

REPORT 2.0 / 01

数学阶段 诊断报告

核对本次作答，选择少量值得复盘的证据，再计划一个短期学习尝试。

完全合成教学样板 | 不代表真人结果 | 不可用于个体决策

本报告不对应任何真实学生、学校、考试、常模或测量模型。所有人物、作答、分数、日期与编号均为虚构。

学生显示编号	S-2042
测评	八年级数学 · 合成阶段诊断卷 C
合成日期	2026-07-18
报告编号	YZ-DEMO-2042-MATH-01
报告版本	2.0.0-demo

这份样板不做什么

不评价性格、天赋或未来；不提供全省排名、录取概率、心理诊断，也不证明 IRT、CDM 或预测模型已经部署。

首页摘要

观察事实 · O + S

68 / 100

完成 19/20 题；这是本卷得分率，不是百分位。

待核证据 · O/H + S

18 / 30

几何相关题本卷相对较低，其中一题未作答；原因不能确定。

共同行动 · A + S

2 周实验

每周 3 次、每次 20 分钟，第二周用新题复查迁移。

先区分四种话，再进入三个阅读层。

STUDENT

我的证据、我的选择

先看本次事实、代表题、反例入口与可选择行动。报告支持复盘，不替学生定义自己。

PARENT

怎样理解、怎样支持

增加题量、覆盖、限制与支持性提问；不把一次结果改写为人格或长期判断。

PROFESSIONAL

为什么能说，为什么不能说

追到字段、版本、方法、误差、质量闸门、抑制原因与责任边界。

四类信息

O + S 已观察	在合成情境中记录的作答与计分事实。	例如：本次合成卷获得 68/100。
M + N 模型估计	依赖 IRT、CDM 或预测模型的输出。	本样板没有获准证据，因此不显示数值。
H + S 待验证假设	对原因、迁移或策略的候选解释。	必须同时写反例、验证任务和撤销条件。
A + S 行动	学生共同选择的短期学习尝试。	不是提分、等级或升学结果保证。

三个词不能混用：得分率 68% 是 $68 \div 100$ ；百分位 PR 要有明确参照群体；掌握概率要有获准认知诊断模型。本样板只有第一个。

SUMMARY / 03

先看三件事。

01 · 发生了什么

68 / 100

O + S 完成 19/20 题。

02 · 保留什么

26 / 35

O/D + S 代数相关题本卷相对较高；两道建模题主要步骤完整。

03 · 先做什么

3 × 20 分钟

A + S 每周 3 次，练习图形—文字—式子转换。

一项先别急着下结论的证据

O + S 几何相关题获得 **18/30**，其中一题未作答。

H + S 这可能与图形关系转换或时间分配有关，**目前不能确定原因**；需要新题和作答记录继续核对。

可以这样问

“几何题中，哪一步最容易失去线索？我们先用新题检查，再决定是否增加练习。”

不要这样说

“报告已经证明你几何不行，所以以后必须多刷题。”

本页当前不提供

N · 不可用 量尺分、置信区间、PR、等级、掌握概率与升学定位，因为本合成样板没有相应获准证据。

OBSERVATION / 04

总体表现与数据完整度。

68%本卷得分率 · $68/100 \cdot O/D + S$

不是：“超过 68% 的人”，也不是“能力只有 68%”。没有适用常模时，得分率不能改写成百分位。

20

应施测题目

19

已作答且可评分

1

未作答

0

数据丢失 / 无效

数据完整度提示

本次有一题未作答。该题按本合成评分规则计 0 分，但“未作答”与“作答错误”分开保留。未作答本身不能证明不会、粗心或时间管理差。

学生 / 家长怎么做

- 先确认身份、试卷与日期一致；
- 若记得作答过 I12，提出数据复核；
- 把“系统记录为空”与“不会”分开。

专业证据状态

- S** 合成计分内部勾稽为 68；
- S** 作答覆盖 $19/20 = 95\%$ ；
- N** CTT 信度、IRT θ 、条件 SE/CI；
- 没有 CI 时，不硬编码固定区间。

DIMENSIONS / 05

三个维度，只回答本卷 发生了什么。



图形的同等文字说明：所有数值均为完全合成 S。

本卷标签	证据	当前允许的读法
代数	26/35 · 7 题	本卷相关题中相对较高；需用新题确认能否迁移。
数据与概率	24/35 · 7 题	本卷有较多正确证据，也有步骤遗漏。
几何	18/30 · 6 题	本卷相对较低；一题未作答，原因待核。

不确定性边界：三个维度题量、分值和题型不同，率差不能自动解释为显著能力差异，也不能变成天赋或知识点掌握概率。

学生可以补充反例

- “课堂或作业中，我在这个维度通常能做到……”
- “这次考试中，有一个条件影响了我的表现……”
- “我认为最值得复查的是……”

五道值得复盘的合成代表题。

为保护题目安全，本样板只显示合成任务类型与步骤，不显示真实题干。

代表题用于生成新证据，不证明稳定掌握或永久短板。

项目	本次记录	观察	不能直接说
I03 代数建模	5/5	方程建立和求解步骤完整	已“永久掌握”
I06 代数条件题	4/5	主要方法正确，最后条件检查缺失	一定是粗心
I09 几何转换	2/5	识别一条关系，另一条未写成式子	没有空间能力
I12 几何综合	未作答 · 0/5	系统记录为空	不会、放弃或时间差
I15 数据解释	5/5	图表读取与理由完整	所有数据题都稳定

复盘顺序

01

回忆先看到了什么

02

指出从哪一步不确定

03

对照答案与评分规则

04-05

新情境重做并记录迁移

支持性提问

“你从哪一步开始不确定？”

需要避免

“这么简单为什么错？”

专业边界：本页只使用 item.score / response_status / rubric_scores 的合成 S 数据。irt_a,b,c / projected_probability / residual_class 全部为 N。

把已观察、候选解释与验证任务分开。

已观察 O + S

- I09 从图形关系转为式子时，只写出一条关系；
- I06 最后一步没有核对条件；
- I12 系统记录为未作答。

待验证假设 H + S

候选假设	为何检查	反例 / 替代解释	怎样验证
H1 图形—文字—式子转换暂不稳定	I09 有步骤缺失	也可能是题意理解或评分表达	做 3 道新情境题并口述依据
H2 多步题结束前检查不足	I06 漏条件	可能是时间、书写或题目特殊	做 4 道多步题，用 60 秒检查表
H3 I12 与时间顺序有关	该题未作答	也可能没看到、不会或记录错误	核对记录，再做一题同构任务

撤销规则：这些都不是结论。新证据不支持时，应删除或改写假设，不能为了维持原解释而忽略学生反例。

家长可以记录

“近两次作业中也出现最后条件未检查。”这是可继续核对的观察。

专业规则：每个 H 都要有证据、反例、验证任务、负责人、复查日期与撤销条件；AI 只能生成候选追问，不能把 H 自动升级成 O 或 M。

家长不能定性

“孩子一直不认真。”这是人格归因，不是本报告支持的证据。

两周，只先验证一件事。

本样板假设学生选择 A1；A2 只是备选。行动来自共同选择，不是自动推荐或效果承诺。

A1 · AGREED · A + S

图形关系转换 | 已选择

- 行为 每周 3 次、每次 20 分钟；每次完成 2 道“图形 → 文字 → 式子”任务。
- 方法 先口述两条关系，再写式子，最后用图形反查。
- 保留证据 解题步骤，以及“独立 / 提示后 / 仍不会”的自评。
- 成功标准 第二周 6 道新题至少 5 道能独立写出完整关系并说明依据。
- 复查日期 2026-08-01
- 暂停条件 任务明显超出课堂范围、持续挫败，或需要未获批准资源。

A2 · PROPOSED · A + S

多步题条件检查 | 备选

- 行为 完成题目后，用 60 秒核对“条件、单位、范围、答案句”。
- 成功标准 4 道新题至少 3 道没有因遗漏条件失分。
- 复查规则 若错误不在条件检查，停止该假设，转查知识或表征。

家长支持边界：提供固定时间与安静环境；每周只问一次“哪种方法更有用”；不把完成量扩大为惩罚性刷题，也不替学生作答。

到期必须记录：继续、调整、暂停或升级。不能无限复制同一模板，也不能把末等值的分数上涨作为唯一结果指标。

为什么本样板不显示 PR、等级或升学定位。

N 全省 / 全国百分位 PR

没有明确、适用、经过审核的参照群体与常模。

N 成就等级

没有获准标准设置、临界值误差与维护记录。

N 七档升学定位

成就等级与升学参考必须分开治理，不能互相替代。

N 录取概率或未来分数

没有外部验证、校准、年度政策与公平审查。

关键限制：本报告是完全合成样板，没有真实、适用、经过审核的参照群体、常模、标准设置或预测验证。68% 只能读作本卷得分率，不能改写成“超过 68% 的人”。

若未来正式报告显示 PR，必须同时回答

谁

与哪个目标总体比较

何时

常模年份与有效期

多少

有效样本与代表性

多准

误差、带宽与适配性

专业呈现动作

`/norm/* = N`; `/level/* = N`; `/admission/* = N`。渲染状态为 `suppress`：显示抑制原因，不渲染空排名条，也不共用一个“`validated`”标志解锁不同用途。

为什么本样板不显示 IRT、CI 或 DINA。

本样板没有真实标定样本、题目参数、模型拟合、条件精度、DIF、公平、Q 矩阵与分类准确度证据，因此所有模型字段均为 N。

当前抑制字段与解锁所需证据摘要。

字段	状态	为什么当前不报告 / 未来需补什么
θ / 量尺分	M + N	需 AM4 模型、参数匹配、拟合、估计器与极端分策略。
SE / CI	M + N	需实时条件精度、方法匹配、覆盖验证与误差传播。
ICC / TIF / Wright	M + N	需由当前题参实时生成，并与模型与 SE 勾稽。
掌握概率	M + N	需 Q 矩阵、可辨识、拟合、分类准确度与外部任务。
模型预期概率	M + N	θ 、题参、题型模型与残差规则需全部通过。

可用替代：透明的原始得分、题目步骤、数据质量状态与验证任务。没有模型数字不代表报告失败；用不存在的模型数字填满页面才是错误。

国际方法成熟

只说明该方法有公共研究传统，不说明它适合这份测评、这个群体或这个用途。

项目字段可发布

必须在当前总体、语言、版本、用途与有效期内完成标定、验证、签署和逐份质量闸门。

禁止：用静态 ICC、TIF、Wright Map、掌握概率或预测图冒充本报告已实算结果。

三次合成原始记录，只能并列看。

三个试卷没有完成共同量尺与等值。分数看起来逐次增加，不等于能力增长了 7 分，也不能预测下一次成绩。



图形的同等数据表。

日期	合成试卷	原始分	可比性
2026-03-12	卷 A	61/100	未等值
2026-05-23	卷 B	65/100	未等值
2026-07-18	卷 C	68/100	未等值

允许的后续假设：在相同蓝图的新任务中，几何关系转换是否更稳定。需要同时记录课程内容、考试时长、身体状态、练习效应与其他解释。

equating_status=unavailable/N; scale_score=null; reliable_change_status=not_reportable。不得画连续能力折线、增长箭头或预测线。

先确认理解，再决定下一步。

请用自己的话完成三句

01 一个观察事实

“我这次……”

02 一个限制

“这份报告不能说明……”

03 一个行动

“接下来两周，我选择……，到……复查。”

你可以提出异议

- 身份、考试或分数记录不对；
- 记得作答过，但系统显示未作答；
- 某个解释与真实情况不一致；
- 不希望某位接收者继续查看；
- 需要学科、数据、隐私或其他支持。

正式产品必须提供

- 真实可用的更正与申诉入口；
- 授权撤回与接收者范围；
- 支持、复测与专业升级路径；
- 更正后的新报告与旧版替代链。

本样板没有虚构服务入口。在正式报告中，联系方式为空会阻断发布；在本合成样板中，只能显示“演示无服务入口”。

家长边界：家长也完成 teach-back，但不能替学生填写学生的选择。各方理解不一致时，记录差异并暂停确定性结论。

升级边界：安全、健康、虐待、自伤或他伤线索不进入一般学习建议，应按本地合规流程升级。

字段知道自己能说到哪里。

本样板整体为 S；模型、常模、等值与预测字段为 N。

模块	关键字段	状态	最终呈现
报告身份	report_id / version / date	S	合成描述
作答 / 评分	raw / max / rate / coverage	S + reconciled	合成描述
维度	descriptive_aggregation / item_count	S	描述 + 限制
题目证据	score / response / rubric_step	S	代表题描述
IRT / 量尺 / 精度	theta / scale / SE / CI	N	抑制
常模 / 等级	norm / percentile / level	N	抑制 + 原因
CDM	q_matrix / mastery_probability	N	抑制 + 代表题替代
趋势	raw_history / equating / reliable_change	S / N	仅历次原始记录
行动	recommendation / action_plan	S	合成行动示意

合成勾稽

- 总分：26 + 18 + 24 = 68；
- 满分：35 + 30 + 35 = 100；
- 题数：7 + 6 + 7 = 20；
- 作答 19；未作答 1；缺失 0；无效 0。

已知限制

- 没有真实作答、题目、评分或用户；
- 没有信度、效度、公平与可访问性研究；
- 没有模型标定、常模、等值或预测；
- 行动计划没有效果数据。

勾稽通过只说明合成样板内部一致，不能把 S 升成 V。本文只验证信息架构、语言、字段状态与行动闭环，不能用来估计正式上线表现。

这份样板到哪里为止。

报告编号	YZ-DEMO-2042-MATH-01	生成目的	信息架构与培训演示
报告版本	2.0.0-demo	禁止用途	真人评估、模型证明、排名、升学、诊断或效果宣传
样板状态	Synthetic / S		

声明执行摘要

候选声明	闸门结果	呈现动作
原始分 68/100	合成数据内部勾稽通过	description_only
代数本卷相对较高	仅限本卷描述，同时给题数	description_only
几何稳定短板	单次、无精度与迁移证据	suppress
PR / 量尺 / CI / DINA	无对应常模或模型证据	suppress
能力成长	历次试卷未等值	改为“历次记录”
两周行动	合成共同决策模板	示意 + 不保证

撤回触发

- 水印被删除或被宣传为真人报告；
- 任一 N 字段填入模型、常模或预测数值；
- 分数或维度合计不再一致；
- 假设被改成原因，行动被改成效果承诺；
- 旧版被误当作当前正式模板。

来源边界

以下标准支持“用途先行、解释有证据、清楚报告限制、处理公平 / 精度与帮助用户理解”的设计原则；它们不证明本样板或 YANZ 字段有效。

- AERA / APA / NCME (2014), *Standards for Educational and Psychological Testing*。
- International Test Commission, *Guidelines on Test Use*, v1.2。
- International Test Commission, *Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*, v1.2。
- W3C (2024), *Web Content Accessibility Guidelines (WCAG) 2.2*。

完全合成 · 非真人 · 不可用于决策

后续更新应把旧版标记为“已被新版本替代”，保留审计链，不能静默覆盖后继续沿用旧截图。