

# PRD - 掌握度策略 - V1.0

## 版本管理

版本号	日期	变更人	变更类型	变更详情
V1.0	2025-04-11	钱晋菲	新建	新建文档
V1.0.1	2025-05-12	钱晋菲	更新	<div>1. 外化掌握度<div>a. 更新计算方式，由星星改成百分比展示</div><div>b. 掌握度更新时机，中途退出学习不更新</div><div>c. 补充对特殊情况的处理</div></div> <div>2. 目标掌握度<div>a. 更新各维度系数</div><div>b. 换算成百分比，按0.05向上取整</div></div>
V1.0.2	2025-05-14	钱晋菲	更新	<div>1. 知识点掌握度计算方式，增加了错题重答</div>

## 关联需求

关联需求名称	所属 PM	需求进度	文档链接
AI课中 V1.0			<a href="#">📄 PRD - AI课中 - V1.0 (WIP)</a>

# 一、背景和目标

## 1. 需求背景

在自适应学习系统中，掌握度是非常重要的功能模块，是未来各种推荐策略的基础输入。

1. 精准评估学习状况：掌握度能够精确反映学生对各个知识点的理解和掌握程度。让学生和教师都能清楚了解学生在学习过程中的优势和不足。
2. 优化学习路径：巩固练习和拓展练习都需要依据掌握度来为学生量身定制学习路径，推送适合该学生的学习内容和练习题目。

## 2. 项目收益

- 1. 提升产品体验，让用户感受到系统的个性化和智能性
- 2. 提高学习效率，避免重复学习已经掌握的内容
- 3. 是巩固练习和拓展练习等推荐策略的前置输入

## 3. 覆盖用户

使用新AI课学习的全部用户

# 二、概念厘清

## 1. 掌握度的定义

掌握度是衡量学生对特定知识点理解与运用程度的关键指标。在新AI课系统中定义了多种类型的掌握度：

- 1. **学科能力（用户分层）**：量化学生在某个学科的能力水平，用于学生分层&学习路径规划。仅基于考试数据更新。
- 2. **知识点掌握度**：量化学生对某个知识点的熟练程度，用于动态调整学习内容和路径，在每次答题后更新。通常和用户的答题表现（正确率、知识点难度、错误模式）、行为数据（答题时长、尝试次数、学习间隔时长）等因素相关。
- 3. **外化掌握度**：用于向学生展示对于单节课程的掌握程度，牵引用户学习。在学生完成AI课、巩固练习、作业、测试等学习任务后更新。

由于知识点掌握度是一个动态变化的指标且和多种因素相关不便于用户理解，因此在向用户展示掌握度时采用简化的和答题正确率&题目难度更相关的计算方式，降低用户理解成本。

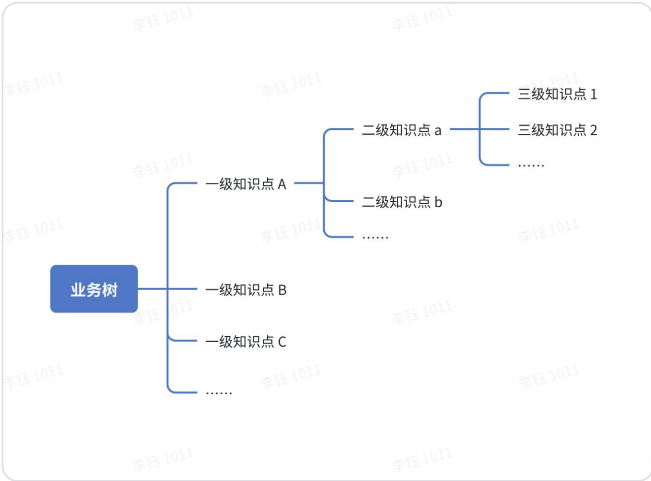
- 4. **目标掌握度**：在目标设定阶段，系统依据学生设定的目标，推导出学生在每节课应达到的外化掌握度。

	学科能力（用户分层）	知识点掌握度	外化掌握度（星星）	目标掌握度
评估对象	学生在某个学科的能力水平	学生对某个知识点的掌握程度	学生对单节课程的掌握程度	根据目标推导出学生对单节课程应该达到的掌握程度
目的			激励手段，牵引用户学习	

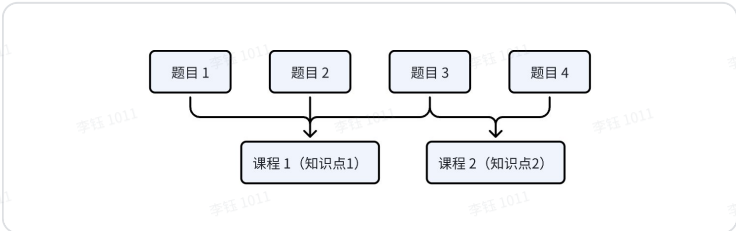
	学生分层&学习路径规划	动态调整学习内容和路径		
更新时间	前期每次考试后更新 后续根据知识点掌握度计算	每次答题后更新	完成AI课、巩固练习、作业、测试等学习任务后更新	用户设置/修改目标后更新
计算方式	冷启动阶段基于策略计算 后续模型直接输出		基于策略计算	

2. 内容结构

掌握度结算仅考虑业务树节点



- 一节课等于一个知识点：业务树的末级节点 = 单节课
- 题目和课关联：1道题目可关联多节课，不存在没有关联课的题目。



三、策略方案

1. 学科能力（用户分层）

冷启动阶段计算思路：

- a. 锚定学校在全国范围内的大致水平
- b. 校准不同届的生源质量
- c. 基于学生各科成绩在学校的排名&学校在全国范围的水平进行分层

1.1 数据准备

--	--	--	--	--

数据类型	字段	字段说明
学生信息	学生ID	和AI课系统一一对应的学生ID
	各学科年级排名百分比 1. 学生单科排名 2. 年级人数	如果拿到的是学生排名信息，需要转换为排名百分比。 示例：年级50名 / 年级共200名学生 → 排名百分比=25%。 。代表这名学生在年级所有学生中处于前 25% 的位置。
学校信息 - 仅高中收集	学校近3年的高考本科录取率	1. 高考年份 2. 对应年份的本科录取率
	学生中考成绩全省/市排名 或 一分一段表	1. 对应高考年份的学生的中考排名 2. 本届学生的中考排名

注：每届新生都需要收集对应信息，重新计算下文中的各类系数

## 1.2 计算方式

### 1.2.1 高中

#### 1. 锚定学校水平

根据学校历届高考的平均本科录取率，对应到不同的校准系数a

本科录取率区间	[90,100]	[80,90)	[60,80)	[40,60)	[20,40)	[0,20)
校准系数a	1	1.2	1.3	1.5	2	3

如果学校未提供录取率，可简化为按学校类型设置系数

学校类型	重点中学	普通中学	较差中学
校准系数a	1	1.3	3

#### 2. 校准生源波动

根据学校中考排名计算校准系数b，如果学校没有提供排名数据，则校准系数默认为1

情况一：学校近3年高考本科录取率的相对变化小于10%

- 校准系数 $b = \sqrt{\text{（历年排名中位数的倒数的平均 / 今年排名中位数的倒数）}}$

示例：

	今年	2024	2023	2022
排名中位数	a	b	b'	b''
	150	210	189	203

$$\text{校准系数 } b = \sqrt{\left(\frac{\frac{1}{b} + \frac{1}{b'} + \frac{1}{b''}}{3}\right) \div \frac{1}{a}} = \sqrt{\left(\frac{\frac{1}{210} + \frac{1}{189} + \frac{1}{203}}{3}\right) \div \frac{1}{150}} \approx 0.82$$

**情况二：**学校近3年高考本科录取率的相对变化大于10%，仅采用一年的分数进行校准（人工指定）

### 3. 确定用户分层

- 学科能力排名 = 用户单科排名百分比 \* 校准系数a \* 校准系数b，取值[0,1]，向上取整

用户分层	学科能力分层区间
S+	[0 , 5]
S	(5 , 10]
A	(10 , 50]
B	(50 , 80]
C	(80 , 100]

注：本科录取率在40%以下的学校 或 较差中学，用户分层中没有S+，落入S+区间的学生分层记为 S

### 示例：

一个高考本科录取率为57%的学校，今年入学全市排名中位数为2187，历届入学排名中位数为2000、1897、1878。

数学在年级排名百分比为38%的学生对应的学科能力为：

$$38 \times 1.3 \times \sqrt{\frac{\left(\frac{1}{2000} + \frac{1}{1897} + \frac{1}{1878}\right) / 3}{1/2187}} \approx 53$$

该学生的数学能力分层为B

1.2.2 初中

1. 锚定学校水平

根据学校类型，对应到不同的校准系数a

学校类型	重点中学	普通中学	较差中学
校准系数a	1	1.3	3

2. 确定用户分层

- 学科能力 = 用户单科排名百分比 \* 校准系数a，取值[0,1]，向上取整

用户分层	学科能力分层区间
S+	[0 , 5]
S	(5 , 10]
A	(10 , 50]
B	(50 , 80]
C	(80 , 100]

注：较差中学的用户分层中没有S+，落入S+区间的学生分层记为 S

示例：

一个较差中学的学生，数学在年级排名百分比为1%，对应的学科能力为： $1 * 3 = 3$

由于较差中学不设置S+分层，该学生的数学能力分层为S

1.3 更新用户分层

每次年级统一考试后，根据学生本次考试的年级排名百分比，更新用户所处分层。

统考成绩数据怎么拿到？

2. 知识点掌握度

2.1 初始掌握度

V1.0不考虑知识点之间的迁移性，将所有知识点视为孤立知识点

### 1. 学科初始值 0.3

V1.0每个学科设置统一的初始掌握度，后续需要支持不同学科设置不同的掌握度初始值

### 2. 根据用户分层设置不同的初始掌握系数，每个用户在每个知识点上的初始掌握度 = 学科初始值 \* 分层掌握系数

用户分层	S+	S	A	B	C
分层掌握系数	1.2	1.0	0.8	0.5	0.4
初始掌握度	0.36	0.3	0.24	0.15	0.12

## 2.2 单次答题的掌握度增量

### 2.2.1 数据准备

#### • 题目难度分级

此处的难度等级为知识点内的题目之间的相对难度，不会出现简单知识点下全部都是L1，复杂知识点下全部都是L5的情况。

难度档位	L1	L2	L3	L4	L5
难度系数	0.2	0.4	0.6	0.8	1.0

### 2.2.2 计算方式

#### • 整体逻辑

通过判断答题结果（正确或错误），结合题目本身的难度系数及用户当前的知识点掌握度，来确定掌握度的增减数值。这意味着不同水平的学生对不同难度题目的作答情况，会以不同幅度影响知识点的掌握度。

#### • 引入答错补偿系数

为避免答错了高难度题而被过度惩罚，在作答错误场景需要引入补偿系数

题目难度档位	L1	L2	L3	L4	L5
答错补偿系数	1.8	0.8	0.5	0.3	0.2

#### • 设置默认增量

- 作答正确：默认掌握度增量为 0.2

o 作答错误：默认掌握度增量为 -0.1

## • 首次答题计算方式

- a. 作答正确：掌握度增量 =  $0.2 * (\text{题目难度系数} - \text{用户当前掌握度})$
- b. 部分正确：掌握度增量 =  $0.2 * (\text{题目难度系数} - \text{用户当前掌握度}) * 0.5$
- c. 作答错误：掌握度增量 =  $-0.1 * (\text{题目难度系数} - \text{用户当前掌握度}) * \text{答错补偿系数}$

### o 示例

- i. 用户当前掌握度为0.24，做对了一道难度为L3的题目，单题掌握度增量的计算过程如下：
  - 掌握度增量 =  $0.2 * (0.6 - 0.24) = 0.07$
- ii. 用户当前掌握度为0.24，做错了一道难度为L5的题目，单题掌握度增量的计算过程如下：
  - 掌握度增量 =  $-0.1 * (1 - 0.24) * 0.2 = -0.02$
- iii. 用户当前掌握度为0.6，做错了一道难度为L1的题目，单题掌握度增量的计算过程如下：
  - 掌握度增量 =  $-0.1 * |0.2 - 0.6| * 1.8 = -0.072$

## • 错题重练计算方式

- a. 引入和尝试次数相关的衰减系数，衰减系数 =  $0.5^x$  (x为重练次数)

示例：第一次重练，系数为0.5；第二次重练，系数为0.25；第三次重练，系数为0.125；

- b. 错题重练增量 = 首次答题增量 \* 衰减系数

示例：第一次重练

- i. 作答正确：掌握度增量 =  $0.2 * (\text{题目难度系数} - \text{用户当前掌握度}) * 0.5$
- ii. 部分正确：掌握度增量 =  $0.2 * (\text{题目难度系数} - \text{用户当前掌握度}) * 0.5 * 0.5$
- iii. 作答错误：掌握度增量 =  $-0.1 * (\text{题目难度系数} - \text{用户当前掌握度}) * \text{答错补偿系数} * 0.5$

## • 特殊情况

- a. 作答正确/部分正确：如果题目难度系数-用户当前掌握度  $\leq 0.01$ ，按0.01计算
- b. 作答错误：
  - i. 如果题目难度系数-用户当前掌握度在  $[-0.01, 0.01]$ ，按0.01计算
  - ii. 如果题目难度系数-用户当前掌握度在  $[-1, -0.01)$ ，取绝对值计算
- c. 自评的主观题，在掌握度计算时需要降低权重。
  - i. 如果作答时间小于5秒，答题结果不进行结算
  - ii. 如果作答时间大于等于5秒，计算增量时需要额外 \* 0.5

## 2.3 累计掌握度更新



**计算方式：**累计掌握度=当前掌握度 + 单题掌握度增量，结果限制在[0,1]区间

**限制：**针对同一用户在同一道题多次作答的情况

a. 在同一节课里多次答同一道题

i. 之前已作答正确的题目，不再更新掌握度变化。

ii. 之前作答错误的题目：按“错题重答”的计算方式更新掌握度变化

b. 在不同的课程里作答同一道题

i. 当前课程中的第一次作答正常结算掌握度

ii. 后续作答按“在同一节课里多次答题”处理

**更新时机：**用户每次作答后更新

**更新范围：**用户作答题目关联的所有知识点。如果一道题关联了多个知识点，则多个知识点的掌握度都会更新。

### 3. 外化掌握度

**应用场景：**用于在报告页展示学生对于本节课的掌握情况

#### 3.1 更新时机

1. 用户学完本节课的「AI课」，中途退出不更新。

2. 用户完成「巩固练习」，中途退出不更新。

3. 用户完成「作业任务」，中途退出不更新。

4. 用户完成「测验任务」，中途退出不更新。

#### 3.2 计算方式

1. 用户在本节课的知识点掌握度，换算成百分比展示。

2. 外化掌握度只增不降，当用户的知识点掌握度数值下降时，外化掌握度保持不变。

##### 3.2.1 最少答题量限制

为避免答题量过少导致的偶然性偏差，提升统计可靠性。需要对最少答题量进行限制。

答题量限制：5题

用户距离上次更新掌握度，需累计答题5道以上，才更新外化掌握度。

举例：

- 1. 用户首次学习课程，在AI课中答了3道题后退出，此时掌握度依然为0颗星。
- 2. 用户当前掌握度为2星，完成4道巩固练习后退出，掌握度不更新，需要累计作答5题才更新。

注：累计答题量为用户在本节课的答题量，不包括用户在别的课中做过本节课的题目。

### 3.2.2 特殊情况

- 1. AI课中有题目，但是题目数量不足5题，则用户完成一节AI课后，需要更新外化掌握度。
- 2. AI课中没有配置题目。用户学完后，外化掌握度统一按20%展示。

## 4. 目标掌握度

在用户设置/修改目标后，离线计算每节课对应的目标掌握度。

**目标设置：**高中阶段选择大学院校为目标，初中阶段选择年级排名为目标。

**整体逻辑：**目标掌握度和每节课的难度、考试频率、目标难度高度相关。单节课难度系数越低，考试频率越高，目标难度越高，要求的目标掌握度就越高。

### 4.1 数据准备

#### 1. 单课难度系数

	高	中	低
文科	0.7	0.8	1
理科	0.8	0.9	1

#### 2. 考试频率

	高	中	低
考频系数	1	0.9	0.8

#### 3. 目标难度系数

- 高中阶段

	C9	985 / 211	一本	二本	二本以下
目标难度系数	1.2	1.1	1	0.9	0.8

- 初中阶段

如果C端设置目标时选择的是年级排名的绝对值，数据收集时需要拿到学校的年级人数，通过年级人数和目标名次计算排名百分比，确定对应的难度系数。

	年级前5%	年级前10%	年级前20%	年级前30%	年级前50%
目标难度系数	1.2	1.1	1	0.9	0.8

## 4.2 计算方式

1. 每节课对应的目标掌握度 = 目标难度系数 \* 单课难度系数 \* 考试频率
2. 换算成百分比，按0.05向上取整

举例：

理科，课程难度高、考频中、用户目标一本

目标掌握度=0.8\*0.9\*1=0.72 ≈ 75%

## 5. 指标跟踪

用于上线后监测异常情况，进行策略优化

- **外化掌握度变化：**记录每个学生，每一次外化掌握度结算后的星级变化情况。是否会出现掌握度突然大幅下降或上升，超出了正常波动范围。
- **用户答题量：**记录每个用户在每次星级变化时的答题数量。

# 四、学校数据收集

## 1. 初始化数据

数据类型	是否必填	字段	目的
学生信息	是	学生姓名/ ID	对应AI课系统的学生ID
		学生单科年级排名	用于计算学生单科成绩的年级排名百分比
		年级人数	
学校信息 - 高中收集	2选1 必填	1. 高考年份 2. 对应年份的本科录取率	根据学校近3年的高考本科录取率 / 学校类型，判断学校在全国范围内的大致水平
		学校类型（重点/普通/较差）	

	2选1 非必填	1. 对应高考年份的学生的中考排名 2. 本届学生的中考排名	用于校准本届学生和历届学生的生源波动
		1. 对应高考年份的学生的一分一段表 2. 本届学生的一分一段表	
学校信息 - 初中收集	是	学校类型（重点/普通/较差）	根据学校类型，大致判断学校在全省/市范围内的水平

## 2. 使用中数据

学生在年级统一考试（期中、期末）中的单科排名，用于更新用户分层。

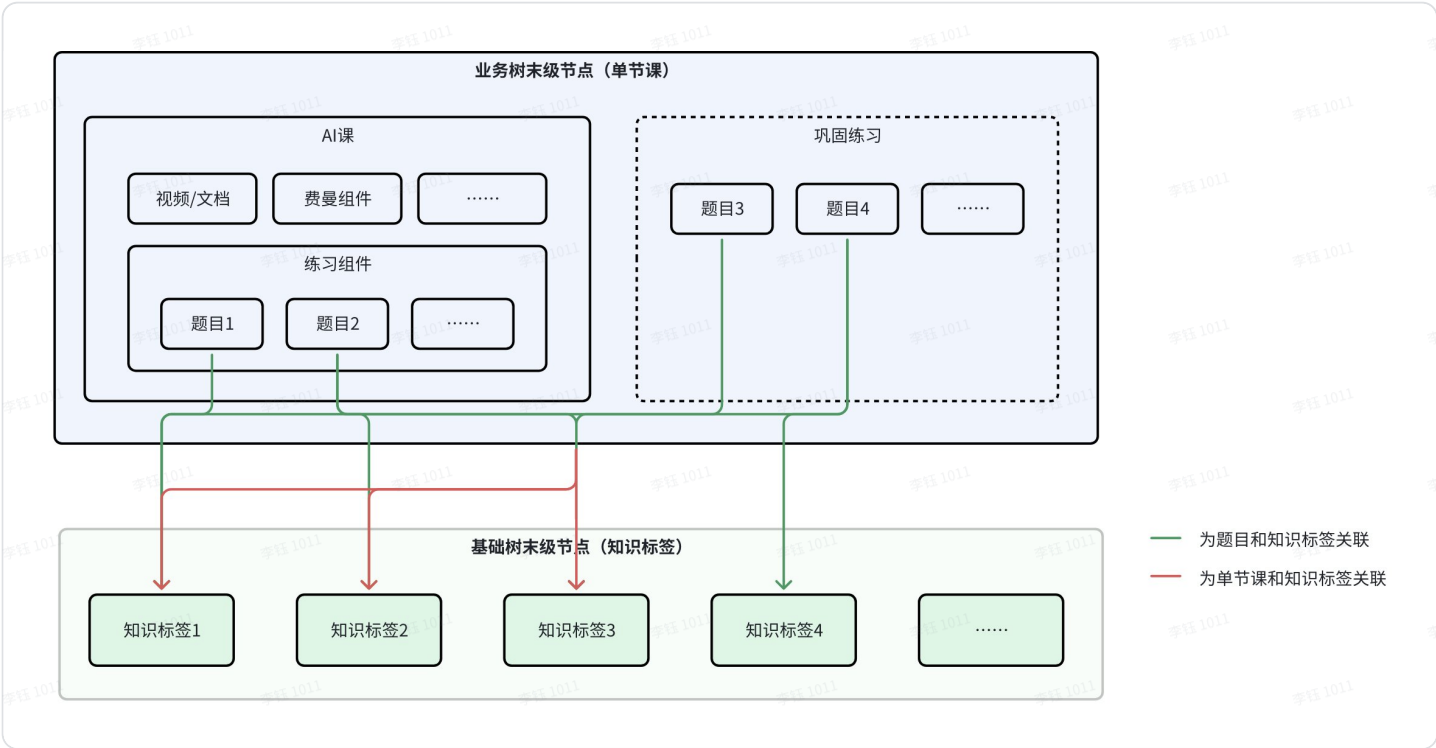
# 历史存档（不用看）

## 内容结构

基础树（知识点）	业务树（教材目录）
基础树的末级节点 = 知识点标签 	业务树的末级节点 = 单节课的名称 

- **基础树和业务树的关联：**业务树的末级节点（单节课）会关联基础树的末级节点（知识点），一节课可以关联多个知识点标签，不存在没有关联知识点标签的课。
- **基础树和题目的关联：**题目需要关联知识点标签，1道题目可关联多个知识点，不存在没有知识点标签的题目。

- 单节课对应的知识标签集合不等于题目对应的知识标签集合，但不存在单节课对应的知识点在巩固练习里没有包括的情况。
- AI课的内容（视频、文档、互动、费曼）不关联知识点标签，看内容不影响掌握度



## 外化掌握度

### 数据准备

- 课程（业务树末级节点）和知识点的对应关系

一节课可对应多个知识点，其中一个知识点为主知识点，其余知识点为次知识点。不同类型的知识点有不同的结算系数。

	主知识点	次知识点
结算系数	1	0.2

注：外化掌握度的结算关系为课程对应的知识点，不是课程中全部题目对应的知识点

### 计算方式

#### 1. 同步课 - 加权平均

外化掌握度 = （知识点1掌握度 \* 知识点1结算系数 + 知识点2掌握度 \* 知识点2结算系数+ ……） / （知识点1结算系数 + 知识点2结算系数+ ……）

#### 2. 复习课 - 算术平均

外化掌握度 = (知识点1掌握度 + 知识点2掌握度 + ……) / 知识点数量

按学生维度计算用户分层

数据类型	字段	字段说明
学生信息	学生ID	和AI课系统一一对应的学生ID
	年级排名百分比 <div>1. 学生排名</div> <div>2. 年级人数</div>	如果拿到的是学生排名信息，需要转换为排名百分比。 <div>示例：年级50名 / 年级共200名学生 → 排名百分比=25%。代表这名学生在年级所有学生中处于前 25% 的位置。</div>
学校的历届高考成绩（近3年）	年份	高考年份
	高考成绩对应不同院校梯队的达线率	从成绩分布角度，看处于某一分数段（对应能报考某类院校梯队）的考生在学校全部考生中的占比。 <div>示例：2023年，学校有2000名学生参加高考，其中100名学生的高考分数达到985 / 211分数线，则2023年的985 / 211达线率为 100/2000=5%</div>
学校的入学平均分 <div>用于校准不同届的生源质量</div>	年份	对应到历届高考年份
	历届入学平均分（近3年）	历届学生入学时中考成绩的平均分
	今年的入学平均分	当前届学生入学时中考成绩的平均分

情况一：有学校的历届高考成绩

1. 计算学校历届高考成绩对应不同院校梯队的平均达线率，向上取整

示例：

对应院校梯队	高考达线率-2024	高考达线率-2023	高考达线率-2022	平均达线率
985头部	a %	a' %	a'' %	(a + a' + a'') / 3
985 / 211	b %	b' %	b'' %	(b + b' + b'') / 3
一本	c %	c' %	c'' %	(c + c' + c'') / 3
二本	d %	d' %	d'' %	(d + d' + d'') / 3

2. 根据学校入学平均分计算校准系数

- a. 校准系数 =  $\sqrt{\text{今年入学平均分} / \text{历年平均分的平均}}$
- b. 如果学校没有提供入学平均分数据，则校准系数默认为1

示例：  
校准系数 =  $\sqrt{[a / (b + b' + b'') / 3]} = \sqrt{[620 / (650 + 655 + 647) / 3]} \approx 0.98$

	今年	2024	2023	2022
入学平均分	a	b	b'	b''
	620	650	655	647

3. 用平均达线率 \* 校准系数作为最终的分层阈值。根据学生在年级排名百分比，确定学生所属分层

- a. 如果院校梯队对应的达线率为0，则该学校没有对应分层的学生

示例：

- 年级共200名学生
- 学生A：年级50名 → 排名百分比=25% → 分层A
- 学生B：年级138名 → 排名百分比=69% → 分层B

用户分层	对应院校梯队	平均达线率	分层阈值	分层区间
S+	985头部	5%	$5 * 0.98$	$[0, 4.9]$
S	985 / 211	10%	$10 * 0.98$	$(4.9, 9.8]$
A	一本	50%	$50 * 0.98$	$(9.8, 49]$
B	二本	80%	$80 * 0.98$	$(49, 78.4]$
C	二本以下	-	-	$(78.4, 100]$

情况二：仅有学生在年级排名百分比

采用基础阈值，根据学校是否为重点中学，设置不同的排名系数。（系数可针对学校设置）

- 分层阈值 = 基础阈值 \* 排名系数
  - 重点中学：1.2
  - 普通中学：1.0

较差中学：0.7

较差中学冷启动时默认没有S+分层

示例：

用户分层	对应院校梯队	排名百分比基础阈值	排名系数	学生在校内排名区间
S+	985头部	1%	0.7	-
S	985 / 211	5%		[0 , 3.5]
A	一本	40%		(3.5 , 28]
B	二本	70%		(28 , 49]
C	二本以下	-		(49 , 100]