



STUDENT ABILITY DIAGNOSTIC REPORT

# Ability Diagnostic Report

## HISTORY

---

Ability estimate  $\theta$  · Actual vs Projected · Dimension profile · Cognitive diagnosis · Score stability · Personalised growth

YANZ Assessment Institute · 2026 Grade-12 Second Joint Mock · 2026-05-08

## STUDENT DETAILS · HOW TO READ

## ■ This report is about you

This report is built from your **item-by-item responses** in this joint mathematics examination. It uses Item Response Theory (IRT) to estimate your ability and Cognitive Diagnosis (DINA) to reconstruct your mastery across knowledge and ability dimensions. It does more than tell you **how many marks you scored** — it answers **where your true level is, and where to push next**.

School	<b>YANZ Demonstration High School</b>	Total Mark	<b>118 / 150</b>
Name	<b>ZHAO Zixuan</b>	School Percentile	<b>top 18%</b>
Class	<b>Grade 12 · Class 2</b>	Ability $\theta$	<b>+0.92</b>
Exam No.	<b>1228</b>	Overall Level	<b>4 / 5</b>
Assessment	<b>2026 Grade-12 Second Joint Mock · History</b>	School Mean	<b>96.5 / 150</b>
Date	<b>2026-05-08</b>		

### How to read this report

- **Actual vs Projected:** the *projected* result is the probability that a student at your current ability  $\theta$  would answer correctly. ■ green = actual **above** projection (over-performance); ■ red = actual **below** projection (the most worthwhile to revisit).
- **Levels 1–5:** 1 = needs work, 3 = competent, 5 = excellent — set by your score rate against the whole-cohort distribution.
- **Difficulty E/M/H:** Easy / Medium / Hard, auto-calibrated from the cohort's success rate. **Norm** = the cohort's score rate on that item.
- **Stability coefficient:** how much your mathematics result swings across repeated tests — the smaller, the more stable.

Methods: IRT 2PL/3PL ability estimation · Classical Test Theory difficulty & discrimination · DINA cognitive diagnosis · weighted dimension synthesis. The algorithms are validated on ~20 million real item-level responses.

OVERALL

### 01 Your overall performance at a glance

On this mathematics assessment your total mark is **118 / 150** (score rate **79%**), placing you in roughly the **top 18%** of your school, at overall level **4 / 5**. The chart below shows your ability level by dimension — the dashed line marks the maximum level.

<b>118/150</b> Total	<b>79%</b> Score rate	<b>top 18%</b> School percentile	<b>+0.92</b> Ability $\theta$
-------------------------	--------------------------	-------------------------------------	----------------------------------

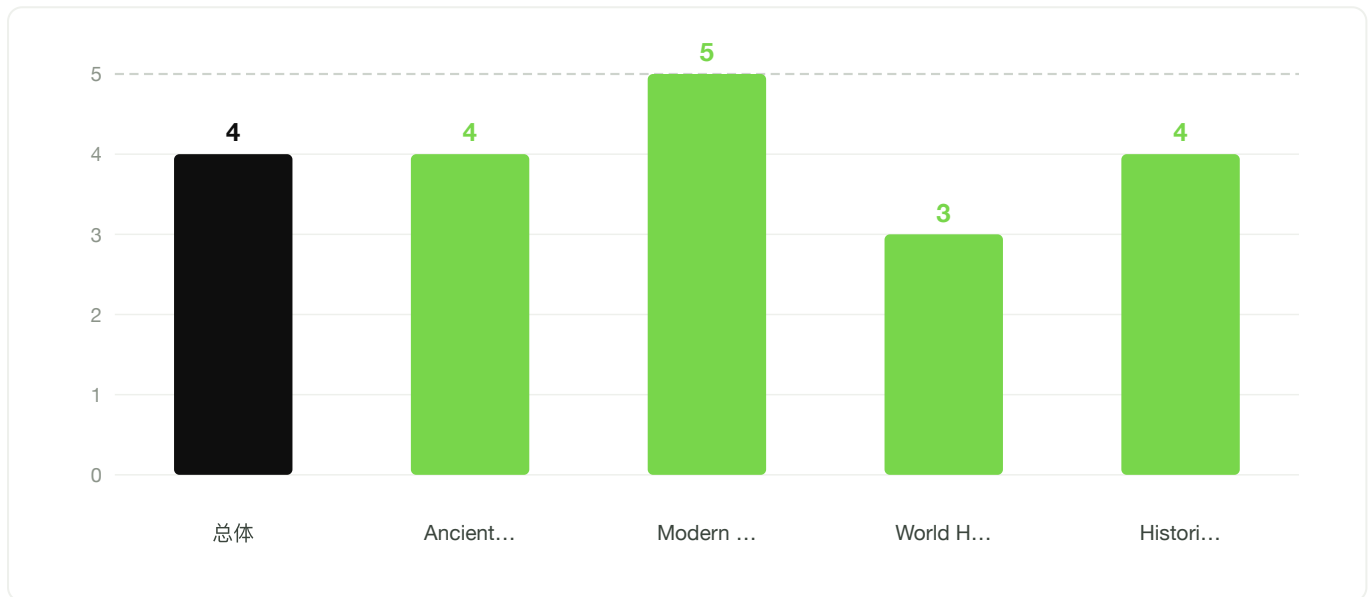


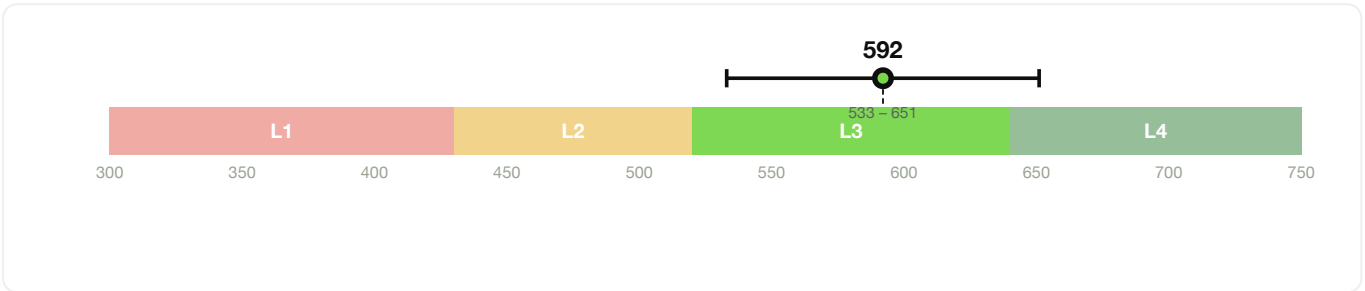
Fig 1 · Ability level by dimension (1–5). Dark = overall, green = content dimensions; dashed = level 5. **How to read:** Taller bar = higher level on that dimension; the dashed line is the maximum (level 5).

**In one line:** overall you are among the top of your year. **Modern China** and **Historical Literacy** are your strengths; **World History** clearly lags, and your scores swing between tests. Closing **World History** and shoring up **stability** are your two fastest routes to a higher score.

ABILITY SCALE · LEVELS · PRECISION

**Your ability scale score & achievement level**

Item Response Theory (IRT) converts each of your responses into a single **ability scale score** (mean 500, SD 100) — more stable than a raw mark and comparable across different papers. Every test carries error, so we also report the **standard error of measurement (SEM)** and a **95% confidence interval**: there is a 95% chance your true ability lies within that band.



**How to read:** The coloured bands are the four achievement levels; the dot is your scale score and the whiskers are the 95% confidence interval — the narrower it is, the more precise this measurement.

**Achievement level**

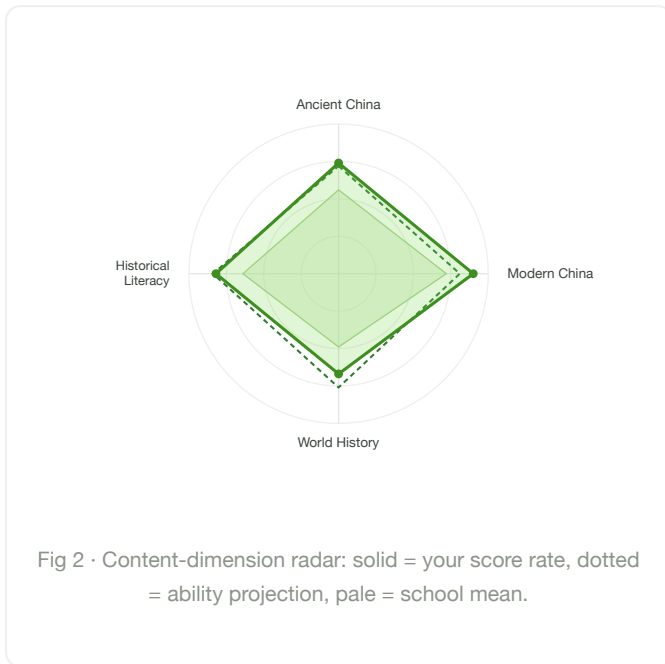
- L1 Emerging** Foundational knowledge and skills are not yet secure; needs systematic strengthening before progressing.
- L2 Approaching** Has grasped some core knowledge and can complete basic tasks; still unsteady on integrated, complex problems.
- L3 Proficient** Has a fluent command of the core curriculum standards, completes medium-difficulty tasks reliably, and already solves some integrated / harder problems.
- L4 Advanced** Excels across most dimensions, transfers knowledge flexibly, and reliably cracks higher-order and novel-context problems.

**Precision:** your SEM this time is about 30 points (high precision); the confidence interval sits mainly in L3 with its upper edge reaching L4 — so the “Proficient” classification is **robust**, and you are right on the threshold of “Advanced”.

DIMENSION PROFILE

## 02 Performance by content dimension

The radar compares your **score rate**, the **school mean** and your **ability projection** across four content dimensions. Where “your performance” beats “your projection”, you over-performed; where it falls short, you have a **high-yield place to improve**.



Dimension	Score / Max	Score %	School %	Projected %	Level
Ancient China	28/38	74%	56%	72%	4
Modern China	27/30	90%	72%	81%	5
World History	18/27	67%	49%	76%	3
Historical Literacy	45/55	82%	64%	84%	4

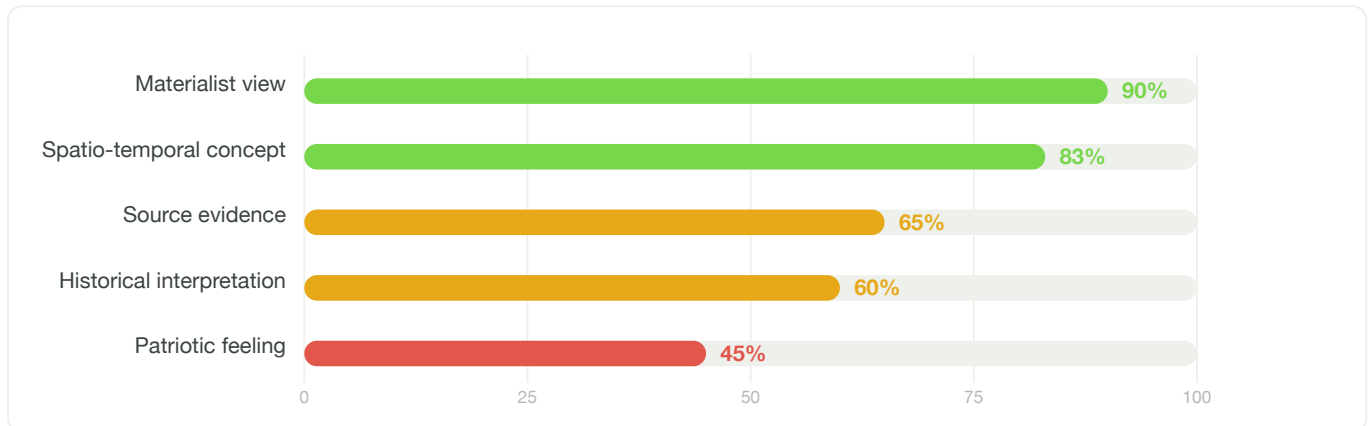
**How to read:** The further the solid line reaches out, the stronger you are; crossing the dotted line (your projection) means over-performance.

**Modern China** sits well above your ability projection (+9 points) — your signature strength. **World History**, by contrast, runs 9 points below projection: the same ability should have earned more marks.

CORE COMPETENCIES

### 03 History core-competency profile

Beyond the score, this report uses cognitive diagnosis (DINA) to estimate your mastery across the **subject's curriculum-standard core competencies** — explaining the *why*: marks are won or lost through the strength of a competency, not a single knowledge point.



Mastery of each subject core competency (0–100%). Green = attained (≥75%), amber = developing (50–75%), red = needs work (<50%).

**How to read:** Longer bar = stronger mastery of that ability; colour shows the mastery band.

Most of your core competencies are attained; the weaker ones are the priority to develop — competencies transfer to new item types far better than single questions.

ACTUAL VS PROJECTED

04 Item-by-item: actual vs projected

The tables compare your **actual marks** with your **ability projection** item by item. Green = over-performance (above projection), red = under-performance (below). The red items are the ones you “could do but didn’t get” — **revisiting them has the highest payoff**.

Selected-response (incl. multi-select, partial credit)

Q	Diff.	Max	Actual	Proj.
1 MC	Easy	5	5	5
2 MC	Easy	5	5	5
3 MC	Medium	5	5	5
4 MC	Medium	5	5	5
5 MC	Medium	5	0	5
6 MC	Medium	5	5	5
7 MC	Hard	5	5	3
8 MC	Hard	5	5	4
9 Multi	Medium	6	6	4
10 Multi	Hard	6	4	4
11 Multi	Hard	6	3	5

Fill-in & free-response

Q	Diff.	Max	Actual	Proj.
12 Fill	Easy	5	5	5
13 Fill	Medium	5	5	5
14 Fill	Hard	5	0	4
15 Fill	Hard	5	5	2
16 Free	Medium	12	11	10
17 Free	Medium	12	8	11
18 Free	Medium	14	11	11
19 Free	Hard	17	14	11
20 Free	Hard	17	11	9

You over-performed on 7 items and under-performed on 4. The shortfalls cluster on **Q5, Q11, Q14, Q17** — the items you “could do but didn’t get”, where revisiting pays off most.

BY DIFFICULTY

## 05 Performance across difficulty levels

Sorting items by difficulty (Easy / Medium / Hard) shows your strength at each level. ★ marks your current ability position — it sits **between Medium and Hard**: medium items should be fully secured, and you already break into hard items reliably.

Stronger (met or above projection)	Diff.	Weaker (below projection)
Q7, Q8, Q10, Q15, Q19, Q20	Hard	Q11, Q14
Q3, Q4, Q6, Q9, Q13, Q16, Q18	Medium ★ Your ability level	Q5, Q17
Q1, Q2, Q12	Easy	—

Your ability already presses into the “Hard” band. **The priority is not harder problems, but plugging the leaks at the Medium level** — each medium item secured lifts your score more reliably than grinding the final problem.

## SCORE STABILITY

## 06 Score stability analysis

Stability reflects how much your mathematics result swings across repeated tests: **the smaller the stability coefficient, the steadier you are**. Below, the x-axis is the stability coefficient and the y-axis is the total mathematics mark, split into four zones by “high vs low score × stable vs not”. ★ is your current position (from 5 tests this term: 109, 124, 103, 121, 118).

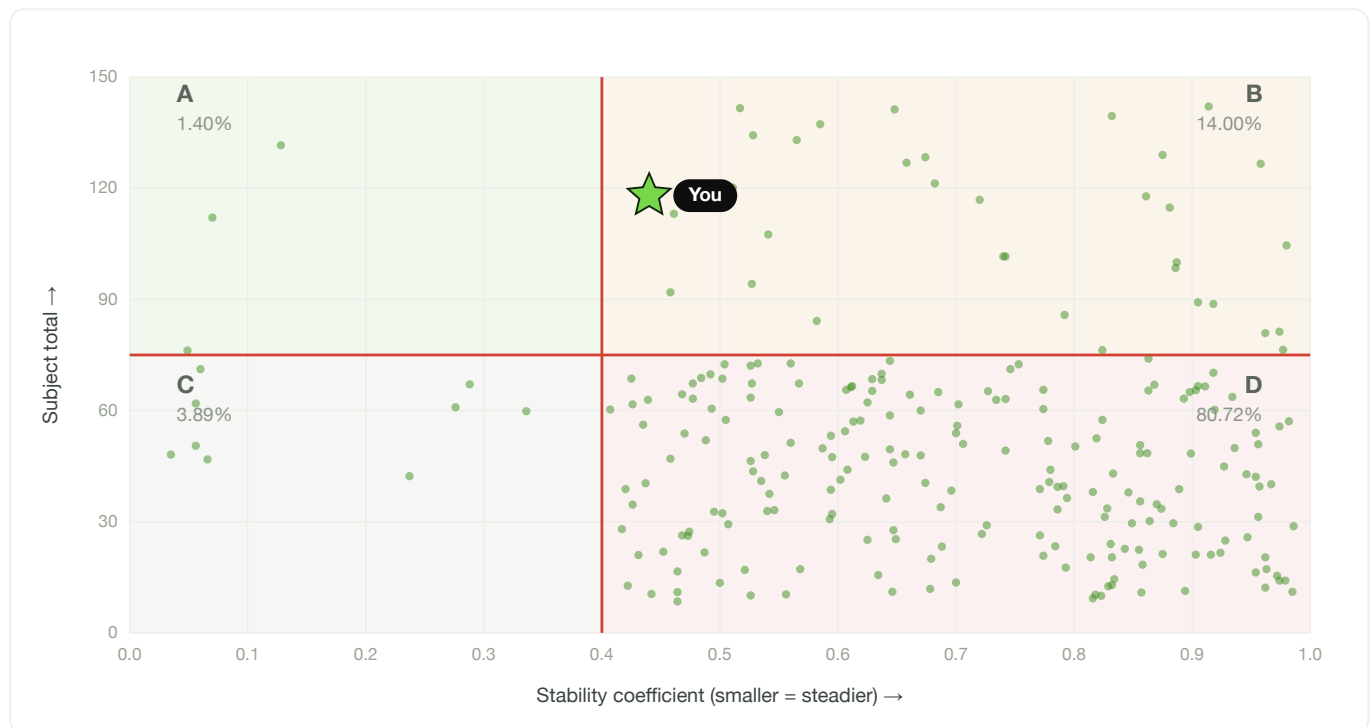


Fig 4 · Score-stability quadrant. Each dot is a candidate; ★ is you. X = stability coefficient (0–1), Y = mathematics total (0–150). **How to read:** Further left = steadier, higher up = higher score; the goal is to move toward Zone A (top-left).

**Zone A · High & Stable**

**Zone B · High but Unstable**

**Zone C · Low but Stable**

**Zone D · Low & Unstable**

You are in **Zone B (high but unstable)**: your marks are already among the year’s best, but they swing (103–124 this term). The swing comes mainly from **occasional slips on medium items**. **Stability is your most certain source of extra marks** — moving from Zone B to Zone A usually just means eliminating the “knew it, got it wrong”.

### Three ways to shrink the swing

- Keep an **error-attribution log**: for every wrong item ask one question — “didn’t know” or “knew it, got it wrong”? The latter is your stability leak.
- **Timed sectional drills**: 40 minutes on selected-response + fill-in, turning medium-item fluency into muscle memory.
- **Pacing discipline**: lock in the medium items (Q1–15) before attacking the final problem, so hard questions don’t eat your checking time.

PEER GROUPS

### 07 Where you stand across groups

The same score can sit differently across reference groups. Below is your position within three groups — **gender / age / region** (the further right the dot, the higher you rank).

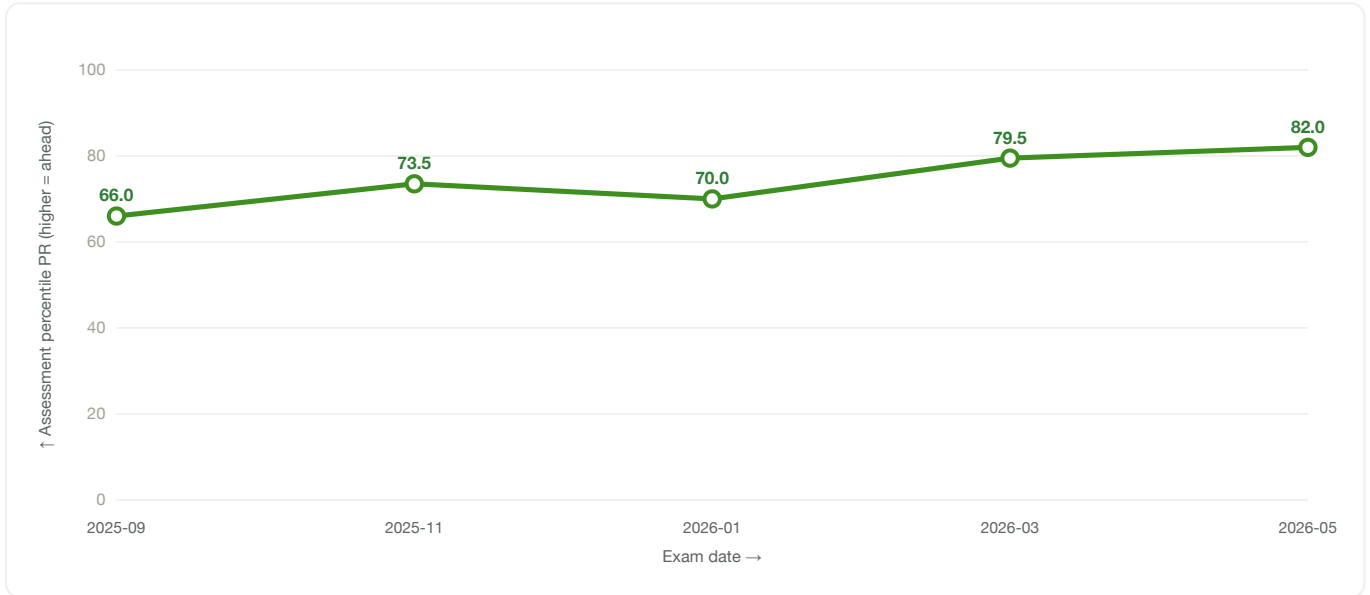


Within your **gender** and **same-age** groups you are already near the top; your **region** is more competitive — room to keep climbing.

TREND

### 08 Assessment-percentile trend over time

The chart shows how your **assessment percentile** changes over successive exams. Because cohort sizes differ each time, absolute rank is not comparable, so we use a **percentile rank PR (0–100, higher = ahead)** — a ratio score; **the higher the line, the higher your standing**, and a steady rise means steady progress.



X = exam date, Y = assessment percentile PR (a ratio score, independent of cohort size); higher on the curve is better. **How to read:** Line up = standing improved; down = slipped. Note the exam at each turning point and review what happened.

Your assessment percentile is on a **steady upward trend** (PR 66 → 82); keep the pace and push for a higher percentile.

## DETAILED DATA

## 09 Exam history

Your exam-by-exam detail. **Click any row** to jump to the full diagnostic report for that exam.

Date	Score	Percentile (PR)	Level	Report
2025-09	101 / 150	66.0	Lv3 Competent	<a href="#">View →</a>
2025-11	108 / 150	73.5	Lv3 Competent	<a href="#">View →</a>
2026-01	104 / 150	70.0	Lv3 Competent	<a href="#">View →</a>
2026-03	112 / 150	79.5	Lv4 Good	<a href="#">View →</a>
2026-05	118 / 150	82.0	Lv4 Good	<a href="#">View →</a>

Note: “Percentile (PR)” is a ratio score (0–100, higher = ahead, independent of cohort size); “Level” matches the 1–5 levels used earlier. Click any row to open that exam's full report (in this sample the links point to the demo report).

## PRIORITIES

**10 Priority improvement list**

These are the items where you had the **ability but didn't score** (actual clearly below projection), ranked by payoff. Clearing them one by one is the **most efficient** revision you can do right now.

Q	Diff.	Yours	Focus point	Diagnosis & advice
Q5 MC	Medium	0/5	Industrial revolution	Industrial revolution is not yet secure; marks within reach were missed — revisit first.
Q11 Multi	Hard	3/6	Modern invasions & resistance	Modern invasions & resistance is not yet secure; marks within reach were missed — revisit first.
Q14 Fill	Hard	0/5	World wars & Cold War	World wars & Cold War is not yet secure; marks within reach were missed — revisit first.
Q17 Free	Medium	8/12	Dynasties & institutions	Dynasties & institutions is not yet secure; marks within reach were missed — revisit first.
Q10 Multi	Hard	4/6	Ancient economy & thought	Ancient economy & thought is solidly mastered and stable.
Q16 Free	Medium	11/12	Chronology & interpretation	Over-performed on Chronology & interpretation — securing marks even on harder items.

## ITEM FOCUS

**11 Per-item focus & feedback (Assessment Focus)**

The full per-item list: difficulty, cohort success rate (Norm), your mark, the assessment focus and personalised feedback. This is your navigation map when you revisit the paper.

Q	Diff.	Norm	Yours	Focus point	Feedback
Q1	Easy	90%	5/5	Dynasties & institutions	Dynasties & institutions is solidly mastered and stable.
Q2	Easy	86%	5/5	Ancient economy & thought	Ancient economy & thought is solidly mastered and stable.
Q3	Medium	66%	5/5	Modern invasions & resistance	Modern invasions & resistance is solidly mastered and stable.
Q4	Medium	71%	5/5	PRC construction & reform	PRC construction & reform is solidly mastered and stable.
Q5	Medium	58%	0/5	Industrial revolution	Industrial revolution is not yet secure; marks within reach were missed — revisit first.
Q6	Medium	62%	5/5	World wars & Cold War	World wars & Cold War is solidly mastered and stable.
Q7	Hard	38%	5/5	Source-based evidence	Over-performed on Source-based evidence — securing marks even on harder items.
Q8	Hard	30%	5/5	Chronology & interpretation	Over-performed on Chronology & interpretation — securing marks even on harder items.
Q9	Medium	55%	6/6	Dynasties & institutions	Over-performed on Dynasties & institutions — securing marks even on harder items.
Q10	Hard	34%	4/6	Ancient economy & thought	Ancient economy & thought is solidly mastered and stable.
Q11	Hard	28%	3/6	Modern invasions & resistance	Modern invasions & resistance is not yet secure; marks within reach were missed — revisit first.
Q12	Easy	77%	5/5	PRC construction & reform	PRC construction & reform is solidly mastered and stable.
Q13	Medium	60%	5/5	Industrial revolution	Industrial revolution is solidly mastered and stable.
Q14	Hard	36%	0/5	World wars & Cold War	World wars & Cold War is not yet secure; marks within reach were missed — revisit first.
Q15	Hard	24%	5/5	Source-based evidence	Over-performed on Source-based evidence — securing marks even on harder items.
Q16	Medium	58%	11/12	Chronology & interpretation	Over-performed on Chronology & interpretation — securing marks even on harder items.
Q17	Medium	52%	8/12	Dynasties & institutions	Dynasties & institutions is not yet secure; marks within reach were missed — revisit first.
Q18	Medium	50%	11/14	Ancient economy & thought	Ancient economy & thought is solidly mastered and stable.
Q19	Hard	30%	14/17	Modern invasions & resistance	Over-performed on Modern invasions & resistance — securing marks even on harder items.

## LEARNING PATH

## 12 A growth path built for you

Turning the diagnosis into action. The recommended order is **steady first, stretch second**: 4 weeks on stability and Trigonometry & Sequences, then 4 weeks reaching for modelling and the final problem.

### 1 Weeks 1–4 · Steady the base, close the gaps

1. Build an error-attribution log and clear the “knew it, got it wrong” losses at the medium level.
2. Daily targeted practice on your weakest dimension (World History), fixing the gaps this paper exposed.
3. One timed mini-paper each week to train pacing.

### 2 Weeks 5–8 · Build modelling, take the final problem

1. A synthesis & application theme: practise the “situation → method” translation.
2. Step-wise training on the hardest items — aim to secure the first parts of each hard question.
3. Turn the workflow of high-frequency item types into a reusable template.

**Stage goal** Holding your strengths, lift **World History** markedly and halve your score swing — your total should rise steadily, moving your quadrant position from Zone B into **Zone A (high & stable)**.

METHOD & HOW TO READ

How this report is computed

This report uses the international mainstream **Item Response Theory (IRT 2PL/3PL)** and **cognitive diagnosis (DINA)**. The three charts below explain, in turn: how an item's difficulty maps to your ability (the item characteristic curve), how precisely this paper measures someone at your level (the test information function), and where you stand relative to all candidates and to the item difficulties (the ability–difficulty map).

① Item Characteristic Curve (ICC)



**How to read:** The x-axis is ability  $\theta$ , the y-axis is the probability of answering the item correctly; a curve further right = a harder item. The vertical line is your  $\theta$ , and its intersection with the curve = your probability of getting that item right.

② Test Information Function (TIF)



**How to read:** A higher curve = more precise measurement at that ability (smaller SEM). This paper's information peaks at upper-medium ability; your  $\theta$  sits in the high-information zone, so the paper measures you quite precisely.

③ Ability–Difficulty Map (Wright Map)



**How to read:** On one shared “logit ruler”: the left is the ability distribution of all candidates; each dot on the right is one item's difficulty. Items **below** your ability line you are more than 50% likely to get right; those **above** are your challenge zone.

Glossary

Ability  $\theta$  (theta)

The latent ability estimated by IRT; 0 is the cohort average, positive is above average.

Scale score

A linear transform of  $\theta$  (mean 500, SD 100), enabling comparison across papers.

SEM

The uncertainty of a single estimate; smaller = more precise, used to build the confidence interval.

Difficulty  $b$  / Discrimination  $a$

$b$  = the ability needed for a 50% chance of success;  $a$  = how sharply the item separates high vs low ability.

Projected

The success probability predicted from your current  $\theta$ ; comparing with actual reveals over/under-performance.

DINA cognitive diagnosis

Infers the mastery probability (0–1) of each cognitive ability from your responses.



#### METHOD & NOTICE

### ■ Every conclusion comes from your real responses

Every metric in this report — ability  $\theta$ , actual vs projected, dimension levels, cognitive mastery, stability — is computed from your item-level responses through statistical models: auditable and explainable. It is not a label; it is a **map for moving upward**.

<b>20 of 20</b> Items scored	<b>IRT 2PL/3PL</b> Ability model	<b>DINA</b> Cognitive model	<b>~20 million</b> Validation set
---------------------------------	---	--------------------------------	--------------------------------------

Note: this is a **sample demonstration**; the student details and data are illustrative. Metric definitions and algorithms are documented in the *YANZ Algorithm & Data Manual*. For instructional use by students and teachers only — please do not distribute.