A-SCALA AI ::

REDEFINING STUDENT SUCCESS ANALYTICS

Creating successful learners is a challenging task for educational institutions. As a rule, the measurement of student learning performance is carried out using tests and exams. However, test and exam data are often subjected to inadequate analysis, leading to incorrect conclusions about the progress of student learning and, therefore, misleading recommendations on improving the learning process.

Using innovative statistical and machine-learning methods and proprietary algorithms, A-SCALA[*] analyzes test and exam data and provides accurate and reliable information about each student's learning performance. A-SCALA creates Student Success Profiles for each course and quantifies its components. A-SCALA identifies gaps in student knowledge and suggests ways to address them, helping educators lead students to success in their field of education.

This paper provides an overview of A-SCALA's capabilities and presents a real-life case study.

[*] The origin of the name A-SCALA is in Hebrew and Latin. Like many Hebrew words, השכלה (hA-SCALA) has multiple meanings, such as "education," "enlightenment," "knowledge," and "erudition." In Latin, scala means "ladder."

Contents

INTRODUCTION 3
The Difficulty of Items (Questions) 3
Foundational Items 3
A-SCALA – OVERVIEW AND BENEFITS 4
1. Ability of Students and Difficulty of Items 4
2. Causal Relationships Among Items 4
3. Student Success Profile 4
4. Student Proficiency Cards 4
5. Students Class Strengths and Gaps 4
CASE STUDY 5
Ability of Students, Difficulty of Items 5
Difficulty of Items 5
Ability of Students 6
Item Characteristic Curve 7
Causal Relationships Among Items 8
Student Success Profile 9
Student Proficiency Cards 10
Item Level Proficiency 10
Exam Level Proficiency 11
Comparison of Students 13
Students Class Strengths and Gaps 13
Strengths and Gaps in the Course 14
Personalized Programs for Students 15
CONCLUSION 16

INTRODUCTION

A series of knowledge exams are commonly used to assess a student's qualifications. However, the raw scores of these exams are often incorrectly analyzed. This leads to incorrect conclusions about students' strengths and shortcomings, leading to wrong recommendations on how to eliminate these shortcomings.

The reasons for incorrect analysis often originate from misusing exam raw scores. When students are evaluated through a series of knowledge exams, it is tempting to manipulate the raw exam data using simple mathematics. However, researchers agree that using exams' raw scores to assess and compare students' achievement is erroneous[1].

The Difficulty of Items (Questions)

Do students exert equal effort to answer each item (question) in a knowledge exam? The answer is "No" since all items are improbable to be of similar or equal complexity. In some cases, educators assign different points (weights) to items, thereby reflecting the varying complexity of the items. However, it is the student's ability that determines the degree of difficulty of the exam item. Measuring students' proficiency (or qualification) using the sum or average of raw exam scores while ignoring the difficulty of various items and students' different ability leads to misleading results. To solve the problem of correctly measuring student achievement, A-SCALA uses the modified Polytomous Rasch Measurement Model, which correctly analyzes the raw exam scores while simultaneously assessing the difficulty of the items and the ability of students.

Foundational Items

In any course, knowledge of various topics taught is interdependent. Thus, we cannot assume that the knowledge required to answer one exam item correctly does not depend on the knowledge necessary to answer other items. Suppose one exam question (item) tests the knowledge of a specific mathematical technique, and two additional questions confirm the use of this technique for solving problems. A lack of the knowledge needed to provide a correct answer to the first question (item) leads to a failure to provide the correct answers to the other two questions.

On the other hand, the correct answer to the first question increases the chances of success in the two related questions – thus, the knowledge needed to provide the correct answer to the first question is considered foundational. Such relations among exam items are not always straightforward and obvious; they may include dependencies on more than one item and, therefore, are not easy to detect. Identification of the foundational items is essential for the continued success of students. A-SCALA uses Relational Bayesian Networks to solve this problem successfully.

[1] Wright BD, Stone MH. Best Test Design. Chicago: MESA Press; 1979

A-SCALA – OVERVIEW AND BENEFITS

A-SCALA methodology, methods, algorithms, and software solutions have been developed to extract actionable insights from knowledge exams. A-SCALA can be easily integrated with existing education management systems through a simple API or a data interface.

A-SCALA provides five core functionalities that are not available in the traditional methods of measuring the effectiveness of learning processes:

1. Ability of Students and Difficulty of Items

A-SCALA's modified Polytomous Rasch Measurement Model (PRMM) can process incomplete data (for example, missing values) and provide a reliable estimate of the difficulty of items and students' ability. A-SCALA provides educators with not only accurate information about the actual achievement of the students but also identifies malfunctioning (or faulty) exam items. Eliminating such items improves the quality of exams.

2. Causal Relationships Among Items

A-SCALA helps educators identify cause-effect relationships among exam items and identify foundational items. This functionality is implemented using Relational Bayesian Networks (RBN) and proprietary algorithms that build the structure of the networks from exam raw scores. Identifying foundational items, the causal relationships among them, and their dependence on students' ability play an essential role in developing the Student Success Profile.

3. Student Success Profile

Using the results of the PRMM and the RBN, the proprietary A-SCALA algorithm creates Student Success Profiles for each course. The most important outcome is the quantitative values for the components of the Student Success Profile. Using Student Success Profiles, universities and colleges can determine the threshold of a student's ability to ensure success in their studies.

4. Student Proficiency Cards

A-SCALA automatically creates Student Proficiency Cards that contain an estimation of individual student's proficiency levels in each item and serve as a basis for determining their overall proficiency in the course. A-SCALA automatically creates recommendations for personalized training programs that improve students' ability, eliminate existing gaps, and increase students' chances of success in the course.

5. Students Class Strengths and Gaps

A-SCALA aggregates individual Student Proficiency Card data to evaluate class competency in the course. Educators can use this information to identify possible gaps in the course and ways to address them.

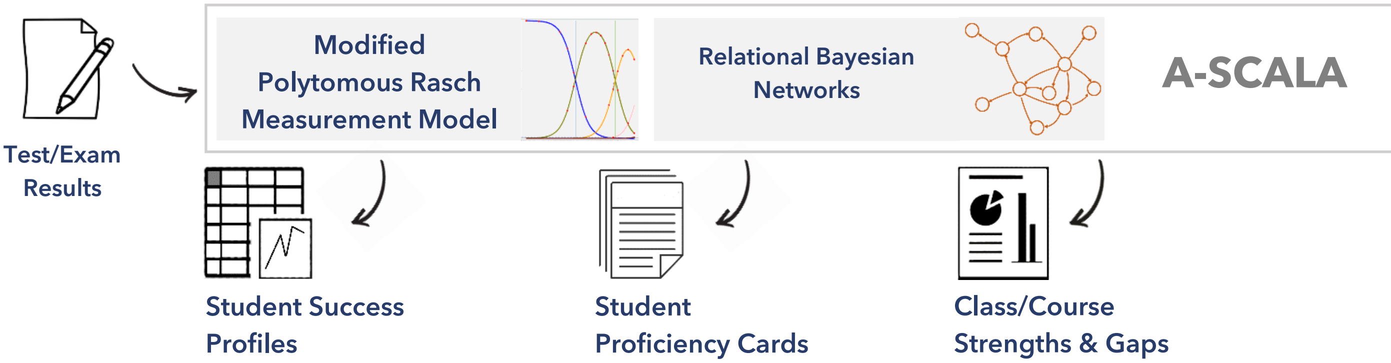


Figure 1. Architecture of A-SCALA

CASE STUDY

The course "Structure and Interpretation of Computer Programs" final exam was held for 2nd-year Electrical Engineering and Computer Science students. The exam was given to forty students. It contained 16 questions, each graded as either 1 (F), 2 (D), 3 (C), 4 (B), or 5 (A).

The traditional grading system based on the averaging of raw exam scores is presented on Figure 2 as a distribution of grades in the course:

- 2 out of 40 (5%) students received a B
- 30 (75%) students received a C
- 8 (20%) students received a D

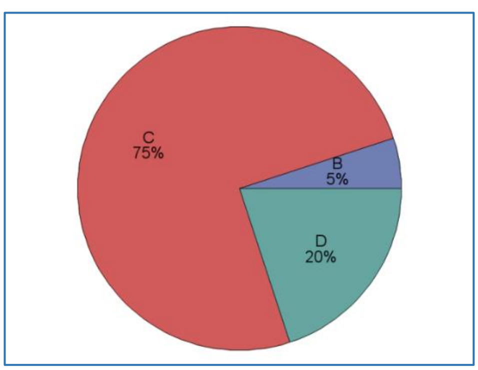


Figure 2. Distribution of Average Grades

The traditional grading method couldn't explain the reasons for the many low grades (C and D).

The course instructors decided to use A-SCALA to analyze the exam data to make informative and actionable conclusions.

Ability of Students, Difficulty of Items

The nature of the examination structure is that exam questions (items) have *different difficulty*, and students have *different ability regarding* the exam items. A-SCALA assesses the difficulty of exam questions (items), which reflects how easy or hard it was for students to answer them, and evaluates each student's ability regarding the exam as a whole. This is done by incorporating the modified Polytomous Rasch Measurement Model (PRMM) that estimates difficulty and ability *simultaneously*. As a result, items are ranked according to their difficulty, and the students are ordered according to their ability.

Difficulty of Items

The difficulty of the exam items, assessed by the PRMM, reflects how easy or hard it was for students to answer each question. Items with lower difficulty are easier to answer for students, and items with higher difficulty are harder to answer for students. Figure 3 demonstrates a substantial difference in the difficulty level of each item. Ignoring this critical information in the analysis creates false conclusions.

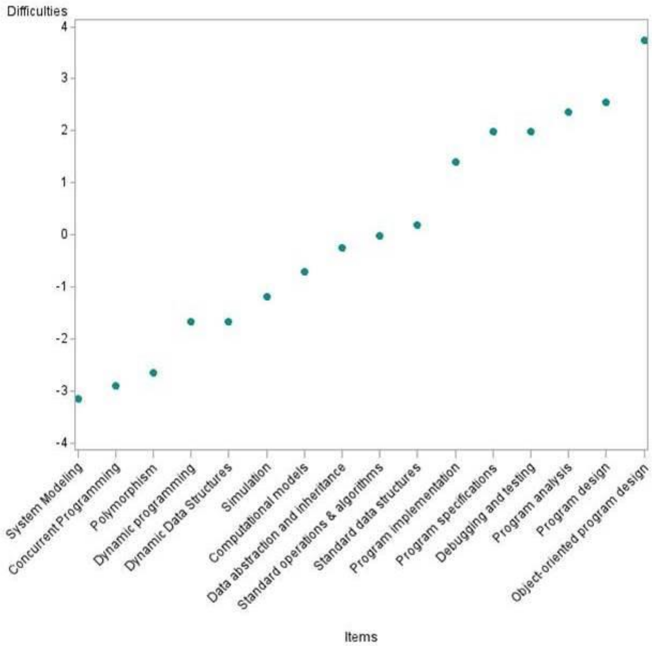


Figure 3. Difficulty of Exam Items (Questions)

The PRMM not only estimates the difficulty of the items but also determines the OutFit value, an outlier-sensitive fit. The OutFit is a mean-square residual summary statistic, which has an expectation of 1.0 and a range from 0 to infinity. An OutFit value greater than 1.0 indicates an underfit to the Rasch model, meaning the data (exam item score) is less predictable than the model expects. An OutFit value of 1.3 (see Table 1) indicates that there is 30% more randomness in the data than modeled and that the item's difficulty level does not always correspond to the student's ability.

There are three items shaded gray in Table 1, for which the OutFit value is greater than 1.3, so they appear malfunctioning (faulty).

Table 1. Items Difficulty

#	Item	Difficulty	OutFit
1	System Modeling	-3.14	3.04
2	Concurrent Programming	-2.90	0.90
3	Polymorphism	-2.65	0.87
4	Dynamic Programming	-1.67	0.65
5	Dynamic Data Structures	-1.67	0.46
6	Simulation	-1.19	1.34
7	Computational Models	-0.71	0.91
8	Data Abstraction and Inheritance	-0.25	0.70
9	Standard Operations & Algorithms	-0.02	0.60
10	Standard Data Structures	0.19	0.58
11	Program Implementation	1.41	0.50
12	Program Specifications	1.98	0.44
13	Debugging and Testing	1.98	0.58
14	Program Analysis	2.36	0.75
15	Program Design	2.55	5.56
16	Object-Oriented Program Design	3.73	1.13

It is possible that the questions were not clearly worded, may contain errors, were not covered sufficiently by the instructor, or may have other causes leading to a misunderstanding. A-SCALA excludes these items from the evaluation of the student’s proficiency.

Ability of Students

While making conclusions about the students’ performances, it is crucial to evaluate their ability, considering the difficulty of the exam questions (items). Such findings reflect *the actual proficiency* of the students, as opposed to the raw exam scores.

The student's ability is estimated by the PRMM, which is conditional on the difficulty of the items: small numbers mean lower student ability, and large numbers indicate higher ability. Figure 4 shows that students have different ability, which must be considered when analyzing.

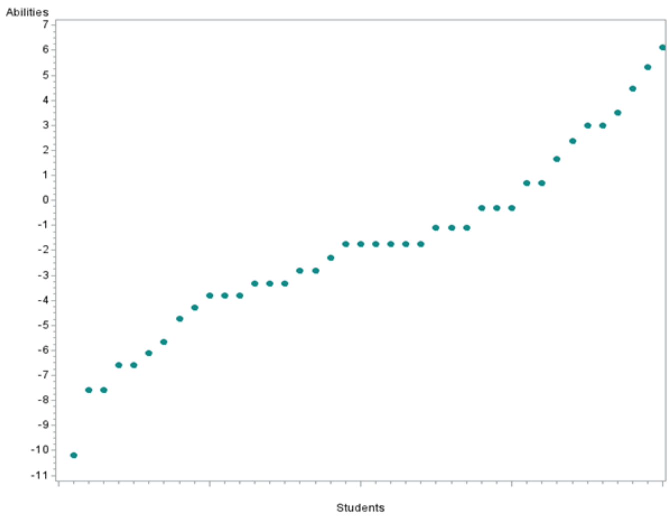


Figure 4. Ability of Students

An OutFit value above 1.3 indicates students for whom some of the observed exam scores are too far from the expected values assessed by the PRMM. This suggests a possible good guess or careless mistake by the students. There are six students in Table 2 (gray-shaded) for which the OutFit value exceeds 1.3. A-SCALA suggests re-examining these six students.

Table 2. Students Ability

#	Student	Ability	OutFit
1	SID001	-10.18	1.61
2	SID008	-7.58	0.43
3	SID031	-7.58	0.27
4	SID006	-6.58	0.24
5	SID028	-6.58	0.40
6	SID036	-6.11	0.28
7	SID022	-5.65	0.46
8	SID033	-4.74	0.28
9	SID015	-4.28	0.29
10	SID009	-3.81	0.27
11	SID026	-3.81	0.47
12	SID038	-3.81	9.75
13	SID014	-3.32	0.47
14	SID021	-3.32	0.22
15	SID025	-3.32	0.38
16	SID039	-2.82	0.57
17	SID040	-2.82	0.78
18	SID007	-2.29	0.48
19	SID004	-1.73	5.25
20	SID018	-1.73	0.32
21	SID019	-1.73	3.56
22	SID029	-1.73	0.32
23	SID030	-1.73	0.27
24	SID037	-1.73	0.32
25	SID010	-1.09	2.20
26	SID032	-1.09	0.53
27	SID034	-1.09	0.43
28	SID016	-0.30	0.41
29	SID020	-0.30	0.14
30	SID024	-0.30	0.14
31	SID002	0.68	0.07
32	SID003	0.68	0.07
33	SID017	1.64	1.37
34	SID013	2.38	0.23
35	SID005	2.99	0.66
36	SID027	2.99	0.22
37	SID012	3.52	0.68
38	SID023	4.46	0.32
39	SID035	5.31	0.38
40	SID011	6.12	0.21

Item Characteristic Curve

The PRMM creates Item Characteristic Curves (ICC) that describe the relationship between student ability and the likelihood (probability) that students will receive a specific score (Grade Category). Each item in the exam has its ICC, and for each item, A-SCALA estimates the probability that each student will answer a particular question (item) correctly.

For example, the following ICC is created for the "Dynamic Data Structures" item (see Figure 5). Each ICC curve represents the probability that students will receive a specific score (Grade Category), depending on their ability. Thresholds (solid vertical lines) determine the ability for which the probabilities of adjacent scores (Grade Categories) are equal. For example, the pink curve represents the probability distribution for this item to get a score of 4. According to Threshold 3 (red solid vertical line), a student with an ability of 3.92 has an equal chance of scoring 3 or 4. Thus, to get a score of 4, the student must have an ability above 3.92. The red dots on the curves denote the students' scores (Grade Categories).

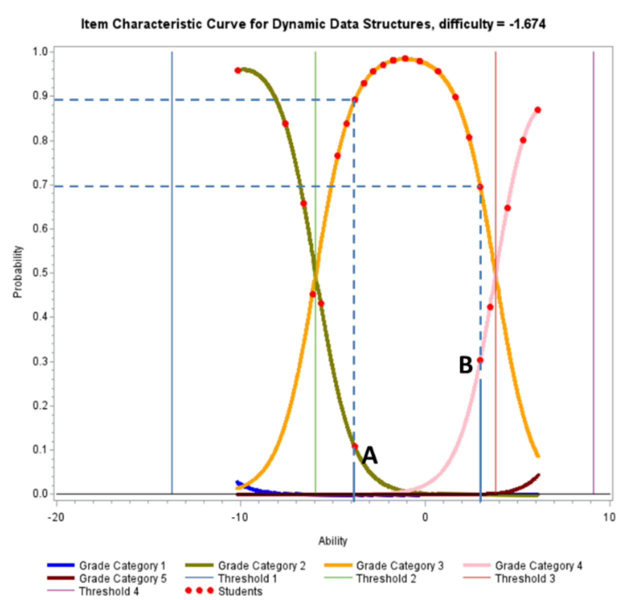


Figure 5. ICC for "Dynamic Data Structures" Item

Point A is located on the green curve (Grade Category 2) and represents student SID038 with an ability of -3.81. This student received a score of 2, while the probability of obtaining this score with their ability is only 0.11. The broken blue vertical line that passes through point A meets the yellow curve (Grade Category 3) at the point corresponding to the probability of 0.89. This indicates that the student SID038 has a probability of 0.89 to get a score of 3 instead of a score of 2.

Point B is located on the pink curve (Grade Category 4) and represents student SID005 with an ability of 2.99. This student received a score of 4, while for their ability, the probability of getting a score of 4 is only 0.30. The broken blue vertical line reaches the yellow curve (Grade Category 3) at the point where the corresponding probability of getting a score of 3 is 0.70. Can a score of 4 for this student be a lucky guess?

The ICC tells a different story for the "Simulation" item in the Figure 6. We already know that this item was identified as malfunctioning due to the high OutFit value. Points C, D, E, and F on the ICC show students who are more likely to receive scores different than what they received (see Table 3).

Table 3. Students Actual and Expected Scores

Point	Student	Ability	Actual Score	Expected Score / Probability
C	SID038	-3.81	2	3 / P = 0.84
D	SID004	-1.73	2	3 / P = 0.97
E	SID005	2.99	4	3 / P = 0.79
F	SID012	3.52	4	3 / P = 0.69

A-SCALA creates an ICC for all items and assesses what level of student ability can ensure success in the course.

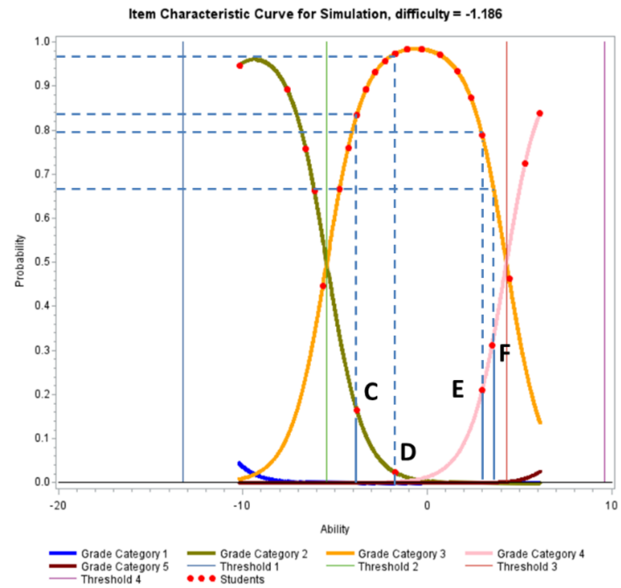


Figure 6. ICC for "Simulation" Item

Causal Relationships Among Items

Identifying the causal relationships among the exam items allows us to determine which *foundational knowledge* contributes to success in the exam. A-SCALA uses Relational Bayesian Networks (RBN) methodology to identify probabilistic *causal relationships* among exam items and students' ability. RBN visualizes the dependence or influence of one item on another in the form of a graph. The arrows in the graph, pointing from one item to another, reflect how the students' knowledge needed to answer one item correctly affects competency in another item.

In this case study, according to the RBN created by A-SCALA (see Figure 7), the ability of students is influenced directly by their knowledge of four items:

- Dynamic Programming (Item 4),
- Computational Models (Item 7),
- Standard Operations and Algorithms (Item 9),
- Standard Data Structures (Item 10).

However, in the educational process, it is crucial to determine which items are foundational. A-SCALA identifies foundational items critical to improving students' ability in the course. The following four items are identified as *foundational*:

- Data Abstraction and Inheritance (item 8),
- Program Implementation (item 11),
- Debugging and Testing (item 12),
- Program Analysis (item 14).

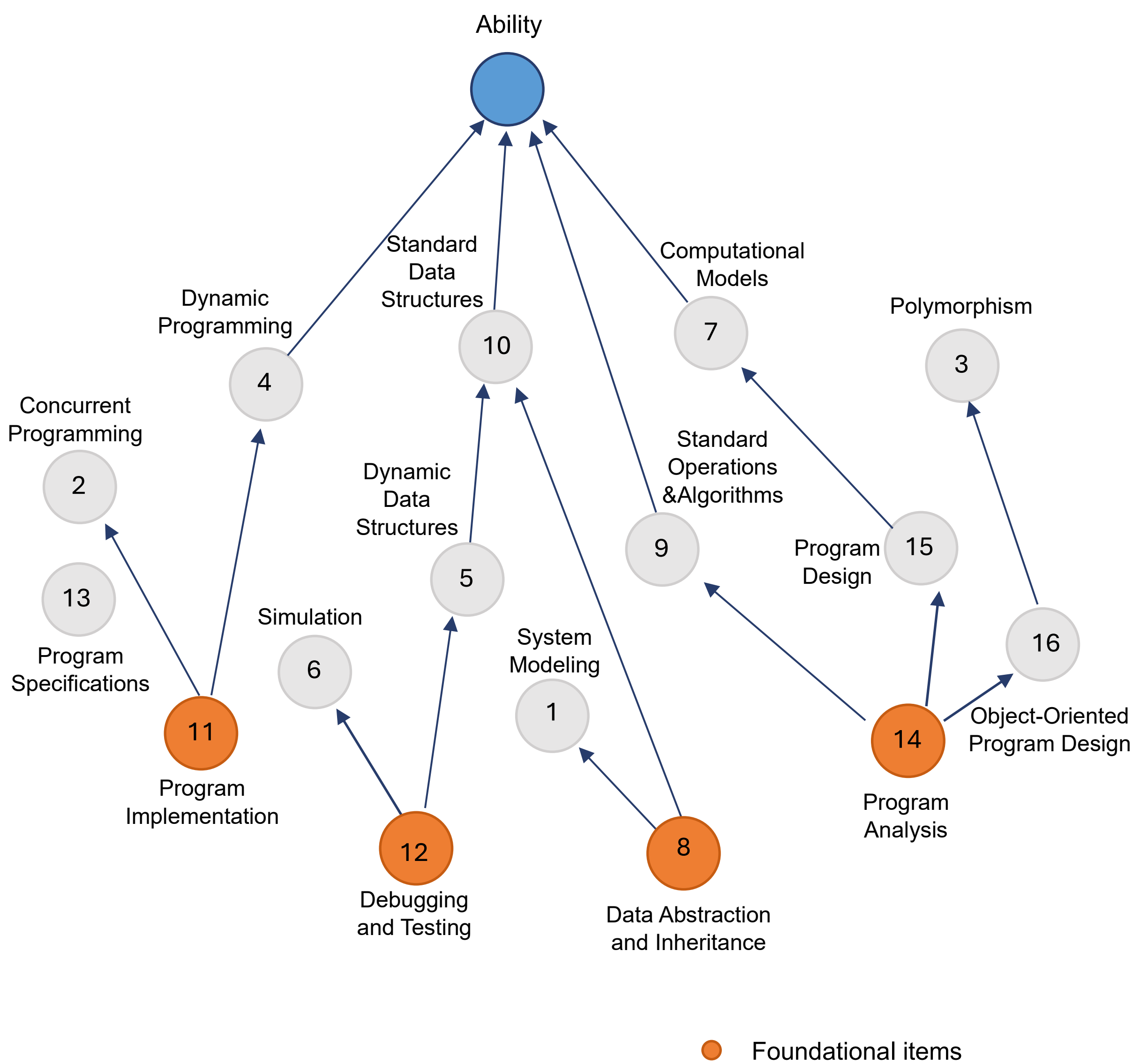


Figure 7. Relational Bayesian Network

Student Success Profile

The results produced by the RBN and the PRMM – a set of foundational and influential items, as well as the probabilities of obtaining specific scores on these items, form the basis of the Student Success Profile for the course. In this case study, A-SCALA creates a Student Success Profile (see Table 4) for the course "Structure and Interpretation of Computer Programs." Components of the Student Success Profile for the course are exam items associated with the lowest scores necessary for students to master the course successfully. The Student Success Profile states:

- Which items should be considered for success, and which should be excluded (items highlighted in gray were identified as faulty and were excluded from the Success Profile).
- What is the lowest score a student should get for each item to be considered proficient and
- Which items tested during the exam are foundational (red-bordered) for a student's success in the course?

Table 4. Student Success Profile

Student Success Profile for "The Structure and Interpretation of Computer Programs"

#	Item	Difficulty	Score	Item Importance
1	System Modeling	-3.14	4	Excluded
2	Concurrent Programming	-2.90	4	
3	Polymorphism	-2.65	4	
4	Dynamic Data Structures	-1.67	3	
5	Dynamic Programming	-1.67	3	
6	Simulation	-1.19	3	Excluded
7	Computational Models	-0.71	3	
8	Data Abstraction and Inheritance	-0.25	3	Foundational
9	Standard Operations & Algorithms	-0.02	3	
10	Standard Data Structures	0.19	3	
11	Program Implementation	1.41	3	Foundational
12	Debugging and Testing	1.98	3	Foundational
13	Program Specifications	1.98	3	
14	Program Analysis	2.36	3	Foundational
15	Program Design	2.55	3	Excluded
16	Object-Oriented Program Design	3.73	3	

Student Proficiency Cards

In this case study, 75% of students were given Grade C by the traditional approach. Is a C a good enough grade to be successful in this course? Do these 75% of students have the same level of proficiency in the course? Are these students on the road to success?

A-SCALA answers these questions using Student Proficiency Cards, the Student Success Profile for the course, and ICCs.

Item Level Proficiency

Student Proficiency Cards determine the level of competency in each exam item and associate it with the importance of the item and the required level of knowledge needed to answer each item correctly:

- Strength – the student exceeds the score requirement for the item in the Success Profile,
- Fit – the student meets the score requirement for the item in the Success Profile,
- Opportunity to Fit – the student has a high probability of meeting the score requirement for the item in the Success Profile,
- Gap – the student's actual and expected scores are lower than required for the item in the Success Profile.

Student Proficiency Cards contain actual and expected scores for each item (expected scores are assessed by the PRMM). This helps to identify *"hidden" gaps or "hidden" opportunities* – for example, cases where a student could guess the answer and get a higher score that was not expected according to the student's ability, or a student could mistakenly give the wrong answer and get a lower score than expected.

Exam Level Proficiency

Student Proficiency Cards are the basis for determining the *proficiency of each student in the exam*:

- Exceeds Proficiency (A) – the student demonstrates Strength in all items of the Success Profile,
- Proficient (B) – the student shows Strength or Fit in all items of the Success Profile,
- Foundational Proficiency (C) – the student shows Strength or Fit in all foundational items of the Success Profile,
- Partially Proficient (D) – the student shows Strength, Fit, or Opportunity to Fit in all foundational items of the Success Profile,
- Insufficient Proficiency (F) – the student was not classified in any of the four abovementioned groups.

The Student Proficiency Cards examples presented below contain the following data:

- Gray-shaded items are excluded from consideration as they were identified as malfunctioning (faulty).
- Red-bordered items were identified as foundational.
- Light-green cells determine the highest probability of scores for each item.
- The "Actual Score" column contains the score obtained by the student on a particular item.
- The "Most Likely Score" column contains the score that, according to the PRMM, is most probable for the student to obtain (See the probabilities in the light-green shaded cell).

Let's look at two examples. Student SID018, with an ability of -1.73, is Partially Proficient (D) in the course, as they exhibit an *Opportunity to Fit* in the foundational item "Debugging and Testing." The Proficiency Card also includes the non-foundational items in which the student demonstrated a *Gap* in proficiency: "Concurrent Programming," "Polymorphism," and "Object-Oriented Program Design."

Proficiency Card for Student SID018, Ability -1.73, Partially Proficient (D)

Item	Actual Score	Prob. Receiving Score 1	Prob. Receiving Score 2	Prob. Receiving Score 3	Prob. Receiving Score 4	Prob. Receiving Score 5	Most Likely Score	Success Profile	Status
System Modeling	3	0.00	0.00	0.98	0.02	0.00	3	4	Gap
Concurrent Programming	3	0.00	0.00	0.98	0.01	0.00	3	4	Gap
Polymorphism	3	0.00	0.01	0.98	0.01	0.00	3	4	Gap
Dynamic Data Structures	3	0.00	0.02	0.98	0.00	0.00	3	3	Fit
Dynamic Programming	3	0.00	0.02	0.98	0.00	0.00	3	3	Fit
Simulation	3	0.00	0.02	0.97	0.00	0.00	3	3	Fit
Computational Models	3	0.00	0.04	0.96	0.00	0.00	3	3	Fit
Data abstraction and Inheritance	3	0.00	0.06	0.94	0.00	0.00	3	3	Fit
Standard Operations & Algorithms	3	0.00	0.07	0.93	0.00	0.00	3	3	Fit
Standard Data Structures	3	0.00	0.09	0.91	0.00	0.00	3	3	Fit
Program Implementation	3	0.00	0.25	0.75	0.00	0.00	3	3	Fit
Debugging and Testing	2	0.00	0.37	0.63	0.00	0.00	3	3	Opportunity
Program Specifications	3	0.00	0.37	0.63	0.00	0.00	3	3	Fit
Program Analysis	3	0.00	0.46	0.54	0.00	0.00	3	3	Fit
Program Design	2	0.00	0.51	0.49	0.00	0.00	2	3	Gap
Object-Oriented Program Design	2	0.00	0.77	0.23	0.00	0.00	2	3	Gap

Another student, SID029, with the same ability -1.73, has *Foundational Proficiency (C)* as they demonstrated *Fit* for all foundational items. This student also shows *Gap* in proficiency: "Concurrent Programming," "Polymorphism," and "Object-Oriented Program Design."

Proficiency Card for Student SID029, Ability -1.73, Foundational Proficiency (C)

Item	Actual Score	Prob. Receiving Score 1	Prob. Receiving Score 2	Prob. Receiving Score 3	Prob. Receiving Score 4	Prob. Receiving Score 5	Most Likely Score	Success Profile	Status
System Modeling	3	0.00	0.00	0.98	0.02	0.00	3	4	Gap
Concurrent Programming	3	0.00	0.00	0.98	0.01	0.00	3	4	Gap
Polymorphism	3	0.00	0.01	0.98	0.01	0.00	3	4	Gap
Dynamic Data Structures	3	0.00	0.02	0.98	0.00	0.00	3	3	Fit
Dynamic Programming	3	0.00	0.02	0.98	0.00	0.00	3	3	Fit
Simulation	3	0.00	0.02	0.97	0.00	0.00	3	3	Fit
Computational Models	3	0.00	0.04	0.96	0.00	0.00	3	3	Fit
Data abstraction and Inheritance	3	0.00	0.06	0.94	0.00	0.00	3	3	Fit
Standard Operations & Algorithms	3	0.00	0.07	0.93	0.00	0.00	3	3	Fit
Standard Data Structures	3	0.00	0.09	0.91	0.00	0.00	3	3	Fit
Program Implementation	3	0.00	0.25	0.75	0.00	0.00	3	3	Fit
Debugging and Testing	3	0.00	0.37	0.63	0.00	0.00	3	3	Fit
Program Specifications	2	0.00	0.37	0.63	0.00	0.00	3	3	Opportunity
Program Analysis	3	0.00	0.46	0.54	0.00	0.00	3	3	Fit
Program Design	2	0.00	0.51	0.49	0.00	0.00	2	3	Gap
Object-Oriented Program Design	2	0.00	0.77	0.23	0.00	0.00	2	3	Gap

Comparison of Students

A-SCALA can distinguish among students who cannot be differentiated using traditional averaging (or summation) of scores. Let's look at the examples of two Student Proficiency Cards mentioned above: using the traditional method, students SID018 and SID029 received the same average value of their scores: 2.81. They each have the same grade in the course, a C (Grade C is associated with a score of 3, and the average value of 2.81 is considered a score of 3).

Table 5. Comparison of Students

#	Exam Items	Student SID018	Student SID029	Success Profile
1.	System Modeling	3	3	4
2.	Concurrent Programming	3	3	4
3.	Polymorphism	3	3	4
4.	Dynamic Data Structures	3	3	3
5.	Dynamic programming	3	3	3
6.	Simulation	3	3	3
7.	Computational models	3	3	3
8.	Data abstraction and inheritance	3	3	3
9.	Standard operations & algorithms	3	3	3
10.	Standard data structures	3	3	3
11.	Program implementation	3	3	3
12.	Debugging and testing	2	3	3
13.	Program Specification	3	2	3
14.	Program analysis	3	3	3
15.	Program design	2	2	3
16.	Object-oriented program design	2	2	3
	TRADITIONAL GRADE (based on average score)	2.81 (C)	2.81(C)	
	A-SCALA GRADE	D	C	

However, A-SCALA revealed that these students are different. Table 5 demonstrates that the student SID029 is in Fit with all the foundational items (blue shaded cells) and thus demonstrates Foundational Proficiency (C) in the course. The student SID026 is in Fit with only three out of four foundational items (the orange shaded cell corresponds to the item for which this student doesn't meet the score requirement) and, therefore, is only Partially Proficient (D). *The traditional method failed to reveal this critical difference.*

Students Class Strengths and Gaps

Students' proficiency in the exam is based on the scores obtained for the foundational items and not on all items, where some may have low importance or just be derived from the foundational items. Figure 8 snows that:

- 4 out of 40 (10%) students are Proficient (B),
- 11 (27.5%) students have Foundational Proficiency (C),
- 8 (20%) students are Partially Proficient (D),
- 17 (42.5%) students have Insufficient Proficiency (F)

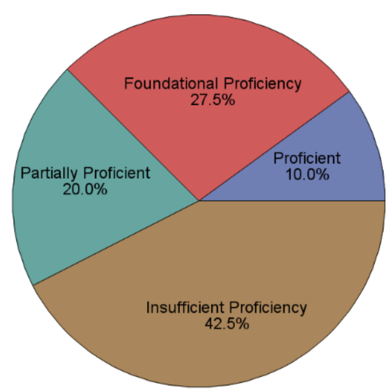


Figure 8. Distribution of Grades Based on A-SCALA Analysis

The exam grades calculated by A-SCALA are significantly different from the exam grades calculated using the traditional averaging method. The traditional approach (see Figure 9) *assigned the same Grade C* to thirty students (75%), while the A-SCALA approach allowed differentiating of these students. As we will show, the traditional method *failed to identify* students with Insufficient Proficiency (F).

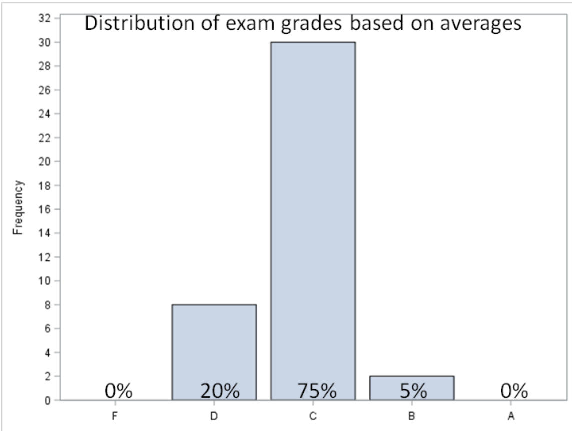


Figure 9. Distribution of Grades Based on Averages

A-SCALA identified 17 students with insufficient knowledge of the foundational items' topics (see Figure 10). Although these students have passed the exam per the traditional approach, they are at risk of attrition or failure in the future.

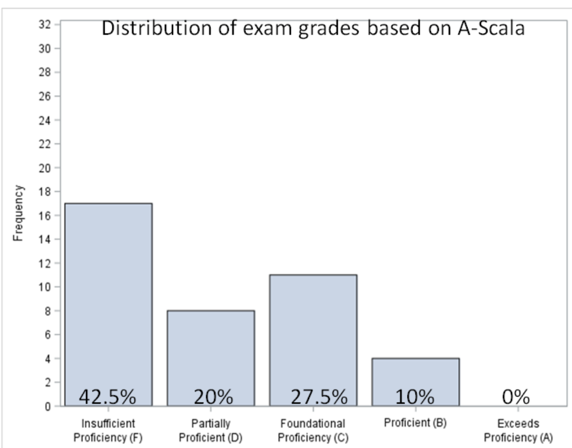


Figure 10. Distribution of Grades Based on A-SCALA

A-SCALA estimates that thirty students who received a Grade C based on traditional averaging of the scores should have *very different grades*. As shown in Figure 11:

- Two students are graded as Proficient (B) as they showed Strength or Fit in all items according to the Success Profile.
- Eleven students demonstrated Foundational Proficiency (C) as they showed Strength or Fit in all foundational items of the Success Profile.
- Eight students are Partially Proficient (D) as they showed Strength, Fit, or Opportunity to Fit in all foundational items of the Success Profile.
- Nine students were identified as having Insufficient Proficiency (F) as they failed at least one of the foundational items of the Success Profile.
- In addition, eight students who received a Grade D based on the traditional averaging method demonstrated Insufficient Proficiency (F) as they failed at least one of the foundational items of the Success Profile.

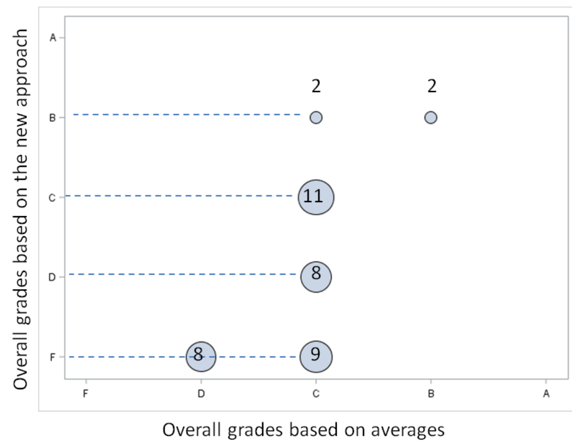


Figure 11. Grades Based on Averages vs. Based on A-SCALA

Strengths and Gaps in the Course

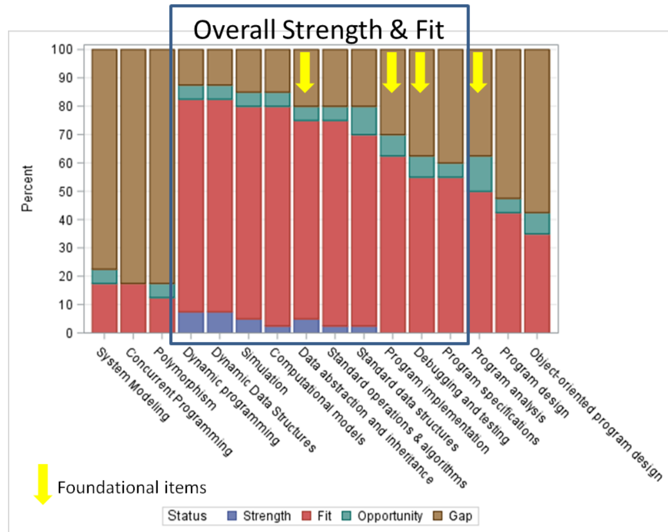
A-SCALA identifies areas of common strengths and gaps in the course, thus providing effective and actionable feedback to the teacher. The bar chart in Figure 12 displays the following insights:

- The first three exam items (from the left), although among the easiest, present Gaps regarding the Success Profile.
- The three most difficult items (from the right) have 50% or less Fit to the Success Profile.
- Foundational items determine the course's most important topics; in one of them, the "Program Analysis" item, less than 50% of students showed Fit.

- Ten items in the blue square outline topics where students mostly demonstrate Strength and Fit.

Personalized Programs for Students

A-SCALA automatically creates recommendations on improving students' ability, eliminating existing gaps, and increasing the chances of success in the course. For each foundational item (items 8, 11, 12, and 14), the students for whom reinforcement is required (red) or optional (green) were identified (see Table 6).



Items are ordered from the lowest difficulty to the highest

Figure 12. Students Strength & Fit to the Course

Table 6. Improvement Program for the Course

Improvement Program for "The Structure and Interpretation of Computer Programs"

Student	Data Abstraction and Inheritance (Item 8)	Program Implementation (Item 11)	Debugging and Testing (Item 12)	Program Analysis (Item 14)
SID001	Required	Required	Required	Required
SID006	Required	Required	Required	Required
SID008	Required	Required	Required	Required
SID022	Required	Required	Required	Required
SID028	Required	Required	Required	Required
SID031	Required	Required	Required	Required
SID033	Required	Required	Required	Required
SID036	Required	Required	Required	Required
SID009		Required	Required	Required
SID015		Required	Required	Required
SID021		Required	Required	Required
SID014		Required		Required
SID025			Required	Required
SID026			Required	Required
SID040	Optional	Optional	Required	
SID007		Optional	Required	
SID039				Required
SID019		Optional	Optional	
SID032			Optional	Optional
SID004	Optional			
SID018			Optional	
SID010				Optional

CONCLUSION

Traditional assessment of students' proficiency is made by summing or averaging the raw scores of the exam. In this case study, we showed that this approach produces misleading results. A-SCALA offers a solution that accurately assesses students' proficiency and gives practical advice on how to help students become more successful learners.

A-SCALA allows educators and institutions to:

1. Evaluate students according to their *ability, depending on the difficulty* of the exam items.
2. Determine the *quality of the exam structure* (in terms of items) and which topics are *foundational* for students' success in the course.
3. Measure the total student proficiencies in the course.
4. Create a *Student Success Profile* for each course, which determines the scores for each item a student must receive to become proficient.
5. Create Student Proficiency Cards that indicate each student's proficiency in each item and overall competency in the exam.
6. Create an improvement plan for each student.

According to the traditional measurement of proficiency, students receive grades that allow them to pass exams and move on. However, the reality of proficiency is much more complex than that. A-SCALA can determine much more accurately the state of student progress and proficiency in the classroom. A-SCALA provides a correct and accurate measurement of student performance and provides informative and actionable recommendations for addressing gaps in student education. A-SCALA helps educational institutions develop better and more successful learners, identify and handle issues in the educational process before they become problems, and ultimately, significantly reduce student attrition.