

# 学校社会资本与初中生教育期望： 基于 CEPS 数据的实证研究报告

社会科学方法论课程组

2026 年 1 月 24 日

## 摘要

本研究利用中国教育追踪调查（CEPS）初中基线数据（ $N \approx 9314$ ），旨在探究学校场域内的两种“人际关系资本”——同伴关系（Bonding Social Capital）与师生关系（Linking Social Capital）对初中生教育期望的影响。研究采用有序 Logit 模型（Ordered Logit Model）及聚类稳健标准误（Cluster-Robust SE），实证结果表明：（1）在控制认知能力与家庭背景后，同伴关系与师生关系均能显著提升学生的教育期望；（2）交互效应分析揭示了同伴关系的“马太效应”，即高家庭社会经济地位（SES）学生从中获益更多，而师生关系对农村学生呈现出微弱的“补偿效应”；（3）探索性机器学习分析进一步验证了各因素的重要性排序。本报告详细阐述了从变量构造、模型选择到结果解释的全过程，并讨论了研究的局限性。

**关键词：**教育期望；社会资本；有序 Logit；马太效应；CEPS

## 目录

<b>1</b>	<b>研究背景与问题提出</b>	<b>3</b>
<b>2</b>	<b>数据来源与样本筛选</b>	<b>3</b>
2.1	数据来源 . . . . .	3
2.2	样本清洗流程 . . . . .	3
<b>3</b>	<b>变量测量与构造</b>	<b>4</b>
3.1	因变量：教育期望 ( $Y$ ) . . . . .	4
3.2	自变量：学校社会资本 . . . . .	4
3.2.1	同伴关系指数 ( $Bonding\_Idx$ ) . . . . .	4
3.2.2	师生关系指数 ( $Linking\_Idx$ ) . . . . .	4
3.3	控制变量 . . . . .	5
<b>4</b>	<b>统计模型与方法</b>	<b>5</b>
4.1	模型设定 . . . . .	5
4.2	估计策略 . . . . .	5
<b>5</b>	<b>实证分析结果</b>	<b>5</b>
5.1	主效应分析 ( $H_1$ ) . . . . .	5
5.2	交互效应与异质性分析 ( $H_2$ ) . . . . .	6
5.3	非线性检验与机器学习探索 . . . . .	7
<b>6</b>	<b>结论与讨论</b>	<b>7</b>
6.1	主要发现 . . . . .	7
6.2	政策建议 . . . . .	7
6.3	研究局限 . . . . .	7

## 1 研究背景与问题提出

初中阶段是青少年教育分流的关键时期，个体的教育期望（Educational Expectation）——即学生对自己未来能达到最高学历的预期——对其学业成就具有重要的导向作用。除了众所周知的家庭社会经济地位（SES）和个人认知能力外，学校环境中的人际关系是否构成了另一种重要的资源？

本研究引入社会资本理论，将学校人际关系划分为两个维度：

1. 同伴关系（Bonding Social Capital / 横向关系）：指同学间的紧密程度、友好互动及班级归属感。
2. 师生关系（Linking Social Capital / 纵向关系）：指掌握制度性资源的教师给予学生的情感支持、表扬与关注。

核心研究问题：

- 主效应 ( $H_1$ )：这两种关系能否独立提升初中生的教育期望？
- 交互效应 ( $H_2$ )：这种提升作用在不同阶层（高低 SES）或户籍（城乡）学生中是否公平？是存在“马太效应”（优势累积），还是“补偿效应”（雪中送炭）？

## 2 数据来源与样本筛选

### 2.1 数据来源

数据来源于中国教育追踪调查（CEPS）初中基线调查数据。该数据具有全国代表性，包含了多阶段概率抽样抽取的初中生样本。

### 2.2 样本清洗流程

1. 初始样本：包含约 9827 名学生。
2. 关键剔除：剔除在教育期望问题（w2b18）中回答“10= 无所谓”的学生（ $N = 434$ ）。
  - 理由：有序 Logit 模型要求因变量具有严格的等级顺序（ $1 < 2 < \dots < 9$ ）。“无所谓”代表退出教育竞争态度，而非更高等级。

- 选择性偏差说明：经  $t$  检验和  $\chi^2$  检验，被剔除组在 SES、认知得分、同伴及师生关系得分上均显著低于保留组。因此，本研究结论仅适用于有明确升学意愿的学生群体。

3. 最终样本：进入主模型分析的有效样本量约为 9314 人。

## 3 变量测量与构造

### 3.1 因变量：教育期望 ( $Y$ )

基于问卷题目 w2b18，将选项编码为 1-9 的定序变量：

1= 不念了, 2= 初中, 3= 中专/技校, 4= 职高, 5= 普高, 6= 大专, 7= 本科, 8= 研究生, 9= 博士。

描述性统计显示，数据呈左偏分布，选择“本科”（7）的人数最多（ $N = 3429$ ），其次为“研究生”（8）和“博士”（9）。

### 3.2 自变量：学校社会资本

所有连续变量在合成后均进行 Z-score 标准化处理。

#### 3.2.1 同伴关系指数 ( $Bonding\_Idx$ )

- 指标来源：w2b0605（同学友好）、w2b0606（班风好）、w2b0607（常参加活动）。
- 构造方法：取三者均值/加总后标准化。

#### 3.2.2 师生关系指数 ( $Linking\_Idx$ )

- 指标来源：
  - 表扬维度 ( $Praise$ ): w2b0507, w2b0508, w2b0509 的均值。
  - 谈心维度 ( $Talk$ ): w2c09（1= 是, 0= 否）。
- 构造方法： $Linking = Z(Z(Praise) + Z(Talk))$ 。

### 3.3 控制变量

- 家庭背景 (*SES\_PCA*): 选取父母学历、家庭经济自评、藏书量、书桌、电脑拥有情况, 通过主成分分析 (PCA) 提取第一主成分并标准化。
- 户籍 (*Hukou*): 二分变量, 0= 城镇, 1= 农村。
- 认知能力 (*Cog\_Score*): 标准化后的认知测试得分 (0 分视为缺失)。

## 4 统计模型与方法

由于因变量  $Y$  为定序变量, 本研究采用 有序 Logit 模型 (Ordered Logit Model)。

### 4.1 模型设定

设  $\theta_k$  为第  $k$  个等级的切点 ( $k = 1, \dots, 8$ ), 模型公式为:

$$\ln \left( \frac{P(Y \leq k | \mathbf{x})}{P(Y > k | \mathbf{x})} \right) = \theta_k - (\beta_1 Bonding + \beta_2 Linking + \beta_3 SES + \beta_4 Cog + \dots) \quad (1)$$

其中,  $\beta$  为回归系数。若  $\beta > 0$ , 表示随着自变量增加, 学生更有可能选择较高的教育期望等级。

### 4.2 估计策略

- 聚类稳健标准误: 考虑到学生嵌套于班级, 使用按班级编号 (*clsids*) 聚类的稳健标准误 (Cluster-Robust SE) 以校正组内相关性。
- 比例优势假设 (PO) 检验: 使用似然比检验 (LR Test) 对比有序 Logit 与多项 Logit (MNLogit)。结果显示  $p \approx 0$ , 即 PO 假设被拒绝。鉴于有序模型解释的简洁性, 仍将其作为主模型, 但以多项 Logit 结果作为稳健性对照。

## 5 实证分析结果

### 5.1 主效应分析 ( $H_1$ )

主模型回归结果显示, 所有核心变量均在  $p < 0.001$  水平上显著。

表 1: 教育期望影响因素的主模型回归结果 (示意)

变量	系数 ( $\beta$ )	标准误 (Robust SE)	显著性
Cog_Score (Z)	0.56	0.015	***
SES_PCA (Z)	0.33	0.018	***
Linking_Idx (Z)	<b>0.22</b>	0.016	***
Bonding_Idx (Z)	<b>0.14</b>	0.014	***
Hukou (农村 =1)	0.11	0.045	*

p<0.01, \* p<0.05。控制了性别等其他变量。

#### 结果解读:

1. **认知主导:** 认知能力 ( $\beta \approx 0.56$ ) 是最强的预测因子。
2. **社会资本有效:** 师生关系 ( $\beta \approx 0.22$ ) 和同伴关系 ( $\beta \approx 0.14$ ) 均显著正向预测教育期望。假设  $H_1$  得到支持。
3. **农村韧性:** 在控制其他变量后, 农村学生的教育期望略高于城镇学生 ( $\beta \approx 0.11$ )。

## 5.2 交互效应与异质性分析 ( $H_2$ )

为了检验社会资本作用的公平性, 我们引入了交互项。

- **同伴关系  $\times$  SES:** 系数显著为正 ( $\beta > 0, p < 0.05$ )。

**结论: 马太效应。** 高 SES 学生能从良好的同伴关系中获得比低 SES 学生更大的期望提升。这表明同伴网络可能在某种程度上放大了阶层优势。

- **师生关系  $\times$  SES:** 系数不显著 ( $p > 0.1$ )。

**结论:** 师生关系对不同家庭背景学生的作用差异不明显。

- **师生关系  $\times$  户籍 (Rural):** 系数边缘显著为正 ( $p \approx 0.05 \sim 0.16$ )。

**结论: 微弱补偿效应。** 相比城镇学生, 农村学生似乎能从师生关系中获  
得略多的边际收益。

### 5.3 非线性检验与机器学习探索

- **非线性检验：**使用样条函数发现，同伴关系呈线性特征，而师生关系存在一定的非线性信号 ( $p \approx 0.005$ )。
- **随机森林验证：**特征重要性排序为  $Cog > SES > Linking > Bonding$ ，与 Logit 模型系数大小一致。部分依赖图 (PDP) 显示了自变量与高教育期望概率的单调正向关系。

## 6 结论与讨论

### 6.1 主要发现

本研究证实，学校不仅是知识传授的场所，更是社会资本积累的重要场域。

1. **社会资本是独立资源：**即便家庭贫困，良好的师生和同伴关系也能为学生提供向上的动力。
2. **同伴关系的“双刃剑”：**同伴网络更倾向于“锦上添花”，高阶层学生更能利用这一资源（马太效应）。
3. **教师的关键角色：**相比同伴，师生关系更具普适性，且对农村学生具有潜在的补偿价值。

### 6.2 政策建议

- **强化导师制：**对于农村及低 SES 学生，单纯的同伴融合可能不足以弥补差距，应优先加强教师对其的一对一支持（Mentoring）。
- **警惕圈层化：**在班级建设中，需注意防止高 SES 学生的同伴圈层化导致的优势垄断。

### 6.3 研究局限

1. **因果推断：**横截面数据无法完全排除反向因果（如高期望学生主动寻求好关系）。
2. **测量主观性：**社会资本基于学生自评，缺乏客观社交网络数据。
3. **样本限制：**剔除了“无所谓”群体，结论外推需谨慎。