

# 汇报讲解稿

## (逐页对应 report\_summary.pdf, 19 页)

面向口播/演示的逐页解说

January 25, 2026

### Contents

<b>1</b>	<b>第 1 页：封面、摘要、假设总览</b>	<b>4</b>
1.1	页面内容 .....	4
1.2	讲解话术 .....	4
<b>2</b>	<b>第 2 页：数据来源与样本清洗</b>	<b>4</b>
2.1	页面内容 .....	4
2.2	讲解话术 .....	4
<b>3</b>	<b>第 3 页：教育期望编码表 (Table 1)</b>	<b>4</b>
3.1	页面内容 .....	4
3.2	讲解话术 .....	5
<b>4</b>	<b>第 4 页：Bonding &amp; Linking 问卷条目表</b>	<b>5</b>
4.1	页面内容 .....	5
4.2	讲解话术 .....	5
<b>5</b>	<b>第 5 页：PCA 方法与碎石图 (Figure 1)</b>	<b>5</b>
5.1	页面内容 .....	5
5.2	讲解话术 .....	5
<b>6</b>	<b>第 6 页：PC1 载荷与 SES 分布 (Figure 2, 3)</b>	<b>5</b>
6.1	页面内容 .....	5
6.2	讲解话术 .....	6
<b>7</b>	<b>第 7 页：PCA 输入相关矩阵 (Figure 4)</b>	<b>6</b>
7.1	页面内容 .....	6
7.2	讲解话术 .....	6

<b>8</b>	<b>第 8-9 页：统计模型与因变量分布（Figure 5）</b>	<b>6</b>
8.1	页面内容 .....	6
8.2	讲解话术 .....	6
<b>9</b>	<b>第 10-11 页：自变量分布与相关矩阵（Figure 6, 7, 8）</b>	<b>6</b>
9.1	页面内容 .....	6
9.2	讲解话术 .....	7
<b>10</b>	<b>第 11-12 页：回归结果（表）</b>	<b>7</b>
10.1	页面内容 .....	7
10.2	讲解话术 .....	7
<b>11</b>	<b>第 12-13 页：交互效应（Figure 9, 10, 11）</b>	<b>7</b>
11.1	页面内容 .....	7
11.2	讲解话术 .....	8
<b>12</b>	<b>第 13 页：非线性与稳健性（Figure 12）</b>	<b>8</b>
12.1	页面内容 .....	8
12.2	讲解话术 .....	8
<b>13</b>	<b>第 14-18 页：机器学习补充（Figure 13, 14）</b>	<b>8</b>
13.1	页面内容 .....	8
13.2	讲解话术 .....	9
<b>14</b>	<b>第 19 页：结论、局限、后续方向</b>	<b>9</b>
14.1	页面内容 .....	9
14.2	讲解话术 .....	9

# 1 第 1 页：封面、摘要、假设总览

## 1.1 页面内容

- 标题、摘要；研究目标与假设列表 H1/H2a-c/H3。

## 1.2 讲解话术

### 开场与摘要

“本报告用 CEPS 八年级数据（有效样本  $\approx 9,000$ ，剔除 ‘10= 无所谓’），看学校社会资本如何影响教育期望。主效应：师生联结（Linking）和同伴纽带（Bonding）都显著正向。事前假设是 ‘补偿效应’，但 Bonding $\times$ SES 反而指向优势累积（高 SES 获益更多， $p \approx 0.052$ ）；Linking $\times$ SES 不显著；Linking $\times$ Rural 弱正向未显著。PO 假设被拒绝；Linking 在低段有非线性（样条  $p \approx 0.005$ ）。H3 探索：比较 Linking vs Bonding 的补偿力度，是否对弱势更有帮助或至少不放大差距。”

# 2 第 2 页：数据来源与样本清洗

## 2.1 页面内容

- 数据：CEPS Wave 2，PPS 抽样，全国代表性。
- 样本处理：剔除 “10= 无所谓”， $N \approx 9,012$ ；保留班级 ID 做聚类稳健 SE。

## 2.2 讲解话术

### 样本与清洗

“我们用 CEPS 第二轮八年级追踪。因变量是 1-9 的期望等级，10= ‘无所谓’ 无法排序，直接剔除。最终样本约 9,012；班级号 clsids 用于聚类稳健标准误，防止班内相关性低估 SE。若 ‘无所谓’ 人群有系统偏差，后续可用选择模型或多项 Logit 做稳健性。”

# 3 第 3 页：教育期望编码表（Table 1）

## 3.1 页面内容

- Table 1: w2b18 选项 1-9（不读/小学…博士），10= 无所谓已剔除。
- 引出核心自变量 Bonding/Linking。

## 3.2 讲解话术

### Table 1 讲解

“问卷问：‘你希望最高读到什么学历？’编码 1-9 从不读/小学到博士，10= ‘无所谓’ 仅列示但不入模。这样保持标签严格有序，支持有序 Logit 和序数分类。”

## 4 第 4 页：Bonding & Linking 问卷条目表

### 4.1 页面内容

- 新增表：w2b0605/06/07, w2b0507/08/09, w2c09 的题干与 1-4 频率选项；反向计分与二元化规则。

### 4.2 讲解话术

#### Bonding/Linking 题目

“Bonding: w2b0605 反向（经常有麻烦/被欺负），0606 友善，0607 归属感，频率 1-4。Linking: w2b0507-0509 老师表扬频率（1-4），w2c09 主动与老师谈话（1=经常，其余归 0）。先反向，再合成，最终 Z-score。”

## 5 第 5 页：PCA 方法与碎石图（Figure 1）

### 5.1 页面内容

- SES PCA 文本解释；Figure 1 碎石图（PC1 44.1%，拐点后仅保留 PC1）。

### 5.2 讲解话术

#### Figure 1 碎石图

“PCA 把父母学历、家庭经济、藏书、书桌、电脑 5 个相关指标压成综合 SES。碎石图显示 PC1 解释 44.1% 方差，PC2 后迅速下跌，拐点在 PC1 之后，因此保留 PC1 作为 SES 综合分。”

## 6 第 6 页：PC1 载荷与 SES 分布（Figure 2, 3）

### 6.1 页面内容

- Figure 2: PC1 载荷条形图（全为正，藏书/书桌最高）。

- Figure 3: SES 得分分布（近似正态，均值 0）。

## 6.2 讲解话术

### Figure 2 & 3

“PC1 载荷全正，说明是综合 SES 轴；藏书和书桌贡献最大，其次父母学历、经济，电脑最小。SES 得分近似正态，便于下游模型线性/树模型使用。”

## 7 第 7 页：PCA 输入相关矩阵（Figure 4）

### 7.1 页面内容

- Figure 4: 五个 SES 指标的相关矩阵，呈中度正相关。

### 7.2 讲解话术

#### Figure 4

“各指标有中度正相关，验证了用 PCA 降维的必要性，避免多重共线性拖累回归。”

## 8 第 8-9 页：统计模型与因变量分布（Figure 5）

### 8.1 页面内容

- Ordered Logit 定义、切点、累积概率公式；交互项写法。
- Figure 5: 教育期望分布（1-9，左偏，本科最多）。

### 8.2 讲解话术

#### 模型与分布

“Ordered Logit 适合有序标签；切点决定等级边界。交互示例：Bonding×SES,  $\beta_3 > 0$  表示高 SES 获益更多。因变量分布左偏，本科占比最高，硕士、大专次之。”

## 9 第 10-11 页：自变量分布与相关矩阵（Figure 6, 7, 8）

### 9.1 页面内容

- Figure 6: Bonding 分布（标准化，多峰源于离散问卷）。

- Figure 7: Linking 分布（标准化，多峰）。
- Figure 8: 核心变量相关矩阵（Bonding-Linking 中度相关；认知-期望最强；户籍最弱）。

## 9.2 讲解话术

Figure 6-8

“Bonding/Linking 多峰是因为原始题目是 4 级频率。核心变量相关：Bonding 与 Linking  $r \approx 0.46$ ，说明有重叠但非同一概念；认知与教育期望  $r \approx 0.30$  最强；户籍相关最弱且略负。”

## 10 第 11-12 页：回归结果（表）

### 10.1 页面内容

- Ordered Logit 主效应表：Linking 0.19, Bonding 0.14, SES 0.33, Cognitive 0.56, Hukou 0.11；MLE，聚类 SE 方向不变。表下注明主效应模型；含交互模型系数仍约 0.19/0.14。

### 10.2 讲解话术

主效应表

“系数为 log-odds：Linking/Bonding 均显著正向；认知 0.56 最大，SES 0.33 次之；户籍 0.11 弱正向。表中为主效应模型，含交互模型下 Linking/Bonding 量级相近且仍显著。”

## 11 第 12-13 页：交互效应（Figure 9, 10, 11）

### 11.1 页面内容

- 文字：Bonding×SES 正向（MLE  $p \approx 0.015$ ，聚类  $p \approx 0.052$ ，趋势）；Linking×SES 不显著；Linking×Rural 正但  $p \approx 0.16$ 。
- 图：Figure 9 Bonding×SES 分组概率；Figure 10 Linking×SES；Figure 11 Linking×Rural。

## 11.2 讲解话术

### 交互解读

“Bonding×SES 显示高 SES 获益更大，违背补偿，接近马太效应；稳健 SE 边界显著。Linking×SES 近似零，说明师生关系对不同 SES 一视同仁。Linking×Rural 正向但不显著，弱补偿迹象有限。”

## 12 第 13 页：非线性与稳健性（Figure 12）

### 12.1 页面内容

- PO 检验被拒；未实现 partial PO，附录有多项 Logit 对照。
- Figure 12: Linking 样条效应（df=4，degree=3，低段更强， $p \approx 0.005$ ）；Bonding 无非线性。

### 12.2 讲解话术

#### 非线性

“PO 被拒，需提醒阈值异质性。样条检验显示 Linking 在低水平的边际效应更大，说明弱联系时提升最明显；Bonding 没有类似非线性。”

## 13 第 14-18 页：机器学习补充（Figure 13, 14）

### 13.1 页面内容

- Figure 13: RF 流程图（数据 → 标准化/编码 → PCA SES → RF；200 树、深度 5、叶子  $\geq 80$ 、random\_state=42；未做 train/val 或 CV，评估节点标记 planned）。
- 变量重要性条形图（Figure 14）：认知 0.409，SES 0.378，Linking 0.133，Bonding 0.070，Hukou 0.010。
- 文本：RF 与回归方向一致，但未评估预测性能，属探索性、非因果；提示可补 PDP/Permutation 与 CV。

## 13.2 讲解话术

### RF 流程与结果

“随机森林用全量数据做模式识别，不做因果推断。参数定死（200 树，深度 5，叶 80）。重要性排序与回归一致：认知/SES 领先，Linking 高于 Bonding，户籍最弱。由于缺少 train/val/CV 和类权重，性能可能乐观；后续需补评估、PDP、Permutation。”

## 14 第 19 页：结论、局限、后续方向

### 14.1 页面内容

- 结论要点：主效应正向；Bonding×SES 优势累积；Linking×SES ns；Linking×Rural 弱；PO 被拒；Linking 低段非线性；RF 对照一致但仅模式参考。
- 局限：剔除“无所谓”潜在偏差；未做 partial PO；样条设定依赖；RF 无 CV/类权重；量表维度有限。
- 后续方向：partial PO/广义有序；对“无所谓”做选择模型或 MI；扩展非线性/高阶交互；补齐 ML 评估与 PDP/Permutation；采集 Bridging/学校层级社会资本。

### 14.2 讲解话术

#### 收束与展望

“主效应稳健，交互揭示 Bonding 的优势累积风险，Linking 更均衡。局限与未来工作已点明：补做 partial PO、选择模型、非线性稳健、ML 评估与可视化，以及补充跨层社会资本指标，才能把‘补偿 vs 累积’的机制讲得更透。”