

SCHOOL DECISION SUPPORT REPORT

# 校级测评 决策支持报告

学力通行证 · 智能决策引擎

五份校级报告 · 全员个性化 IRT 报告 · 样板

南粤实验中学

2026 级高一 · 起点测试 (示例 · 脱敏)

研值 YANZ 测评研究院

9 次联考 · 137 校 · 三届追踪实证 · V2.0

## 目录 · Contents

本报告共六部分，逐项覆盖《学力通行证》两份产品文档应展示的全部内容。

### 第一部分 · 核心资产

编制说明 · 本报告定位 .....	03
我们有什么 · 联考数据资产 .....	04
八大核心学校 · 纵向追踪 .....	05
数据—算法—报告 对应关系 .....	06

### 第二部分 · 技术底座

智能决策引擎 · 五大算法 .....	07
引擎① LGCM 潜在增长曲线 .....	08
引擎② E 值矩阵 $EA \times EB$ .....	09
引擎③ CDM/DINA 认知诊断 .....	10
引擎④ XGBoost 上线预测 .....	11
引擎⑤ HLM 多层线性模型 .....	12
数据挖掘层级 一 · 二 (CDM/GRM) .....	13
数据挖掘层级 三 · 四 (等值/推荐) .....	14
数据挖掘层级 五 + 综合架构 .....	15

### 第三部分 · 五份校级报告

五份校级报告 · 总览 .....	16
报告一 · 学校生源起点定位 .....	17
报告一 · 升学潜力预测 .....	18
报告二 · 班级分层地图 .....	19
报告三 · 制度落地参数 .....	20
报告四 · 分层干预名单 .....	21
报告五 · 教师效能基线 .....	22

### 第四部分 · 学生个性化报告

学生 IRT 报告 · 五个本质问题 .....	23
报告获取与隐私保护 .....	24

### 第五部分 · 服务与落地

学校合作 · 三个服务层级 .....	25
高一 vs 高三 差异化服务 .....	26
交付标准与售后 .....	27
试点设计 · 反事实推断 .....	28

### 第六部分 · 附录

常见问题解答 (FAQ) .....	29
方法与术语表 .....	30

## ■ 编制说明 · 本报告定位

本文件是《“学力通行证”课程创新模式说明书》（完整版）的**配套补充文档**，聚焦一个核心问题——

“我们有 9 次联考、137 所学校、3 个完整年级、连续 4 次追踪的个体逐题得分数据，如何用这套数据资产，让「学力通行证」制度在新学校精准落地？”

### 完整版已阐述（本文件不再重复）

- 课程制度设计（达标获通行证 · 自主安排课时申请制、教师审核机制）
- 项目式学习与多元评价设计
- YANZ 平台六大功能模块说明
- 国家课程政策依据
- 育人价值与实施规划

### 本补充文档聚焦

- **核心资产**：联考大数据有什么、价值何在
- **技术底座**：五大决策引擎 + 五大数据挖掘层级
- **五份校级报告**：面向不同管理角色的样板
- **学生个性化 IRT 报告**：定位、内容与隐私
- **服务与落地**：三个层级、差异化场景、交付售后、试点验证

数据来源：广东省高考研究会高考测评研究院 9 次联考数据实证（137 所学校，18/19/20 三个完整年级，连续追踪 4 次，个体层面逐题得分）。版本 V2.0（最终版）。本样板学校与个体数据均为**脱敏示意**。

## 第一部分 · 核心资产

## ■ 我们有什么 · 联考数据资产

本项目的算法模型并非“实验室产品”，而是基于真实教育场景的大样本数据训练而成。

9 次 全省联考	137 所 覆盖学校	≈20 亿 逐题数据条	3 届 · 4 追 踪 纵向同生追踪
-------------	---------------	----------------	--------------------------

## 数据资产明细

维度	数据规模
考试次数	9 次全省联考
数据规模	≈20 亿逐题数据条
覆盖年级	18 / 19 / 20 级 (连续 3 届)
追踪深度	19 级同一批学生 高一→高三 连续 4 次追踪
覆盖学校	137 所
数据粒度	个体逐题得分 · 试卷原题 · 双向细目表 · 评卷要求

数据粒度是关键：我们掌握的不止是总分，而是个体层面的逐题得分 + 试卷原题 + 双向细目表 + 评卷要求——这使得知识点级诊断（CDM）、能力结构分析（GRM）、跨年等值（IRT）等高级算法成为可能。

## ■ 八大核心学校 · 纵向追踪样本

以下 8 所学校横跨 18 / 19 / 20 三个年级，是纵向追踪价值最高的核心样本——它们贡献了同一批学生从高一到高三连续 4 次的逐题数据。

1 汕头市潮阳实验学校

2 湛江第一中学

3 东莞市东华高级中学（东城校区）

4 东莞市东华高级中学（生态园校区）

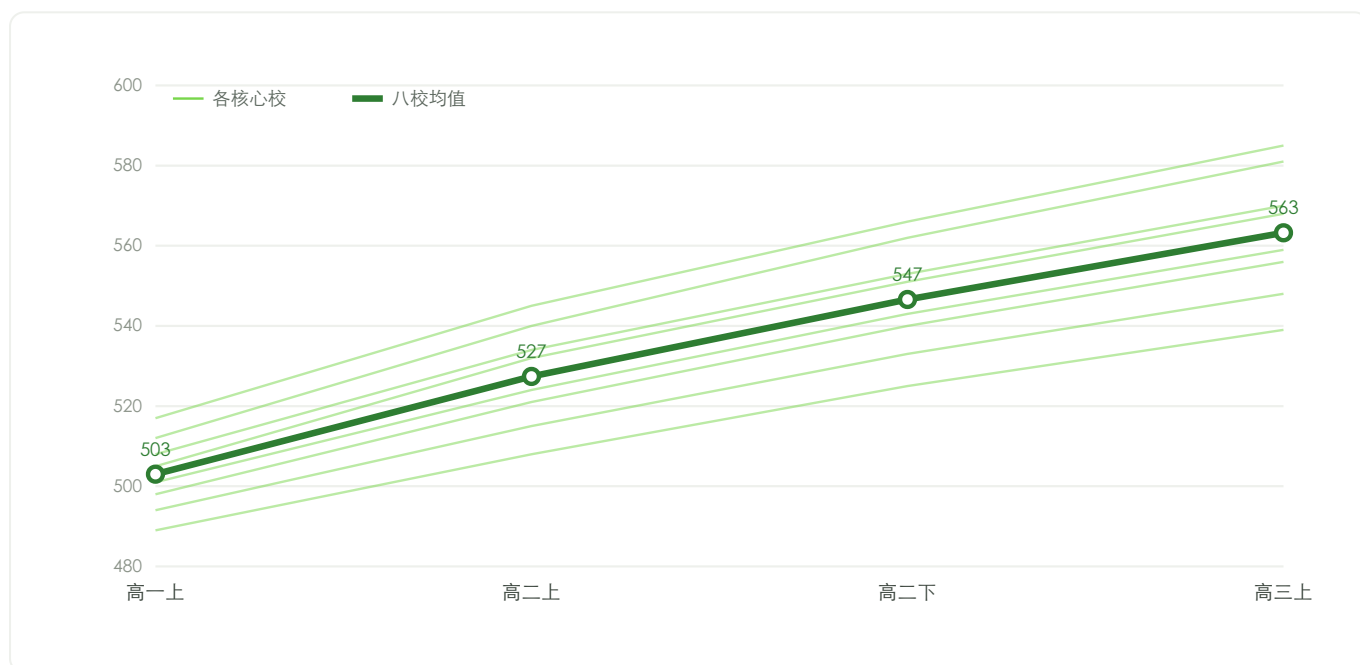
5 鹤山市第一中学

6 中山市中山纪念中学

7 广东高州中学

8 广东北江中学

### 八校 19 级纵向能力追踪（量表分）



细线=各核心校 19 级群体能力轨迹；粗线=八校均值。三年净增约 +70 量表分，增长稳定可预测。

### 这一资产的核心价值

#### 未来可预测

对“不同起点的学生未来大概率如何发展”建立了稳定认知。

#### 增值可量化

对“什么样的学校能实现高增值”有可量化的判断标准。

#### 预测已验证

预测模型经真实高考/模拟考结果回溯验证，误差  $\pm 2\%$  以内。

#### 算法旧 · 结论新

算法已稳定可靠，结论是贵校独有的专属预测。

## 数据 — 算法 — 报告 对应关系

每一条结论都可追溯到具体的数据资产与算法模型，形成完整的可解释链路。



### 对应关系总表（附录）

旧联考数据资源	数据挖掘 / 算法	产出成果	最终交付
19级4次追踪个体总分/排名	潜在增长曲线模型 (LGCM)	升学潜力预测参数 · 个性化达标线	报告一、三
18/19级各校各科均分/标准差	E 值矩阵算法 (EA×EB)	班级四象限分型标准 · 常模阈值	报告二
8核心校逐题得分 + 双向细目表	认知诊断模型 (CDM/DINA)	达标基准线 · 易错知识点预警 · 能力画像	报告二、三、IRT
18/19级各校上线率 + 临界生	XGBoost 二分类预测	临界生识别阈值 · 概率模型	报告四
同师不同起点班级出口数据	多层线性模型 (HLM)	教师增值基线 · 班级预期目标	报告五
全体学生逐题作答记录	IRT 垂直等值 + 协同过滤	成长轨迹 · 错题本与变式推荐	IRT 报告

## 第二部分 · 技术底座

## 智能决策引擎 · 五大算法

智能决策引擎由五大算法模块构成，共同支撑「学力通行证」制度在新学校的智能化运行。

算法模块	技术原理	训练数据来源	在「学力通行证」中的应用
LGCM 潜在增长曲线模型	计算「起始分→增长率」的协方差矩阵	19级4次追踪数据	为不同起点学生设定个性化达标线；预测升学出口；生成增值证明
E-Matrix E值矩阵 (EA×EB)	四象限分型：尖子带动 / 均衡稳健 / 偏科分化	18/19级各校各科均分与标准差	诊断班级学情类型；差异化审核自主申请权限
CDM/DINA 认知诊断模型	将答题模式映射到知识点掌握概率	8所核心校逐题得分 + 双向细目表	精确判定是否真正掌握当周知识点；防止「刷分通关」
XGBoost 梯度提升二分类	用高一/高二数据预测高三各层次上线概率	各校上线率 + 个体临界生数据	锁定临界生；动态调整自主申请权限
HLM 多层线性模型	剥离生源差异，计算教师教学净效应	同一教师所带不同起点班级的出口数据	生成教师增值评价基线；为评优评先提供公平依据

<b>LGCM</b> 潜在增长曲线模型 为不同起点学生设定个性化达标线；预测升学出口；生成增值证明	<b>E-Matrix</b> E值矩阵 (EA×EB) 诊断班级学情类型；差异化审核自主申请权限	<b>CDM/DINA</b> 认知诊断模型 精确判定是否真正掌握当周知识点；防止「刷分通关」	<b>XGBoost</b> 梯度提升二分类 锁定临界生；动态调整自主申请权限	<b>HLM</b> 多层线性模型 生成教师增值评价基线；为评优评先提供公平依据
--	---	---	---	--

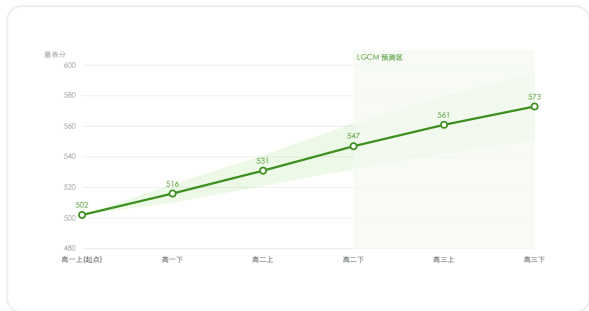
“算法是旧的，结论是新的。”——算法基于过去几万名学生的大数据训练而成，已稳定可靠；结论是把新学校的起点数据代入经过验证的算法框架后，算出的独有预测结果。后续5页将逐一详解每个引擎的原理与应用。

决策引擎 ① · LGCM

## 潜在增长曲线模型

LGCM 把每个学生的多次测评看作一条个人增长轨迹，估计「起始水平」与「增长速率」两个潜变量及其协方差——起点低但斜率高的学生，未来反超起点高但停滞的学生，模型能提前识别。

输入数据	19 级同生 4 次追踪的等值能力值（IRT 垂直等值后）
核心参数	截距(起点)、斜率(增长率) 的均值与协方差矩阵
输出	个人增长曲线 + 95% 预测区间 + 达标概率
方法出处	结构方程/多层模型框架 (Bollen & Curran, HLM)



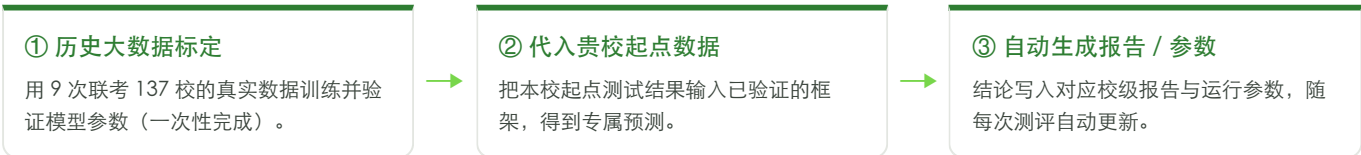
以起点为锚按真实增长协方差外推；阴影为 95% 预测区间。

### 在「学力通行证」中的应用

- 为不同起点学生设定个性化达标线
- 预测三年后升学出口（特控/本科上线人数）
- 运行一年后生成增值证明报告（实际 vs 预测）

**动态达标线（自适应）：**引入个人增长斜率后，系统可要求“本次能力值须高于本人前 3 次均值”方算达标，从机制上鼓励真实增长、抑制“刷分”。

### 三步落地



**一句话要点：**不必再用「一刀切」的固定达标线——让每个学生与自己的成长曲线赛跑。

## 决策引擎 ② · E-Matrix

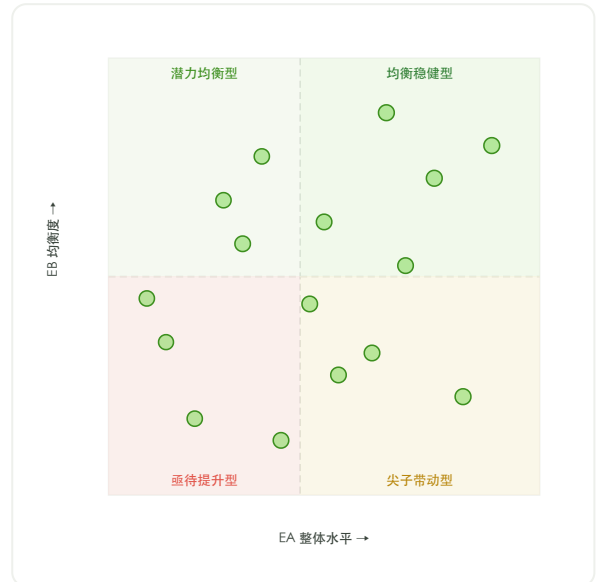
## E 值矩阵 (EA×EB)

E 值矩阵用两个维度刻画班级学情：EA（整体效能）与 EB（均衡效能）。把全部班级投到 EA×EB 平面上，按四象限自动分型，为差异化管理提供依据。

输入数据	18/19 级各校各科均分与标准差
EA 整体效能	班级均分相对全省常模的标准化位置
EB 均衡效能	班内离散度的倒数（越大越均衡）
输出	四象限分型标签 + 常模参照阈值

## 在「学力通行证」中的应用

- **均衡稳健型**（高 EA · 高 EB）：放宽自主申请权限
- **尖子带动型**（高 EA · 低 EB）：补齐尾部、防两极分化
- **潜力均衡型**（低 EA · 高 EB）：整体培优、稳步推进
- **亟待提升型**（低 EA · 低 EB）：强化基础、暂缓离堂



EA×EB 的含义：EA 衡量班级整体水平（均分相对常模），EB 衡量均衡度（标准差越小越均衡）。两维交叉自动分型，使每个班的自主申请审核策略差异化，而非一刀切。

## 三步落地

## ① 历史大数据标定

用 9 次联考 137 校的真实数据训练并验证模型参数（一次性完成）。

## ② 代入贵校起点数据

把本校起点测试结果输入已验证的框架，得到专属预测。

## ③ 自动生成报告 / 参数

结论写入对应校级报告与运行参数，随每次测评自动更新。

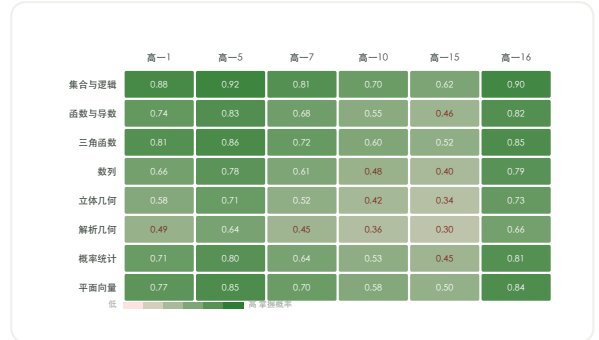
**一句话要点：**一张图看懂每个班是「尖子带动」还是「亟待提升」，审核策略不再凭经验。

## 决策引擎③ · CDM/DINA

## ■ 认知诊断模型

认知诊断 (DINA/GDINA) 借助双向细目表 (题目 × 知识点的 Q 矩阵), 把学生的答题模式反演为每个知识点的掌握概率; 多维 IRT (MIRT) 进一步给出「力学/电磁学/代数/几何」等细分维度的能力值。

输入数据	8 核心校逐题得分 + 双向细目表 (Q 矩阵)
DINA 参数	每题 slip(失误)/guess(猜测) + 学生属性掌握向量
MIRT	细分维度潜在能力 $\theta$ (了解/理解/掌握/运用)
输出	知识点掌握概率矩阵 + 层级判定



数学知识点 × 班级 掌握概率热力图: 越深越掌握, 浅红=待达标预警。

## 在「学力通行证」中的应用

- 实现真正的“达标”验证: 精确判定每个知识点的掌握概率
- 为 1 / 2 / 3 / 4 / 5 / 5+ / 5++ 能力等级提供精准锚点
- 易错知识点红色预警, 防止“刷分通关”

典型场景: 学生 A 物理卷面 85 分, 但“电磁感应”知识点掌握概率仅 0.3——系统应拒绝其申请该章节的自主安排课时, 并推送该知识点的巩固任务。这正是“真达标”而非“分数达标”。

## 三步落地

## ① 历史大数据标定

用 9 次联考 137 校的真实数据训练并验证模型参数 (一次性完成)。

## ② 代入贵校起点数据

把本校起点测试结果输入已验证的框架, 得到专属预测。

## ③ 自动生成报告 / 参数

结论写入对应校级报告与运行参数, 随每次测评自动更新。

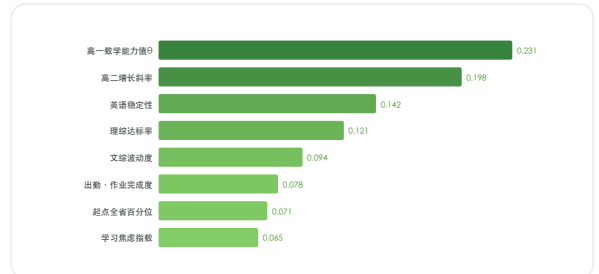
一句话要点: 把「分数达标」升级为「知识点真达标」, 这是弹性管理制度的诚信基石。

## 决策引擎 ④ · XGBoost

## ■ 梯度提升二分类预测

用高一/高二的多维特征，训练**梯度提升树**预测高三各层次**上线概率**。模型同时输出每个学生的上线概率与全局特征重要度，既能锁定临界生，又能解释「什么最影响上线」。

输入特征	能力值 $\theta$ 、增长斜率、稳定性、达标率、学习品质...
目标	高三特控/本科上线（二分类概率）
训练数据	18/19 级各校上线率 + 个体临界生数据
输出	个体上线概率 + 特征重要度排序



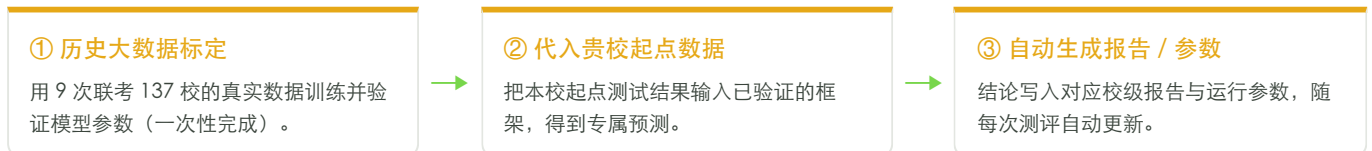
上线概率预测的特征重要度（Top 8）：高一数学能力值与高二增长斜率最关键。

## 在「学力通行证」中的应用

- 锁定临界生：识别“距上线仅差 5 - 15% 概率”的学生
- 动态调整自主申请权限（潜力型放宽、平台期收紧）
- 三届回溯验证，上线率预测误差  $\pm 2\%$  以内

**为什么用 XGBoost：**上线与否受多因素非线性交互影响（能力值、增长斜率、稳定性、学习品质...）。梯度提升树能自动捕捉这些交互，并输出可解释的特征重要度，让“抓哪科、盯哪生”有据可依。

## 三步落地



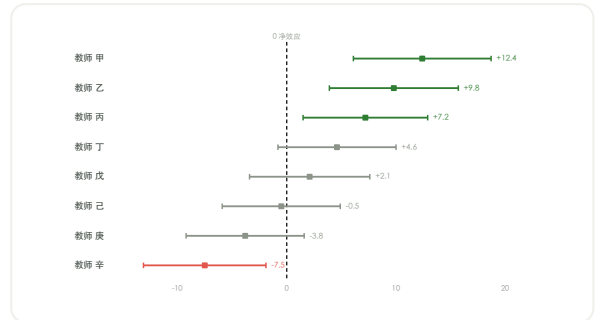
**一句话要点：**把有限的师资精力，精准投到「最可能被改变结果」的临界生身上。

决策引擎 ⑤ · HLM

## ■ 多层线性模型

教学质量不能只看出口分数，否则带「好生源」的老师天然占优。HLM（多层/分层线性模型）把数据的嵌套结构（学生 ∈ 班级 ∈ 教师）显式建模，剥离生源差异后得到教师的「教学净效应」。

模型结构	学生层（起点）嵌套于 班级/教师层
固定效应	起点对出口的总体回归
随机效应	各教师偏离基线的净效应 + 95%CI
输出	教师增值基线 + 班级预期目标



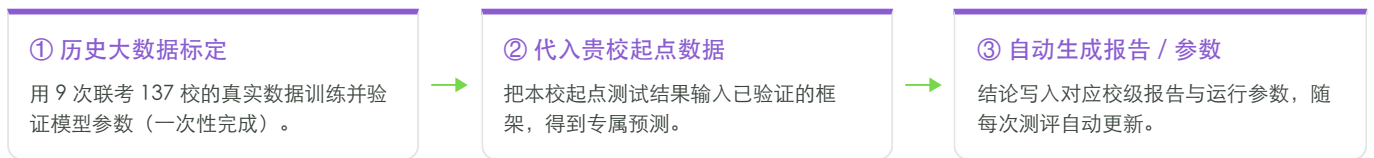
教师净效应森林图：方块=点估计，横线=95%CI；CI 不跨 0 即显著（绿正/红负/灰不显著）。

### 在「学力通行证」中的应用

- 生成教师增值评价基线，为评优评先提供公平依据
- 只与自身比较，不参与跨校排名评比
- 结合班级起点设定预期期末目标值

为何要剥离生源：直接比“班级均分”对接手弱班的老师不公平。HLM 把“学生层”嵌套进“班级/教师层”，控制起点后估计教师的净效应及其置信区间——CI 不跨 0 才算统计显著，避免误判。

### 三步落地



一句话要点：评价老师看「净增值」而非「生源运气」，让接手弱班的老师也被公平看见。

五大数据挖掘层级 · 之一二

### 三维认知诊断 + 能力结构诊断

#### 层级一 · 三维认知诊断 (CDM)

**对应数据** 逐题得分 + 双向细目表 (知识点 + 认知水平: 了解/理解/掌握/运用)

**挖掘算法** DINA / GDINA (知识点掌握概率); MIRT (力学/电磁/代数/几何 维度  $\theta$ )

**通行证应用** 真「达标」验证; 为 1 - 5++ 能力等级提供精准锚点

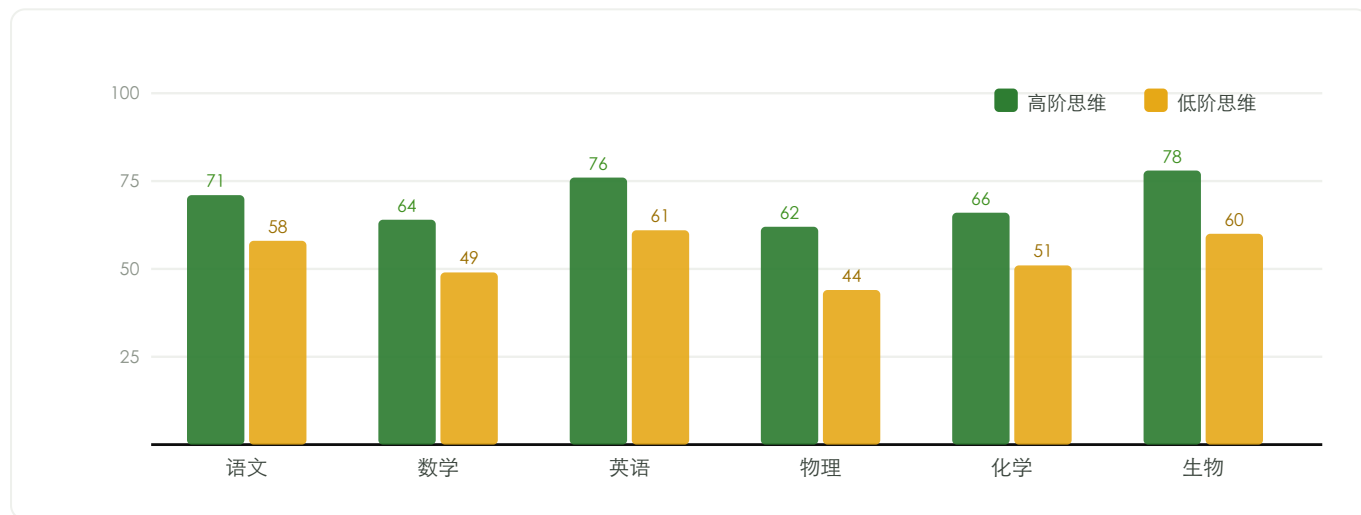
#### 层级二 · 能力结构诊断 (高阶 vs 低阶)

**对应数据** 评卷要求 (识记/理解/分析综合/论证探究等层级标注)

**挖掘算法** GRM 等级反应模型 (主观题); 认知负荷与错误归因编码

**通行证应用** 差异化审核: 高阶强 → 探究性自主; 低阶弱 → 强制基础巩固

#### 高阶 vs 低阶思维得分率 (GRM 估计)



深绿=高阶思维 (分析/论证/探究); 琥珀=低阶思维 (识记/理解)。物理高阶-低阶落差最大, 提示思维训练重点。

**赋能项目式学习 (PBL)**: 把 PBL 的“研究设计/过程投入/成果质量/反思成长”维度与联考主观题评分标准对标, 使“非执笔考核”与笔试能力描述一致, 错误归因还能建立学生的“错误诊断档案”。

## 五大数据挖掘层级 · 之三·四

## ■ 纵向等值追踪 + 个性化学习路径

## 层级三 · IRT 等值纵向追踪与增值预测

对应数据	19 级同生 4 次联考逐题得分 + 试卷原题
挖掘算法	垂直等值（锚题/共同知识点，统一能力量尺）；HLM/LGCM 增长斜率
通行证应用	动态达标线（须高于本人前 3 次均值）；预测下一阶段达标概率

## 层级四 · 个性化学习路径推荐

对应数据	试卷原题 + 双向细目表（考点/难度/区分度）+ 个体得分
挖掘算法	矩阵分解（SVD/FunkSVD）识别隐式偏好；基于内容的协同过滤
通行证应用	自主安排期精准推送 3-5 道变式题，让自习「学有所依」

## "测—评—练—拓"闭环



从联考错题到巩固资源：矩阵分解能识别学生更擅长“图像分析类”还是“计算推导类”，据此把历次联考错题直接转化为通行证达标后的个性化任务包，实现闭环。

## 五大数据挖掘层级 · 之五

## 诚信与效能分析 + 综合实施架构

## 层级五 · 多模态过程数据的诚信与效能分析

对应数据	评卷要求（含步骤分等过程评分规则）+ 客观题选项选择
挖掘算法	异常检测（Isolation Forest）； 解题路径复原（基于步骤分）
通行证应用	反作弊：识别「刷达标/虚假达标」触发二次复核（如随机抽题面谈）； 效能评估：对比获大量自主课时者与未获者后续步骤分/规范分变化

## 综合实施架构：数据层 → 算法层 → 应用层

数据层	算法层	「学力通行证」应用层
双向细目表 + 试卷	认知诊断 (DINA/MIRT)	自动生成学科知识图谱； 打 1 - 5++ 能力等级标签
逐题得分 + 评卷标准	多水平 IRT (GRM) + 错误归因	生成高阶/低阶思维雷达； PBL 设计能力基线
纵向 4 次等值分数	增长曲线模型 (LGCM)	计算能力增长率； 动态设定个性化达标线
学生 × 试题矩阵 + 错题标签	矩阵分解 + 协同过滤	自主安排期精准推送「最需练习的 3 道题」
选项 + 步骤得分 + 申请日志	异常检测 (Isolation Forest)	监测刷分行为触发复核； 评估对答题规范性的影响

诚信是弹性管理制度的生命线：越是给学生自主权，越要确保“达标”真实。异常检测在学生频繁申请初期介入，配合随机面谈，守住制度公信力。

## 第三部分 · 五份校级报告

## ■ 五份校级报告 · 总览

层级二及以上学校，在完成起点测试后获得以下**五份校级报告**，由智能决策引擎自动生成，数据分析师到校逐页解读。以下为各报告样板。

**报告一 · 学校生源起点定位与升学潜力预测**

阅读对象：校长 / 教学处主任 | 模块：起点雷达 · 等级分布 · LGCM 出口预测 · 同侪对标

**报告二 · 学科教学质量诊断与班级分层地图**

阅读对象：级部主任 / 教研组长 | 模块：E 值四象限 · 学科贡献率 · 班级 ANOVA

**报告三 · 制度落地适应性评估与运行参数建议**

阅读对象：教学处 / 教务主任 | 模块：放行比例 · 监控学科 · 分层达标线

**报告四 · 学生个人能力画像与分层干预名单**

阅读对象：班主任 / 年级德育主任 | 模块：临界生预警 · 偏科名单 · 学习品质

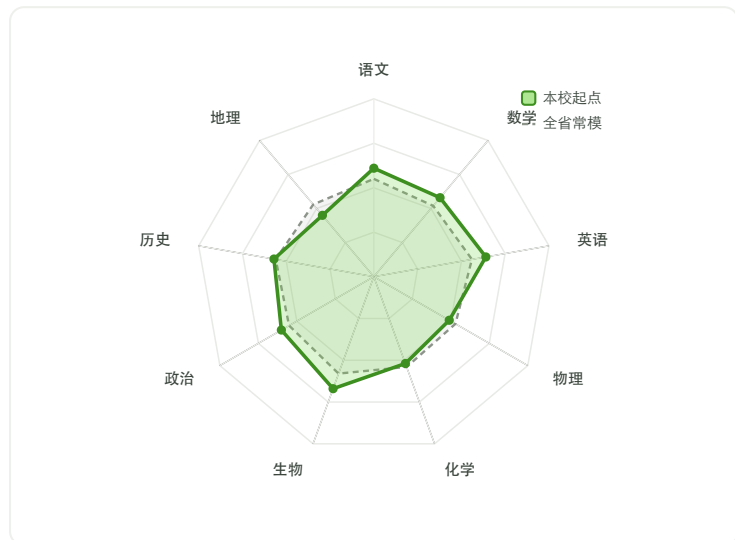
**报告五 · 教师教学效能基线报告**

阅读对象：教学处（评价存档） | 模块：HLM 教师净效应 · 班级起点-目标匹配

报告一 · 阅读对象：校长 / 教学处主任

## 学校生源起点定位

### 九大学科起点能力雷达



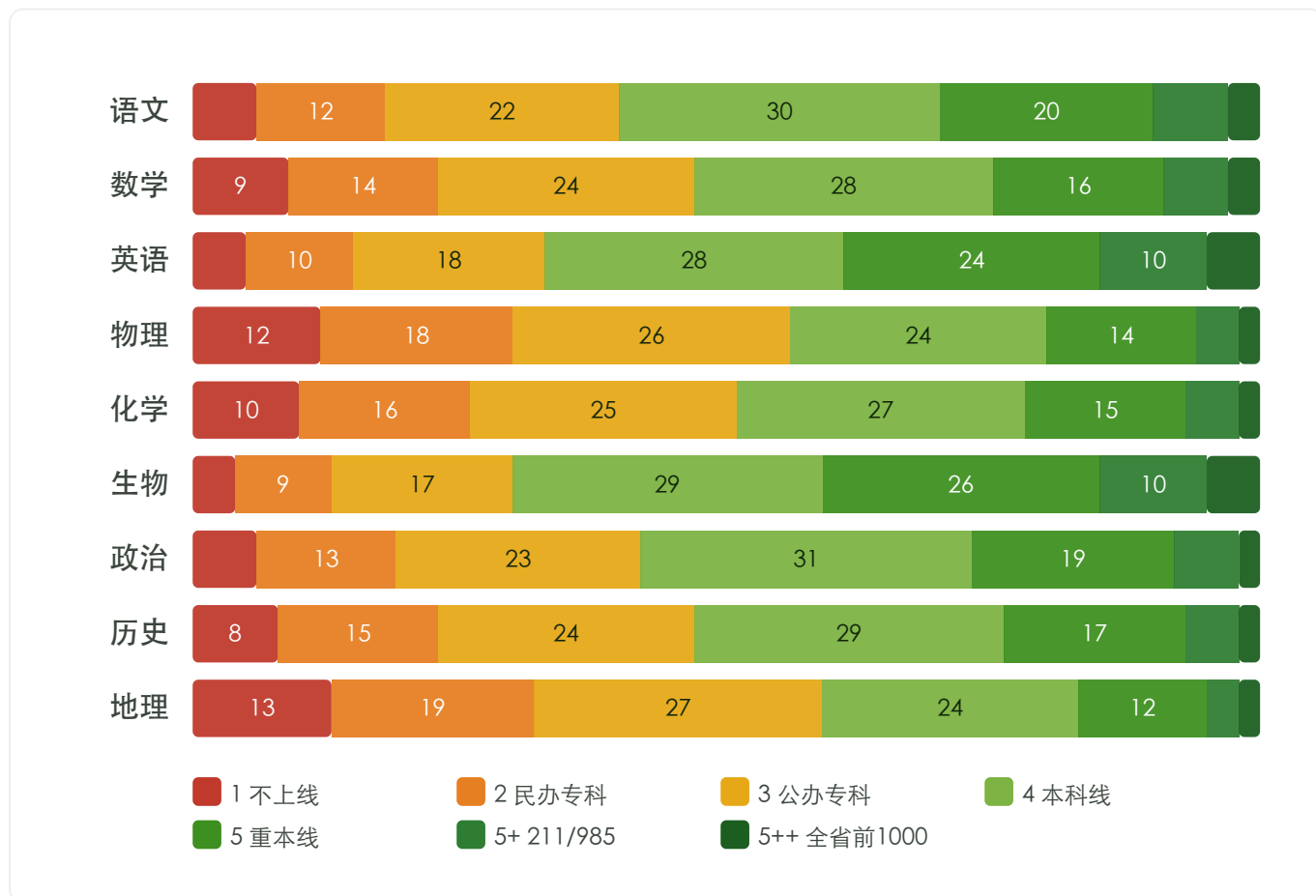
绿=本校起点达标率；灰=全省常模。生物/英语领先，地理/物理低于常模。

### 起点定位结论

<b>中上</b> 全省位次	<b>+3.1</b> 综合相对常模	<b>2科</b> 显著优势	<b>2科</b> 红色预警
-------------------	-----------------------	-------------------	-------------------

贵校 2026 级高一起点综合水平略高于全省常模。优势学科为生物 (+9)、英语 (+8)；地理 (-8)、物理 (-4) 需重点关注。建议高一阶段把资源向理科基础与地理倾斜。

### YANZ 能力等级人数分布 (各学科占比 · 1 - 5++)

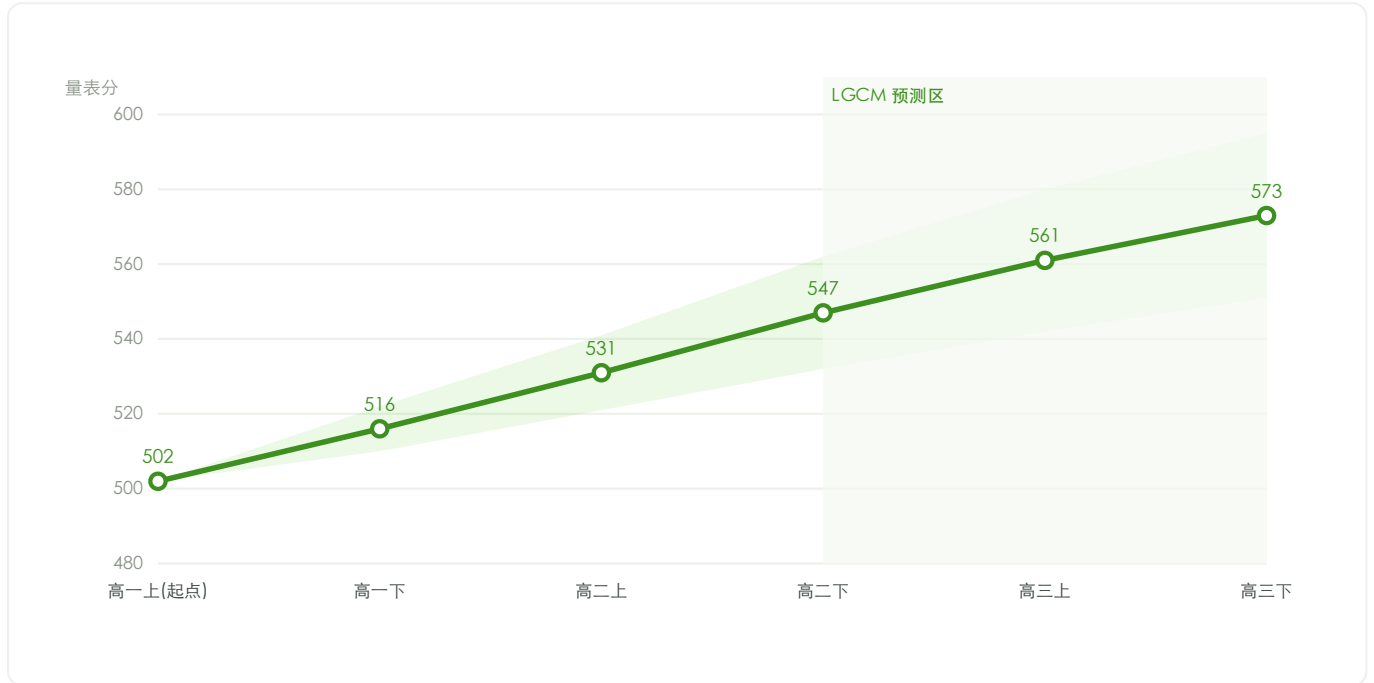


1 不上线 / 2 民办专科 / 3 公办专科 / 4 本科线 / 5 重本线 / 5+ 211 · 985 / 5++ 全省前1000——与个人报告同一套等级，由 IRT 能力值 + 全省常模 (CDM 知识点掌握) 综合判级。

报告一 · 阅读对象：校长 / 教学处主任

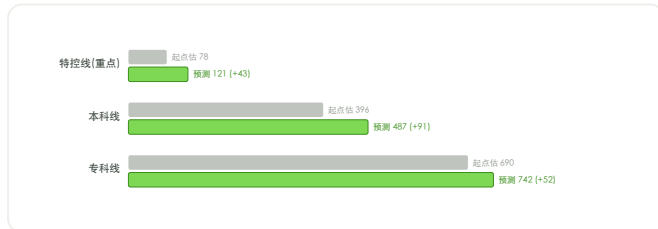
## ■ 高考出口潜力预测 (LGCM)

### 群体能力增长曲线与预测区间



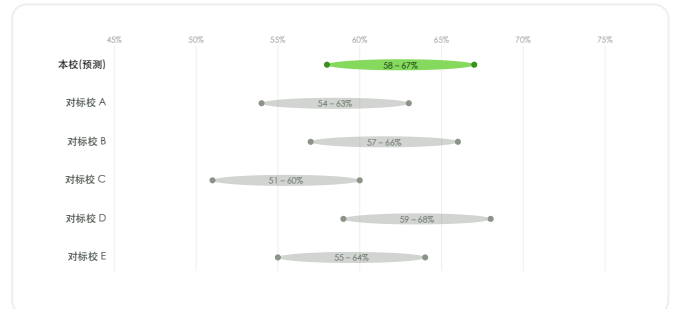
以起点为锚，按 19 级真实增长协方差外推；阴影为 95% 预测区间。预计三年提升约 +71 量表分。

### 各线预测上线人数



灰=按起点静态估计；绿=LGCM 动态预测。本科线预计可增 +91 人。

### 同侪对标定位 (本科上线率)



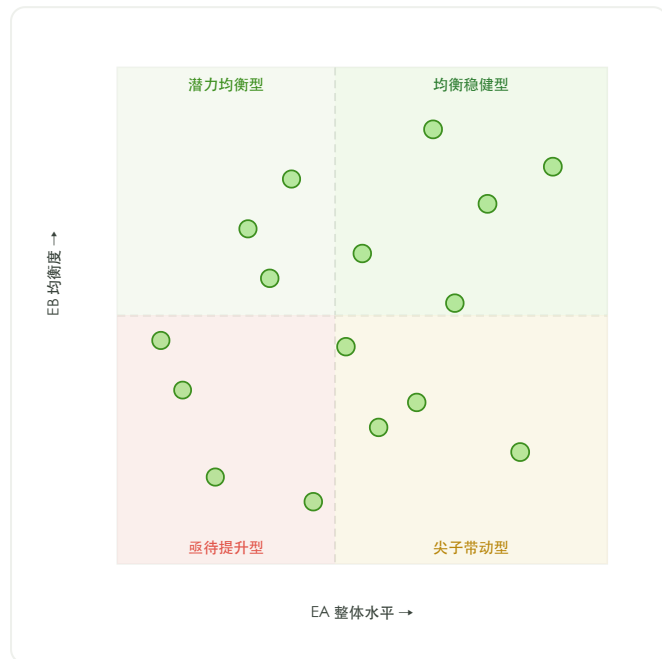
与 5 所起点相近学校的出口区间对比；本校（绿）预测处于对标群中上游。

预测校验：模型在 18/19/20 三届回溯验证，上线率预测误差 ±2% 以内，可提供完整验证数据。本预测为起点基线，每次新测评汇入后自动更新。

报告二 · 阅读对象：级部主任 / 教研组长

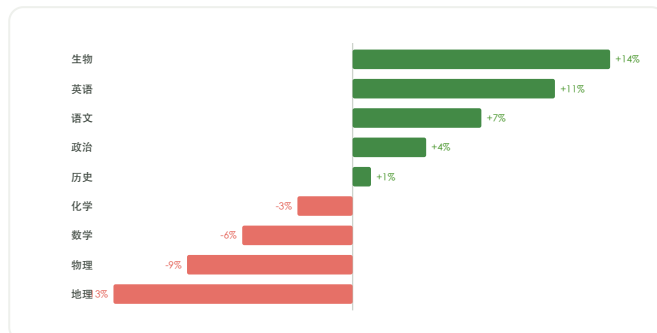
## 学科教学质量诊断与班级分层地图

### E 值矩阵四象限 (16 个班)



EA=整体水平、EB=均衡度；气泡=班级规模。自动分型：尖子带动/均衡稳健/潜力均衡/亟待提升。

### 学科贡献率 (对总分排名)

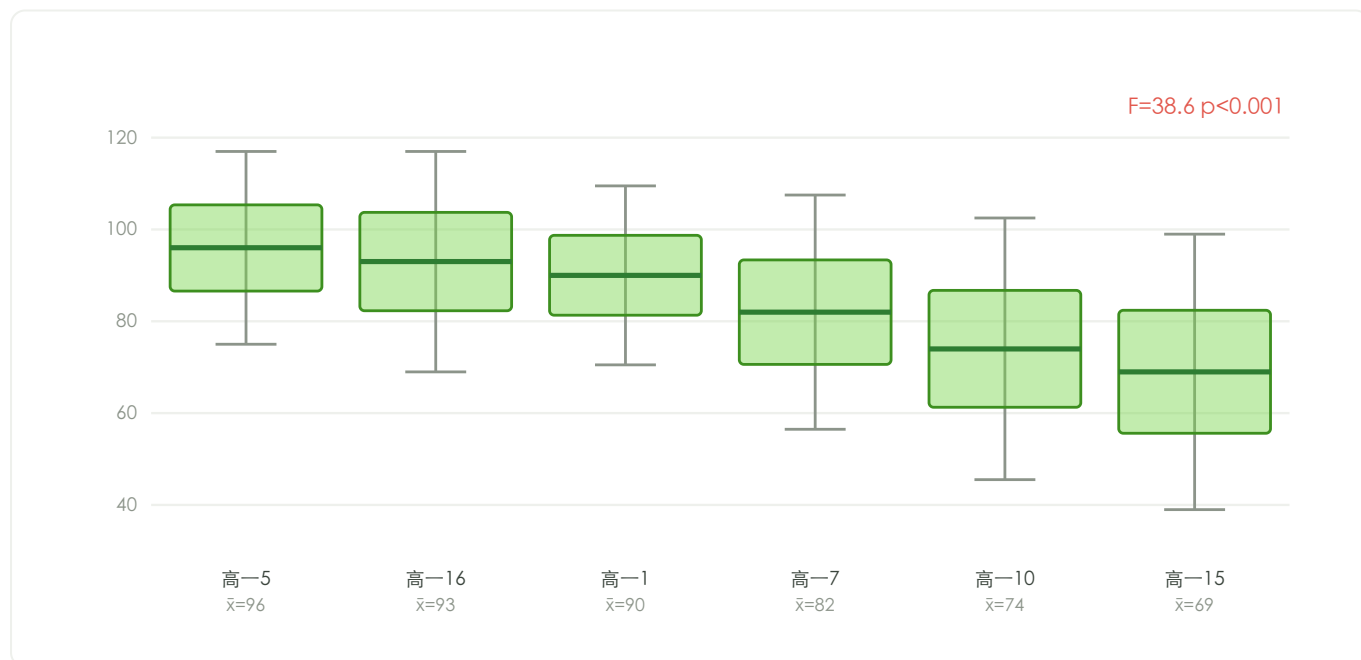


正=拉升项，负=拖累项。生物/英语贡献最大，地理/物理拖累最重。

### 班级分型行动

- 均衡稳健型：放宽自主申请权限
- 尖子带动型：补齐尾部、防两极分化
- 亟待提升型：强化基础、暂缓离堂

### 班级间差异检验 (数学 · 单因素 ANOVA)

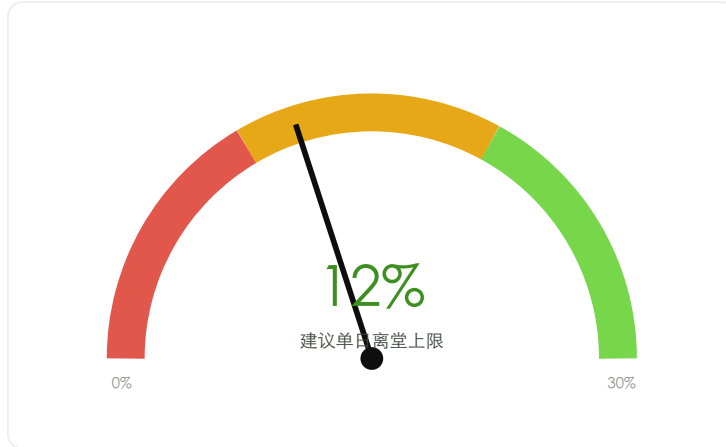


箱体=四分位，横线=均值。班级间差异极显著 ( $F=38.6, p<0.001$ )，存在系统性教学质量差异，需教研干预。

报告三 · 阅读对象：教学处 / 教务主任

## "学力通行证"制度落地参数建议

### 初始放行比例建议



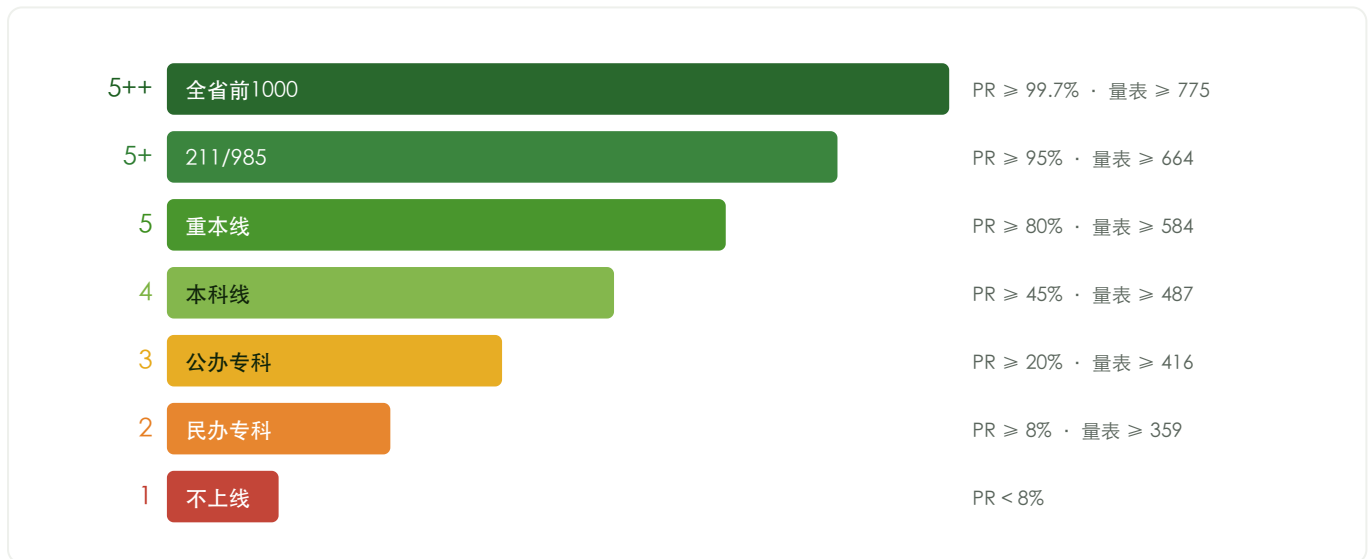
依达到 5 级（重本线）及以上占比（约 23%）建议初期单日离堂上限 12%，运行 4 周后据达标数据上调。

### 重点监控学科清单

学科	得分率	常模	差距	预警
地理	45%	53%	-8	红色预警
物理	49%	53%	-4	橙色关注
化学	52%	54%	-2	黄色留意
数学	58%	52%	+6	正常

红色预警学科：即便总分达标，系统强制推送基础巩固、暂缓自主申请。

### 分层达标线 · 能力等级对照（与个人报告 1 - 5++ 一致）



达标判定不仅看总分：须各核心知识点掌握概率达阈值（CDM），方可认证对应层级、开放相应自主权限。

**动态达标线（自适应）：**引入个人增长斜率后，达标线可设为“本次能力值须高于本人前 3 次均值”，防止“刷分”、鼓励真实增长。

报告四 · 阅读对象：班主任 / 年级德育主任

## 学生个人能力画像与分层干预名单

### 临界生预警名单 (XGBoost 上线概率)

学号	班级	总分	上线概率	缺口%	诊断与建议
S1042	高一5	612	46%	8%	本科线临界 · 偏科物理
S0317	高一1	598	41%	12%	本科线临界 · 英语待提
S0926	高一7	605	44%	10%	本科线临界 · 数学波动
S1188	高一12	619	49%	6%	特控线临界 · 稳态偏弱
S0455	高一9	589	38%	15%	本科线临界 · 文综薄弱
S0733	高一3	601	43%	11%	本科线临界 · 化学失分

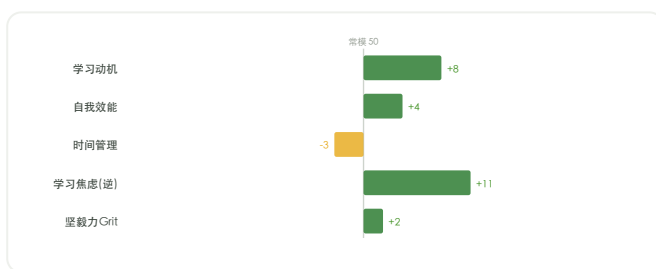
"距上线仅差 5-15% 概率"的临界生，是提分性价比最高的群体；建议班主任一对一跟进。

### 高潜力偏科生名单

学号	班	总分	跛腿科	该科
S0210	高一16	658	地理	38
S0588	高一2	641	语文	44
S0991	高一5	672	政治	41
S0124	高一9	633	物理	46
S0846	高一12	649	英语	49

总分优秀但单科仅 1-2 级——补齐短板即可冲击更高等级。

### 学习品质指标 (vs 常模 50)



时间管理略低于常模，是群体性可干预点；学习动机良好。

### 分层干预策略 (按能力等级 1-5++)

#### 1-2 级 · 未上本科

夯实基础 · 暂缓离堂

每周基础巩固包 + 一对一答疑，达标前不开放自主申请。

#### 3 级 · 专科段

稳基提速 · 限额申请

知识点补缺 + 限额离堂；红色预警学科强制基础任务。

#### 4-5 级 · 本科-重本

适度放权 · 变式拓展

开放常规自主申请，推送相似变式题，向更高等级冲刺。

#### 5+ 级及以上 ·

211/985+

充分自主 · 探究项目

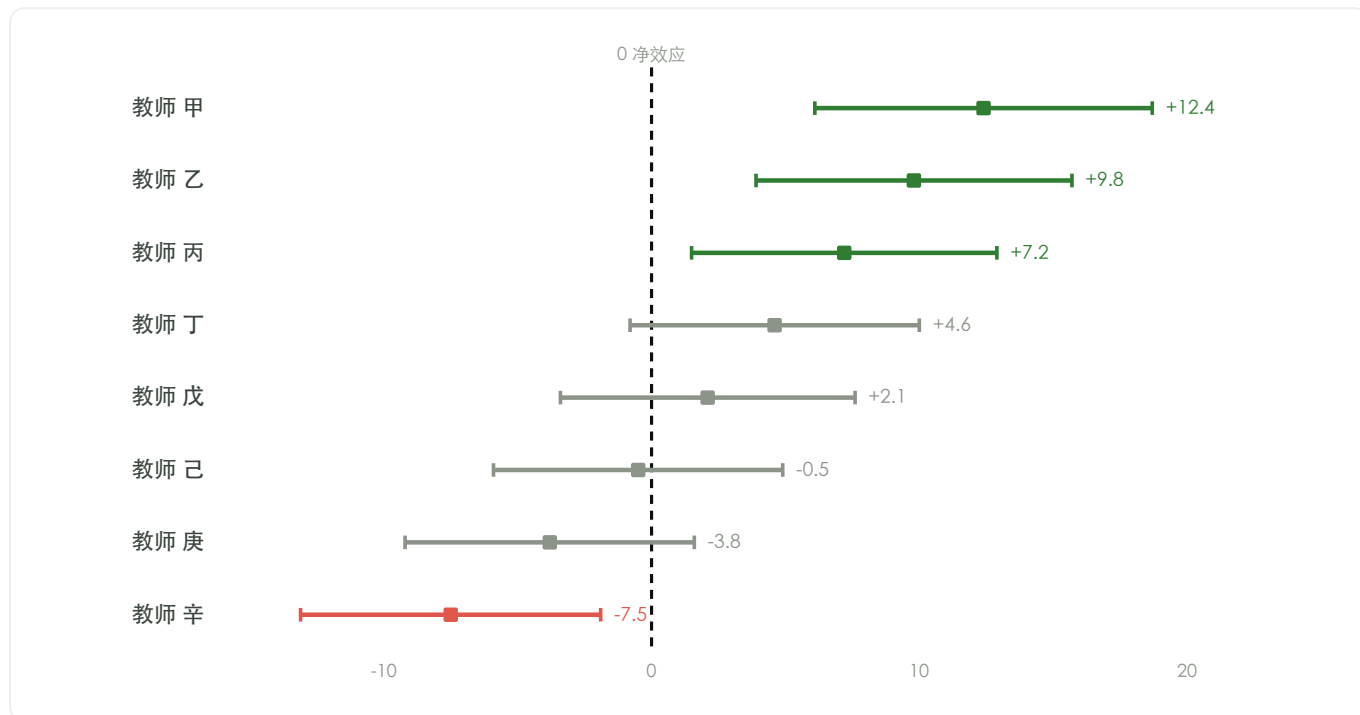
高自主权限 + 项目式学习/竞赛拓展，防止「吃不饱」。

报告五 · 阅读对象：教学处（教师评价存档）

## 教师教学效能基线（HLM）

多层线性模型剥离生源差异，计算每位教师的“教学净效应”，作为增值评价基线封存，为评优评先提供公平依据——只与自身比较，不参与跨校排名。

### 教师个人增值净效应（点估计 + 95% 置信区间）



方块=净效应点估计，横线=95%CI；CI 不跨 0 即统计显著（绿=显著为正，红=显著为负，灰=不显著）。

### 班级起点—目标匹配表

班级	起点均分	预期期末目标	预期增值	基线分组
高一1	90	108	+18	A 基线
高一5	96	112	+16	A 基线
高一7	82	101	+19	B 基线
高一10	74	95	+21	B 基线
高一15	69	92	+23	C 基线
高一16	93	110	+17	A 基线

**基线封存协议：**起点数据一次性封存，仅用于“自我增值比较”，运行一年后出具增值证明报告（含效应量与显著性检验），实证验证制度效果。

## 第四部分 · 学生个性化 IRT 报告

## 从「考了多少分」到「五个本质问题」

传统成绩单回答“你考了多少分”；IRT 个性化报告回答五个更本质的问题：

## 1. 你学会了什么？

知识点层面的掌握概率（CDM）

## 2. 你的能力在什么水平？

1-5++ 能力等级定位（IRT  $\theta$  + 全省常模）

## 3. 你在群体中处于什么位置？

与全省常模的百分位对比

## 4. 你的优势短板在哪里？

九学科及细分能力维度雷达图

## 5. 接下来该做什么？

基于诊断的个性化学习建议

## 报告内容（六大模块）

模块	呈现内容	技术依据
YANZ 能力等级总览	各学科 1-5++ 等级 + 趋势箭头 + 常模百分位	IRT $\theta$ + 常模参照
学科能力结构雷达图	核心素养维度轮廓；与同层级常模叠图	CDM/DINA
知识点掌握概率矩阵	当周/当月知识点掌握热力图；已/待达标标识	双向细目表 + 认知诊断
个人成长轨迹	历次能力值折线 + 预测区间 + 同起点平均对比	IRT 垂直等值 + LGCM
错题本与变式推荐	历次错题（原题/答案/知识点）+ 3-5 题变式	协同过滤
学习建议与行动计划	本周优先知识点 + 选科/升学 + PBL 选题	诊断 + 预测模型

## 报告获取与隐私保护

### 获取方式

- 每次完成达标测评后，报告10分钟内自动更新
- 形式：线上交互式报告 + PDF 导出
- 历史数据永久留存，学生可随时回溯完整成长轨迹

### 隐私与权限

- 个体报告仅学生本人及法定监护人可查看
- 科任教师仅见班级群体统计，无法查看个体详情
- 班主任仅见本班"需要关注的学生"名单及简要说明

**合规承诺：**报告以能力层级（5++ ~ 1）呈现，**不排名、不公布**，完全符合教育部「考试成绩不排名、不公布」的政策要求。本套模板的学生级母版已落地（简/繁/英三语 15 - 17 页），可与本校级报告配套交付。

### 学生报告与校级报告的关系

#### 学生个性化报告

每人一份，回答“我学会了什么、下一步做什么”，驱动学生自主学习与通行证申请。



#### 校级五份报告

群体统计聚合，回答“学校如何决策、教学如何改进、制度如何落地”。

## 第五部分 · 服务与落地

## ■ 学校合作 · 三个服务层级

所有层级均不收取任何费用；永久免费依托于百万量级测评数据库——数据脱敏后用于常模建立与学术研究。

## 层级一 · 基础服务

永久免费 · 无需提供数据

YANZ 平台六大功能模块全部使用权：每周知识点达标测评、在线申请与审核、项目式学习提交与评价、教师工作台、个人能力画像报告、学校/区域分析报告。完成起点测试即可开通。

## 层级二 · 数据增值服务

免费 · 需授权脱敏数据用于常模研究

在层级一基础上额外获得：五份校级报告、增值证明报告（运行一年后）、校长专属数据驾驶舱、与全省常模的年度对比报告、参与学术研究并署名。

## 层级三 · 深度合作

免费 · 需连续 3 次以上测评数据

在层级二基础上额外获得：专属常模定制、教师个人增值档案（每年更新）、学生三年完整成长轨迹报告、学校教学管理改进建议书、制度效果评估报告、参与课题研究并获成果署名权。

**重要说明：**学校提供的测评数据越多、质量越高，专属分析就越精细。学校有权选择数据共享范围，不共享数据不影响核心测评功能的使用；数据仅用于平台内部算法迭代，不向任何第三方提供。

## 高一 vs 高三 · 差异化服务

同一套引擎，按合作时间窗口提供不同侧重的服务组合。

### 高一开学（长周期）

报告侧重：报告一（升学潜力预测）+ 选科建议

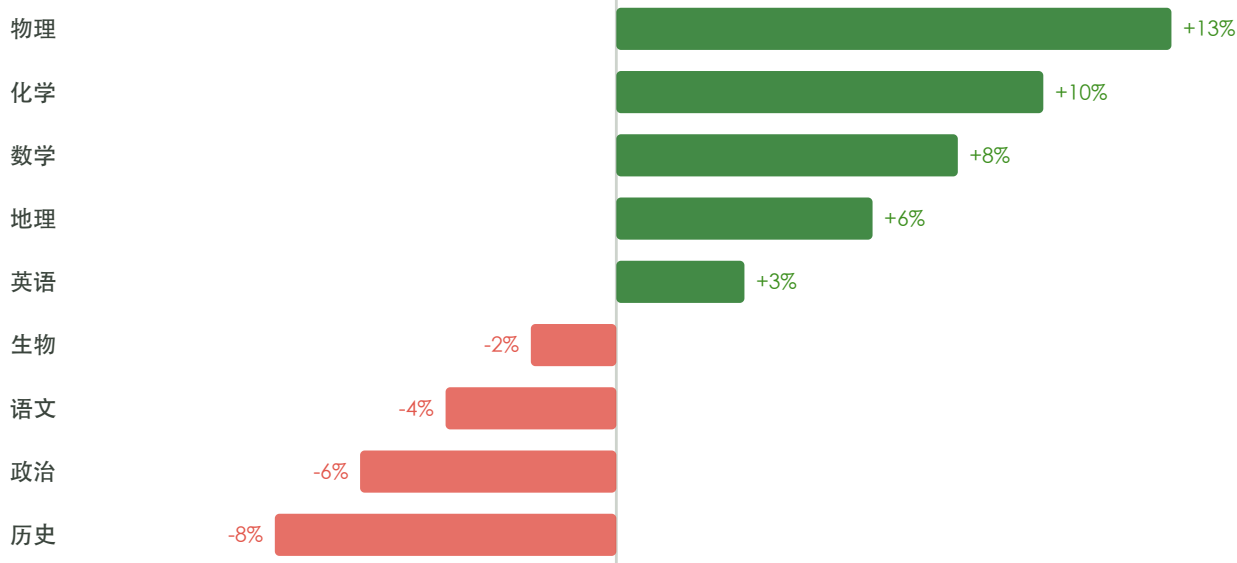
额外提供：基于 Q 值分析的 3+1+2 选科组合模拟，为高一下学期选科走班提供数据支持。

### 高三开学（短周期冲刺）

报告侧重：报告二（薄弱学科诊断）+ 报告四（临界生名单）

额外提供：各学科“提分空间排序”——告诉学校“有限时间内抓哪科性价比最高”。

### 各学科“提分空间排序”示意（高三场景）



正向越大=单位投入提分潜力越高。该排序由当前得分率距常模的差距 × 区分度 × 可训练性综合给出。

选科模拟（高一）：结合 Q 值（达 A/达 AB 率）与各组合历史出口，模拟 12 种 3+1+2 组合的上线期望，帮助学校与学生科学选科、避免盲目跟风。

## ■ 交付标准与售后 · 完整服务闭环

我们给学校的不只是报告文件，而是一个看得懂、用得上、可追踪的服务闭环。

1

### 五份校级报告

PDF + 线上驾驶舱；含全部可视化图表及文字解读

2

### 全员个性化 IRT 报告

线上 + PDF 导出；每人一份，含能力定位/知识诊断/学习建议

3

### 线下报告解读会（1 次 · 半天）

数据分析师到校，面向校长/教学处/级部/教研/备课组逐页解读

4

### 校长专属数据驾驶舱

按学科/班级/时间筛选，动态更新，多维交叉分析

5

### 「基线数据封存」协议

起点数据封存，仅用于自我增值比较，不参与跨校排名

6

### 增值证明报告（1 年 / 3 年后）

实际增长曲线 vs LGCM 预测的统计对比（效应量+显著性）

7

### 持续数据更新

后续每次测评数据汇入后，报告自动更新，实现动态追踪

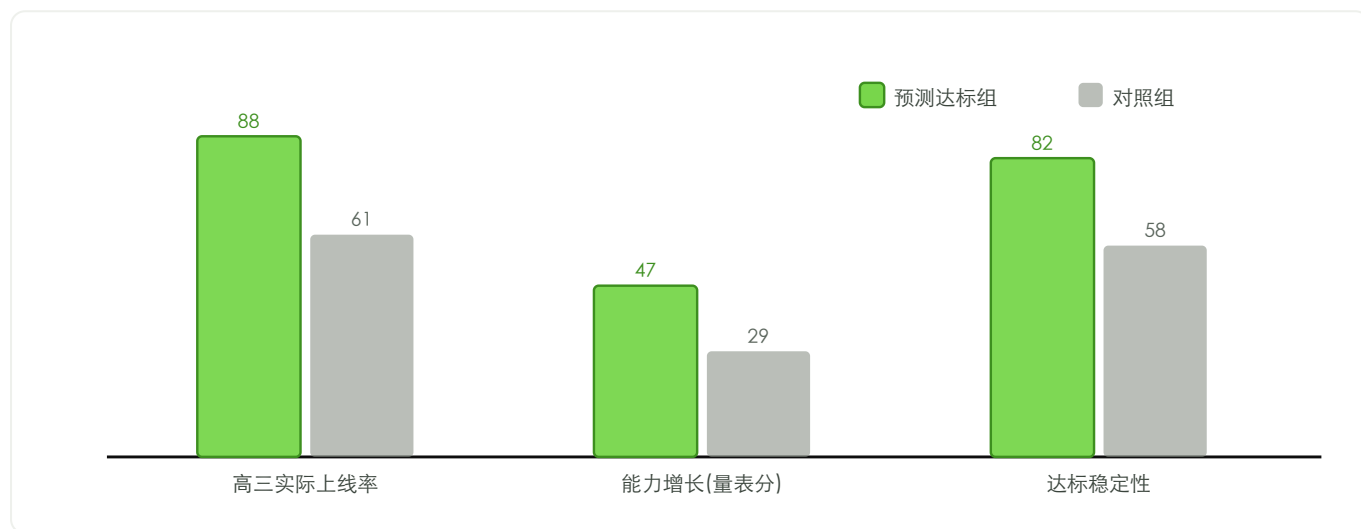
**动态追踪是核心价值：**报告不是一次性的“体检单”，而是随每次测评自动刷新的“实时仪表盘”，让学校的每一步教学决策都有最新数据支撑。

## ■ 试点设计 · 反事实推断

利用 19 级 8 所核心学校（追踪 4 次）的历史数据，可做一个极具说服力的“反事实推断”，用历史数据模拟验证制度的预测效率与干预价值。

**设计思路：**假设在高三上学期（2021 年 12 月）实行了「学力通行证」制度，那么根据高二下学期（2021 年 6 月）的能力诊断，哪些学生应获批准？再看他们高三上学期的实际表现，是否显著优于“未被预测达标”的学生——这相当于用历史数据模拟验证了制度的预测效率。

### 反事实对比：预测达标组 vs 对照组（历史回溯）



绿=被模型预测应达标的学生组；灰=对照组。三项后续指标上，预测达标组均显著领先，验证了「预测—干预」链路有效。

### 这说明什么

- 模型的“达标预测”真实可信，不是事后解释
- 获得自主权的学生确有更好发展，制度有正向干预价值
- 为新学校落地提供最有力的实证论据

### 预测精度校验

<b>±2%</b> 上线率预测 误差	<b>3 届</b> 回溯验证	<b>8 校</b> 核心追踪样 本	<b>4 次</b> 纵向追踪
---------------------------	--------------------	--------------------------	--------------------

## 第六部分 · 附录

## ■ 常见问题解答 (FAQ)

### Q1 学校必须使用 YANZ 提供的测试卷吗？

不一定。若已有近期标准化联考/统考数据（含个体逐题得分及双向细目表），可直接导入平台。但强烈建议至少参加一次 YANZ 统一测评，以便与全省常模对标。

### Q2 起点测试需要多长时间？

约 2-3 小时，覆盖语数英 + 学生所选其他科目，可分 2-3 天完成。

### Q3 报告真的能预测高考成绩吗？

模型基于同省份、同层次学校过去三届（18/19/20 级）真实联考数据训练，经回溯验证预测误差在  $\pm 2\%$  以内，可展示完整验证数据。

### Q4 学校中途退出，数据如何处理？

可随时申请退出。退出后平台停止服务；已脱敏纳入常模库的数据无法单独移除（常模为群体统计量，不涉及个体识别）；未脱敏原始数据 30 天内删除。

### Q5 个性化 IRT 报告会排名吗？

不排名、不公布。报告以能力层级（5++ ~ 1）呈现，完全符合教育部「考试结果不排名、不公布」的政策要求。

### Q6 教师能看到具体学生的报告吗？

科任教师仅见班级群体统计（如「本班 5 级及以上占比 28%」），无法查看个体报告；班主任可见本班「需要关注的学生」名单及简要说明；个体报告仅学生本人及监护人可见。

## 方法与术语表 · Glossary

本报告所有结论均有成熟的心理计量与统计学方法支撑，方法可复现、可解释。

### IRT 项目反应理论

用潜在能力  $\theta$  解释作答；同分不同能力可被区分，跨卷可比。

### MIRT 多维项目反应

估计力学/电磁/代数/几何等细分维度的潜在能力。

### IRT 垂直等值

用锚题把不同年级试卷放到同一量尺，使历次成绩可比。

### HLM 多层线性模型

建模嵌套结构，剥离生源得到教师净效应。

### E 值矩阵 $EA \times EB$

整体效能  $\times$  均衡效能的四象限班级分型。

### 异常检测 Isolation Forest

识别刷分/虚假达标的异常作答模式。

### CDM / DINA · GDINA

认知诊断模型：由 Q 矩阵把答题模式反演为知识点掌握概率。

### GRM 等级反应模型

处理主观题多级评分，刻画高阶/低阶思维结构。

### LGCM 潜在增长曲线

估计个人起点与增长斜率，预测发展轨迹。

### XGBoost 梯度提升树

非线性集成模型，预测上线概率并给特征重要度。

### 矩阵分解 + 协同过滤

分解学生  $\times$  试题矩阵识别偏好，推荐变式题。

### Cronbach $\alpha$ / 信度

测验内部一致性指标， $\alpha$  越高越可靠。

"算法是旧的，结论是新的。" 所有方法均为国际成熟范式；新的是把贵校起点数据代入这些验证过的框架后，算出的专属预测与决策建议。

yanz



# 让每一次测评， 成为向上的跑道

数据驱动决策 · 算法赋能育人

脱颖而出 · Stand Out · Reach Your Zenith