



中国 教育 追踪 调 查  
CHINA EDUCATION PANEL SURVEY

# **2014-2015 学年追访 数据使用手册**

中国人民大学中国调查与数据中心  
中国教育追踪调查(CEPS)项目组

2017 年 12 月



# 目录

一、	研究设计 .....	3
二、	追访情况说明 .....	5
(一)	追访学生情况 .....	5
(二)	追访班级情况 .....	6
(三)	追访教师情况 .....	6
三、	测量工具 .....	8
(一)	学生问卷 .....	8
(二)	家长问卷 .....	10
(三)	班主任问卷 .....	12
(四)	任课教师问卷 .....	13
(五)	校领导问卷 .....	14
(六)	认知能力测试 .....	16
(七)	主课（语数外）期中考试成绩 .....	17
四、	数据结构 .....	18
(一)	相关变量说明 .....	18
(二)	数据合并匹配 .....	20



# 一、 研究设计

中国教育追踪调查(China Education Panel Survey, 简称为 CEPS)是由中国人民大学中国调查与数据中心设计与实施的、具有全国代表性的大型追踪调查项目,旨在揭示家庭、学校、社区以及宏观社会结构对于个人教育产出的影响,并进一步探究教育产出在个人生命历程中发生作用的过程。

中国教育追踪调查(CEPS)以 2013-2014 学年为基线,以初中一年级(7 年级)和初中三年级(9 年级)两个同期群为调查起点,以人口平均受教育水平和流动人口比例为分层变量从全国随机抽取了 28 个县级单位(县、区、市)作为调查点。调查的执行以学校为基础,在入选的县级单位随机抽取了 112 所学校、438 个班级进行调查,被抽中班级的学生全体入样,基线调查共调查了约 2 万名学生。

中国教育追踪调查(CEPS)以问卷调查为主要手段,对全体被调查学生及其家长或监护人、班主任老师、主课任课老师以及学校负责人进行问卷调查。问卷的内容包括:学生的基本信息、户籍与流动、成长经历、身心健康、亲子互动、在校学习、课外活动、与老师/同学的关系、社会行为发展、教育期望、以及家庭成员基本信息,家长的基本信息、生活习惯、亲子互动、家庭教育环境、家庭教育投入、社区环境、对学校教育的看法、与老师的互动、对孩子的教育期望,班主任老师对学生行为的评价、与学生家长的互动、对本地与外地户籍学生的比较,包括班主任在内的任课老师的个人基本信息、教育理念、日常教学工作、工作压力与满意度,学校负责人基本信息、教育理念,学校的基本信息、教学设施、招生入学、在校师生情况以及日常教学管理。

中国教育追踪调查(CEPS)同时还对学生进行综合认知能力测试、基本人格测试,并收集学生的重要考试(期中考试、中考、高考等)成绩,并组织对学生的健康与体格检查,采集生物医学指标,综合利用各种手段和技术全面采集高质量的数据。

中国教育追踪调查(CEPS)在初中阶段逐年进行追踪调查,然后在学生初中毕业后的第 1 年、第 3 年、第 4 年、第 7 年、第 8 年、第 17 年、第 27 年进行追踪调查,整个调查周期长达 30 年。按计划,最后一次的调查将于 2043 进行。

中国教育追踪调查 (CEPS) 所产出的内容广泛的多层次追踪数据向学术界全面开放，为教育学、经济学、管理学、社会学等相关学科的研究者提供基础数据资源，同时也服务于教育政策的制定者和学校的管理者，为其决策提供可靠的实证数据依据。

中国教育追踪调查 (CEPS) 的项目网址为: <http://ceps.ruc.edu.cn>

中国教育追踪调查的微信公众号为: china\_eps



## 二、 追访情况说明

中国教育追踪调查(CEPS)以 2013-2014 学年为基线,初中一年级(7 年级)和初中三年级(9 年级)两个同期群为调查起点,其中初中三年级(9 年级)为实验性测试样本。2014-2015 学年的追踪调查中,追访对象为基线调查时的初中一年级(7 年级)的全部 10279 名学生。

### (一) 追访学生情况

参与了基线调查的所有学生,无论是留级、跳级或个别转入其他班级的学生,只要学生还在原学校就读,均列为 2014-2015 学年追访调查对象。班级中新转入的插班生,或该班在基线调查时由于请假等原因未参加基线调查的学生,也纳入了 2014-2015 学年的追踪调查,并在数据中标注为“新入样学生”。如果基线调查时原入样班级从七年级到八年级已经重新打乱分班,只调查仍在该校就读的被追踪学生(包括留级、跳级的学生),不增加新入样学生。

基线调查中,初中一年级(7 年级)入样学生有 10279 人,在 2014-2015 学年追踪调查中,成功追访学生人数 9449 人,追访率为 91.9%。新入样学生人数 471 人,失访学生人数 830 人。其中,学生失访原因具体情况如下表 1 所示。

表 1 2014-2015 学年学生失访原因

失访原因	数量	百分比(%)
已转校	589	70.96
退学	121	14.58
不详	51	6.14
调查当天请假	31	3.73
长期治病	14	1.69
转入本校其他班级	8	0.96
休学	4	0.48
本校留级	3	0.36
读技校	3	0.36
开除	2	0.24
死亡	2	0.24
参加艺体特长训练比赛	1	0.12
去外校寄读	1	0.12
合计	830	100.00

## （二） 追访班级情况

对于基线调查样本中的 112 所学校，有 20 所学校存在重新分班的情况，1 所学校存在班级合并的情况，91 所学校的班级没有发生变化。基线调查入样班级共有 221 个班，在 2014-2015 学年追踪调查中，共调查了 301 个班级，其中 184 个班级上在基线调查时就已经存在，另外 117 个班级是在追踪调查时由于重新分班或者是由于基线调查时的学生转入而被纳入调查范围的追访班级变动情况如图 1 所示：

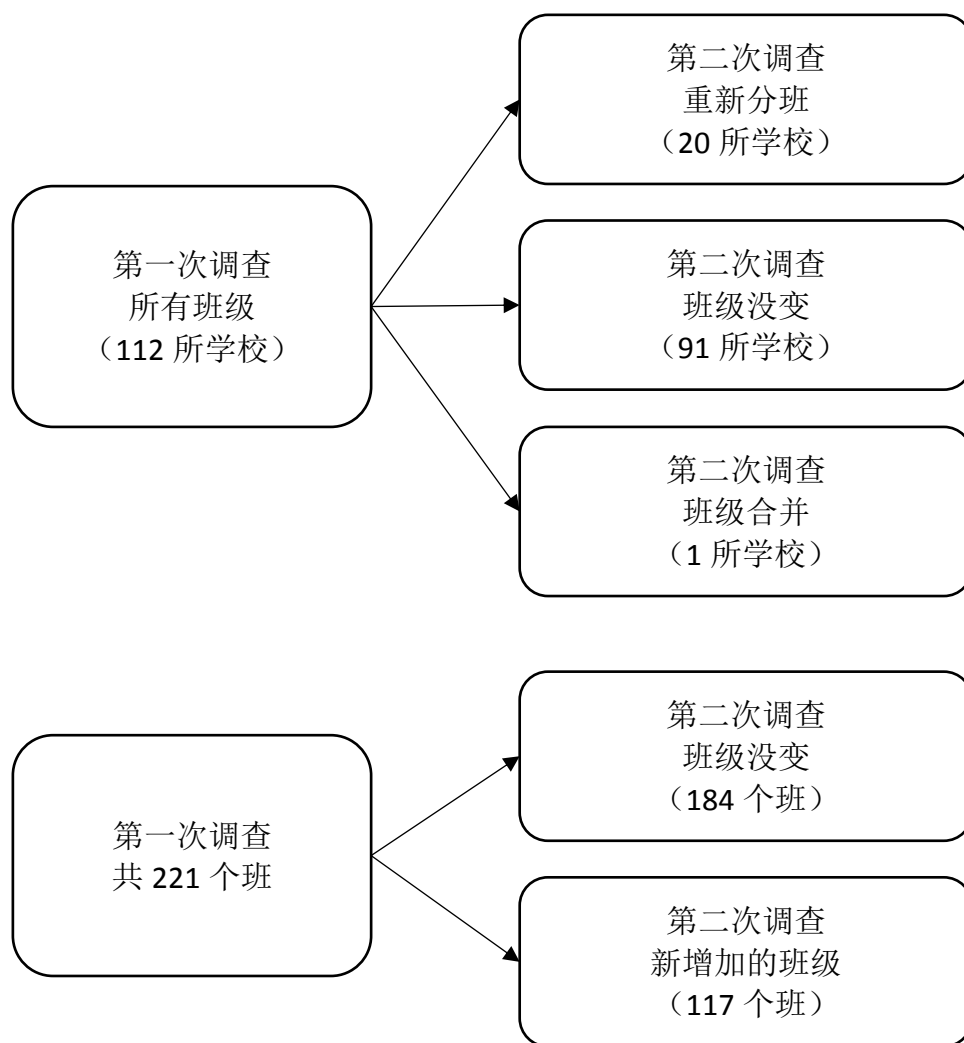


图 1 2014-2015 学年班级追访情况统计

## （三） 追访教师情况

2014-2015 学年追踪调查中，对入样学生的负责班主任老师和主课老师（语文老师、数学老师、英语老师）进行了问卷调查。如果基线调查的七年级班级已



经分班，则所有被追踪学生所在班级的班主任均需要填写班主任问卷，这些班级的主科目（语、数、英）任课教师同样均需要填写任课教师问卷；如果班主任所教科目为主科目，该班另外 2 名主科目任课教师填写任课教师问卷；如果班主任所教科目不是主科目，该班另外 3 名主科目任课教师填写任课教师问卷；如果 2 个被调查班级有相同的主科老师，该老师针对这 2 个班级分别填写问卷。2014-2015 学年追踪调查一共调查了 301 个班级，但个别班级只有一两位学生纳入了调查范围，应调查的班主任老理由及任课教师有一定的缺失。

**表 2 2014-2015 学年负责教师调查样本统计**

类别	W1 原样本	W2 新入样	缺失样本
班主任	183	107	11
数学老师	160	87	54
语文老师	162	91	48
英语老师	162	80	59

### 三、测量工具

2014-2015 学年追访问卷沿用了与基线调查一样的学生问卷、家长问卷、班主任问卷、任课教师问卷、以及校领导问卷共 5 套问卷。但是，为了获得更准确的调查信息，追访问卷在题设和选项方面做了部分修改和增删。

#### （一）学生问卷

2014-2015 学年追访学生问卷主要包括：个人基本信息、家庭情况、亲子互动、教育期望、在校学习、课外活动、身心健康、社会行为发展等模块。与基线问卷相比，追访学生问卷中修改问题以及新增问题如表 3 和表 4 所示。

另外，与 2014 年秋季学生问卷相比，2015 年春季学生问卷做了以下修改：

- （1） A8 题的选项 2 后面对应的小题“b. 你被判给谁来抚养？”添加了一个选项：3. 其他（请注明\_\_\_\_\_）；
- （2） B16 题添加两个选项：10. 地铁 11. 其他（请注明\_\_\_\_\_）

表 3 2014-2015 学年追访学生问卷修改题目列表

题号对比		变量对比		修改类别
W2 题号	W1 题号	W2 变量	W1 变量	
A13 A14	B8	w2a13 w2a14	b08b b08a	追访问卷对选项进行了重新的定义
A30	B33	w2a30	b33	
A31	B34	w2a31	b34	
B19	C23	w2b19	c23	
B7	B14 B15	w2b07a- w2b07f	b14a1 b14a2 b15a1 b15a2 b15b1 b15b2 b15e1 b15e2 b15f1 b15f2	追访问卷是选择题；基线问卷是填空题
B8	B14 B16	w2b08a- w2b08f	b14b1 b14b2 b16a1 b16a2 b16b1 b16b2	

			b16e1 b16e2 b16f1 b16f2	
B9 B10	B17	w2b09 w2b10	b1701 b1702	追访问卷多了一个选项： 6.每周一次以上
A21	B24	w2a2102a w2a2102b	b24b2 b24a2	追访问卷问的是你与同 学的关系；基线问卷问的 是你与朋友的关系

表 4 2014-2015 学年追访学生问卷新增题目列表

模块	新增题号	新增问题	新增变量
个人基本信息	A4	你会本地方言吗	w2a04
家庭情况	A7	在你现在的家里,和你一起住的都有：（继父/继母）	w2a0703 w2a0704
	A8	你的亲生父母现在的婚姻状况是	w2a0801-w2a0804a
在校学习	B1	小学六年级时,你们班总共有多少个同学?	w2b01
	B5	关于主课程,你是否同意下列说法:(数学老师在课堂上常常注意我/语文老师在课堂上常常注意我/英语老师在课堂上常常注意我)	w2b0501 w2b0502 w2b0503
	B6	关于学校生活,你是否同意下列说法-我的父母经常收到老师对我的表扬	w2b0601
	B16	从家里到学校,你通常使用什么方式	w2b16
	B17	使用上面选的交通方式,你从家到学校通常要花多长时间	w2b17
课外活动	B7	周一到周五,你通常每天的课外活动时间安排是-上兴趣班(与课业无关)	w2b07d
	B8	在周末,你通常每天的课外活动时间安排是-上兴趣班(与课业无关)	w2b08d
	B13	除了上题中提到的兴趣爱好,你平时还喜欢做哪些事情	w2b1300- w2b1305

身心健康	C3 C23	对自己身材/相貌的评价	w2c03 w2c23
	C5-C8	与身体健康相关问题	w2c05-w2c08b
	C10-C12		w2c10-w2c12
	C13-C17	与体育锻炼/生理健康/心理特征等相关问题	w2c13a-w2c17a
	C19-C22		w2c1900-w2c22b
	C24-C26		w2c2401-w2c2607
社会行为发展	D1	过去一年中,你能够做到以下几点吗:帮助老人做事情/遵守秩序,自觉排队/待人真诚友善	w2d0101-w2d0103
	D2	过去一年中,你有没有以下行为:骂人,说脏话/吵架/打架/欺负弱小同学/脾气暴躁/注意力不集中/逃课,旷课,逃学/抄袭作业,考试作弊/抽烟,喝酒/上网吧,游戏厅	w2d0201-w2d0210
	D3	你是否同意下列说法: 我很害羞/我常自己一个人坐着,而不愿与别人在一起/与同学或同伴在一起时,我不常讲话,多数时间听他们说话/有一些我尊敬和崇拜的成年人/我能和大人轻松地交谈/当我不小心伤害或得罪人时,我会道歉/如果我处理事情的方法不对,我会尽量想别的办法解决/即使情况很糟糕,我也能保持镇静/对于需要完成的任务,我通常很有信心	w2d0301-w2d0309
	D4-D8	与异性同学或朋友的感情及对恋爱的态度等问题	w2d04-w2d08

## (二) 家长问卷

2014-2015 学年追访家长问卷主要包括: 家长基本信息、生活习惯、亲子互动、家庭教育投入、家庭教育环境、社区环境、对学校教育的看法、与老师的互动、对孩子的教育期望等模块。与基线问卷相比, 追访家长问卷中修改问题以及新增问题如表 5 和表 6 所示。

表 5 2014-2015 学年追访家长问卷修改题目列表

题号对比		变量对比		修改类别
W2 题号	W1 题号	W2 变量	W1 变量	
A6 A7	E10	w2ba06 w2ba07	be1001 be1003	追访问卷对选项进行了重新的定义
A30	A19	w2ba30	ba19	
A31	A20	w2ba31	ba20	
E22	E8	w2be22	be08	
E38	E33	w2be38	be3301	
E39		w2be39	be3302	
E40		w2be40	be3303	
C9	C14	w2bc09	bc14	追访问卷减少了一个选项：5.到其他地方读职业高中/中专/技校
D9	D14	w2bd09	bd14	追访问卷多了一个选项：2. 近视,不清楚具体度数

表 6 2014-2015 学年追访家长问卷新增题目列表

模块	新增题号	新增问题	新增变量
亲子互动	A2	孩子的哪些亲属目前在家中与孩子同住： (孩子的继父/孩子的继母)	w2ba0203 w2ba0204
	A14	平时直接负责照顾这个孩子日常生活起居的家庭成员是：	w2ba1400-w2ba1405a
	A15	平时直接负责辅导这个孩子功课,学习的家庭成员是：	w2ba1500-w2ba1505a
	A16	平时直接负责陪同这个孩子娱乐,玩耍的家庭成员是：	w2ba1600-w2ba1605a
	A27	过去一年中,孩子能够做到以下几点吗:帮助老人做事情/遵守秩序,自觉排队/待人真诚友善	w2ba2701-w2ba2703

	A28	过去一年中,孩子有没有以下行为: 骂人,说脏话/吵架/打架/欺负弱小同学/脾气暴躁/注意力不集中/逃课,旷课,逃学/抄袭作业,考试作弊/抽烟,喝酒/上网吧,游戏厅	w2ba2805-w2ba2810
学校教育	C1	孩子上学期交给学校的费用: (教辅材料费/校服费/保险费)	w2bc01d w2bc01e w2bc01g
孩子基本情况	D5	根据出生证,孩子父母生育这个孩子时的年龄	w2bd05a w2bd05b
	D6	到目前为止,孩子生过下面这几种大病吗:肾病/肺部疾病/心脏病/脑部疾病/上肢骨折/下肢骨折	w2bd0601a-w2bd0606c
	D10	孩子的近视是先天的吗	w2bd10
	D11	孩子现在是否有下列障碍-癫痫	w2bd1107
	D12	孩子有什么样的医疗保险-校内外意外伤害保险	w2bd1205
家长基本信息	A9	和孩子住在一起的家庭成员中,有没有人几乎每周都喝酒?	w2ba09 w2ba09a
	A10	和孩子住在一起的家庭成员中,有没有人抽烟?	w2ba10 w2ba10 a
	E1-E18	孩子父母的出生日期/民族/户口/政治面貌/教育程度/职业等相关问题	w2be01-w2be18
社区环境	E35	您家周围的邻居大多数是	w2be35
	E36	您家周围的邻居大多数位于社会的什么阶层	w2be36

### (三) 班主任问卷

2014-2015 学年追访班主任问卷主要包括: 个人基本信息、教育理念、日常教学工作、工作压力与满意度、对学生行为的评价、与学生家长的互动以及对本

地与外地户籍学生的比较等模块。与基线问卷相比，追访班主任问卷中修改问题以及新增问题如表 7 和表 8 所示。

表 7 2014-2015 学年追访班主任问卷修改题目列表

题号对比		变量对比		修改类别
W2 题号	W1 题号	W2 变量	W1 变量	
A12	A12	w2tcha1201 w2tcha1202 w2tcha1203 w2tcha1204	h1a12	追访问卷是多项选择题； 基线问卷是单项选择题

表 8 2014-2015 学年追访班主任问卷新增题目列表

模块	新增题号	新增问题	新增变量
日常教学管理	A22	假设八年级学生小明，上课不专心，经常边学习边玩，总是坐不住，活动过多，情绪容易冲。遇到这种学生，您会怎么处理？	w2tcha2201-w2tcha2207
	A23	假设八年级学生小红，最近变得情绪低落、思维缓慢、行动迟缓、提不起劲来。遇到这种学生，您会怎么处理？	w2tcha2301-w2tcha2307
与学生家长的互动	B2	您和学生家长联系时通常使用哪些方式	w2tchb0201-w2tchb0207
个人基本信息	C7	您是否接受过心理健康培训	w2tchc07

#### （四）任课教师问卷

2014-2015 学年追访任课教师问卷主要包括：个人基本信息、教育理念、日常教学工作、工作压力与满意度等模块。与基线问卷相比，追访任课教师问卷中修改问题以及新增问题如表 9 和表 10 所示。

表 9 2014-2015 学年追访任课教师问卷修改题目列表

题号对比		变量对比		修改类别
W2 题号	W1 题号	W2 变量	W1 变量	
A12	A12	w2tcha1201 w2tcha1202	ch1a12 / mata12 / enga12	追访问卷是多项选择题；

		w2tcha1203 w2tcha1204		基线问卷是单项选择题
--	--	--------------------------	--	------------

表 10 2014-2015 学年追访任课教师问卷新增题目列表

模块	新增题号	新增问题	新增变量
日常教学管理	A22	假设八年级学生小明，上课不专心，经常边学习边玩，总是坐不住，活动过多，情绪容易冲。遇到这种学生，您会怎么处理？	w2tcha2201-w2tcha2207
	A23	假设八年级学生小红，最近变得情绪低落、思维缓慢、行动迟缓、提不起劲来。遇到这种学生，您会怎么处理？	w2tcha2301-w2tcha2307
个人基本信息	B7	您是否接受过心理健康培训	w2tchc07

## （五）校领导问卷

2014-2015 学年追访校领导问卷主要包括：个人基本信息、教育理念、学校基本信息、教学设施、招生入学、在校师生情况和日常教学管理等模块。与基线问卷相比，追访校领导问卷中修改问题以及新增问题如表 11 和表 12 所示。

表 11 2014-2015 学年追访校领导问卷修改题目列表

题号对比		变量对比		修改类别
W2 题号	W1 题号	W2 变量	W1 变量	
A4	A7	w2pla04	pla07	题干的时间点不一样：追访问卷问的是在过去一年里；基线问卷问的是近五年来
A5	A8	w2pla05	pla08	
C3	C3	w2plc0301	plc0301	题干的时间点不一样：追访问卷问的是 0-4 年；基线问卷问的是 1-4 年
B3	B4	w2plb03	plb04	题干的时间点不一样：追访问卷问的是在过去一年里；基线问卷问的



				是近五年来。同时追访问卷对选项进行了重新定义
A19	A23	w2pla19	pla23	追访问卷对选项进行了重新定义
B7	B10	w2plb0701 w2plb0702 w2plb0703	plb1001 plb1002 plb1003	
B8	B11	w2plb0801 w2plb0802 w2plb0803	plb1101 plb1102 plb1103	
E13	E13	w2ple13	ple13	

表 12 2014-2015 学年追访校领导问卷新增题目列表

模块	新增题号	新增问题	新增变量
在校师生情况	C13	学校初中部有没有对教师进行心理健康培训	w2plc13 w2plc13a
招生入学	D1	对每个本县（区）户籍的八年级学生，学校上学期收取的费用是：（教辅材料费/校服费/保险费）	w2pld0103 w2pld0104 w2pld0106
	D2	上学期，在对八年级学生收取这些费用时，学校初中部对于本县（区）学生和外县（区）学生实行不同收费标准的有哪些项目？	w2pld0200- w2pld0209
日常教学管理	E23	学校初中部哪些年级开设健康教育课之类的课程	w2ple2300-w2ple2303
	E24	学校是否安排初中部所有学生体检	w2ple24
	E25	假设八年级学生小明，上课不专心，经常边学习边玩，总是坐不住，活动过多，情绪容易冲。遇到这种学生，您会怎么处理？	w2ple2501-w2ple2507
	E26	假设八年级学生小红，最近变得情绪低落、思维缓慢、行动迟缓、提不起劲来。遇到这种学生，您会怎么处理？	w2ple2601-w2ple2607

## （六） 认知能力测试

除了调查问卷以外，CEPS 还为八年级学生设计了一套认知能力测试题，该测试题的内容不涉及学校课程所教授的具体识记性知识，而是测量学生的逻辑思维与问题解决能力，并且具有国际可比性、全国标准化的特点。

为了使测试题具有更高的区分度，更加精确地测量学生的认知能力发展水平，本次追踪调查的八年级学生认知能力测试题分为 A、B、C 三种类型，不同类型试题的题目数量相同，但题目内容和难度有一定差别。根据基线调查的认知能力的得分，给每位被追踪学生分配好相应的试题类型，试题共 35 题，时间为 30 分钟。

测试题共分 3 个维度、11 个构念，包括：

- （1）语言：词组类比、语言文字推理
- （2）图形：图形规律分析、折纸类题目、几何图形应用
- （3）计算与逻辑：数学应用、自定义运算规则、数列应用、抽象规律分析、概率、数值大小逆向思维。

变量 **w2coglmh** 表示根据基线调查时学生的认知能力测试结果，将样本学生划分为“低”、“中”、“高”三组。

变量 **w2cogtype** 表示学生在认知能力测试中使用的试卷类型，相对应变量 **w2coglmh** 的“低”、“中”、“高”三组分别使用“A”、“B”、“C”三种不同类型问卷，新入样学生统一要求使用 B 类问卷。但是因为现场执行时的特殊原因，个别同学使用了非指定的问卷类型，具体的问卷测试类型见表 13。

表 13 认知能力测试问卷使用情况统计

		w2cogtype			合计
		A	B	C	
w2coglmh	低	2975	5	0	2980
	中	1	3147	2	3150
	高	4	0	3232	3236
	合计	2980	3152	3234	9366

变量 **w2cogscore** 表示学生在认知能力测试中答对的题目数量，取值为

[0, 35]。

变量w2cog3pl表示用三参数的IRT模型估计出的学生认知能力测试标准化总分。

### **（七）主课（语数外）期中考试成绩**

学生在 2014 年秋季学期的期中考试成绩由被调查学校直接提供，非学生自填。变量 w2chn/w2mat/w2eng 分别表示语文、数学、英语的原始得分。变量 w2upchn/w2upmat/w2upeng 分别表示语文、数学和英语的满分。

## 四、 数据结构

2014-2015 学年追访数据文件包括学校数据、班级数据、学生数据和家长数据 4 部分。学校数据包括校领导问卷变量（变量名前缀为 w2p1）。班级数据包括班主任/语文老师/数学老师/英语老师问卷变量（变量名前缀为 w2tch）。学生数据包括学生问卷变量、认知能力变量、主课（语数外）期中考试成绩（变量名前缀为 w2）。家长数据包括家长问卷变量（变量名前缀为 w2b）。

其中，2014-2015 学年追访班级数据文件的排版变化较大：基线班级数据将班主任问卷变量和任课教师问卷（语文老师、数学老师、英语老师）变量横向合并，依据变量名进行区分（班主任问卷变量名前缀为 hr，语文老师问卷变量名前缀为 chn，数学老师问卷变量名前缀为 mat，英语老师问卷变量名前缀为 eng）；而追访数据将班主任问卷变量和任课教师问卷（语文老师、数学老师、英语老师）变量纵向合并，依据变量 w2tchhr（w2tchhr=0 表示不是班主任，1 表示是班主任）和变量 w2tchsubject（w2tchsubject=0 表示其他科目老师，1 表示是数学老师，2 表示是语文老师，3 表示是英语老师）进行区分。

### （一） 相关变量说明

与基线数据相比，2014-2015 学年追访学生数据库新增变量有 w2clsids, w2status, w2mtype, w2mreason, w2clsra 等；追访班级数据库新增变量有 tchids, w2tchstatus 等。具体变量说明见表 14 和表 15。

本轮数据对所有教师（包括班主任和主课教师）进行编码，为每一位教师生成唯一的 tchids, 2014-2015 学年追访到的所有教师仍使用基线调查时的 tchids, 对于变动教师依据基线编码规则赋值唯一的 tchids。

表 14 2014-2015 学年追访数据库连续变量说明表

变量名	观测值数量	最小值	最大值	变量标签
ids	10750	1	19958	学生个人代码
clsids	10750	1	436	班级代码
w2clsids	9920	1	555	W2 班级代码
schids	10750	1	112	学校代码
ctyids	10750	1	28	县（区）代码
tchids	791	6	1536	教师代码

表 15 2014-2015 学年追访数据库分类变量说明表

变量名	变量标签	取值标签	观测值数量
w2frame	PSU 所属抽样框	1 全国核心样本 f1	6230
		2 上海样本 f2	876
		3 全国补充样本 f3	3644
w2subsample	PSU 所属子样本	1 全国非上海非流动子样本 s1	5900
		2 上海子样本 s2	1170
		3 全国非上海流动子样本 s3	3680
w2fall	调查时间	0 2015 年春季	4252
		1 2014 年秋季	6498
w2status	学生追访情况	1 成功追访	9449
		2 未追访到	830
		3 新入样	471
w2mtype	学生失访类型	1 仍在原初中就读	43
		2 转入其他初中学校	590
		3 读技校	3
		4 目前不在学校读书	141
		5 死亡	2
		6 不详	51
w2mreason	学生失访原因	1 调查当天请假	31

变量名	变量标签	取值标签	观测值数量
		2 艺体特长训练比赛	1
		3 转入本校其他班级	8
		4 本校留级	3
		5 已转校	589
		6 外校寄读	1
		7 读技校	3
		8 休学	4
		9 长期治病	14
		10 退学	121
		11 开除	2
		12 死亡	2
		13 不详	51
w2clsra	七年级升八年级是	0 否	8897
	否重新分班	1 是	1853
w2tchstatus	教师追访情况	1 成功追访	504
		2 新入样	287

## （二）数据合并匹配

### （1）2014-2015 学年不同层次追访数据的合并匹配

因为追访数据库的班级数据是班主任问卷变量和任课教师问卷变量纵向合并起来的，所以变量 w2clsids 在数据库中存在重复值，使用班级数据库与其他三个数据库进行合并时，需将班主任问卷观测值、语文老师问卷观测值、数学老师问卷观测值以及英语老师问卷观测值从班级数据库中分别保留出来再进行一一合并匹配。

具体的合并匹配方法，如图 2：

- 1) W2 学校数据和其他三个数据库（W2 班级数据、W2 家庭数据、W2 学生数据）进行合并时，用唯一的学校代码 schids；

2) W2 班级数据和 W2 家庭数据以及 W2 学生数据进行合并时, 用唯一的班级代码 w2clsids;

3) W2 家庭数据和 W2 学生数据进行合并时, 用唯一的学生个人代码 ids。

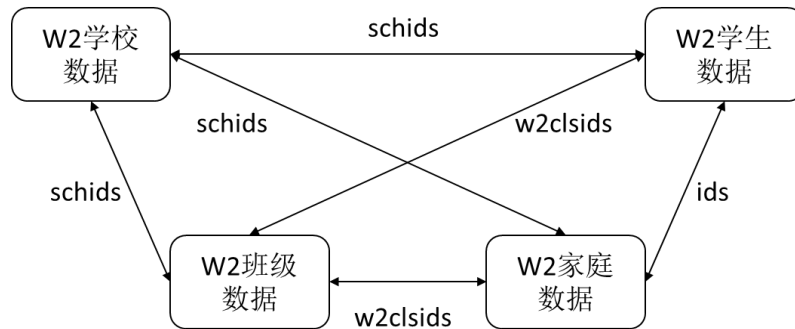


图 2 2014-2015 学年追访数据合并流程图

## (2) 不同学年相同层次数据的合并匹配

通过 ids 系列变量对基线数据和追访数据进行横向匹配, 具体的合并匹配方法:

1) 对不同学年的学校数据进行合并时, 用唯一的学校代码 schids;

2) 对不同学年的班级数据进行合并时, 先将班主任问卷观测值、语文老师问卷观测值、数学老师问卷观测值以及英语老师问卷观测值从班级数据库中分别保留出来, 用追访数据库的第二轮班级代码 w2clsids 和基线数据库的班级代码 clsids 进行一一对应匹配;

3) 对不同学年的家庭数据进行合并时, 用唯一的个人学生代码 ids;

4) 对不同学年的学生数据进行合并时, 用唯一的个人学生代码 ids。