum_{j \epsilon KC} (\boldsymbol{\beta}_j + \boldsymbol{\gamma}_j \boldsymbol{n}_{ij}) \text{......................} (2.5)$$

$$\mathbf{P(m)} = \frac{1}{1+e^{-m}} \text{ ....................................................} (2.6)$$

Where $m$ is a logit value representing the accumulated learning for student $i$ (ability captured by α parameter) using one or more KCs $j$. The β parameters capture the easiness of these KCs, and the benefit of the frequency of prior practice for each KC is a function of the $n$ of prior observations for student $i$ with KC $j$ (captured by the addition of $\gamma$ for each observation). Equation 6 is the logistic function used to convert $m$ strength values to predictions of observed probability. It captures the KC easiness with a single parameter for each KC and the learning rate for each KC with a single parameter for each KC.

However, the model states that all students accumulate learning in an identical fashion and ignores the correct and incorrect responses produced by the student. It therefore has very little power to dynamically differentiate between individual KCs for particular students. So, the LFA model is unsuitable for adaptive learning algorithms.

Because of the possible advantages of LFA, it has been formulated to a version that could be used adaptively. This reconfigured version is termed the Performance Factors Analysis (PFA).

### 2.6.6 Performance Factors Analysis (PFA)

PFA is a new alternative student modelling approach presented by Pavlik et al. (2009). It is a variant of learning decomposition and is based on reconfiguring the Learning Factor Analysis (LFA) (Cen, Koedinger & Junker, 2006). PFA provides a flexible adaptability to create the required model to be used in any tutoring system. In addition, it provides information regarding the student's performance, which is considered to be the strongest indicator of a student's learning process. Performance is an essential characteristic of student learning because correct responses strongly indicate the current ability of the student is already high. Furthermore, correct responses may simply lead to more learning than incorrect responses, since the production of a correct response may increase during the answering process, or perhaps may be due to ineffective review procedures after incorrect responses.

However, although it is a good start to enable the model to be sensitive to correct answers, it also seems useful to make the model specifically sensitive to incorrect answers.

Adapting sensitivity to incorrectness allows incorrectness to measure student learning in an inverse to correctness. Together, the matching of both correctness and incorrectness in the model will cause it to be sensitive to not only the quantity of each but also the relative ratio of correct to incorrect. This will achieve a better estimation of the learning ratio even when the student fails to answer any item correctly.

Briefly speaking, this takes the form of a standard logistic regression model with the student scores being independent variables and the performance being the dependent variable. It recomposes LFA by dropping the student proficiency variable and replacing the skill variable with the question identity (i.e. one parameter per question). The model estimates a parameter for each item representing the item's difficulty, and also two parameters for each skill reflecting the effects of the prior successes and prior failures achieved for that skill.

Equation (2.7) shows the PFA logistic form; its variable, α, (which was in LFA) has been removed from the model as it does not provide an estimation ahead of time in adaptive situations. However, as noted by Corbett, models that do track subject level learning variability can greatly improve model adequacy. $\boldsymbol{\beta_j}$ has been previously explained, $\boldsymbol{succ_{i,j}}$ tracks the prior successes for the KC for the student, $\boldsymbol{fail_{i,j}}$ tracks the prior failures of the given skill for the student, and the $\boldsymbol{\gamma_j}$ and

$\boldsymbol{\rho_j}$ scale the effect of these observation counts. Equation 2 is still applied for conversion to probability predictions. (Again, the model can be used in a compensatory fashion for observations requiring multiple KCs by summing the $\boldsymbol{\beta_j}$ and $\boldsymbol{\gamma_j}$ and $\boldsymbol{\rho_j}$ frequency components for all $j$ KCs needed). We call this model the Performance Factors Analysis (PFA).

$$\mathbf{Performance}(\boldsymbol{i}, \boldsymbol{j} \in \mathbf{KCs}, \mathbf{succ}, \mathbf{fail}) =$$

$$\boldsymbol{m_i} = \sum_{\boldsymbol{j \in KCs}}^{\boldsymbol{n}} \boldsymbol{\beta_j} + \boldsymbol{\gamma_j succ_{i,j}} + \boldsymbol{\rho_j fail_{i,j}} \ldots (2.7)$$

The probability of the performance:

$$\mathrm{P}(\boldsymbol{m_i}) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e^{-m_i}}} \ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$

The below table summarises all the explained models and compares these with the developed model – the Cognitive Factor Analysis (CFA), which will be explained in detail in chapter three.

**Table2. 5**: Summary of the cognitive models

| Model | Type | Parameter | Output | Limitations |
|---|---|---|---|---|
| IRT | Statistical | Student proficiency and item difficulty | Determines the student's current performance | Does not give performance future estimation, does not consider the cognitive factors (slipping/ guessing factors and students skills vs. item skills) |
| KT | Probabilistic | Initial learning, learning rate, guess and slip parameters | The probability of having the correct answer for the next item | Does not support multiple skills items and does not consider the prior incorrect scores |
| NIDA | Probabilistic | Guessing/slipping parameters of student's current skills. | The probability of providing the correct answer based on slipping/guessing parameters for a single skill rather than the whole item | Only gives an estimation of the single skill rather than the whole item. Moreover, it does not consider the item difficulty level and the student's incorrect scores. This may lead to an incorrect estimation of the guessing and slipping. |

| | | | | |
|---|---|---|---|---|
| DINA | Probabilistic | Guessing/slipping parameters of current student skills | The probability of having the correct answer based on slipping/guessing parameters for the whole item with multiple skills. | It is a conjugative model that requires all the necessary skills to be mastered in order for the examinee to have a high probability of responding correctly, regardless of the student's proficiency and learning rates. |
| LFA | Statistical | Student proficiency, prior correct scores, item difficulty level and success rate, supports the multiple skill items. | More accurate estimation of the student's performance for the next item | Does not support guessing/slipping parameters, does not support multiple skills items and does not determine whether the student has achieved the answer based on their own knowledge or has slipped the answer. Furthermore, it ignores the prior incorrect scores. |
| PFA | Statistical | Item difficulty, prior correct/ incorrect scores, success/failure rate and supports multiple skills items. | More accurate estimation of the student's performance for the next item as it considers the prior incorrect score and failure rate for the student. | Does not support slipping/guessing factors, does not support skills correlations. As these are two skills highly correlated to each other, the student who answers one of them correctly will most likely also answer the other item correctly. However, this may affect the performance prediction in the case where the student attempts one of these skills incorrectly and gets the other one correct, as the model does not decide whether this behaviour is learning from failures or performance from successes. |

Table 2.5 lists all the models that are useful tools in ITS and which have been used extensively in the literature. As has been explained, each model has its own disadvantages which negatively affect the estimation of student performance and the student learning process. Therefore, the developed model of this research, called the Cognitive Factor Analysis (CFA), overcomes all the challenges listed in the table above. CFA is based on the dynamic assessment technique and it uses multiple-choice items, i.e. it supports the guessing and slipping parameters. Furthermore, it supports multiple skills items with item correlation. It

first determines the student's current skills and compares these with the item skills. Moreover, it considers the prior correct scores as well as the incorrect scores towards a better estimation of the student's latent performance. Since the model uses the dynamic assessment type, it adopts the concept of providing hints/help for the student to demonstrate the function of having a tutor. Hence, the estimation of the student's correct answer will be based on guessing or receiving a hint. Likewise, the incorrect answer estimation is based on slipping or the lack of knowledge/skill. This ensures a more accurate performance prediction for the next item answer. Moreover, chapter three explains in detail the model structure, factors and its algorithm.

## 2.7 Examples of using DINA and PFA models

The significant impact of both DINA and PFA models in the field of ITS does not only feature in terms of student performance estimation but also by way of the analyses of students' strengths and weaknesses.

Despite the fact that these two models have certain limitations, which are listed in table 2.5, the developed model of this research, namely CFA, is based on the designs of DINA and PFA. CFA will expand the concept of slipping and guessing parameters of the DINA model and the concept of using a student's prior PFA success/failure scores to produce an improved model for an accurate latent performance inference.

Although the DINA and PFA models will be fully explained in chapter three, this section has briefly demonstrated how both the PFA and DINA model work to offer an understanding to other researchers regarding how cognitive factors are excluded and applied.

Since both models use multiple-choice assessments, they are binary representation models that use binary data (0 or 1). Both models use two sets of data; one is the Q matrix, which was explained earlier in section 2.5, and the second is the student's prior scores data. The Q matrix consists of four skills items, while the student's data set includes the student's ID, the item's ID and the answer (which is either 1 for a correct answer or 0 for an incorrect answer). This section used a sample of the data set used to develop the CFA model in chapter 3, section 3.5.

**Table2. 6**: Q matrix sample

| Item ID | Skill 1 | Skill 2 | Skill 3 | Skill 4 |
|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 |

**Table2. 7:** Student A's prior score

| Student ID | Item ID | Answer |
|---|---|---|
| A | 1 | 1 |

The **DINA** model will implement the following steps in its analyses:

- Firstly, analysing the prior scores of the student to determine their current skills $\alpha_{ik}$. Several approaches can be used to determine the student's current skills from prior scores. Such approaches are explained in detail in chapter 3, section 3.4.2. However, the DINA model uses Maximum Likelihood Estimation (MLE).

**Table2. 8**: Student A's current skills

| Student ID | Skill 1 | Skill 2 | Skill 3 | Skill 4 |
|---|---|---|---|---|
| A | 1 | 1 | 0 | 1 |

Therefore, Table 2.8 shows that student A has not mastered all the given skills and lacks skill 3.

- Secondly, determining the student proficiency level $\eta_{ij}$: By applying equation 2.1:

$\eta_{ij}=\prod_{k=1}^{k} \alpha_{ik}^{q_{jk}}$

Since :

$\alpha_{ik} = 1\ 1\ 0\ 1$

and, $q_{jk}= 1\ 0\ 1\ 1$

Therefore, the proficiency level $\eta_{ij}= 1^{1}*1^{0} * 0^{1}*1^{1}=0$

i.e. the proficiency level of student A answering item T1 is 0.

However, the student has attempted this item correctly, therefore, the DINA model analysed the guessing and slipping parameters for item T1:

**Table2. 9**: Guessing and slipping parameter values

| Item ID | Slipping $s_j$ | Guessing $g_j$ |
| --- | --- | --- |
| T1 | 0.24 | 0.53 |

- By applying the DINA model equation in 2.2,

$\mathbf{P}\ (Y_{ij} = 1|\alpha_{ik}) = (1 - s_j)^{\eta_{ij}} {g_j}^{1-\eta_{ij}}$

$= (1 - 0.24)^{0} * 0.53^{\ 1-0}$

=0.53, which means the probability of this student answering an item with the same given skills as item T1 is 58%. According to table 2.7, this student has attempted the answer correctly, despite the fact that he/she lacks in one skill and the proficiency level is 0. This means that there is a probability that the student has guessed the answer as he/she does not possess sufficient skills to answer the item.

**PFA :**

PFA simply applies the logistic regression formula to the Q matrix and student data set to estimate the coefficients associated with equations 2.7 and 2.8:

$$\mathbf{Performance}(\boldsymbol{i},\boldsymbol{j}\in\mathbf{KCs},\mathbf{succ},\mathbf{fail}) = \boldsymbol{m_i} = \sum_{\boldsymbol{j}\in\boldsymbol{KCs}}^{\boldsymbol{n}} \boldsymbol{\beta_j} + \boldsymbol{\gamma_j succ_{i,j}} + \boldsymbol{\rho_j fail_{i,j}} \text{.......(2.7)}$$

$$\mathrm{P}(\mathbf{m}) = \frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{-\boldsymbol{m_i}}} \text{.......(2.8)}$$

The coefficient estimates are as follows (the process of theses estimations will be explained later in chapter four, table 4.4):

**Table2. 10**: PFA coefficient estimates

| Skill | $\boldsymbol{\beta_j}$ | $\boldsymbol{\gamma_j}$ | $\boldsymbol{\rho_j}$ |
|---|---|---|---|
| Skill 1 | 0.45 | 0.38 | 0.28 |
| Skill 2 | -0.11 | 0.27 | 0.42 |
| Skill 3 | 0.37 | 0.33 | 0.43 |
| Skill 4 | 0.17 | 0.38 | 0.104 |

To apply these estimated coefficients to the PFA formula and estimate the probability of achieving the correct answer, PFA supposes that student A has no prior knowledge for answering the first item, therefore the success($\boldsymbol{succ_{i,j}}$)/failure($\boldsymbol{fail_{i,j}}$) scores= 0.

Let us suppose that student A is attempting to answer item T1, which has skills 1, 3 and 4. According to formula 2.7, the performance for item T1 can be calculated as follows:

$$[\boldsymbol{\beta_1} + \boldsymbol{\gamma_1 succ_{A,1}} + \boldsymbol{\rho_1 fail_{A,1}}] + [\boldsymbol{\beta_3} + \boldsymbol{\gamma_3 succ_{A,3}} + \boldsymbol{\rho_3 fail_{A,3}}] + [\boldsymbol{\beta_4} + \boldsymbol{\gamma_4 succ_{A,4}} + \boldsymbol{\rho_4 fail_{A,4}}] = 0.99$$

The probability of student A answering item T1 correctly is: P $(\boldsymbol{m_i})$= $\frac{\mathbf{1}}{\mathbf{1}+e^{-0.99}}$=0.72

According to table 2.11, the student has scored the item correctly, therefore the prior success score $(\boldsymbol{succ_{i,j}})$ is 0 and prior failure score $(\boldsymbol{fail_{i,j}})$ is 1. The same process of calculation has been made for item T2 and the probability of having the correct answer P$(\boldsymbol{m_i})$ is 0.74; hence, the probability value is increasing. However, the student has answered the item incorrectly and scored 0 (as shown in table 2.11). According to PFA theory, having the score zero does not represent a failure but a lack in knowledge. In such cases, the probability of a student's performance is increasing as the student's learning is becoming higher by the addition of the failure rates of the given skill.

This section has used the parameters, utilised later in chapter four, which have been estimated after applying the data to the formula and running the code. However, all the explanations of estimating the parameters and calculating the success and failure rates are fully described in chapter three, as this example simply serves to demonstrate how the PFA and DINA models work with the given data.

**Table2. 11**: PFA predictions' values

| Student ID | Item ID | Skills | Score |
|---|---|---|---|
| A | T1 | 1011 | 1 |
| A | T2 | 1001 | 0 |

## 2.8 Conclusion

This chapter lists the related students' estimation models, the students' cognitive parameters and the scoring procedure employed. It has shown that a students' cognitive model is the main factor in the Intelligent Tutoring System and highlighted the essential cognitive factors, such as the item difficulty, student skills and guessing/slipping parameters and how the Q matrix is applied. A literature review has been presented for the techniques used in ITS. Finally, the approximation methods used to estimate the required probability of knowledge level and learning have been described.

# Chapter Three: Cognitive Factor Analysis (CFA)

## 3.1 Introduction

This chapter presents the structure of the proposed model "Cognitive Factor Analysis" (CFA) and its mathematical background. It begins by exploring the student behaviour analyses and their impact on the quality of performance estimation, then moves on to consider the research problem, including the issues with the PFA and DINA models, and ways of achieving better student performance estimations, depending on multiple factors, to attain the correct item answer including the guessing and slipping parameters. Later, the section explains the types of parameters used in CFA in terms of whether they are descriptive or predictive, followed by consideration of all the components which contribute to the CFA framework, including the model block diagram and its developed algorithm. Finally, the data used to fit the developed model parameters will be discussed.

## 3.2 Analysing student behaviour

In academic settings, specifying a proper definition of student behaviour could help to define their goals. Since student behaviour is comprehensive, as reflected by their past behaviours, carrying out proper analyses therefore leads to an accurate performance estimation because such analyses may seek to define the following:

(1) The cognitive processes that underlie students' actions;

(2) The differences between the students' current skills and expert skills;

(3) The students' behavioural patterns or preferences; or

(4) The students' characteristics.

As discussed in the previous chapters, existing frameworks have been developed for estimating students' performance based on their prior scores/behaviours. However, most of these lack true efficacy in estimating student behaviour as they do not adopt cognitive factors. As a result, a student's performance estimation will be negatively affected. Therefore, in order to achieve an accurate latent performance inference, student behaviour should be defined carefully. To achieve this goal, the student's cognitive factors, which play an essential role in improving such analyses, are included in the design. The following points explain the cognitive factors and their importance:

- Importance of guessing and slipping parameters: Most cognitive frameworks, such as the DINA/NIDA models, have used the concept of guessing/slipping parameters. In MC assessment, the student has to choose the correct answer from multiple options. However, it is still possible that the student might guess the correct answer even without possessing the required skills which should match the item skill. Likewise, the student might slip and give the incorrect answer even though s/he has acquired the necessary skills. Therefore, without considering the slipping/guessing parameters, the estimation of student performance might give a false indication about the student's knowledge level and this might negatively affect the improvement of students' skills and the overall learning progress.

- Importance of hints/help parameter: Hinting is an important teaching factor in one-on-one tutoring, used when the student gives an unexpected answer to an assessment item. The tutor therefore needs to align the hinting strategy with the student's need while making the strategy fit the high level tutoring plan and the tutoring context. In many student-oriented tutoring systems, the machine tutor will offer hints when the student asks for help, e.g. Andes (Gertner et al., 1998). On the other hand, and as explained earlier, the developed model (CFA) uses MC items embedded in the dynamic assessment. In the educational environment, dynamic assessment refers to assessment that is specifically intended to improve and accelerate the student learning process through teaching students within the assessment process. Therefore, most models use hints as a part of the teaching process. This empowers the students to be self-regulated learners which is the core goal of the ITS (Pintrich & Zusho, 2002). Further, the intelligent tutor should provide an automated hint (help) attached to each item which can be requested whenever needed, so that the model can perform the role of the teacher in the classroom. Therefore, CFA provides one constructive hint per item which appears when the student requests it or when s/he attempts an incorrect choice. In addition to featuring hints and based on comparing the current student's skills against the item skills, the model can determine whether the student has really learnt from the hints or s/he needs further training. If the student has asked for a hint and attempts the item correctly, then the system will consider this as learning and that the student does not therefore require any further training. However if s/he missed the correct answer after receiving a hint, then this student is deemed to need further practice regarding this skill.

  Although making hints available ensures that the student is directed towards attempting the correct answer, this might affect the student's latent performance estimation. With hints, the estimation depends on determining whether the student attempts the correct answer based on learning from hints (without acquiring the necessary skills) or depending on his skills (without asking for a hint). Therefore, CFA has attached the hint factor to the correct answer to estimate the student's performance. This extension will ensure a better learning (from hints) and more accurate knowledge estimation.

- Prior correct/incorrect scores: The students' prior knowledge (scores) provides an indication of the alternative conceptions as well as the scientific conceptions possessed by students. Most tutoring models consider only the number of correct scores of the student to make the decision regarding performance. However, research has shown that the number of incorrect scores can also provide a significant impact on the student's knowledge estimation, especially when using the hinting strategy. The Performance Factor Analysis (PFA) model considers the prior incorrect scores as well as the prior correct scores for its estimation and has shown promising results. However, PFA has one disadvantage, it does not consider the hints parameters as well as the guessing/slipping parameters, although it uses MC assessments. Therefore, CFA extends PFA by adding the hint/guessing/slipping parameters to make the model capable of determining the student knowledge skill effectively with multiple choice assessment items and to make the prediction more cognitively diagnostic to be applied in multiple applications.

- Number of choices provided in MC items: Multiple choice items consist of a stem, the correct answer, the keyed alternative and distractors. The *stem* is the initial part of the item that presents the item as a problem to be solved, a question to be asked of the respondent or an incomplete statement to be completed, as well as any other relevant information. The options are the possible answers that the examinee can choose from, with the correct answer referred to as the *key* and the incorrect answers termed *distractors*. Only one answer can be keyed as correct. The goal of MC items in ITS is to improve the student's learning process by measuring precisely the student's performance against his/her acquired skills; therefore, the number of choices provided per item depends on the number of keyed alternatives and the number of distractors. The more keyed alternative choices that are provided, the more accurate the estimation of the student's knowledge will be. The keyed alternative choices help the students to make a logical guess based on the choices available; therefore, the likelihood of the random guess is very small. In such cases, the student can also learn from the choices. Hence, it is important to provide fewer distractors and more alternatives and thus offering more alternative choices reflects a stronger assessment item.

Therefore, proper adaptation of the factors explained above in the design phase of ITS will help to properly analyse student behaviour and subsequently improve the overall latent performance estimation of students.

## 3.3 Research problem

A student's performance is an essential factor in the learning process; it reflects several behaviours, such as the student's behaviours, learning curves and the quality of learning resources. Furthermore, student performance is a major part of Intelligent Tutoring Systems, since it depends on the observation mechanisms which have been captured by the student's performance (e.g. the correctness of the student response compared to the given item skills) and then uses those parameters to estimate the student's underlying hidden attributes, such as goals, preferences and knowledge inference.

Recently, ITS have attracted major interest among researchers and therefore several satisfactory estimation models have been developed towards better inference prediction. However, there still remain certain challenges that could be addressed, such as the inability to feature the proper types of predictors in the model against the number of the item skills being used. In order to make a better decision on model usefulness, throughout the design phase process certain questions need to be considered for practical model selection:

1. Which factors of the data are estimated as predictive parameters and which are estimated as descriptive parameters?

2. How is the student's latent knowledge inference estimated according to the given guessing and slipping parameters along with prior score records?

In this section, we focus our efforts on the following two aspects:

- We compare two competing student cognitive models: the DINA Model and the PFA model. We list the issues relating to each of them and, based on this, our developed model is presented which will tackle these issues.
- Furthermore, for PFA, a new modified version is proposed by the addition of the hint factor in order to increase the performance estimation accuracy.

### 3.3.1 DINA Model:

The DINA model divides each item into those that acquire all the required attributes and those that do not. It can be represented as a latent response model in which the slips and guesses appear at the item level, rather than at the subtask level.

#### 3.3.1.1 Issue with the DINA Model

The issue with the DINA model is that it is a conjugative model which requires all the necessary skills to be mastered in order for the examinee to have a high probability of responding correctly, regardless of the student's proficiency and learning rates. The student might achieve the correct answer, depending on h/is/er learning from failures or hints, regardless of mastering the given skills. In this case, the model should not consider the guessing probability for this student.

### 3.3.2 PFA Model

PFA is a parameterisation of the Linear Logistic Test Model that predicts performance on the current item using the entire history of success and failure on previous items addressing the same skill.

#### 3.3.2.1 Issues with PFA Model

- Identifying the skill correlations:

  As explained earlier, the classic PFA model predicts student performance depending on the item difficulty and the student‘s prior success or failure on a number of skills required for this item. Although this approach supports multiple skills, it ignores the correlations between other skills.

This assumption is reasonable and easily understandable since, if there are two skills that are highly correlated to each other, the student who answers one of them correctly will most likely answer the other correctly as well. However, this may affect the performance prediction in the case where the student attempts one of these skills incorrectly and answers the other one correctly, so the model needs to determine whether this behaviour can be viewed as learning or as improving performance. Therefore, there must exist a way by which to identify student behaviour as to whether he/she might have slipped or guessed the answer. In this case, the model assumes that the probability that a student successfully solves a problem might also depend on his proficiencies in other skills. However, there is no easy way by which to identify which other skills are important to a given skill; therefore, in this study, we used all skills.

- Student's correctness predictions:

  Since PFA depends on a student's prior records, there are therefore other factors that need to be considered in the student's historic records, such as the prior used hints. Hints are considered to be an essential factor that helps students in learning and directing him/ her to the correct answer. On the other hand, it also enhances the student's overall performance. In this case, by considering the hints factor, the performance accuracy might be changed according to the student's prior answer with or without using hints. In a binary performance model, a student would receive a '1' if they solved the problem correctly on their attempt with or without asking for a hint. Therefore, the performance estimation may differ, as the student requesting a hint is less likely to understand the skill. To solve this issue with PFA, we created a scoring method that would split the success rate into a prior correct answer and prior used hints.

By applying this method, CFA will estimate whether the correct answer is achieved based on the student's current knowledge or by using a hint. This will enable the model to estimate the strength level of the student knowledge when answering the item correctly as the student might have learnt from the given hint and made a logical guess.

In this case, achieving the correct answer based on a logical guess will increase the student skill level even if s/he previously lacked in this skill. Therefore, the level of complexity of learning materials may change to reduce the number of practices the student might need. This will have a huge impact on a reduction in the student's study time.

The next section explains the way of calculating the prior correct/failure scores and the prior used hints according to descriptive or predictive manner.

## 3.4 Prediction of the outcomes versus description/evaluation of the inputs

The estimation of model parameters, either in a predictive or descriptive manner, is an important aspect of model building; however, it is often overlooked or ignored.

Deciding the way in which the data is analysed and the parameters are constructed could have an important effect on the structure of the model. To clarify this, consider the following examples:

A new student has been described by h/is/er teacher as intelligent, self-confident and hard working. Consider these two types of questions:

a- Description/evaluation: how does the given description impress you in terms of academic ability?

b- Prediction: what is your estimate of the grades point average that this student will obtain?

There is an important difference between these two questions. In the first, the input is evaluated while, in the second, the output is predicted. An evaluation is the description of quality for which data is given. Since prediction is the estimation of future performance, the prediction should therefore be more regressive than evaluation. The second question has greater uncertainty than the first. However, in the statistical theory of prediction, the description is more likely to be inaccurate or the prediction will be invalid. On the other hand, the observed equivalence between prediction and evaluation would be justified if prediction accuracy was perfect, which depends on the prior records/observation of each student (Yudelson, M., Pavlik Jr, P. I., & Koedinger, K. R. (2011, July).

**Table3.1 Table 3.1: Predictive and Descriptive coding example**

| Student ID | Item No. | Answer | Pred1 | Pred2 | Pred3 |
|---|---|---|---|---|---|
| A | 1 | 1 | 1.00 | 0 | 0.60 |
| A | 2 | 0 | 0.50 | 1.00 | 0.60 |
| A | 3 | 0 | 0.33 | 0.50 | 0.60 |

| | | | | | |
|---|---|---|---|---|---|
| A | 4 | 1 | 0.50 | 0.33 | 0.60 |
| A | 5 | 1 | 0.60 | 0.50 | 0.60 |

Table 3.1 presents an explanation of the descriptive and predictive parameters. It gives the prior scores of student A and three ways to determine the student's behaviours (Pred1, Pred2 and Pred3). Pred1 is the mean success rate - mean of correct - over prior user/item including the current one. Pred2 is the mean success rate over the items which are strictly prior to the current one. Pred3 is the percent correct over all user trials. An example of strictly predictive coding of the data is Pred2 because it considers that there is no information regarding the user and gives an estimation of his/her performance. Therefore, a model that estimates a parameter for Pred2 would capture the predictive nature of this value without any prior information. While Pred1 and Pred3 are descriptive codings since they give overall information regarding the user's behaviours and there is no universal recipe for deciding when to include predictive or descriptive parameters into the model. However, for most user modelling and student performance estimations, predictive coding is more suitable and reasonable as the output of the model should be estimated/predicted more than observed.

For our developed model, these parameters can define the student skills and be combined with the item skills excluded from the Q-matrix. As described in chapter two, the construction of the weight matrix $Q$ (usually called Q-matrix) involves the qualitative preliminary work of experts:

Firstly, the tested overall data is subdivided into a few skills according to a well-established qualitative relationship between the skills. In our model, these skills are termed as $\boldsymbol{\alpha_k}$ , $k = 1, \ldots, n$ *(where n is the number of skills per each item)*. Secondly, based on the relationship between the skills, these skills are determined per item j which is denoted in a binary $J * K$ weight matrix $Q$, in which $\boldsymbol{q_{jk}}$ expresses whether skill $k$ is needed ($\boldsymbol{q_{jk}}$ = 1) or not ($\boldsymbol{q_{jk}}$ = 0) for enabling examinees to positively respond to item $j$.

Thus, the Q-matrix reflects the essential theory of how skills contribute to responding to each item and enables the estimation model to infer the examinees' possession of the *skill K* from their response vectors. Therefore, the success/failure rates, including the prior used hints, are determined as predictive parameters to estimate the performance produced by PFA.

The next section demonstrates how the predictive parameters approach will be applied to the prior scores of the student (prior correct, incorrect + prior used hints).

### 3.4.1 Calculating the prior success/failure/hint scores

As explained earlier, hinting is considered to be an essential part of teaching and, since the original PFA model formula does not consider a hint parameter, therefore a modified version of PFA is introduced. The PFA formula was extended to include the hint parameter to be added to the prior correct scores. This extension is a part of the CFA model which will be described later in this section. In order to implement the ModPFA formula, the prior correct/incorrect and hints scores will be determined as predictive parameters.

An algorithm has therefore been developed to describe the counting process of splitting the prior success scores using the prior hints (used in Mod PFA), besides the counting for prior failure scores. This algorithm has been applied to an item with two skills and the values have been determined as shown in table 3.2.

**Table3.2: Coding prior success scores and prior used hints as descriptive parameters**

| Student ID | Item ID | Skill1 | Skill2 | $A_k$ | $succ_{i\,1}$ | $succ_{i\,2}$ | hint | $H_{i1}$ | $H_{i2}$ | $succ\_total_{i,1}$ | $succ\_total_{i,2}$ | $f_{i1}$ | $f_{i2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 1 | 3 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | 3 | 0 | 1 | 0 |
| 1 | 4 | 1 | 0 | 0 | 2 | 0 | 1 | 2 | 0 | 4 | 0 | 1 | 0 |
| 1 | 5 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 3 | 0 | 3 | 0 | 0 |
| 1 | 6 | 1 | 0 | 0 | 2 | 0 | 1 | 3 | 0 | 5 | 0 | 2 | 0 |
| 1 | 7 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 4 | 0 | 5 | 0 | 0 |
| 1 | 8 | 1 | 0 | 1 | 2 | 0 | 1 | 4 | 0 | 6 | 0 | 3 | 0 |
| 1 | 9 | 1 | 1 | 0 | 3 | 2 | 1 | 5 | 5 | 8 | 7 | 3 | 0 |

**Algorithm 1: Counting the prior scores/hints**

*Input:* for student i, and item j:

**given the prior answers** $\boldsymbol{A_j}$**,** $\boldsymbol{where\ j\ is\ the\ number\ of\ the\ item}$**,**

**A is the prior student's answer/score which is either 0 or 1.**

Each **item** $_{\mathbf{j}}$ is provided with a number of skills: $\boldsymbol{skill_k}$ where k is the number of skills. Further, each item j is provided with a hint for is counted for each given skill k: **hint** $_{\mathbf{k}}$.

*Output:* counting the prior success scores for each student i and each $\boldsymbol{skill_k}$ (split into prior correct answers $\boldsymbol{succ_{ik}}$+Prior used hints $\boldsymbol{h_{ik}}$) and counting of total in incorrect/failure scores is $\boldsymbol{f_{ik}}$ .

For j=1 : m , j++ *# m= no. of given items*

{

if k=1 # checking for the skill $_k$ availability

then,

{

for j=j-1:j *#counting for the prior scores where first prior score is zero*

{

# counting the prior success /correct scores

If $A_j$ =1 and hint $_k$=1 # checking whether the answer/hint of the prior item is 1

Then {

$succ_{ik}$ =$A_j$ + $succ_{ik}$,

$H_{ik} = H_{ik} + hint_k$

$succ\_total_{i,k}$ =$H_{ik}$ + $succ_{ik}$

# counting for the predictive total success scores for $skill_k$ **with the hint parameters**

**}**

# Counting for incorrect/failure prior scores

{

If $A_j$ =0 # checking whether the answer of the prior item is 0

Then {

$fail_{ik}$=$A_j$ + $fail_{ik}$,# counting for the predictive prior failure scores

}

else 0, # when $skill_k$ is not available , everything else is zero

end if,

}}

Table 3.2 shows how the success/failure scores and the prior used hints are calculated as predictive parameters to be used later in the CFA model. The following example demonstrates the algorithm and how it can be applied to the data shown in table 3.2. Starting with item 1, which has only one skill available (skill1), the prior scores for success/failure are zero, as well as the prior used hint (because they are predictive parameters where there are no prior knowledge) regardless of the answer $A_k$ which is 1.

The counting starts from item 2, since it also has skill 1 available and the prior answer $A_k = 0$, therefore, the prior success scores for skill 1 ( $succ_{i1}$) is 0, and for skill 2 is also zero as the skill is not available while the prior used hints for skill1( $H_{i1}$) =1 while it is zero for skill 2. The failure/incorrect prior scores for skill1 ($H_{i1}$) =1, whereas it equals 0 with skill 2 .

Here, the correct scores are going to be extended to also include the prior used hints. This extension will enable the model to predict whether the correct answer has been performed from the first attempt or by using hints. In other words, it will determine if this student has answered the item correctly depending on his/her current skills or by learning from the provided hints. Having such factors (i.e, determining hints with the correct answer) improves the prediction accuracy. Having the correct answer without acquiring the necessary skills will not be considered as random guess. Instead, it will be defined by the model as a logical guess (learning from the hint). This modified model will improve the student's skills and reduce study time by recommending more advanced learning resources rather than basic ones.

## 3.5 Modified PFA formula

In order to explain the mathematical form of the modified version of PFA, a formula has been created to be included in the CFA full formula which will be explained later in this chapter.

The formulae are as follows:

$$\boldsymbol{H_{ik}} = \sum_{1 \le k \le j-1} \boldsymbol{h_k} \quad \text{......................(3.4)}$$

$$\boldsymbol{succ_{i,k}} = \sum_{1 \le k \le j-1} \boldsymbol{A_j} \quad \text{.................(3.5)}$$

$$\boldsymbol{succ\_total_{i,k}} = \boldsymbol{H_{i,k}} + \boldsymbol{Succ_{i,k}} \quad \text{.....(3.6)}$$

$$\boldsymbol{fail_{i,j}} = \sum_{1 \le k \le j-1} |\boldsymbol{1 - A_j}| \quad \text{...........(3.7)}$$

Where,

$\boldsymbol{H_{i,k}}, \boldsymbol{succ_{i,k}}, \boldsymbol{succ\_total_{i,k}}, \boldsymbol{fail_{i,k}}$ are the counts for prior hints used in the given skill, prior correct scores for the given skill, total success counts and prior incorrect/failure scores counts in the given skill , respectively.

$$h_k \ = prior\ used\ hints\ for\ skill\ k\ in\ \ item\ j\,,$$

$$A_j \ = prior\ student's\ score\ (\ 0\ or\ 1\ for\ item\ j)$$

Besides, i, j and k, represent the student i for item j with skill k, and it is the order of counting the skills (which have been counted using the predictive manner as explained earlier). This part of updated/modified PFA with the hint parameter will be added to the CFA model; therefore, the overall ModPFA equation would be as:

$$\mathbf{Performance}(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \in \mathbf{skill}, \mathbf{succ}, \mathbf{fail}, \mathbf{hint}) =$$

$$\boldsymbol{Modm_i} = \sum_{j \in KCs}^{n} \boldsymbol{\beta_k} + \boldsymbol{\gamma_k succ\_total_{i,k}} + \boldsymbol{\rho_k fail_{i,k}} \ldots\ldots(3.8)$$

Where $\boldsymbol{succ\_total_{i,k}}$ is the form of equation (3.6), $\boldsymbol{\beta_k}$ is the difficulty level of the skill k, $\boldsymbol{\gamma_k}$ and $\boldsymbol{\rho_k}$ are the success and failure rates of skill k respectively.

The other issues with the DINA model and PFA mentioned earlier will be tackled with our proposed main model: Cognitive Factor Analysis (CFA).

## 3.6 Cognitive Factor Analysis Model (CFA)

The problems mentioned earlier regarding DINA and PFA mainly concern the accurate estimation of the student's performance. To tackle this issue, and as has been mentioned in the hypothesis of this research which stated that "including the cognitive factors will improve the performance estimation", CFA was developed based on the inclusion of the student's cognitive factors in which the prediction of the latent performance inference will be improved. Therefore, based on the above drawbacks, CFA attempts to:

- Infer a single aspect of the data (predicted variable) from a certain combination of other aspects of the data (predictor variables) which is to make inferences about the student's present latent knowledge. The model uses cognitive factors as predictors, such as the success/failure rates, the guessing/slipping parameters as well as the item difficulty level and the number of prior used hints.

- Design the given data (item/skills data and the student's prior data as predictive parameters (this will be explained in detail in section 3.3.5).
- Measure how much latent skill a student possesses and the next time the skill is encountered while they are learning in terms of probability of correctness in the given items.

To achieve the above requirements, the CFA model has been designed according to the following characteristics:

- The model involves dichotomous items which may involve multiple knowledge components/skills.
- The number of choices for each item is designed to have fewer distractors and more alternatives so that the student can learn from the given choices.
- The model determines whether the student has acquired all the necessary knowledge to answer the given items.
- Depending on the student responses in the given multiple choices item, the model determines the probability of guessing and slipping per given item.
- The model considers the hint parameter by including the prior used hint with the prior correct answer to check whether the student answered the item correctly depending on the acquired skills or using the provided hints.

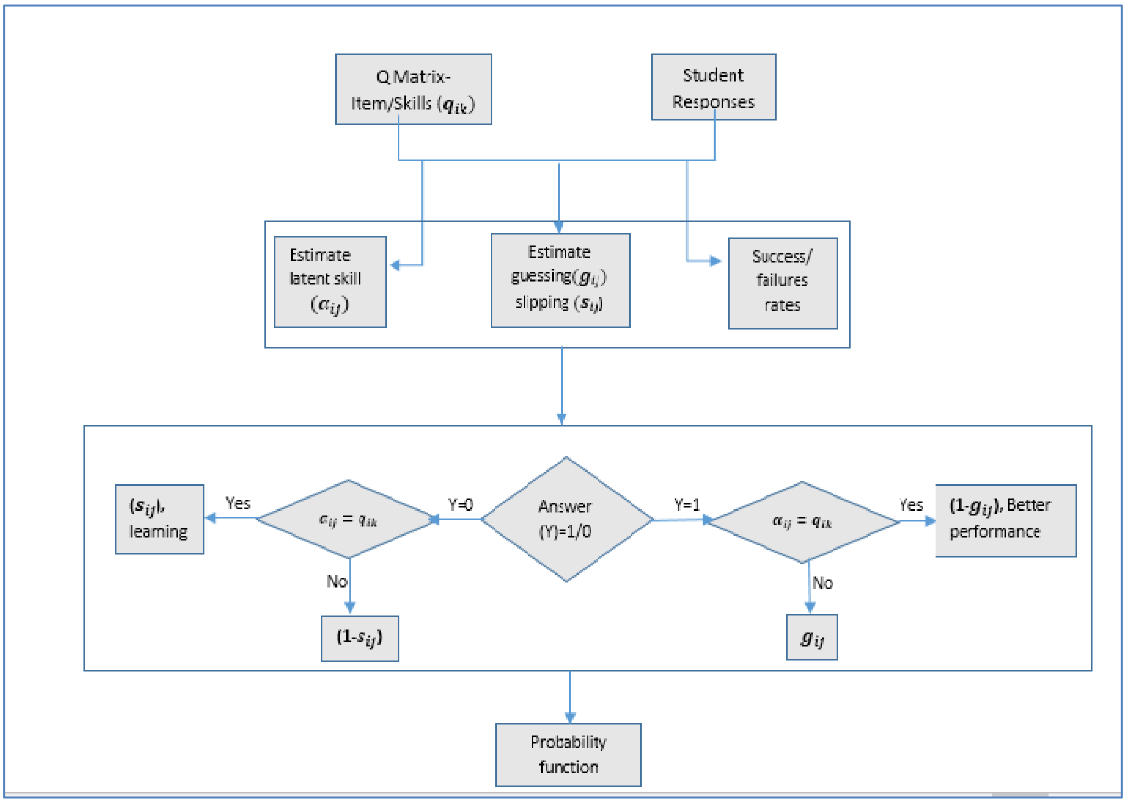



**Figure3. 1: The proposed model flow diagram**

Figure 3.1 shows the main factors' determination based model process. There are three essential cognitive factors in the calculation process:

- latent skill ($\boldsymbol{\alpha}_{ij}$) estimation to estimate the
- guessing($\boldsymbol{g}_{ij}$) /slipping($\boldsymbol{s}_{ij}$) estimation
- success rate $\gamma_k$ and failure rate $\rho_k$ determination

In order to create a better cognitive model, CFA explores incorporating the information in the existing Q matrix ($\boldsymbol{q}_{ik}$) which includes the items j, their skills k and the student's prior records which include the student's prior scores of the given items. This step enables a comparison of the student's current skills with the item skills to estimate the probability of achieving the correct answer for the next item. Therefore, the following steps should be implemented:

- Include the prior used hints with the prior correct scores and calculate these as predictive parameters. In addition, calculate the student's prior incorrect scores as a predictive parameter as well.
- Estimate the item skill's success rate $\gamma_k$ and failure rate $\rho_k$ as shown in equation (3.8)
- Determine the item difficulty level $\beta_k$ as in equation (3.8)
- Estimate the student's current skills $(\alpha_{ij})$ from his/her prior correct/incorrect scores.
- Determine whether the student has acquired the necessary skills to answer the item (the proficiency level $\eta_{ij}$), by comparing the student's latent skills with the item, i.e. if $\alpha_{ij} = q_{ik}$ then $\boldsymbol{\eta_{ij}}$=1, otherwise $\eta_{ij} = 0$.

Based on the above parameters, CFA will conduct some comparisons to estimate the guessing /slipping parameters as explained below:

- If the student has answered the item correctly (Y=1, where Y represents the score) and his/her latent skills match the item skills (i.**e** $\eta_{ij} = 1$) then the decision will be better performance with no chance of guessing (1-$g_{ij}$), otherwise, the model will calculate the probability of guessing the answer $(g_{ij})$ i.e ($\eta_{ij}$=0 and Y=1)
- If the student has not answered the item correctly (i.e, Y=0) and his/her latent skills match the item skills $(\eta_{ij} = 1)$ then this has slipped the answer and the model will calculate the probability of slipping $(s_{ij})$, and failure means the student is still learning, otherwise, (1-$s_{ij}$) i.e, ($\eta_{ij}$=0 and Y=0)
- Finally, a probability function is designed to estimate the probability of the students attempting the next item correctly given the cognitive factors: $(\gamma_k, \rho_k, \beta_k, \eta_{ij}, s_{ij}, g_{ij})$. The next section explains the mathematical process of developing the CFA equation form.

### 3.6.1 The mathematical form of the CFA model

Performance Factors Analysis (PFA) and the DINA model are two popular models of student learning that employ logistic regression and probability to estimate parameters and to predict performance. However, they differ in their parameterisation of student learning and the estimation methods. One key difference is that the DINA model has parameters for the slipping and guessing rates of learned skills, whereas the logistic model/ PFA does not.

Thus, the logistic models assume that, as students gain more practice, their probability of correctly answering the item will increase as the incorrect prior score does not mean failure but represents a lack of knowledge whereas the DINA model allows the determination of the probability of having the correct answer depending on the individual skill and item skill to determine the guessing and slipping factors. Based on this, our presented model is built on a novel modification of logistic regression that allows it to account for situations resulting in false negative student actions (e.g. slipping/guessing on known/unknown skills).

We applied this novel regression approach to create a new method PFA+Slip/guess and compare the performance of this new model with the traditional PFA and DINA models. Also, we further extended the PFA with a new parameter by splitting the success rate into (prior correct answer + number of used hints per each knowledge component).

The proposed model uses a logistic regression form to estimate student cognitive factors and to predict the student's latent performance. Therefore, the probability function that student i will answer the next item j correctly takes the following form:

$$\mathbf{P_{ij}} = \frac{\mathbf{1}}{\mathbf{1}+e^{-\mathbf{z_i}}} \quad \text{................ (3.9)}$$

Where $\mathbf{z_i}$ is the logistic regression function form of student and item parameters for item i. $\mathbf{P_{ij}}$ is the probability function student I answering item j.

There is no easy way to incorporate explicit slipping/guessing parameters into the logistic models, e.g., the prediction probability cannot be bounded by an additional term to the logistic function. In order to add these parameters, we modified the underlying logistic model to create the following form *(some of the parameters in the following equations are just to give an example which is used to represent the general logistic regression form only and they won't be applied to the CFA form in the later sections)*:

$$\mathbf{P_{ij}}\,(\mathbf{x_{i1}}, \mathbf{x_{i2}}, \mathbf{x_{i3}}, \mathbf{x_{i4}}, \ldots, \mathbf{x_{ij}} | \boldsymbol{\alpha_l}, \boldsymbol{\beta_j}, \boldsymbol{\rho_{ij}}, \boldsymbol{\gamma}_{\mathbf{ij}}) = \prod_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{p_{ij}}(\mathbf{z_i})^{\mathbf{x_{ij}}}\,)(\mathbf{1} - \mathbf{p_{ij}}(\mathbf{z_i}))^{\mathbf{1}-\mathbf{x_{ij}}} \quad \text{................(3.10)}$$

Where, $\mathbf{x_{ij}} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \ldots, \mathbf{2^k}, \boldsymbol{\alpha_l} = \boldsymbol{students\ skills}$,
$\boldsymbol{\beta_j} = \mathbf{item\ difficulty}$, $\boldsymbol{\rho_{ij}} = \mathbf{failure\ rate}, \boldsymbol{\gamma}_{\mathbf{ij}}$= success rate, k = number of the given skill, and
$\mathbf{x_{ij}} = \boldsymbol{zero\ or\ one\ depends\ on\ the\ student's\ scores}$.

The probability function is:

$$\mathbf{P_{ij}}\,(\mathbf{x_{i1}}, \mathbf{x_{i2}}, \mathbf{x_{i3}}, \mathbf{x_{i4}}, \ldots, \mathbf{x_{ij}} | \boldsymbol{\alpha_l}) = \prod_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{p_{ij}}(\boldsymbol{\alpha_l})^{x_{ij}}\,(\mathbf{1} - \mathbf{p_{ij}}(\boldsymbol{\alpha_l}))^{\mathbf{1}-x_{ij}}$$

By taking the log of the above equation,

$$\mathbf{P_{ij}} = \mathrm{Log}\,(\prod_{\mathbf{j=1}}^{\mathbf{J}} \mathbf{p_{ij}}(\boldsymbol{\alpha_l})^{x_{ij}}\,(\mathbf{1} - \mathbf{p_{ij}}(\boldsymbol{\alpha_l}))^{\mathbf{1}-x_{ij}})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad \frac{1}{1+e^{-zi}} \quad + (1-x_{ij})\log\,(1-\frac{1}{1+e^{-zi}})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad \frac{1}{1+e^{-zi}} \quad + (1-x_{ij})\log\,(\frac{1+e^{-zi}-1}{1+e^{-zi}})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad \frac{1}{1+e^{-zi}} \quad + (1-x_{ij})\log\,(\frac{e^{-zi}}{1+e^{-zi}})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad \frac{e^{-zi}}{(1+e^{-zi})e^{-zi}} \quad + (1-x_{ij})\log \quad \frac{e^{-zi}.e^{-zi}}{(1+e^{-zi})e^{-zi}}$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad \frac{e^{-zi}}{1+e^{-zi}} \quad + (1-x_{ij})\log\,(\frac{1}{(1+e^{-zi})})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad e^{-zi} \quad - x_{ij} \log 1 - (1-x_{ij})\ \log e^{-zi})$$

$$\mathbf{P_{ij}} = \sum_{j=1}^{n} x_{ij} \log \quad e^{-zi} \quad - x_{ij} \log 1 - (1-x_{ij})\ \log e^{-zi})$$

$$\mathbf{P_{ij}} = \textstyle\sum_{j=1}^{n} x_{ij}\, z_{ij-} \ \log(\,1+e^{-zi})$$

The probability of the multiple regression model can be calculated as:

$\mathbf{P_{ij=}}\ \prod_{\mathrm{j=1}}^{\mathrm{J}} \frac{1}{(1+e^{-z_i})}$............ (3.11)

Where $z_i$ refers to any multiple logistics models and in CFA, it represents the modified PFA and DINA model.

According to the PFA model, the probability of having the correct answer can be calculated as:

$$\frac{1}{1+e^{-Mod\,m_i}}$$

Where $Modm_i$ is the modified PFA model with the addition of the hint parameters to the correct answer, shown as equation (3.8)

And, based on DINA model, $P_{ij} = (1 - s_{ij})^{\eta_{ij}} \quad g_{ij}^{1-\eta_{ij}}$

Therefore, the modified logistic/probability function for CFA according to equation (3.9) can be derived as:

$$P_{ij} = \frac{1}{1+e^{-Modm_i}} \times \left[\frac{1}{1+e^{-(1-s_{ij})}}\right]^{\eta_{ij}} \times \left[\frac{1}{1+e^{-g_{ij}}}\right]^{(1-\eta_{ij})} \quad \ldots\ldots\ldots\ldots\ldots \quad (3.12)$$

Where $s_{ij}$, $g_{ij}$ are the parameters that impose the guessing and slipping probability for the student i/ item j, and , $\eta_{ij}$ is the proficiency level of student i corresponding to item j

## 3.7 The data

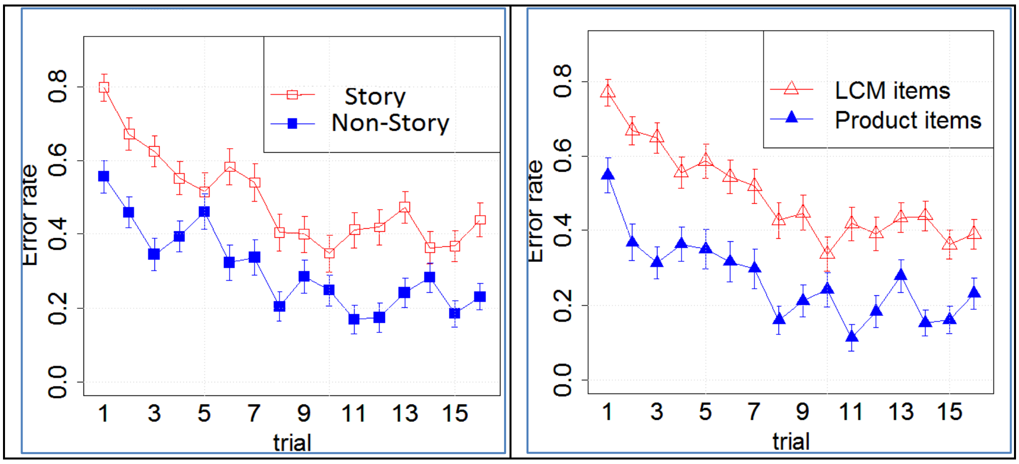
The data used in this thesis was obtained from student activity recorded by a modified Bridge to Algebra (BTA) tutor by Carnegie Learning (http://www.carnegielearning.com) and https://pslcdatashop.web.cmu.edu/). This data contains 255 students who completed all 16 assigned problems. This type of assessment is multiple choice, each item having four options. The number of options ensures that the student can learn from the given choices to reach the correct answer as there are more alternatives than distractors. Each item is provided with one hint which the student can request. The texts of two of the problems are given below in Table 3.2 as examples. The students' scores are represented as either a correct or an incorrect action. The students' records consists of a unique anonymous identifier for each student (student id), item id, number of the used skills required for each item and the student's scores (i.e. whether the student attempted the item correctly or not, 1 meaning success and 0, failure).

**Table3. 3: problem items / skill types**

| Problem | Skill |
|---|---|
| 1: Sally visits her grandfather every 4 days and Molly visits him every 6 days. If they are visiting him together today, in how many days will they visit together again? | Story- Product/ LCM |
| 2: What is the least common multiple of 4 and 9? | Non Story-Product/LCM |

According to the table 3.3, each problem has two important properties. Firstly, problem 1 is a so-called story problem and problem 2 is a non-story problem. Story problems require the use of a concrete strategy. Figures 3.2 (a) and (b) show the difficulty level of the four skills for the given 16 items. The difficulty level is measured through the error rate per each trial that the students might have. As shown in figure 3.2 (a), the story problems are generally harder. Each given item, whether story or non-story, can be further divided into two extra skills, which are LCM/Product. The least common multiple (LCM) could be correctly obtained by multiplying the two inputs. In this case, the problem can be solved by applying a partial problem-solving strategy. As shown in figure 3.2 (b), LCM problems were harder.

As shown in Figures 3.2 (a) and 3.2 (b), the error rate curves respectively of LCM problems and story problems are reliably higher. When these two properties are crossed, the LCM/story problem represents the harder combination of the properties and the Product/non-story, the easier one. These two figures explain how item skills can be combined and affect the success/failure rates and student performance estimation. However, this will be further discussed in chapter four where certain experiments are conducted to show how the students' factors estimations can be affected by items of two skills and four skills.




(a) (b)

**Figure 3. 1: Error rate comparison**

**Table3. 4: Q matrix for the data**

| Item<br>Item No. | Skill 1<br>Non Story/Product | Skill2<br>Non Story/LCM | Skill3<br>Story/LCM | Skill 4<br>Story/Product |
|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 |
| 2 | 0 | 0 | 1 | 0 |
| 3 | 0 | 1 | 0 | 1 |
| 4 | 1 | 1 | 0 | 1 |
| 5 | 0 | 0 | 1 | 1 |
| 6 | 0 | 0 | 1 | 0 |
| 7 | 1 | 0 | 0 | 0 |
| 8 | 0 | 1 | 1 | 0 |
| 9 | 1 | 0 | 0 | 1 |
| 10 | 1 | 1 | 0 | 1 |
| 11 | 1 | 1 | 0 | 1 |
| 12 | 1 | 1 | 1 | 1 |
| 13 | 0 | 1 | 0 | 0 |
| 14 | 1 | 0 | 1 | 1 |
| 15 | 1 | 0 | 1 | 0 |
| 16 | 1 | 0 | 0 | 1 |

Table 3.4 describes the data Q matrix of the given data which consists of 16 items and their four skills. This Q matrix was used to estimate the model parameters. Each assessment item could contain one or more skills or knowledge components (story-LCM, story/Product, non-story/LCM, non-story/Product).

### 3.7.1 Calculating the student's latent skills

To infer the examinees' possession of the $K$ skills, the pattern $\mathbf{2}^{\boldsymbol{K}}$ builds all possible skill combinations.

These patterns are called binary skill vectors: $\boldsymbol{a}_{\boldsymbol{l}} = [\boldsymbol{a}_{\boldsymbol{j1}}, \ldots \boldsymbol{a}_{\boldsymbol{jk}},]$, $j$= 1, . . . , $\mathbf{2}^{\boldsymbol{K}}$. Each element $\boldsymbol{a}_{\boldsymbol{lk}}$ denotes whether or not members of skill class $l$ possess the skill $k$ (i. e., $\boldsymbol{a}_{\boldsymbol{lk}}$=1 or $\boldsymbol{a}_{\boldsymbol{lk}}$=0, respectively).

It therefore allows the allocation of students into $\mathbf{2}^{\boldsymbol{K}}=\mathbf{2}^{\mathbf{4}}$=16 different skill classes: $\boldsymbol{a}_{\mathbf{1}} = [0,0,0,0]$, $\boldsymbol{a}_{\mathbf{2}} = [1,0,0,0]$, $\boldsymbol{a}_{\mathbf{3}} = [0,1,0,0]$, $\boldsymbol{a}_{\mathbf{4}} = [0,0,1,0]$, ..., $\boldsymbol{a}_{\mathbf{16}} = [\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}]$.

CFA formulation will answer the following questions, which are addressed in the model's output:

(Q1) "What is the proportion of examinees acquiring a specific skill $\boldsymbol{a}_{\boldsymbol{k}}$?"

The skill distribution $P(\boldsymbol{a}_{\boldsymbol{k}})$, $k$ =1, . . . ,$K$, quantifies this question and refers to the *population skill possession* question.

(Q2) "Which skills does the $i$ -th individual examinee possess?"

The $i$ -th examinee's skill profile $\boldsymbol{a}_{\boldsymbol{ij}}$ =[$\boldsymbol{a}_{\boldsymbol{i1}}$, . . . , $\boldsymbol{a}_{\boldsymbol{ik}}$ ], $i$ =1, . . . , $I$ , provides this information and refers to the *individual skill possession* question.

As previously discussed in chapter two, the DINA model classifies each item into those who possess all the required attributes/item skills and those who do not have sufficient skills to possess the required attributes. This can be viewed as the latent student skills. Let $\boldsymbol{\eta}_{ij}$ denote whether the $\mathbf{i}_{th}$ student possesses the attributes required for the $\mathbf{j}_{th}$ item. This can be expressed by equation 3.11:

$$\boldsymbol{\eta}_{ij} = \prod_{k=1}^{k} \boldsymbol{\alpha}_{ik}^{qik} \text{.............. (3.11)}$$

Where: $\boldsymbol{\alpha}_{ik}$ represents the binary vector for the current student skills.

$\boldsymbol{q}^{ik}$ is the binary vector for the skills per item in the Q matrix.

Note that there are binary indicators signifying whether the *i*th examinee possesses all the required skills for item *j*. For example, consider a 16-item exam diagnosing four skills (as has been explained earlier). Suppose item 1 requires skills 1, 2, and 3 (i.e. the binary skills vector $q^{ik} = 1110$, and student 1 possesses all four skills (i.e. 1111). Then, $\eta_{ij}$ will be calculated as:

$\prod_{k=1}^{k} \alpha_{ik}^{qik} = 1^1 * 1^1 * 1^1 * 1^0 = 1$, indicating that the examinee possesses all the required skills. In contrast, suppose student 2 possesses skills 1and 2. Then, for item 1,

$\eta_{ij} = 1^1 * 1^1 * 0^1 * 0^0 = 0$, indicating that the examinee is lacking at least one required skill.

### 3.7.2 Parameter Estimation for Cognitive Diagnosis Models

Any successful CDM implementation depends heavily on the ability to accurately classify students' skills as mastered or non-mastered. Obviously, true skills for each examinee are unknown. According to literature, there are three main methods by which to estimate and classify these skills correctly. Examinees are often classified via maximum likelihood estimation (MLE), maximum a posteriori (MAP) or expected a posteriori (EAP). Below is a brief description of all three methods with their mathematical forms.

- **Maximum Likelihood Estimation (MLE) and Maximum A Posterior (MAP):**

MLE and MAP are well-known methods for examinee scoring in IRT models and are similar concepts when viewed in the CDM setting.

To compute the probability of student i's performance for each item j given the student skills vector:

Let vector x= $(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}...\boldsymbol{x_{ij}})$ denote the student (i ) skills for each given item (j).

The likelihood of the responses to J items for the ith examinee is given by (Cheng, 2009) and is determined as following:

$$P(\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \boldsymbol{x_{i3}}, \boldsymbol{x_{i4}}, \dots, \boldsymbol{x_{ij}} | \boldsymbol{\alpha_l}) = \prod_{j=1}^{J} \boldsymbol{p_j}(\boldsymbol{\alpha_l})^{x_{ij}} )(\mathbf{1} - \boldsymbol{p_j}(\boldsymbol{\alpha_l}))^{1-x_{ij}}$$

Where, $\boldsymbol{l} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \dots, \mathbf{2}^{\boldsymbol{k}}$.

$$L(\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \boldsymbol{x_{i3}}, \boldsymbol{x_{i4}}, \dots, \boldsymbol{x_{ij}} | \boldsymbol{\alpha_l}) = \boldsymbol{p_j}(\boldsymbol{\alpha_l})^{x_{ij}} )(\mathbf{1} - \boldsymbol{p_j}(\boldsymbol{\alpha_l}))^{1-x_{ij}}$$

For MLE classification, the likelihood is computed at each $\boldsymbol{\alpha_l}$, the true examinee a is unknown, and the examinee is assigned the estimated skill pattern $\boldsymbol{\alpha}$. The examinee is classified by assigning to him or her the estimated skill pattern $\boldsymbol{\alpha_{mle}}$ that maximises the likelihood, as shown below:

$$\boldsymbol{\alpha_{mle}} = \arg\max L((\boldsymbol{x_{i1}}, \boldsymbol{x_{i2}}, \boldsymbol{x_{i3}}, \boldsymbol{x_{i4}}, \dots, \boldsymbol{x_{ij}} | \boldsymbol{\alpha_l})$$

On the other hand, there are certain cases that determine the proportion of examinees possessing each skill pattern. These proportions can be regarded as prior probabilities, so then the posterior probability can be computed according to Bayes' Theorem:

$$\frac{\mathbf{P}(\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},\boldsymbol{x_{i4}},\ \dots,\boldsymbol{x_{ij}}|\boldsymbol{\alpha_l})\mathbf{P}(\boldsymbol{\alpha_l})}{\mathbf{P}(\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},\boldsymbol{x_{i4}},\ \dots,\boldsymbol{x_{ij}}|\boldsymbol{\alpha_l})\mathbf{P}(\boldsymbol{\alpha_1})+..+\mathbf{P}(\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},\boldsymbol{x_{i4}},\ \dots,\boldsymbol{x_{ij}}|\boldsymbol{\alpha_l})\mathbf{P}(\boldsymbol{\alpha_k})}$$

$$= \frac{\mathbf{P}(\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},\boldsymbol{x_{i4}},\dots,\boldsymbol{x_{ij}}|\boldsymbol{\alpha_l})\mathbf{P}(\boldsymbol{\alpha_l})}{\sum_{m}^{2^k}\mathbf{P}(\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},\boldsymbol{x_{i4}},\dots,\boldsymbol{x_{ij}}|\boldsymbol{\alpha_l})\mathbf{P}(\boldsymbol{\alpha_l})} \dots\dots\dots\dots(3.12)$$

And the estimator will be denoted as: $\boldsymbol{\alpha_{mle}}$=arg max P(($\boldsymbol{\alpha_{l|}}\boldsymbol{x_{i1}},\ \boldsymbol{x_{i2}},\boldsymbol{x_{i3}},$ $\boldsymbol{x_{i4}},\ \dots,\boldsymbol{x_{ij}}$).

Generally, it is true that MLE is equivalent to MAP estimation. See Henson and Douglas (2005); Henson, Roussos, Douglas; and Cheng (2009) for examples of studies using the MLE/MAP method of classification.

- **Expected A Posterior (EAP)**

An alternative to the MLE/MAP classification is EAP. For the EAP approach in the CDM context, the probabilities of the mastery of each individual skill (the marginal skill probabilities) are calculated for an examinee and rounded at .50 to obtain binary mastery classifications. The posterior probabilities P($\boldsymbol{\alpha_l}|\boldsymbol{x_{ij}}$) are computed for l = 1,2,3, . . . , L as in MAP/MLE, but these posterior probabilities take the sum of all estimated $\boldsymbol{\alpha_l}$ corresponding to the mastery of skill k, that is, the sum of the posterior probabilities of all the skill patterns having a 1 as the kth element, as shown below:

EAP= $\sum_{l=1}^{L} \mathbf{P}(\boldsymbol{\alpha_l}|\boldsymbol{x_{ij}})\ \mathbf{I}(\boldsymbol{\alpha_{lk}} = \mathbf{1}\ )$, Here, $\boldsymbol{I}(\boldsymbol{\alpha_{lk}} = \mathbf{1}\ )$ is a binary indicator denoting if the kth element of the lth skill pattern is 1.

Many studies have been conducted to investigate which estimator is more accurate, and it has been shown over all conditions that applying the MLE/MAP method of classification resulted in higher numbers of examinees classified correctly on all K skills, whereas the EAP method of classification resulted in higher total skills classified correctly and fewer severe misclassifications (Huebner, A., & Wang, C. ,2011).

For example, a situation is described in Jang (2008) in which a teacher who administers a diagnostic exam to English as a Second Language (ESL) students to measure progress on the completion of a teaching unit may prefer the use of EAP; it may be desirable to classify ''most'' students ''mostly'' correctly rather than to classify a maximum number of students exactly correctly while having a higher number of severe misclassifications. As a result, we have used the EAP method to estimate the student's binary vector in order to estimate his mastery level of the given skill. However, while it is not the case that one classification method may be judged as better than the other, one method may be preferred over the other, depending on the purpose of the diagnostic assessment. When implementing CDM methodology for operational use, practitioners must decide which method is most consistent with the aims of the assessment in question. For some testing conditions, the reported differences between the estimation methods may be too small to make any practical difference in the classification of examinees in real life situations.

## 3.8 CFA example

In the previous sections, we presented a novel modification of logistic regression (CFA) which combined the parameters of both the PFA and DINA models. From one aspect, it further extended the PFA with a new parameter (which is the hint parameter) by splitting the success rate into (prior correct answer + number of used hints per each knowledge component). From another aspect, the modification accounts for situations that result in false negative student actions (e.g., slipping/guessing on known/unknown skills).

This section illustrates an example of how the CFA algorithm can be applied and how the output is demonstrated. Since this section adopts the same parameter values explained in chapter two, section 2.7, the same calculation processes will be used to apply the numbers to the given equations.

Furthermore, this section will also demonstrate the difference between the CFA, PFA and DINA models and how CFA has been developed to improve the performance estimation method. As explained in chapter two, CFA will be applied to two items (T1 and T2) with four different skills and one student.

Table 3.4 shows the details of student A attempting two items, T1 and T2 with each item having one or more skills out of four skills. These details are the same details presented in chapter two which were used to demonstrate the process of applying the DINA and PFA models in section 2.7. Each item includes its own guessing/slipping parameters which have been determined as will be shown later in chapter four section 4.4. Furthermore, this table shows the scores of and the hints given to student A for both items.

**Table3. 5**: CFA estimated parameters

| Student ID | Item ID | Skills | Slipping | Guessing | Hint | Score |
|---|---|---|---|---|---|---|
| A | T1 | 1011 | 0.24 | 0.53 | 1 | 1 |
| A | T2 | 1001 | 0.35 | 0.42 | 0 | 0 |

**To begin with, the parameters of CFA will be determined as shown in table 3.5:**

- Calculate the guessing/slipping parameters for both items T1 and T2;
- Determine the student current skill $\alpha_{ik}$, which has been estimated using the EAP method, as previously explained in section **3.5.2.** In this example, student A's current skills: $\alpha_{Ak}$ = 1 1 0 1;
- Determine the item skills $q_{jk}$, which have been excluded from the Q matrix. In this example, each item has four skills: for item T1, $q_{T1k}$= 1 0 1 0; for item T2, $q_{T2k}$=1 0 0 1;
- Calculate the student proficiency level ($\eta_{ij=} \prod_{k=1}^{k} \alpha_{ik}^{q_{jk}}$) for items T1 and T2: While this has been conducted previously in section 2.7, the same process is repeated below, as follows:
  -For item T1, $\eta_{AT1}$ =$1^1$*$1^0 * 0^1$*$1^0$=0. The student lacked the third skill;
  -For item T2, $\eta_{AT2}$ =$1^1$*$1^0 * 0^0$*$1^1$=1. The student mastered all the skills;
- Estimate the item difficulty and success/failure rates (as calculated previously from the data set explained in this chapter);

**To estimate the probability of student A having a correct answer for item T1, the CFA model equation is called:**

$$\boldsymbol{P}_{ij}=\frac{1}{1+e^{-Modm_i}} \times \left[\frac{1}{1+e^{-(1-s_{ij})}}\right]^{\eta_{ij}} \times \left[\frac{1}{1+e^{-g_{ij}}}\right]^{(1-\eta_{ij})} \qquad \text{...............(3.12);}$$

therefore, the following two steps will be applied:

- Firstly, $Modm_i$ for item T1 (which has skills 1, 2 and 3) is calculated using the same values provided in Table 2.10. Furthermore, CFA considers the initial knowledge for student A to be zero (i.e. the student has no prior skills), so that the values of prior success ($\boldsymbol{succ}_{i,j}$)/failure$\boldsymbol{fail}_{i,j}$/hints($\boldsymbol{H}_{i,j}$)=0 for all the three skills), therefore:

$$\boldsymbol{Modm}_i = \sum_{j \in KCs}^{n} \boldsymbol{\beta}_j + \boldsymbol{\gamma}_j \boldsymbol{succ\_total}_{i,j} + \boldsymbol{\rho}_j \boldsymbol{fail}_{i,j} \quad \text{.........(3.8)}$$

$$[\boldsymbol{\beta}_1 + \boldsymbol{\gamma}_1(\boldsymbol{succ}_{A,1} + \boldsymbol{H}_{A,1}) + \boldsymbol{\rho}_1 \boldsymbol{fail}_{A,1}] + [\boldsymbol{\beta}_3 + \boldsymbol{\gamma}_3(\boldsymbol{succ}_{A,3} + \boldsymbol{H}_{A,3}) + \boldsymbol{\rho}_3 \boldsymbol{fail}_{A,3}] + [\boldsymbol{\beta}_4 + \boldsymbol{\gamma}_4(\boldsymbol{succ}_{A,4} + \boldsymbol{H}_{A,4}) + \boldsymbol{\rho}_4 \boldsymbol{fail}_{A,4}] = \mathbf{0.99}$$

- Secondly, to estimate the performance probability, all parameters will be applied to the CFA exponential equation in (3.12); therefore, the probability of student A choosing the next item correctly given the student skills $\boldsymbol{\eta}_{ij}$, item difficulty $\beta_j$, success rate $\boldsymbol{\gamma}_j$, failure rate $\boldsymbol{\rho}_j$, guessing parameter$\boldsymbol{g}_{ij}$ and slipping parameter $\boldsymbol{s}_{ij}$ is: $\boldsymbol{P}_{ij} = 0.45$.

Based on the predicted probability value, CFA estimates that student A has guessed the answer for two reasons:

- Student A has insufficient knowledge to attempt the question correctly (lacks the third skill);
- CFA estimated that the probability of student A attempting the answer correctly is 0.45, which is a small value, but the student attempted the answer correctly. This offers strong evidence that the student has guessed the answer and CFA produced more accurate probability compared to other models (PFA =0.72, DINA= 0.53)
- According to table 3.5, student A asked for a hint; therefore, s/he is more likely to guess the answer, as s/he might have learned from the hint and made a logical guess.

Table 3.5 shows that the student scored the item correctly and asked for a hint; therefore, the prior success score is 1 and prior failure score is 0 for item T2 (which has skills 1 and 4), therefore, ($\boldsymbol{succ_{A,1}}$=**0**, $\boldsymbol{succ_{A,4}}$=**0**/$\boldsymbol{fail_{A,1}}$=1, $\boldsymbol{fail_{A,4}}$=1/$\boldsymbol{H_{A,1}} = \mathbf{1}, \boldsymbol{H_{A,4}} = \mathbf{1}$).

However, since this student has asked for a hint, this means that s/he might have developed some knowledge for the third skill (it is likely that s/he made a logical guess to achieve the correct answer). Therefore, the level of learning materials that might be recommended for this student could be more advanced, as s/he is able to learn faster, meaning that s/he does not need to waste time studying at the basic level.

The same calculations have been made for item T2, this time with the prior correct score = 1 and prior used hints = 1. The probability of having the correct answer is 0.82; hence, the probability value is high as the student has mastered all the given skills in item T2 and he is supposed to answer the item correctly. Besides this, item T2 is easier and skills 1 and 4 are repeated. However, the student scored the item incorrectly without asking for a hint. From the perspective of the CFA model, answering item T2 incorrectly means:

- The student has slipped the answer as s/he had mastered all the given skills (having the efficiency level =1).
- The student has learnt from prior success/failure scores, as item T2 shares two skills with item T1;
- Since CFA included the prior hint parameter, the student might have a chance of learning from the provided hints, hence the student's performance can be enhanced and this will contribute to a reduction in the student's study time.
- The student does not require further learning materials for the given skills as s/he already has acquired the skills but s/he slipped the answer.

Based on that which has been explained in the examples of PFA, DINA and CFA models, the CFA model aims not only to predict the student's performance, but also helps the student to target strong and weak points at their own knowledge level. By including cognitive factors (current student skills, slipping and guessing parameters) along with the student's prior hints, the accuracy of the student's performance estimation will be increased; hence, accurately detecting the knowledge level will save the student learning time and will result in a better learning curve representation.

Moreover, unlike the DINA and PFA models, the CFA model does not consider the student's correct answer to be positive evidence that they have mastered all the required skills, as the student might have guessed the answer or might have used a hint to make a logical guess. Similarly, an incorrect answer is not proof of failure, as the student might have made a mistake with this answer despite having mastered all the required skills. However, CFA considers the hint parameters to be a learning tool since the student's knowledge level might improve when they request the hint. This will have an adverse impact on the student's learning curve and will not waste the student's time as they have practiced the skill several times.

**Table3. 6**: Parameters of DINA, PFA and CFA models

| **Model** | **Parameters used** | **Parameters not used** |
|---|---|---|
| **DINA** | Guessing/slipping/student proficiency level | Prior used hints, item difficulty, success/failure rates, prior correct/incorrect scores |
| **PFA** | Item difficulty, success/failure rates, prior correct/incorrect scores | Prior used hints, guessing/ slipping/student proficiency level |

| | | |
|---|---|---|
| **CFA** | Guessing/slipping/student proficiency level, item difficulty, prior used hints, prior correct/incorrect scores, success/ failure rates | None |

## 3.9 Conclusion

This chapter has presented two novel models for estimating the performance of students. The first model extends the previous work of PFA by splitting the success rates into (the prior correct answer + prior used hints). Based on the first model and the considerations of students' multiple prior factors, this chapter presented a second model to estimate the probability of achieving the correct answer by adding the slipping/guessing parameter in the form of logistic regression. This model is called the Cognitive Factor Analysis (CFA) and has been developed as a new alternative cognitive model that can be used in Intelligent Tutoring Systems. CFA's flow diagram was shown to reveal the overall hierarchy of the factors. Furthermore, the theoretical part of this model was presented to explain the full model formula.

The data used to apply this model was discussed with all its components and items.

# Chapter Four: Model Evaluation

## 4.1 Introduction

This chapter presents the analysis of the data explained in the previous chapter according to our two novel models, modified PFA with hints factor and CFA, and compares these to the DINA and PFA models. The results were evaluated using statistical methods to assess the presented prediction accuracy and significance of the cognitive model. This evaluation used various methods, e.g. Log Likelihood (LL), Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and Cross Validation (CV). Evidence has been foundwhich confirms that the slip/guess parameters enable the logistic models to better fit learning rates. In addition, the combination with the DINA model offers a better probabilistic indication for each individual skill and enhanced feedback to improve learning materials.

The first part of this chapter will explain the statistical methods used to evaluate the models, while the later sections will focus on the implementation of the extended version of the PFA with the hints factor. Experiments were also conducted to compare the original PFA and our modified version. The later sections run the DINA model using the same dataset as used in the PFA and provides a full explanation of the obtained results. Finally, our developed model (CFA) is applied to the dataset and all the factors are extracted and analysed to show the significance of this model compared to the other models. Two demonstration scenarios are also described to explain how CFA can be applied as an Intelligent Tutoring System.

## 4.2 Assessment tools

Fitting a mathematical model of user behaviour to data is difficult (Yudelson, Pavlik & Koedinger, 2011). An effective statistical model balances model fit and complexity, while at the same time minimises prediction risk. It should capture sufficient variation in the data but should not be overly complicated. Unfortunately, inexperienced modellers often transfer basic knowledge of statistics directly to their cognitive models, which typically results in confusion (Wasserman, 2004).

Certain statistical methods, such as the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978) are considered to be good estimators for prediction risk. Although such tools are routinely used in many statistical packages, they sometimes lead to inappropriate inferences, particularly when the user data is not almost entirely independent (a requirement of the AIC and the BIC). In reality, user observations (prior user data) are often dependent on and related to users, skills or content items that the user has previously accessed.

Therefore, in addition to the BIC and the AIC, the evaluation of prediction performance can be assessed by computing an error measure, such as the root-mean-square-error (RSME) and residual analysis (Chai & Draxler, 2014). In addition to log likelihood (LL), which measures the fit of the data, there is k-fold cross validation (CV) which, although time-consuming, is a more accurate estimator of prediction errors that measures over-fitting (Kohavi, 1995). These models are shown in the following equations:

AIC=-2*Logliklihood+2*number of parameters ……. (4.1)

BIC=-2*Loglikelihood+ number of parameters* number of observations …… (4.2)

RMSE=$(\sqrt{\sum_{i=1}^{n} \boldsymbol{e^2}})/\boldsymbol{n}$ …… (4.3)

where e is the difference between the actual and estimated value and n is the number of observations.

CV=$\sum_{i=1}^{n}(\boldsymbol{Y}-\hat{\boldsymbol{f}}^{-k(i)}(\boldsymbol{x_i}))$^2 ………. (4.4)

where k is the number of folds, n is the number of observations and (Y, f) are the actual and estimated values of X.

To measure the effectiveness of the CFA model, AIC, BIC and LL have been used as assessment tools to penalise the model for containing a larger numbers of parameters.

In addition, CV has been applied to assess the model parameters which involve feature selection. This mechanism assesses a model by rotating the testing and training datasets. The dataset is separated into K subsets and training is performed using k-1 subsets. The remaining set is used for testing. This mechanism is repeated until each subset has been used once for testing. The mean performance across all the testing subsets is then assumed as a measurement of the effectiveness of the system and an indication of the ability of a model to generalise across all situations. K-fold CV has been shown to be the most effective way of estimating the accuracy of a predictive model.

When evaluating cognitive models, the different fit statistics often agree with each other but, if they do not, it is useful to understand how their conclusions differ. The AIC and BIC punish a model for including too many skills, particularly the BIC. The CV values are arguably a more rigorous measure of fit because they evaluate predictive ability against over-fitting and they are more useful for evaluating the ability of a model to predict new items.

Additionally, item stratification gives a sense of how well a skill transfers between items within a tutor or between levels in a game.

## 4.3 PFA and ModPFA

This section demonstrates the fitting evaluations of the data that were applied to the original PFA and ModPFA (hints parameter). This fitting was deployed by running experiments and assessing the results using the tools explained in the previous section. These experiments were conducted on two types of items. The first experiment measures the estimates of items with two skills (product-ns (non-story) and LCM-ns (non-story)). The second experiment was applied to items after splitting them into four skills for each item (product-non story (ns), LCM-ns (non-story), product–s (story) and LCM-s (story). These experiments examine student behaviours and whether the ModPFA could enhance the prediction of the cognitive model.

### 4.3.1 First experiment

This experiment attempts to answer the question: how can student learning behaviour be described in terms of an existing cognitive model? Specifically, the aim was to discover the learning rate (success and failure) and initial difficulty level of each item to estimate the initial performance of students. The question was answered by setting an experiment using the data previously explained in chapter 3. This data relates to the students' prior scores/used hints as well as the item/skills Q-matrix. The prior scores/hints were calculated using the predicted parameters process as illustrated earlier in chapter three. The data was fitted by the logistic regression models for both the original PFA and ModPFA (which was developed using hint parameters). By applying the data to the formulae, coefficient estimations for the skills and students and overall model statistics were obtained (see Tables 4.2 and 4.3).

**Table 4. 1: Fit statistics of given PFAns and ModPFA**

| Model | LL | BIC |
|---|---|---|
| PFAns | -2133 | 4307 |
| ModPFAns | -1563 | 3173 |

Table 4.1 provides a summary of the statistical fit for the original and the modified PFA models (here the PFA and ModPFA models were termed PFAns and ModPFAns as they were applied only to two skills items which were non-story skills/ns). The fit evaluation parameters used were the LL and the BIC. The low value of the BIC indicates that the estimated values of

the model are more likely to be true, i.e. the probability function of the logistic model (ModPFAns). The low value of the BIC means the error rate seems to be less in ModPFAns when compared to PFAns. This indicates a better performance estimation when the hint parameter is adopted in the ModPFA formula.

**Table 4. 2: Parameter estimation for both PFAns and ModPFAns models**

| PFA ns | Est. | Std. Err | p-value | Mod PFA | Est. | Std. Err | p-value |
|---|---|---|---|---|---|---|---|
| Difficulty/Skill1 | 0.452 | 0.079 | 0.000 | Difficulty/Skill1 | 0.676 | 0.073 | 0.000 |
| Difficulty/Skill2 | 0.347 | 0.070 | 0.000 | Difficulty/Skill2 | 0.132 | 0.006 | 0.000 |
| Succ. prod/Skill1 | 0.118 | 0.087 | 0.000 | Succ. prod/Skill1 | 0.110 | 0.062 | 0.033 |
| Fail. prod/Skill1 | -0.110 | 0.037 | 0.003 | Fail. prod/Skill1 | 0.046 | 0.112 | 0.173 |
| Succ .lcm/Skill2 | 0.254 | 0.026 | 0.000 | Succ .lcm/Skill2 | -0.007 | 0.048 | 0.009 |
| Fail. lcm/Skill2 | -0.028 | 0.021 | 0.189 | Fail. lcm/Skill2 | 0.667 | 0.036 | 0.047 |

In Table 4.2 the estimated values for both models with two skills items are shown. These parameters are defined as the difficulty level for skills 1 and 2, the success rate for skills 1 and 2 and the failure rate for skills 1 and 2. These parameters were conducted for both the PFA and ModPFA models.

According to table 4.2 PFAns is acceptable; however, the negative value of the failure rate indicates negative learning or no effect on student future performance. Also, the item difficulty has a positive value, which means that the item difficulty value is average. Therefore users' performance becomes worse after failing the product item (skill 1).

Moreover, the success rate seems to be very limited although the students have used hints to attempt the item but PFA does not consider hints in its performance estimation.

An explanation for this is that the PFAns model is not complex enough since the only way to distinguish higher achieving students (with fewer errors and hints) from lower achieving students (with more errors and hints) is to place more weight on the success rate factor by determining the number of used hints for each knowledge component/skill. Unlike the PFAns, the ModPFA estimated values are more accurate in terms of success/failure and there are no issues with the failure rates. One aspect that does cause concern is the negative value of the success rate of skill 2 which infers that the level of success is low as skill 2 is slightly difficult.

Because of the extension of the correct scores to include the hint parameter, ModPFAns estimated the success rates more accurately in skill 2. Therefore, even when the students used hints, the success level was still low since the students might depend on the hints to achieve the correct answer. Thus, in addition to offering an improved performance estimation in terms of the adoption of hints parameters, ModPFA can also provide useful feedback based on the knowledge component/skill, item difficulty and success/failure rate parameters. The success rate was split into the prior number of correct answers and the prior number of used hints. This is the concept of dynamic assessment upon which this research is based (as explained in chapter two).

### 4.3.2 Second experiment

In the first experiment, ModPFA improved the original model by splitting the success rate into prior correct scores and prior used hints. It was implemented with two skills items. However, the second experiment further addresses model improvement by splitting the given skills into more detailed skills. This then raises a further question: are some combined skills better than if they are split in terms of the success and failure rates and performance estimations?

Table 4.4 shows that the difficulty level in the LCM (skills 3 and 4) item type is higher than in the product (skills 1 and 2) item type. However, extending the model to include more skills adds additional differentiation. Within both the LCM and product types, the skill story intercept (LCM-s/prod-s) is always lower than the non-story (LCM-ns/prod-ns), which indicates that this skill is more difficult. This phenomenon can be explained for the other skills of the story and non-story significant parameters, such as the success and failure rates for the knowledge component. The success rate for a product item type shows no significant difference in the formula. However, after splitting the model into four skills, the success level for product story/skill 2 shows a significant value and is an indication of student performance (see Table 4.3).

**Table 4. 3: Parameter estimation for ModPFA Models/four skills**

| Skills | Par | Std. Err | p Value |
|---|---|---|---|
| Difficulty/skill1/prod-ns | 0.45 | 0.077 | 0.000 |
| Difficulty/skill2/prod-s | -0.11 | 0.087 | 0.000 |
| Difficulty/skill3/lcm-ns | 0.37 | 0.072 | 0.000 |
| Difficulty/skill4/lcm-s | 0.17 | 0.094 | 0.103 |
| Success-ns/skill1/ prod-ns | 0.38 | 0.086 | 0.000 |
| Success/skill2 /prod-s | 0.27 | 0.0370 | 0.0213 |
| Success/skill3/lcm-ns | 0.33 | 0.098 | 0.124 |
| Success/skill4/lcm-s | 0.38 | 0.141 | 0.000 |
| Fail/skill1/ prod-ns | 0.28 | 0.067 | 0.089 |
| Fail/skill2/prod-s | 0.42 | 0.087 | 0.000 |
| Fail/skill3/lcm-ns | 0.43 | 0.075 | 0.007 |
| Fail/skill4/lcm-s | 0.104 | 0.030 | 0.000 |

Furthermore, the success rate parameters for the LCM item types (skills 3 and 4) remain positive. However, skill 4 has a smaller value than skill 3 due to the item having greater difficulty. The failure rates are positive which reflects an increase in learning. This indicates that having multiple skills in items increases the learning processes of the student as some skills might be highly correlated to each other. Furthermore, considering the hint will improve the student's current skills and s/he might begin to learn from it. In summary, the complexity of our modified PFA models improves the fit and better explains student learning with different knowledge components.

As shown in Table 4.4, the ModPFA model exhibits a better fit than the original model. The higher value of the LL indicates better estimation of the performance probability as it is closer to the true value of the students' current skills, learning/failure rates and the item difficulty. This value was obtained by extending the ModPFA model to include hint parameters and gives strength to the cognitive analyses. The AIC and BIC are both penalised-likelihood criteria and are sometimes used for selecting the best predictor subsets. They are also often used to compare non-nested models which ordinary statistical tests cannot do.

Since, the AIC or BIC for a model is usually written in the form [$-2\log L + KP$], where $L$ is the likelihood function, $P$ is the number of parameters in the model and $K$ is 2 for the AIC and $\log(n)$ for the BIC. Therefore, a lower value for the BIC and the AIC reflects better estimation of the posterior probability, which is the probability of having the correct answer for the next item for each student. The estimations are very close to the truth and the model reflects an effective understanding of the situations in which the hint factor improves model fit. This indicates that the model can capture true student learning behaviour (improving in performance or learning from failures) based on details of the used skills and the difficulty level.

**Table 4. 4: Statistical comparison for both PFA models**

| Model | LL | BIC | AIC |
|---|---|---|---|
| PFA/four skills | -2025 | 4149 | 4240.0 |
| ModPFA/four skills | -2055 | 4018 | 4137.5 |

## 4.4 DINA Model

This section demonstrates how to extract items and skills related to the data. According to the DINA model, the guessing and slipping parameters can be estimated using the following equations, developed by De la Torre (2008):

Let P $(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)$ equal the expected number of examinees with attribute pattern $\boldsymbol{\alpha}_l$, and

$\boldsymbol{R}_{jl} = \sum_{i=1}^{I} \mathbf{P}\ (\boldsymbol{\alpha}_l|\boldsymbol{X}_i)\ \boldsymbol{X}_{ij}$ the expected number of examinees with attribute pattern $\boldsymbol{\alpha}_l$ and answering item j. $\boldsymbol{T}_{j\eta}$ is the examinee's proficiency level $(\boldsymbol{\eta})$ when answering item j,

and it is $\boldsymbol{T}_{j0}$ = $\boldsymbol{g}_j$ (guess), $\boldsymbol{T}_{j1}$= $\boldsymbol{s}_j$.

Therefore

$$\frac{\partial l(X)}{\partial T_{j\eta}}=\sum_{\{\alpha_l:\, \acute{\alpha}_l q_j<\acute{q}_j q_j\}}\frac{\partial P_j(\alpha_l)}{\partial T_{j\eta}}\left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)}\right]\left[\ \boldsymbol{R}_{jl}\text{-}\boldsymbol{P}_j(\boldsymbol{\alpha}_l)\ \boldsymbol{I}_l\right]$$

$$+\sum_{\{\ \alpha_l:\, \acute{\alpha}_l q_j=\acute{q}_j q_j\}}\frac{\partial P_j(\alpha_l)}{\partial T_{j\eta}}\left[\frac{1}{P_j(\alpha_l)[1-P_j(\alpha_l)}\right]\left[\ \boldsymbol{R}_{jl}\ \text{-}\boldsymbol{P}_j(\boldsymbol{\alpha}_l)\ \boldsymbol{I}_l\right.$$

$$=\frac{\partial g_j}{\partial T_{j\eta}}\left[\frac{1}{g_j[1-g_j]}\right]\sum_{\{\ \alpha_l:\, \acute{\alpha}_l q_j<\acute{q}_j q_j\}}\left[\ \boldsymbol{R}_{jl}-\boldsymbol{g}_j\boldsymbol{I}_l\right]$$

$$+\ \frac{\partial(1-s_i)}{\partial T_{j\eta}}\left[\frac{1}{(1-s_i)s_i}\right]\sum_{\{\ \alpha_l:\, \acute{\alpha}_l q_j=\acute{q}_j q_j\}}\left[\boldsymbol{R}_{jl}-(\mathbf{1}-\boldsymbol{s}_i)\boldsymbol{I}_l\right]\text{.............}\ (4.5)$$

$$=\qquad \frac{\partial g_j}{\partial T_{j\eta}}\left[\frac{1}{g_j[1-g_j]}\right]\left[\boldsymbol{R}_{jl}{}^{(\mathbf{0})}\text{-}\boldsymbol{g}_j\boldsymbol{I}_{jl}{}^{(\mathbf{0})}\right]+\ \frac{\partial(1-s_i)}{\partial T_{j\eta}}\ \left[\frac{1}{(1-s_i)s_i}\right]\left[\boldsymbol{R}_{jl}{}^{(1)}-(\mathbf{1}-\boldsymbol{s}_i)\boldsymbol{I}_{jl}{}^{(1)}\right]$$

............... (4.6)

where, $\boldsymbol{I}_{jl}{}^{(\mathbf{0})}$ is the expected number of examinees lacking at least one of the required attributes for the item j and where $[\boldsymbol{R}_{jl}{}^{(\mathbf{0})}\text{-}\boldsymbol{g}_j\boldsymbol{I}_{jl}{}^{(\mathbf{0})}]$ is the expected number of examinees among $\boldsymbol{I}_{jl}{}^{(\mathbf{0})}$ correctly answering item j. $\boldsymbol{R}_{jl}{}^{(\mathbf{1})}$ and $\boldsymbol{I}_{jl}{}^{(\mathbf{1})}$ have the same interpretation and pertain to the examinees with all the required attributes for item j. $\boldsymbol{I}_{jl}{}^{(\mathbf{0})} + \boldsymbol{I}_{jl}{}^{(\mathbf{1})}$ is equal to $\boldsymbol{I}_l$ for all j.

When $\boldsymbol{\eta} = \mathbf{0}$ (i.e., $\boldsymbol{T}_{j0}\ =\ \boldsymbol{g}_j$), $\frac{\partial P_j(\alpha_l)}{\partial T_{j\eta}}$ is 1 for the first term of the equation and 0 for the second term. Therefore, to obtain the maximisation of $\boldsymbol{\partial l}(\boldsymbol{X})$ with respect to $\boldsymbol{T}_{j0}$ to solving for $\boldsymbol{s}_j$ is shown in the equation (4.7):

$\frac{1}{g_j[1-g_j]}[\boldsymbol{R_{jl}}^{(0)}\text{-}\boldsymbol{g_j}\boldsymbol{I_{jl}}^{(0)}] = 0$, which gives the estimator $\widehat{\boldsymbol{g_j}} = \boldsymbol{R_{jl}}^{(0)}/\boldsymbol{I_{jl}}^{(0)}$. ……………… (4.7)

Similarly, maximisation of $\boldsymbol{\partial l(X)}$ with respect to $\boldsymbol{T_{j1}}$ is equivalent for $\boldsymbol{s_j}$ to be solved, therefore

$[\frac{1}{(1-s_i)s_i}][\boldsymbol{R_{jl}}^{(1)} - (\boldsymbol{1} - \boldsymbol{s_i})\boldsymbol{I_{jl}}^{(1)}] = 0$

and the estimator $\widehat{\boldsymbol{s_j}}$ will be $[\boldsymbol{I_{jl}}^{(1)}\text{-}\boldsymbol{R_{jl}}^{(1)}] / \boldsymbol{I_{jl}}^{(1)}$ ………………..(4.8)

To calculate the estimated values of s and g, the first step of the algorithm starts with the initial values of $\boldsymbol{s_i}$ and $\boldsymbol{g_i}$, then in step 2, $\boldsymbol{R_{jl}}^{(0)}$,$\boldsymbol{I_{jl}}^{(0)}$, $\boldsymbol{R_{jl}}^{(1)}$ and $\boldsymbol{I_{jl}}^{(1)}$ are computed based on the values of s and g. Step 3 finds s and g by applying equations 3 and 4 . Steps 2 and 3 are repeated until convergence.

The following statistics represent the item parameters (guessing and slipping), p-value and difficulty level. Analysis carried out by the DINA model was conducted by applying the DINA model to 16 items with four skills and the guessing /slipping parameters related to the items in the assessment tool were obtained. The 16 assessment items with a 4 skills Q matrix and 225 students' profiles with prior (correct/incorrect) scores were used to estimate the guessing and slipping parameters. These estimations were conducted using equation (4.7) for the guessing estimator and equation (4.8) for the slipping estimator. As shown in Tables 4.6 and 4.7, the error rates of the estimates for the slipping/guessing parameters are low values which indicates an accurate estimation.

The DINA model parameters pertaining to the assessment tool are provided in Tables 4.5 and 4.6.

**Table 4. 5: Guessing parameters**

| Item | Est. | Std. Err. |
|---|---|---|
| 1 | 0.2776359 | 0.03037773 |
| 2 | 0.3629012 | 0.03266779 |

| | | |
|---|---|---|
| 3 | 0.3247046 | 0.07286604 |
| 4 | 0.4315772 | 0.05603003 |
| 5 | 0.4648594 | 0.03539947 |
| 6 | 0.4404193 | 0.03469817 |
| 7 | 0.4283343 | 0.07621875 |
| 8 | 0.5307982 | 0.03645218 |
| 9 | 0.4250704 | 0.03349788 |
| 10 | 0.4498827 | 0.03349452 |
| 11 | 0.4225016 | 0.03282569 |
| 12 | 0.4010789 | 0.07585528 |
| 13 | 0.3287575 | 0.02786006 |
| 14 | 0.3807637 | 0.05572437 |
| 15 | 0.3610092 | 0.02896941 |
| 16 | 0.4256268 | 0.01059550 |

**Table 4. 6: Slipping parameters**

| Item | Est. | Std. Err. |
|---|---|---|
| 1 | 0.63527483 | 0.04660357 |
| 2 | 0.40521393 | 0.04785718 |
| 3 | 0.38707382 | 0.20787248 |
| 4 | 0.43430383 | 0.05603003 |
| 5 | 0.4648594 | 0.25066590 |
| 6 | 0.33607003 | 0.04412488 |
| 7 | 0.31522588 | 0.16367478 |
| 8 | 0.19430448 | 0.12819587 |
| 9 | 0. 45580484 | 0.14301299 |
| 10 | 0. 76464171 | 0.12223925 |

| | | |
|---|---|---|
| 11 | 0.14779050 | 0.09107638 |
| 12 | 0.14049957 | 0.03118276 |
| 13 | 0.17809368 | 0.13127262 |
| 14 | 0.11149622 | 0.05572437 |
| 15 | 0.16306285 | 0.03448476 |
| 16 | 0.26589370 | 0.01959945 |

The slipping and guessing parameters of the DINA model reveal that the guessing value varies between 0.10 and 0.53 and the slipping value varies between 0.16 and 0.76. The mean values of the slipping and guessing parameters are 0.26 and 0.35, respectively. Wenmin (2006) noted that higher ***slipping*** and lower ***guessing*** values indicate a difficult test. The test was therefore found to be more difficult than average.

Figure 4.1 offers a further explanation regarding the distribution of the DINA model parameters among the 225 students in the given data set. This figure includes the probability of guessing when the students attempt the answer correctly without having acquired the necessary skills, besides the probability of non-slipping when the student attempts the answer incorrectly without having acquired the skills in the item. This figure is necessary as it reflects the distribution of the number of students who do not master the given skills in the assessment items to indicate who has guessed and who has non-slipped the answer. It can be observed that the majority of the 225 students have non-slipped the items as the probability distribution of guessing is less than non-slipping. This matches the findings in Tables 4.5 and 4.6 and presents evidence that the assessment is difficult as the students could not even hazard a guess for the correct answer.

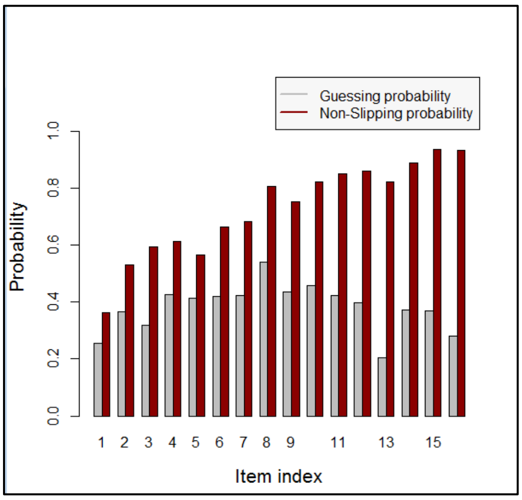



**Figure4. 1: Guessing probability distribution**

**Table 4. 7: Summary of overall item characteristics/DINA model**

| | Min | Max | Mean |
|---|---|---|---|
| p-value | 0.458 | 0.673 | 0.601 |
| Guessing | 0.204 | 0.539 | 0.381 |
| Slipping | 0.163 | 0.635 | 0.270 |

Table 4.7 shows a summary of the item-related values. By obtaining the p-values of the items we can determine the percentage of students solving each item. For example, 46% of the students solved the most difficult item, while 68% solved the easiest item. On average, 60% of the students solved the items, indicating that the difficulty of the items is average. As the guessing parameters $\boldsymbol{g}_i$ reflect the probability of correct responses without acquiring the skills, it can be said that 1- $\boldsymbol{s}_i$ *(slipping) determines the correct responses while mastering the skills.* The item guessing parameters range from 0.2 to 0.54 and have a maximum standard error of 0.02. The item slipping parameters range from 0.16 to 0.62 and have a maximum standard error of 0.01. These ranges for guessing and slipping show that the items have been answered reasonably and that the students have acquired most of the skills.

**Table 4. 8: Skill distribution and respective standard error**

| | Probability | Std. Err |
|---|---|---|
| Skill 1 | 0.68 | 0.02 |
| Skill 2 | 0.73 | 0.02 |
| Skill 3 | 0.65 | 0.01 |
| Skill 4 | 0.70 | 0.03 |

Table 4.8 and Figure 4.2 show the percentages of the distributions for the four skills among the 16 assessment items. The highest percentage skill appearing in the assessment is skill 2, while the least is skill 3. By comparing these findings with Table 4.3, it can be noticed that the skill 2 difficulty level is (-0.11) which means this skill is slightly difficult. Since the majority of the given 16 item assessment consists of skill 2, this indicates that the assessment is more difficult than usual and this explains why the distribution of the probability of slipping is higher than the guessing as shown in Tables 4.5 and 4.6.

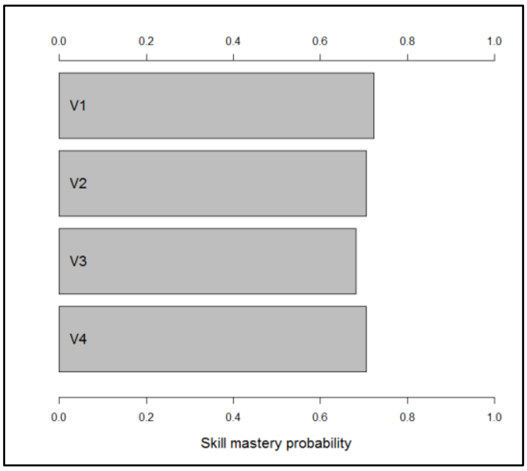



**Figure4. 2: Skill mastery probability**

**Table 4. 9: Individual skill mastering for the first five students**

| Student ID | Skill 1 | Skill 2 | Skill 3 | Skill4 |
|---|---|---|---|---|
| Stud 1 | 1 | 1 | 1 | 1 |
| Stud 2 | 0 | 1 | 0 | 1 |
| Stud 3 | 0 | 1 | 0 | 0 |
| Stud 4 | 1 | 1 | 0 | 1 |
| Stud 5 | 1 | 0 | 1 | 1 |

Table 4.9 shows the individual skill probabilities P($\boldsymbol{\alpha}_k$) and their standard errors. There was little difference in the values of the skills that were distributed among the students. The results show that most students acquired all the skills equally as these values match the findings of Tables 4.5 and 4.6 for both the guessing and slipping parameters. It can be observed that 73% of the students possessed skill 1, 71% of the students obtained skill 2 and 65% acquired skill 3. The latter result indicates that skill 3 is slightly more difficult or requires further explanation.

**Table 4. 10: Correlations between the four skills**

| | Skill1 | Skill2 | Skill3 | Skill4 |
|---|---|---|---|---|
| Skill1 | 1 | | | |
| Skill2 | 0.74 | 1 | | |
| Skill3 | 0.829 | 0.93 | 1 | |
| Skill4 | 0.556 | 0.494 | 0.691 | 1 |

Table 4.10 further examines the correlation between the four given skills in the assessment items. Therefore, when two skills are highly correlated to each other, the student who answers one correctly is more likely to answer the other one correctly too.

## 4.6 Model evaluation

The accuracy level of the proposed counting approach is a measure of how accurately the predicted results follow the true value. For this experiment, the same 16 item assessment with four skills and 225 students was applied to the four models: PDA, ModPFA, DINA and CFA. This experiment will evaluate the prediction accuracy of these models in terms of the latent student performance estimation. The evaluation models used for this purpose are LL, AIC, BIC and RMSE, all of which were explained in section 4.2. Table 4.11 summarises the estimated effectiveness for the selected dataset of the proposed system, CFA, against the other models, DINA, PFA and ModPFA.

**Table 4. 11: Summary of statistical evaluations for the four models**

| Dataset | Model | LL | AIC | BIC | RMSE |
|---|---|---|---|---|---|
| Product/LCM | PFA | -2029 | 4083.3 | 4157 | 0.63 |
| - | ModPFA | -2055 | 4137 | 4218 | 0.454 |
| - | DINA | -1999.8 | 4091 | 4247 | 0.069 |
| - | CFA | -2264 | 4027 | 4016 | 0.0894 |

The logistic models PFA, ModPFA and CFA shown in Table 4.11 reflect better evaluations than the probabilistic model DINA in terms of the LL. Compared to PFA, there is a very slight difference and this is not surprising since PFA holds an advantage because it includes success and failure counts that include information regarding performance on held out data. Moreover, CFA has the best log likelihood values which are indicative of a better ability to fit the data that does not suffer from this discrepancy.

Furthermore, the models have been evaluated using AIC and BIC. As explained earlier, since AIC is an estimate of the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, a lower AIC therefore indicates that a model is considered to be closer to the truth. BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC indicates that a model is considered to be more likely to be the true model.

The findings of the AIC and BIC suggest that the CFA model fits the data better than the other given models (i.e. low values for both the AIC and the BIC). This indicates that the estimated values of the probability of achieving the correct answer for the next item for each student is more likely to be true compared to the actual values taken from the students' prior scores. This is reasonable since the DINA model does not have separate learning rates for success and failure.

The reason the PFA and ModPFA models are less effective (i.e. lower LL and high values for BIC, AIC and RMSE) is because neither model supports the negative learning behaviours of students such as slipping/guessing, particularly with regard to multiple choice assessments. The traditional DINA model was still included as a baseline model however as it is widely used and has explicit parameters for guess and slip. The RMSE shows that the DINA model has better accuracy than the CFA model which could be due to the diversity in the predictors used in the CFA.

In Table 4.12, CV is used to analyse the efficiency of the different features in the proposed model and to show the differences with the original PFA model. CV can be used to check whether a model has been over-fitted. It can be used to predict the performance of a model on unavailable data (predicted parameters) using numerical computation in place of theoretical analysis.
CV divides the data into five folds, ensuring that no data from a level is split between folds. Once the data has been divided, the model is trained on four folds and then used to predict the values of the remaining fold. The AIC, BIC

and LL are predicted and actual values are then reported, with smaller values indicating a more accurate model.

**Table 4. 12: CV evaluations**

| Model | CV-LL | CV-BIC | CV-AIC |
|---|---|---|---|
| PFA | -1530.2 | 3263.6 | 3212.2 |
| DINA | -1545 | 3238 | 3208 |
| CFA | -1599 | 3132.4 | 3173.5 |

The CFA fit results show that the slip/guess model has better predictive accuracy (i.e. CV performance) and LL/BIC and AIC fit than its traditional counterparts across the selected datasets. The AIC/BIC scores also mirror this finding and suggest that the addition of the slip/guess and hint parameters to the logistic model led to improved model fit and an increased ability to predict behaviour.

## 4.7 Prediction scenarios

To illustrate how the proposed model CFA works and to evaluate its efficiency compared to the other models, a prediction scenario study was designed. The LCM/product data (explained in chapter three), with 16 items and four attributes/skills was used to examine the performance of the proposed method. The given data was tested to represent the past performance of the learner for the first eight items. The learner was then measured from that point onwards (i.e. from item 9 to item 16) and the probability of the next item being correct assessed for the four given models.

All factors in the first set of data, such as the guessing/slipping parameters, success and failure rates, item difficulty and the students' current skills were estimated earlier. Based on all the given factors, we predicted the performance of correctly answering the next item for two students for items 9 and 10. Two students were chosen for this scenario, one of whom had fully mastered the given skills and one who had partially mastered the skills (see Table 4.13).

**Table 4. 13: Students' acquired skills**

| Student ID | Skill 1 | Skill 2 | Skill 3 | Skill 4 |
|---|---|---|---|---|
| Student 1 | 1 | 1 | 1 | 1 |
| Student 2 | 1 | 1 | 0 | 0 |

This scenario study was based on the assumption that, if one of the students mastered all the necessary skills but had slips for some items, the developed model needed to be measured to discern whether it would pick up these slips and estimate the performance of that student better than the other models. To demonstrate such a scenario, the students' skills are shown in Table 4.13 and the item/skills are shown in Table 4.14. Tables 4.15 and 4.16 summarise the simulation results for the four models for the two students with two items.

**Table 4. 14: Item/skills**

| Item ID | Skill 1 | Skill 2 | Skill 3 | Skill 4 |
|---|---|---|---|---|
| Item 9 | 1 | 0 | 0 | 1 |
| Item 10 | 1 | 1 | 0 | 1 |

- **Performance scenario for student 1**

For student 1, who has mastered all the skills, item 9 has two skills. This means that student 1 should answer this item correctly with no guessing parameters. According to Table 4.15, however, this student asked for a hint, while according to Table 4.6, item 9 has a high percentage of slipping. Therefore, this student was more likely to slip the correct answer for this item. The CFA model estimated the probability of student 1 answering item 9 correctly as 73%, while the PFA gave an estimation of

65%. CFA considers the factors of both hints and slipping, whereas PFA does not.

**Table 4. 15: Actual/current student responses**

| Student ID | Item9 | Hint? | Answer | Item 10 | Hint? | Answer |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 |

**Table 4. 16: Probability of the next item being correct for the four models/student 1**

| Model | Student 1/item9 | Student 1/item 10 |
|---|---|---|
| PFA | 0.65 | 0.42 |
| PFA Mod | 0.69 | 0.54 |
| DINA | 0.58 | 0.62 |
| CFA | 0.73 | 0.78 |

**Table 4. 17: Probability of the next item being correct for the four models/student 2**

| Model | Student 2 /item9 | Student 2/item 10 |
|---|---|---|
| PFA | 0.72 | 0.75 |
| PFA Mod | 0.54 | 0.54 |
| DINA | 0.56 | 0.61 |
| CFA | 0.42 | 0.47 |

Table 4.15 shows that student 1 attempted to answer item 9 correctly. Table 4.16 presents the performance estimations for all the four models.

Now, student 1 attempts to answer item 10. According to table 4.14, item 10 has three skills (slightly more difficult) and, since the student has mastered all the skills, therefore s/he is supposed to score this correctly. CFA estimated the probability of student 1 having the correct answer for item 10 as 0.78, according to table 4.15, student 1 scored it incorrectly. This means that there is a slipping probability due to the item difficulty, especially as no hint was requested in order to answer this item. CFA (with the factors of hint and slipping) offers the most accurate estimation, since the student has mastered all the skills, but the item is slightly difficult and the student has not requested a hint. Therefore there is a strong probability of slipping and not attempting the answer correctly. According to table 4.16, the DINA model estimated the probability as 0.62 as the student has acquired the skills but s/he might slip. PFA has the least accurate prediction as it assumes that the student does not possess the ability to answer the item, without considering that this student has a strong skills level but there is a probability that the student might slip the answer. Further, this student slipped this item although s/he had requested a hint. This means that CFA predicted that this student might slip this item even though he possessed the knowledge required to answer it. Furthermore, CFA introduced a very important concept which is that attempting an item incorrectly does not represent a lack of knowledge but it might be simply be a slip. This student cannot therefore be treated in the same way as any other student who scores the item incorrectly without mastering the necessary skills. This conclusion depends on the analyses of prior scores with the factors combined in CFA and cannot be found in the other models.

- **Performance scenario for student 2**

For student 2, Table 4.13 indicates that this student mastered only two skills (2 and 3), therefore, the guessing parameter is considered. As item 9 only has two skills (1 and 2), student 2 should either guess the correct answer or answer it incorrectly. According to Table 4.17, CFA estimated the probability of the correct answer as 42%, whereas the DINA model gave an estimate of 56%. PFA estimated that this student would answer item 9 correctly. According to Table 4.15, this student asked for a hint and answered the item correctly. Student 2 probably guessed the answer (or made a logical guess) and therefore CFA demonstrated the best prediction as it considered the hint and guessing parameters alongside other factors.

For item 10, however, which has three skills (two of which match the student's skills), student 2 should not provide an answer and, if he does so, it will be a guess. The estimation values for the PFA and the DINA models indicated that this student would answer this item correctly and the DINA model showed a high percentage of guessing. CFA, however, gave a lower probability of student 2 giving the correct answer. Table 4.15 shows that this student answered item 10 incorrectly and did not ask for a hint, which aligns with the estimation given by CFA.

In addition to the evaluation tools used earlier to assess the efficiency of the models, Figure 4.3 illustrates the probability accuracy of the three models using CV with five folds across 250 students. CFA was found to have the highest accuracy out of the three models. The increase in accuracy among the three models is obvious, as it can be noticed that PFA has a low rate accuracy and, although stable, is still the lowest rate. This is due to the limited parameters used to estimate the students' performance level. With the DINA model, particularly in multiple choices assessments, it can be noticed that the accuracy level of estimation is slightly higher, and this is because of the use of the cognitive parameters (slipping and guessing) which provide better predictions of the students' current skills against the items' skills. By using CFA, the prediction has reached a different level and the estimation accuracy reaches its highest level compared to the PFA and DINA. As explained earlier, CFA uses more cognitive factors of students in its predictions. The adoption of guessing/slipping parameters and the inclusion of the hint parameters in the formula have proved the hypothesis of this study which stated that the inclusion of cognitive factors in ITS will increase the performance estimation and optimise a better knowledge learning process.

Using CFA not only improved the performance estimation level, but also contributed to a reduction in student study time. CFA successfully addressed the strengths and weaknesses of each student. Furthermore it accurately estimated the current knowledge level and this can decrease the number of practices required by each student to reach the mastery level.

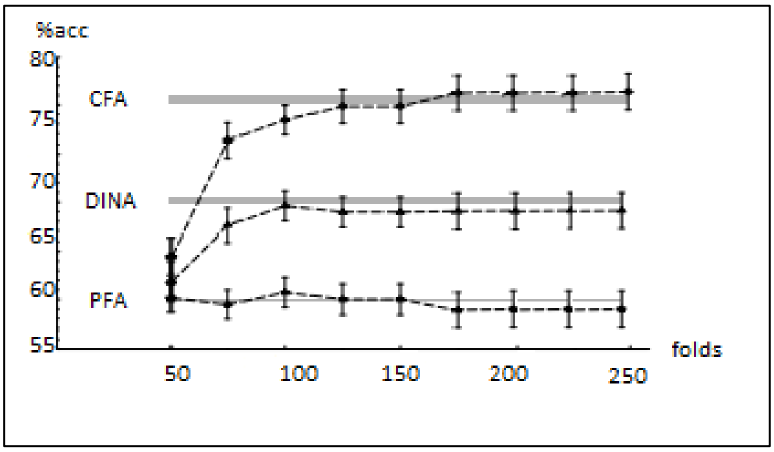



**Figure4. 3: Prediction accuracy curve for three models**

## 4.8 Defining the student's learning curve

This section explains how CFA can influence the student's learning curve and reduce study time. This has been proven by applying CFA on a selected data set from a tutoring system in DataShop (based on techniques of the DINA model and PFA formulas). The data set has been analysed and used to evaluate the effect of our model on the student's learning curve.

CFA uses human input to identify model improvements from student log data, which are then evaluated by a statistical fit with the data. Our formula extends the PFA model, which is a statistical algorithm for modelling learning and performance that uses logistical regression performed over the 'error rate' learning curve data. This formula estimates the probability of having the correct answer across certain factors, which are item difficulty, success/failure rates, and prior used hints, in addition to the slipping and guessing probability.

To describe the student model improvement, we define three ways to describe the learning curve according to the number of errors, and these curves are provided by the visualisation and analysis tools in DataShop:

**1) Smooth learning curves:** a reasonably smooth learning curve is expected to happen when the learning progress of the students is going well, and it indicates no problem with learning materials. On the other hand, when the learning curve of a given skill is noisy, with upward or downward 'blips', the student model is suspect.

**2) No learning progress:** if the student model is accurate, we expect the error rate to decline over the number of opportunities a student has to learn and apply a skill. A flat learning curve is another indication of a potentially flawed student model.

**3) Learning curve associated with item difficulty:** this computes the learning process against the item difficulty. Usually, an easy item does not require many practices. These phenomena will cause over-practice, and this will affect the performance and student time.

According to the given learning curve types, we applied CFA formula to a private data set from DataShop called "Cog Model Discovery Experiment Spring 2010/Control". This data was generated from student interactions with a cognitive tutor for learning Geometry. The assessment items consist of five skills (sometimes they are known as knowledge component KC) distributed over 30 questions for around 1,000 students. The skills are: find rectangular area, find trapezoid area, find added area, find individual area, and enter the given measurements.

A subset of the learning curves for these skills is shown in Figure 4.4. The lines represent the error rate (y-axis) averaged over all students for the first 20 practice opportunities for each skill (e.g. on the fifth opportunity on the trapezoid area skill, about 50 per cent of students made an error). By applying this data to CFA model, we found some interesting phenomena, the first of which was finding a rectangle perimeter curve which is particularly jagged with upward blips in error rate. At opportunities 12 and 15–18, the curve jumps up, from about 25 per cent to above 55 per cent.

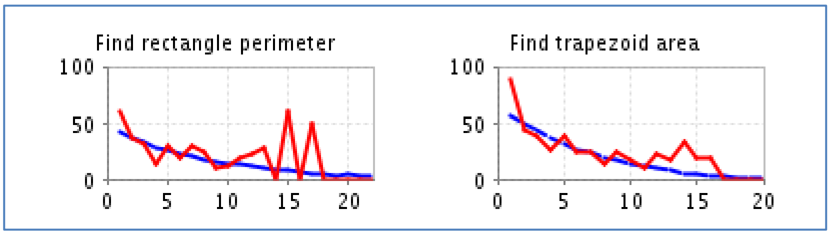



**Figure4. 4: Students' learning curves for some skills**

Figure 4.4 shows the learning curve of the skills (Find added area and Find individual area) given the base and height. Students had an initial error rate around 50 per cent. After 20 and 50 practices, the error rate remained high. However, the item difficulty for Find added area and Find individual area is 0.3 and 0.5 respectively, which means they are not very difficult items. Furthermore, the used hints rate was about 30 per cent. Many practices and hints for an easy/average skill are not a good use of student time. Reducing the number of practices for this skill may reduce a student's time without compromising their performance. Therefore, we want to distribute the blame appropriately across all four skills depending on prior estimates of the skill difficulties skills with a higher prior probability of being known should receive less blame than skills with a lower probability.

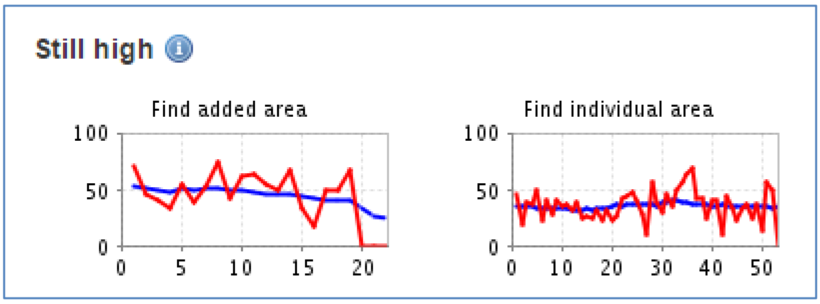



**Figure4. 5: skill over-practice**

There are some factors affecting the learning performance and the findings shown above. These factors can be related to the quality of learning, such as the learning materials and the hints which are provided with each skill, and the student's performance, whether the student maintains the skill or not. This could be determined by the guessing and slipping parameters. On the other hand, if the student guesses the answer without mastering the current skill, they are more likely to make another mistake or answer the next item containing the **same skills** incorrectly. This can explain having some blips with the learning curve or increasing the number of practices per easy item, which may cause the over-practice.

### 4.8.1 Discovering student learning time through CFA

One application of our CFA model is to manage the student's learning time by measuring the over-/under-practice with the goal of improving learning efficiency, and to discover whether the student is really making progress while practising items.

Accordingly, an important question has been considered here: What causes wasting of the student's learning time? The answer according to the data set we have here is that most tutors apply assessment items without considering the individual skill needs compared to the given item skills. Moreover, if the tutor updates its estimates according to the student's proficiency level, it uses the same parameters predicted for each given item. Furthermore, we have discovered some other points that might affect the learning time, such as:

1- More practices have been used on an easy problem/skill, which is shown by the learning curve still having high error rates along with the increasing number of practices.
2- More hints have been requested by the students on an easy skill with a high error rate. Does this mean the students do not have the knowledge to answer this item correctly or that they have mastered the skills but they have slipped?
3- In the case of multiple skills per item and based on what has been stated above: Do the incorrect scores obtained by the students imply learning (if they have not mastered some or all of the given skills) or slipping (if they have not acquired the knowledge)?
4- If slipping occurs consistently, does this mean the provided hints might be designed ineffectively and this may reflect the quality of the learning materials?
5- Again, in the case of having multiple skills for a single item, if the student answers the item incorrectly, it could be that the student does not know one or all of the given skills in that item. In that case, it does not seem appropriate to reduce the probability of having the correct answer without knowing which skill has not been mastered by the student.
6- When the student requires a hint to answer the question and gets the answer right: Does this mean the student has learned or that the student guessed the answer? Likewise, when the student continues to ask for a hint and gets the answer wrong, does it imply slipping or

learning from prior items? By considering in both cases the item difficulty level?

7- Does the probability of having the correct answer increase when the student does not ask for a hint (i.e. their knowledge level is higher)?

To address the above points of the given data set, a demonstration of the CFA estimated values is presented in the next section.

### 4.8.2 Model demonstrations

This section demonstrates how the fully developed CFA formula estimates the student's future performance for the next exam's item ( for the private dataset) by considering all the factors that were used in addition to the item skill excluded from the Q Matrix and the student skills which have been estimated previously. We have chosen two samples of students with different skills vectors and applied the formula to the same items. Table 4.18 describes the five skills provided in the data set assessments items, while table 4.19 gives the estimates of the skills parameters that excluded after applying CFA model to the dataset such as, the difficulty level, success and failure rates for all the given five skills.

**Table 4. 18: Skills/KC representations**

| Skill | Description |
|---|---|
| Skill 1 | Find rectangular area |
| Skill 2 | Find trapezoid area |
| Skill 3 | Find added area |
| Skill 4 | Find individual area |
| Skill 5 | Enter the given measurements |

**Table 4. 19: Estimations of given skills parameters**

| ModPFA | |
|---|---|
| Difficulty/Skill1 | 0.14 |
| Difficulty/Skill2 | -0.21 |
| Difficulty/Skill3 | 0.327 |
| Difficulty/Skill4 | 0.531 |
| Difficulty/Skill5 | -1.2 |
| Success/skill1 | 0.14 |
| Success/skill2 | 0.02 |
| Success/skill3 | 0.12 |
| Success/skill4 | 0.32 |
| Success/skill5 | 0.014 |
| Failure/skill1 | 0.075 |
| Failure/skill2 | 0.053 |
| Failure/skill3 | 0. 17 |
| Failure/skill4 | -0.081 |
| Failure/skill5 | -0.081 |

Table 4.20 below demonstrates the behaviour of one selected student from the data set. This table reveals the necessary parameters of CFA model to predict the students performance level. Starting with the item j the students' scores, the prior used hints by the student, the items' skills (the skills number differs from item to another, some item might have one skill or more). Further, the student's current skill and the student's proficiency level. The

guessing and slipping parameters of the given item for this data set ( these parameters have been calculated previously using the same process shown in chapter four/section 4.4), and finally, the student's performance of having the next item correctly is estimated using CFA model.

As it has been shown in Table 4.20, the student has mastered all the skills required for the whole assessment items . However, the skill difficulty level for each given item is high. It can be noticed that the probability of getting the answer right is increasing although the student got the next item wrong. This probably happened as the item difficulty level is high, and the student is most likely to slip the answer.

**Table 4. 20: Demonstration that student mastered all the given skills**

| J | Score | Difficulty | Hint | ItemSkill 1,2,3,4,5 | Student Skill | Proficiency Level | Guess | Slip | Performance |
|---|---|---|---|---|---|---|---|---|---|
| 1 | - | -0.25 | 1 | 11000 | 11111 | 1 | 0.094 | 0.204 | 0.56 |
| 2 | 0 | -0.23 | 0 | 10101 | 11111 | 1 | 0.083 | 0.208 | 0.62 |
| 3 | 0 | 0.53 | 0 | 10100 | 11111 | 1 | 0.102 | 0.079 | 0.81 |
| | 1 | | | | | | | | |

**Table 4. 21: Demonstration that student mastered only one of the given skills**

| J | Score | Difficulty | Hint | ItemSkill 1,2,3,4,5 | Student Skill | Proficiency level | Guess | Slip | Performance |
|---|---|---|---|---|---|---|---|---|---|
| 1 | - | -0.25 | 1 | 11000 | 01000 | 0 | 0.094 | 0.104 | 0.20 |
| 2 | 0 | -0.23 | 1 | 10101 | 01000 | 0 | 0.083 | 0.108 | 0.32 |
| 3 | 0 | 0.53 | 0 | 10100 | 01000 | 0 | 0.102 | 0.079 | 0.45 |
| 4 | 1 | | | | | | | | |

Table 4.21 provides a different example of a student who has mastered only one of the five given skills. This table has used the same parameters explained in Table 4.20, and this example is applied to the same exam items given earlier in Table 4.20. The student's proficiency value for the given item skills is zero, which means the student might guess the next item. Moreover, the formula shows that the probability of having a correct answer is low. However, the student got the next item wrong. Please note that the second item contains skills 2 and 3, and from the skills correlation table shown earlier, the correlation between those two skills is very high, which means if the student does not know one of them, they are unlikely to get the other skill right. This probably explains the incorrect responses for the next item. The probability of having the correct answer for the third item is getting slightly higher and shows some learning from failure. For the first two items, the student has asked for a hint, and this may give them some knowledge. Although the student got the third item wrong, according to CFA formula, they guessed the answer this time as they did not request a hint. However, the item level is easy and there is some increase in learning.

In both cases, the system shows the incorrect scores for both students, and there is slipping (if all the required knowledge has been mastered) or learning from failure (if not all the required knowledge has been mastered). Furthermore, it is possible to monitor the learning progress and quality of hints and learning materials through the interaction with guessing and slipping parameters. Moreover, it can be beneficial to reduce the over-practice and time wasted when the student has reached the mastery level for all given skills.

## 4.9 Conclusion

This chapter has described an approach for building individualised intelligent tutoring models that are capable of accounting for student differences with respect to initial mastery probabilities and skill learning probabilities. Our approach was based on two cognitive models: probabilistic (DINA model) and logistic regression (PFA). Our implementation improved the prediction accuracy of the success of students along with the learning process (in case of failure). An interesting finding was that splitting student-specific probability of performance into two factors of prior success scores and prior used hints was more beneficial for the accuracy of the model.

On the other hand, we included the probability of the slipping and guessing parameters associated with each item type in the form of a logistic regression. This combination increased the individual learning rate rather than individual proficiency. To summarise, the excluded findings showed a better fit of the data using CFA in terms of the statistical evaluations tools compared to the previously presented models..

# Chapter Five: Conclusions and Future work:

## 5.1 Conclusions

This thesis has discussed the issue of improving the performance estimation of students in ITS. As has been shown, the existing ITS models generally provide less accurate assessments of a student's performance and thus reduce learning efficiency due to an incomplete representation of the student's knowledge. This is as a result of shortcomings in the design phase of ITS.

Therefore, this thesis has presented a novel tutoring model (CFA) that offers a significant contribution to the design phase of ITS. It differs from the other models by adapting student cognitive factors, such as guessing/slipping parameters and student proficiency levels, together with each item's parameters (prior correct/incorrect score, prior used hints, item difficulty level and item skills) to produce a better estimation of student latent performance. CFA works by extracting the student's skills and comparing these with the learning items. Each learning item involves multiple skills, which are not explicitly stated in textbooks, and students are therefore expected to acquire such skills through problem-solving. Therefore, mastering the skills can be achieved by the students performing tasks that require such skills.

The ultimate achievement of CFA is to help students to target their strengths and weaknesses within their knowledge level, and to provide accurate feedback from the assessment item's difficulty level as to whether this is sufficient to improve the student learning process.

CFA has made several significant contributions that stem directly from its development. Specifically, this work has made the following contributions:

- It aims to have an adverse impact on the student's learning curve and reduce the student's learning time by controlling the amount of time spent practicing the skill several times. It assumes the role of modelling the student's learning by making inferences about their latent performance with multiple skills assessments. CFA does not consider the correct answer by the student as positive evidence of mastering all the required skills, as the student might have guessed the answer. Similarly, an incorrect answer is not deemed to be proof of failure, as the student might have slipped the answer despite having mastered all the required skills.
- It estimates the student's proficiency level and matches this with the given skills of the assessment items. This will assist the tutor to predict the accurate skill level of the student rather than completely relying on the student's scores.
- CFA has a positive impact on the field of cognitive learning psychology. It attempts to show how data generated from tutoring systems can be analysed and modelled to create (and improve) a unified computational theory of human learning. Furthermore, it encapsulates psychological findings in a format that can be used by instructional designers and educational scientists to support the development of tutoring systems. Therefore, this work has been inspired by the DINA model through the consideration of psychological factors (slipping and guessing parameters) to achieve better individual knowledge estimation.

## 5.2 Future Work

Although the developed model has produced good results, it is possible that further research could be conducted in order to strengthen certain aspects. These areas include:

- The developed model could consider other factors of the students' records/data, such as the time spent on solving each item. However, since our model is in the form of logistic regression, all the factors are required to be presented in a binary form so, likewise, the time factor should be determined in binary form.
- The developed model predicts the student performance of each given item and records this in the student profile database. Therefore, further analysis can be performed on this data and the student knowledge level could be further classified/organised as novice, medium and professional for each skill. This would improve the learning process and enhance the discovery of knowledge relating to each student.

Since the presented model is generic and flexible, it can be combined with other models within the educational environment. One suggestion is that a recommender engine could be added to the ITS system which would then recommend the learning material/assessment items which match the student proficiency level for each given skill. This will save students a great deal of time by directing them to the appropriate learning resources. The aim of the recommender system is to organise the vast amount of items available by determining user preferences and applying these preferences to items previously unknown to the user. In this way, the ability to recommend what has a high likelihood of being interesting to the target user is developed. We suggest that two recommender system techniques should be employed with the developed CFA model:

1. *Collaborative Filtering*: this is based on the assumption that similar users like similar things and, being content-agnostic, it focuses only on the past ratings that have been assigned. In the early days of recommender systems, content was deemed to be extremely useful as training data and research data sets contained large amounts of attribute information for the purpose of algorithm training. However, since the late 1990s, the so-called collaborative filtering approach has prevailed.

2. *Matrix Factorisation*: this is known to be one of the most successful methods for rating prediction, outperforming other state-of-the-art methods (Koren and Bell, 2009). It is based on approximating the matrix X by the product of two smaller matrices W and H, i.e. X $\approx WH^T$.

   In the context of recommender systems, matrix X is the partially observed ratings matrix, W $\in R^{I\times K}$ , a matrix where each row u is a vector containing the K latent features describing the user u, and H ∈ RI×K is a matrix where each row i is a vector containing the K features describing the item i.

   Let $w_{uk}$ and $h_{ik}$ be the elements of W and H respectively, then the rating given by a user u to an item i is predicted by:

   $r_{ui}=\sum_{k=1}^{k} w_{uk}\ \ h_{ik}=WH^T{}_{u,i}$

   (6.1)

   where W and H are the model parameters and can be learned by optimising a given criterion using stochastic gradient descent.

Therefore, by classifying the students' knowledge level and organising this into groups, as has been described earlier, the recommender system could work on recommending items associated with each group level.

In conclusion, this work has met its hypothesis; "using cognitive factors (the student's current skills, the guessing/slipping parameters) will improve the prediction of the student's performance and optimise their learning experience". The CFA model offers two major key components, the student's cognitive factors and the adoption of the hinting parameter to assist the student to answer correctly. The combination of these two components creates a novel framework within which an accurate latent inference of the students' knowledge is produced. This further supports the building and testing phases of the intelligent tutoring systems by focusing on producing a solid design phase without the need for classroom study.

# References


Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, *228*(4698), 456-462.

Atkinson, R. C. (1972). Ingredients for a theory of instruction. *American Psychologist*, *27*(10), 921.

Baffes, P., & Mooney, R. (1996). Refinement-based student modeling and automated bug library construction. *Journal of Interactive Learning Research*, *7*(1), 75.

Baayen, R.H., Davidson, D.J. and Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), pp.390-412.). (pp. 453-488). Hillsdale, NJ: Erlbaum.

Beck, J. (2007, July). Difficulties in inferring student knowledge from observations (and why you should care). In *Educational Data Mining: Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education* (pp. 21-30).

Beck, J. E., & Chang, K. M. (2007, July). Identifiability: A fundamental problem of student modeling. In *International Conference on User Modeling* (pp. 137-146). Springer Berlin Heidelberg.

Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). On-line Tutoring for Math Achievement Testing: A Controlled Evaluation. Journal of Interactive Online Learning, 6, 1–13.

Beck, J. E., Jia, P., Sison, J., & Mostow, J. (2003, June). Predicting student help-request behavior in an intelligent tutor for reading. In *International Conference on User Modeling* (pp. 303-312). Springer Berlin Heidelberg.

Birenbaum, M., Tatsuoka, C., & Xin, T. (2005). Large-scale diagnostic assessment: Comparison of eighth graders'

Bransford, J. D., Brown, A. L., & Cocking, R. R.( 1989), How people learn, .M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science,.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., & Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron*, *42*(2), 323-334.

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, *23*(7), 396-410.

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent tutoring systems* (pp. 164-175). Springer Berlin/Heidelberg.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika, 74, 619-632.

Clark, R. E., Feldon, D. F., van Merrienboer, J. J. G., Yates, K., & Early, S. (2008). 1, 2 ¨ In J. M. Spector, M. G. Sosman, M. D. van Merrienboer, & M. P. Driscoll (Eds.), ¨ Handbook of research on educational communications and technology (pp. 578–591). Mahwah, NJ: International Journal of Educational Research.

Clark, H.H., 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior*, *12*(4), pp.335-359).

Corbett, A. T., & Anderson, J. R. (1997). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, *4*(4), 253-278.

Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, *2005*(2).

De la Torre, J. (2006, June). Skills profile comparisons at the state level: An application and extension of cognitive diagnosis modeling in NAEP. Paper presented at the International Meeting of the Psychometric Society, Montreal, Canada.

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. Psychometrika, 73, 595-624.

De La Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353.

Doignon, J.P, and Falmagne,J.C. (1999).knowledge spaces. newyork: springer-verlag

Derek Lomas, Kishan Patel, Jodi L Forlizzi, and Kenneth R Koedinger.(2013).Optimizing challenge in an educational game using large-scale design experiments. ACM, New York, New York, USA.

Draney, K. L., Pirolli, P., & Wilson, M. (1995). A measurement model for a complex cognitive skill. *Cognitively diagnostic assessment*, 103-125.

Embretson, S. E. (1997). Multicomponent response models. In *Handbook of modern item response theory* (pp. 305-321). Springer New York.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. User Modeling and User-Adapted Interaction 19(3), 243-266

Frederiksen, N., Glaser, R., Lesgold, A., & Shafto, M. G. (2013). *Diagnostic monitoring of skill and knowledge acquisition*. Routledge.

Frederiksen, N., Mislevy, R. J., & Bejar, I. I. (Eds.). (2012). *Test theory for a new generation of tests*. Routledge.

Gong, Y., Beck, J., & Heffernan, N. (2010). Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In *Intelligent tutoring systems* (pp. 35-44). Springer Berlin/Heidelberg.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI magazine*, *22*(4), 39.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*(1), 75.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301-321.

Haertel, E. H., & Wiley, D. E. (1993). Representations of ability structures: Implications for testing.

Han, K. T. (2012). Fixing the c parameter in the three-parameter logistic model. *Practical Assessment, Research & Evaluation*, *17*(1), 1-24.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. Applied Psychological Measurement, 29, 262-277.

Henson, R., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. Applied Psychological Measurement, 32, 275-288

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, *71*(2), 407-419.

Jang, E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), Towards adaptive CALL:Natural language processing for diagnostic language assessment (pp. 117-131). Ames: Iowa State University Press.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Applied Psychological Measurement, 25, 258-272.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, *53*(3), 83-93.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. Educational psychologist, 38(1), 23-31.

Kenneth R Koedinger, John Stamper. (2013). Using Data-Driven Discovery of Better Student Models to Improve Student Learning. In H C Lane, K Yacef, Jack Mostow, and P Pavlik, editors, Artificial Intelligence in Education, pages 421–430, Memphis, TN, July. (document), 1.1, 1.4

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935-937.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2010). The Knowledge-Learning-Instruction (KLI) framework: Toward bridging the science-practice chasm to enhance robust student learning (CMU-HCII Tech Rep. No. 10-102). Available from http://reportsarchive. adm. cs. cmu. edu/hcii. html. In *In*.

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013, July). Using data-driven discovery of better student models to improve student learning. In *International Conference on Artificial Intelligence in Education* (pp. 421-430). Springer Berlin Heidelberg.

Kyung T. Han. (*2012*). Fixing the c Parameter in the Three-Parameter Logistic Model, Graduate *Management Admission Council.*

Laouris, Y. & Eteokleous, N. (2005b) Successful implementation of m-learning depends on well-defined requirements. IADIS International conference Mobile Learning.

Lovett, M., Meyer, O., & Thille, C. (2008). JIME-The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education*.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8), 30-37.

M L Gick and K J Holyoak. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15(1):1–38, 1.2

M. Feng, N. Heffernan, K. Koedinger.(2009) .Addressing the assessment challenge with an online system that tutors as it assesses, User Modeling and User-Adapted Interaction 19 (3), 243–266, ISSN 0924-1868.

Maris, E. (1999). Estimating multiple classification latent class models. Psychometrika, 64, 187-212.

mathematics performance in the United States, Singapore and Israel. Assessment in Education Principles Policy

Mislevy, R. (1995). Probability-based inference in cognitive diagnosis. In P. D. Nichols, S. F. Chipman, & (pp. 453-488). Hillsdale, NJ: Erlbaum.

Mislevy, R. (1996). Probability-based inference in cognitive diagnosis. In P. D. Nichols, S. F. Chipman, &

Mok, M. M. C. (2012). Assessment Reform in the Asia-Pacific Region: The Theory and Practice of Self-Directed Learning Oriented Assessment. In *Self-directed Learning Oriented Assessments in the Asia-Pacific* (pp. 3-22). Springer Netherlands.

Murray, T. (1996). Having it all, maybe: Design tradeoffs in ITS authoring tools. In *Intelligent Tutoring Systems* (pp. 93-101). Springer Berlin/Heidelberg.

Murray, T. (1996). Having it all, maybe: Design tradeoffs in ITS authoring tools. In *Intelligent Tutoring Systems* (pp. 93-101). Springer Berlin/Heidelberg.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 98-129.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In *Authoring tools for advanced technology learning environments* (pp. 491-544). Springer Netherlands.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In *Authoring tools for advanced technology learning environments* (pp. 491-544). Springer Netherlands.

N. Frederiksen, R. Glaser, A. Lesgold, & Safto, M. (Eds.), Monitoring skills and knowledge acquisition

Nichols, P., S. Chipman, and R. Brennan.( 1995). *Cognitively diagnostic assessment*, Hillsdal , NJ: Erlbaum

Nitko, A. J. (2001). Educational assessment of students (3rd ed.). Columbus, OH: Merrill Prentice Hall.

NRC: National Research Council. (2001). Knowing what students know: The science and design of educational

Osterlind, S. J. (1998). Constructing test items: Multiple choice, constructed-response, performance and other formats (2nd ed.). Boston: Kluwer Academic.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127-144.

Pardos, Z. A., & Heffernan, N. T. (2010, June). Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 255-266). Springer Berlin Heidelberg.

Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis--A New Alternative to Knowledge Tracing. *Online Submission*.

R. M. Bell, Y. Koren. (2007).Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights, in: Proceedings of the 7th IEEE International Conference on Data Mining (ICDM'07), IEEE CS, Washington, USA, ISBN 0-7695-3018-4, 43–52.

R. L. Brennan (Eds.), Cognitively diagnostic assessment (pp. 43-71). Hillsdale, NJ: Erlbaum.

Rai, D., Gong, Y., & Beck, J. (2009, July). Using Dirichlet priors to improve model parameter plausibility. In *Educational Data Mining*.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish Institute for Educational Research.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review*, *14*(2), 249-255.

Robert A Sottilare and Heather K Holden. (2013) Motivations for a Generalized Intelligent Framework for Tutoring (GIFT) for Authoring, Instruction, and Analysis. In Robert A Sottilare and Heather K Holden, editors, AIED .

Sandelin, A., Wasserman, W. W., & Lenhard, B. (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic acids research*, *32*(suppl_2), W249-W252.

Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies

San Martin, E. del Pino, G., & de Boeck, P. (2006). IRT models for ability-based guessing. *Applied*

Shepard, L. A. (2001). Using assessment to help students think about learning. In *Keynote address at the Assessment Training Institute Summer Conference, Portland, OR*.

Sottilare, R. A., & Holden, H. K. (2013, July). Motivations for a generalized intelligent framework for tutoring (gift) for authoring, instruction and analysis. In *AIED 2013 Workshops Proceedings* (Vol. 7, p. 1).

Stiggins. R. J. (2002). Assessment crisis: The absence of assessment for learning. Phi Delta Kappan, 83, 758-765.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*(3), 337-350.

Tatsuoka, K. ( 1983). *Rule space: An approach for dealing with misconceptions based on item response theory.* Journal of Educational Measurement, 20(4): p. 345- 354.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In

Tatsuoka, K. K., Corter, J., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. American Educational Research Journal, 41, 901-906.

Tatsuoka, K., *Rule space: An approach for dealing with misconceptions based on item response theory.* Journal of Educational Measurement, 1983. 20(4): p. 345- 354.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 11*, 1–13.

Ur, S., & Van Lehn, K. (1995). Steps: A simulated, tutorable physics student. *Journal of Interactive Learning Research*, *6*(4), 405.

User modeling–a notoriously black art. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 317-328). Springer Berlin Heidelberg.

Vincent Aleven, Bruce M McLaren, Jonathan Sewall, and Kenneth R Koedinger. (2006).The cognitive tutor authoring tools (CTAT): Preliminary evaluation of efficiency gains. In Mitsuru Ikeda, Kevin D Ashley, and Chan Tak-Wai, editors, The 8th International Conference on Intelligent Tutoring Systems, pages 61–70. Springer. (document), 1.1, 1.4, 3.3, 3.3.1, 5.3.2

Wiggins, G. (1998). Educative assessment: Designing assessment to inform and improve performance. San Francisco: Jossey-Bass.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied measurement in education*, *13*(2), 181-208.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4), 197–221.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). Handbook of modern item response theory. New York, NY: Springer Verlag.

Yudelson, M., Pavlik Jr, P.I. and Koedinger, K.R., 2011, July. User modeling–A notoriously black art. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 317-328). Springer Berlin Heidelberg).assessment. Washington, DC: National Academies Press.

UNIVERSITY OF PORTSMOUTH

# EThOS DEPOSIT AGREEMENT FOR UNIVERSITY OF PORTSMOUTH THESES

COVERED WORK
I, ***[Name: Samara Banno***

***Address: 31 Sark Tower, Erebus Drive, SE28 0GG]*** "the Depositor", would like to deposit

[***Automated Performance Prediction for Personalised Learning]***, hereafter referred to as the "Work", in the University of Portsmouth Library and agree to the following:

NON-EXCLUSIVE RIGHTS
Rights granted to the University of Portsmouth through this agreement are entirely non-exclusive and royalty free. I am free to publish the Work in its present version or future versions elsewhere. I agree that the University of Portsmouth or any third party with whom the University of Portsmouth has an agreement to do so may, without changing content, translate the Work to any medium or format for the purpose of future preservation and accessibility.

DEPOSIT IN THE UNIVERSITY OF PORTSMOUTH LIBRARY
I understand that work deposited in the University of Portsmouth Library will be accessible to a wide variety of people and institutions - including automated agents - via the World Wide Web (University's Institutional Repository (IR)). An electronic copy of my thesis may also be included in the British Library Electronic Theses On-line System (EThOS).*

I understand that once the Work is deposited, a citation to the Work will always remain visible. Removal of the Work can be made after discussion with the University of Portsmouth Library, who shall make reasonable efforts to ensure removal of the Work from any third party with whom the University of Portsmouth has an agreement.

I AGREE AS FOLLOWS:
- That I am the author or co-author of the work and have the authority on behalf of the author or authors to make this agreement and to hereby give the University of Portsmouth the right to make available the Work in the way described above.
- That I have exercised reasonable care to ensure that the Work is original, and does not to the best of my knowledge break any applicable law or infringe any third party's copyright or other intellectual property right.
- The University of Portsmouth do not hold any obligation to take legal action on my behalf, or other rights holders, in the event of breach of intellectual property rights, or any other right, in the Work.

*Please strikethrough this sentence if you do NOT wish your thesis to be deposited in EThOS but please be aware that EThOS may, at some future date, harvest thesis details automatically (including the full text) from the University's Institutional Repository available at http://eprints.port.ac.uk

Signature:

..................Samara Banno..........

Date: ....................17.03.2019...........................................

Return to: Faculty Advisor (Research), Research Degree Team, Student Life, Department of Student & Academic Administration, University of Portsmouth, Floor 5, Mercantile House, Hampshire Terrace, Portsmouth, PO1 2EG