

中国教育追踪调查 (CEPS) 基线调查 认知能力测试心理计量报告

王卫东 李佩华

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

2015 年 10 月

一、研究概况

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,并收集学生的重要考试(期中考试、中考、高考等)成绩,并计划组织对学生的健康与体格检查,采集生物学指标,综合利用各种手段和技术全面采集高质量的数据。

二、测验的编制原则、结构与过程

（一）编制原则

认知能力测试不同于学业成绩测量，在测量编制的内容与形式上，遵循以下基本原则：

1. 测量学生的基本逻辑思维与问题解决能力，而非某种需要死记硬背的知识；
2. 测试题的内容不涉及学校课程所教授的具体识记性知识，不牵涉某些专门的知识领域，是课程无关的；
3. 测验中所提及的问题和场景都是该年龄段青少年非常熟悉的一般性问题，而不会因问题所涉信息或场景的特殊性对被试造成干扰；
4. 绝大多数被试在规定的时间内都能回答完所有测验题；
5. 测验题的采取选择题的形式，每道测验题有 4 个选项，只有一个选项为正确答案；
6. 测验题应生动活泼、由易入难，充分调动被试的答题兴趣与意愿。

（二）测验结构

测验题的编制在测量结构上主要参考了台湾教育长期追踪资料库（TEPS）的基本结构^[1]，从语言、图形与空间、计算与逻辑三个维度上对初中生的认知能力进行测试。测试题共分 3 个维度、11 个类型，具体包括：

（1）语言：

1. 词组类比
2. 语言文字推理

（2）图形与空间：

1. 图形规律分析
2. 折纸类题目
3. 几何图形应用

（3）计算与逻辑

1. 数学应用
2. 自定义运算规则

3. 数列应用
4. 抽象规律分析
5. 概率
6. 数值大小逆向思维

(三) 编制过程

根据设计的原则与基本结构，项目组组成专门的出题小组，在上海、郑州两城市各两所中学进行了试测验，经历 3 轮修改，最后确定了七年级组与九年级组的认知能力测试题，其中七年级组测验 20 个问题，九年级组 22 个问题，答题限时均为 15 分钟。采用在课堂上集中填答的方式测验。

七年级组与九年级组的各题项对应的题目类型见表 1：

表 1：各题号对应测验题类型

七年级		九年级	
题号	题目类型	题号	题目类型
1	图形规律分析	1	图形规律分析
2	图形规律分析	2	图形规律分析
3	语言文字推理	3	数学应用
4	语言文字推理	4	自定义运算规则
5	数学应用	5	自定义运算规则
6	数学应用	6	抽象规律分析
7	词组类比	7	数学应用
8	自定义运算规则	8	词组类比
9	自定义运算规则	9	词组类比
10	数值大小逆向思维	10	数学应用
11	概率	11	数学应用
12	折纸类题目	12	抽象规律分析
13	数学应用	13	数学应用
14	数学应用	14	数列应用
15	数学应用	15	数学应用
16	数列应用	16	数学应用
17	数学应用	17	几何图形应用
18	几何图形应用	18	抽象规律分析
19	抽象规律分析	19	抽象规律分析
20	抽象规律分析	20	数学应用
		21	概率
		22	数学应用

三、古典测验理论指标

（一）信度

以克隆巴赫（信度）系数（Cronbach's alpha）来计量测验的信度，七年级组的 Cronbach alpha 系数为 0.6892，九年级组的 Cronbach alpha 系数为 0.7215。

（二）原始总分分布情况

七年级组的原始总分最小值 0，最大值为 20；均值 3.45，标准差 10.73；九年级组的原始总分最小值 0，最大值为 22；均值 3.89，标准差 9.06；原始总分的分布见图 1。

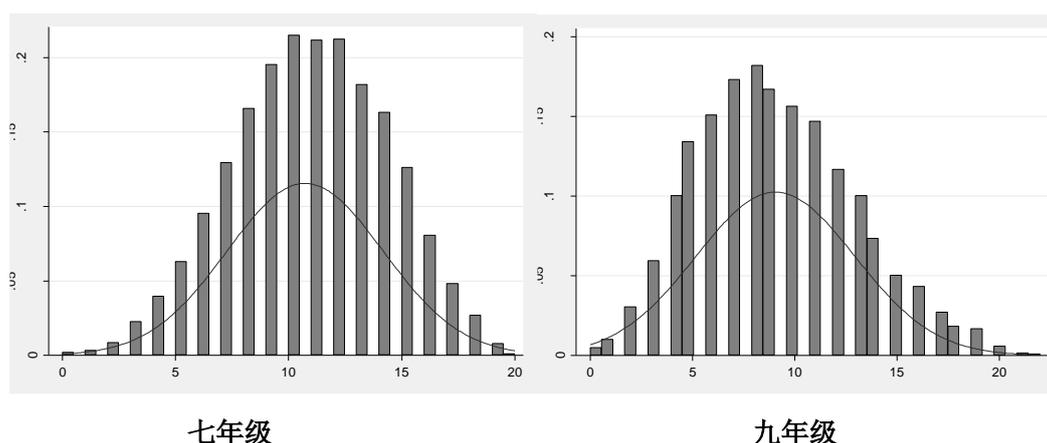


图 1：分年级组原始总分分布

（三）难度与区分度

用各题项的正确率来测量测验题的难度，用各题项与原始总分的点二列相关系数（Point Biserial Correlation）来测量各题项的区分度，七年级与九年级组各题项对应的难度与区分度见表 2。

表 2：各题项的正确率与点二列相关系数

七年级			九年级		
题号	正确率	点二列相关系数	题号	正确率	点二列相关系数
1	61.77%	0.29634926	1	65.80%	0.303807
2	56.71%	0.33701868	2	69.24%	0.287855
3	81.36%	0.21375449	3	71.66%	0.355673
4	81.04%	0.15458508	4	37.90%	0.366114
5	59.38%	0.40758734	5	70.23%	0.451286
6	31.49%	0.22819253	6	51.53%	0.278867
7	24.21%	0.28281031	7	46.99%	0.359986
8	47.37%	0.48590449	8	34.98%	0.408394
9	55.91%	0.52362211	9	30.77%	0.410081
10	69.79%	0.43325444	10	47.91%	0.331912
11	72.94%	0.39230242	11	43.48%	0.285468
12	84.45%	0.39833443	12	43.44%	0.43332
13	81.75%	0.50717417	13	39.44%	0.459531
14	50.63%	0.4127494	14	53.46%	0.563843
15	48.21%	0.52260786	15	34.36%	0.388369
16	58.08%	0.58635686	16	35.41%	0.463698
17	39.59%	0.47301629	17	28.64%	0.402202
18	23.81%	0.26432407	18	26.22%	0.466329
19	19.77%	0.3522887	19	20.31%	0.35688
20	24.37%	0.26680325	20	21.05%	0.377412
			21	20.86%	0.332167
			22	12.23%	0.335677

四、测验的项目反应理论（IRT）分析

不同于古典理论，IRT 根据受测者回答问题的情况，通过对题目特征函数的运算，来推测受测者的能力。IRT 模型的参数有：难度（difficulty index）、区分度（discriminative powder index）和猜测系数（guessing index）。根据参数的不同，特征函数可分为单参数模型（难度）、双参数模型（难度、区分度）和三参数模型（难度、区分度、猜测参数）等^[2]。

在模型中,各符号分别代表:

θ : 受测者能力估计值;

- a: 题目的区分度，它的值越大说明题目对受测者的区分程度越高；
- b: 题目的难度；
- c: 题目的猜测系数，它的值越大，说明不论受测者能力高低，都容易猜对；
- $P(\theta)$: 能力为 θ 的人答对此题目的概率

如果将模型的形态设置为 Logistic 模型，参数的数量不同，则分别得到 1PL、2PL、3PL 模型。

(一) 1PL 模型

1PL 模型只估计难度的一个参数，基于 1PL 计算出来的七年级与九年级组各题项对应的题目难度见表 3:

表 3: 1PL 模型参数

七年级		九年级	
题号	难度	题号	难度
1	-0.69682	1	-0.94899
2	-0.39224	2	-1.17377
3	-2.11674	3	-1.33881
4	-2.08723	4	0.711198
5	-0.55196	5	-1.24056
6	1.127858	6	-0.09276
7	1.647676	7	0.170744
8	0.154221	8	0.893269
9	-0.34507	9	1.168089
10	-1.21333	10	0.117063
11	-1.43452	11	0.376084
12	-2.42223	12	0.378646
13	-2.153	13	0.617203
14	-0.03579	14	-0.2051
15	0.105388	15	0.932641
16	-0.47381	16	0.865817
17	0.615697	17	1.314116
18	1.679458	18	1.487821
19	2.014726	19	1.9583
20	1.635657	20	1.894909
		21	1.910612
		22	2.797225

基于 1PL 得出来的七年级与九年级组的测验特征曲线见图 2:

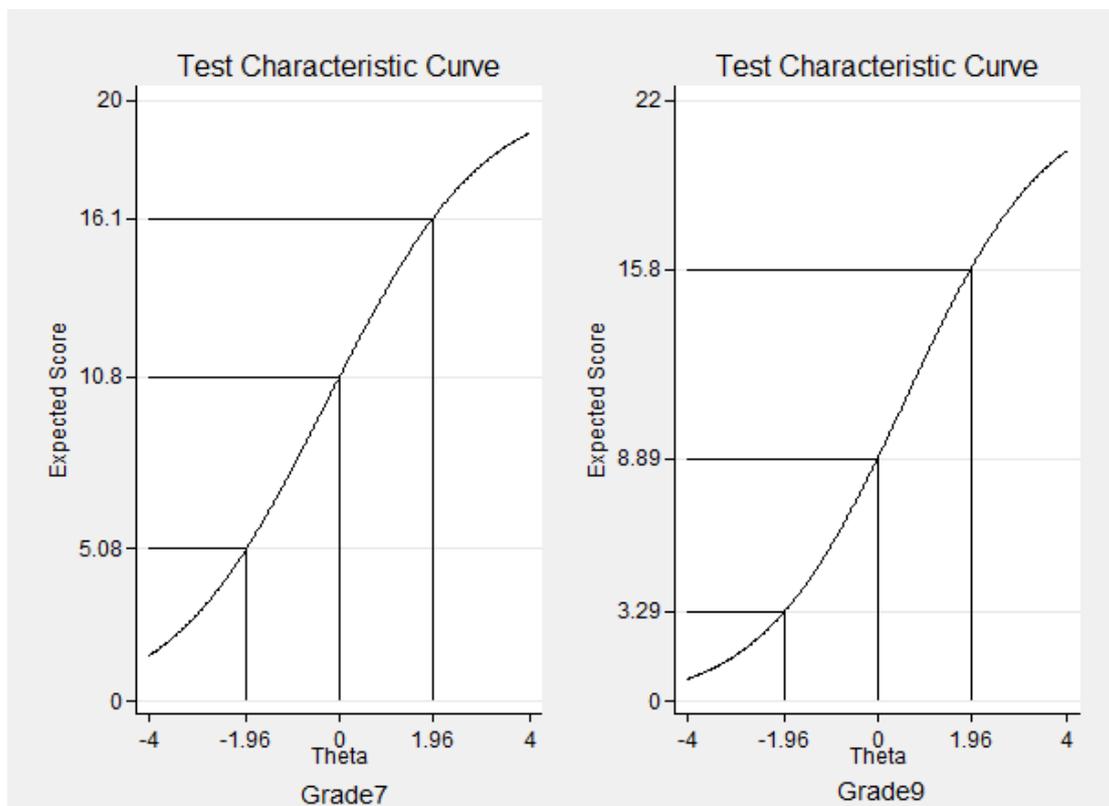


图 2：1PL 模型测验特征曲线

(二) 2PL 模型

2PL 模型估计难度与区分度两个参数，基于 2PL 计算出来的七年级与九年级组各题项对应的题目难度与区分度见表 4。从表 4 可以看出，利用 2PL 计算出来的难度与区分度与表 2 利用古典理论得出的难度与区分度的大小顺序是高度一致的。

表 4：2PL 模型参数

七年级			九年级		
题号	难度	区分度	题号	难度	区分度
1	-1.35167	0.366329	1	-1.62632	0.418906
2	-0.62018	0.457501	2	-2.06669	0.407677
3	-4.94862	0.303526	3	-1.58438	0.636942
4	-10.842	0.13449	4	0.903114	0.589291
5	-0.59284	0.716017	5	-1.04055	0.987639
6	3.324079	0.236895	6	-0.19323	0.325802

七年级			九年级		
7	2.520997	0.474653	7	0.243036	0.526126
8	0.118677	1.0528	8	0.974422	0.704748
9	-0.25224	1.252699	9	1.232244	0.733587
10	-1.0463	0.94825	10	0.1774	0.495691
11	-1.31828	0.869005	11	0.718764	0.376927
12	-1.71285	1.25269	12	0.35416	0.859449
13	-1.17916	2.058689	13	0.489176	1.076575
14	-0.03838	0.802337	14	-0.12004	1.896878
15	0.06283	1.4317	15	0.895246	0.826815
16	-0.27265	2.05578	16	0.614048	1.27567
17	0.452744	1.176232	17	1.095446	0.997536
18	2.719713	0.446211	18	1.00079	1.387312
19	1.909155	0.833337	19	1.701236	0.939593
20	2.684589	0.439618	20	1.56564	1.006955
			21	1.78067	0.856365
			22	2.251208	1.038564

基于 2PL 得出来的七年级与九年级组的测验特征曲线见图 3:

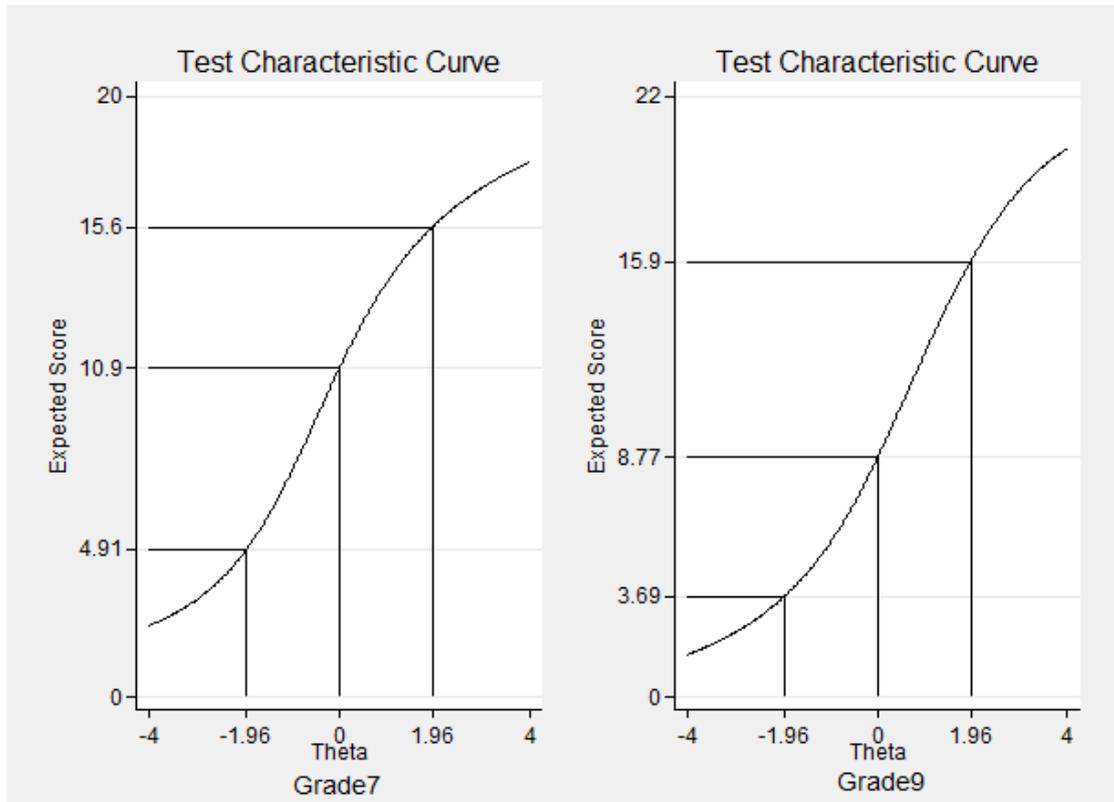


图 3: 2PL 模型测验特征曲线

(三) 3PL 模型

3PL 模型除了估计难度与区分度的两个参数外，还加上了对答题者不看内容而猜测答案这个参数的估计。由于 3PL 的参数众多，对模型参数进行估计时经常难以收敛，所以基于表 4 中 2PL 模型参数中的区分度，依次去掉区分度过小的题项进行估计，直至得到收敛的结果。最终，七年级组去掉了第 3、4、6 题，九年级组去掉了第 6、11 题。目基于 3PL 计算出来的七年级与九年级组各题项对应的题目难度、区分度以及猜测系数见表 5。由从表 5 可以看出，七年级与九年级组的猜测系数都很小，这说明完全基于猜测对被试得到正确答案概率提升的贡献是很小的。

表 5: 3PL 模型参数

七年级				九年级			
题号	难度	区分度	猜测	题号	难度	区分度	猜测
1	-0.86603	0.390882	0.087494	1	-1.43294	0.431686	0.03916
2	-0.21782	0.495177		2	-1.845	0.425234	
5	-0.31216	0.793699		3	-1.43551	0.664734	
7	2.524089	0.686253		4	1.030498	0.638315	
8	0.325089	1.26812		5	-0.92733	1.052449	
9	-0.05998	1.458692		7	0.401503	0.555581	
10	-0.85159	0.994851		8	1.084032	0.76834	
11	-1.1188	0.9029		9	1.331433	0.807445	
12	-1.59877	1.268491		10	0.352951	0.508029	
13	-1.07292	2.18453		12	0.465497	0.906184	
14	0.214053	0.891226		13	0.586952	1.143454	
15	0.23759	1.746712		14	-0.04743	2.029005	
16	-0.11884	2.53162		15	1.003295	0.891305	
17	0.629049	1.477846		16	0.692651	1.423485	
18	2.663288	0.661669		17	1.167014	1.128959	
19	1.745291	1.563785		18	1.029577	1.753127	
20	2.514116	0.689873		19	1.671244	1.1827	
				20	1.541128	1.281932	
				21	1.753072	1.062701	
				22	2.00573	1.618309	

基于 3PL 得出来的七年级与九年级组的测验特征曲线见图 4:

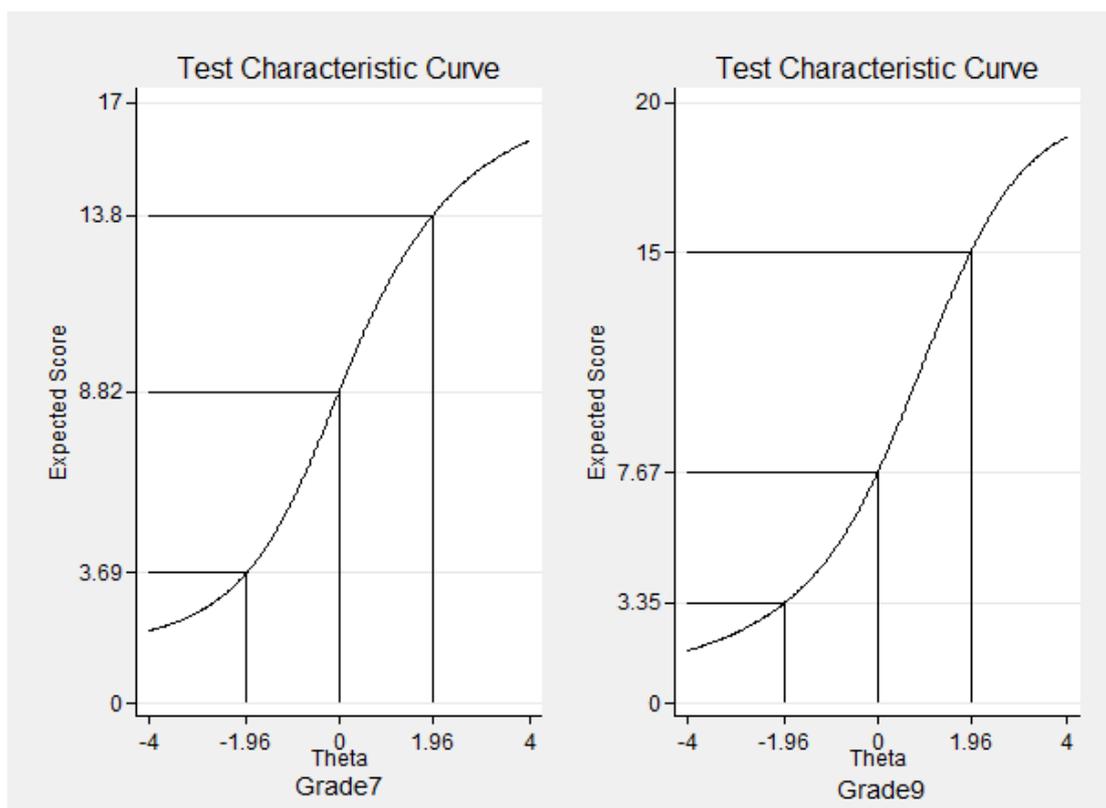


图 4：3PL 模型测验特征曲线

基于 EAP 法分别对 1PL、2PL、3PL 模型下的被试者能力值 θ 进行估计。计算这 3 个得分与原始总分的 Pearson 相关矩阵，从表 6 中可以看出，这四个分数在七年级组和九年级组中的相关系数都是非常大的。尽管 3PL 得分在两个年级组分别消减了几个题目项，但与 1PL 得分的相关系数都在 0.95 左右。并且，从图 3 中可以看出，3PL 得分的正态性非常好，而不同于图 1 所示的原始总分的分布那样陡峭。所以，在正式发布的基线调查数据中，对于认知能力测试只提供了原始总分和消减后的 3PL 得分这两个变量。

表 6：原始总分、1PL、2PL、3PL 模型得分相关矩阵

	七年级			九年级		
	原始总分	1PL 分	2PL 分	原始总分	1PL 分	2PL 分
1PL 分	0.9997			0.9991		
2PL 分	0.9574	0.9575		0.9662	0.967	
3PL 分	0.9459	0.9458	0.9966	0.9559	0.9561	0.9951

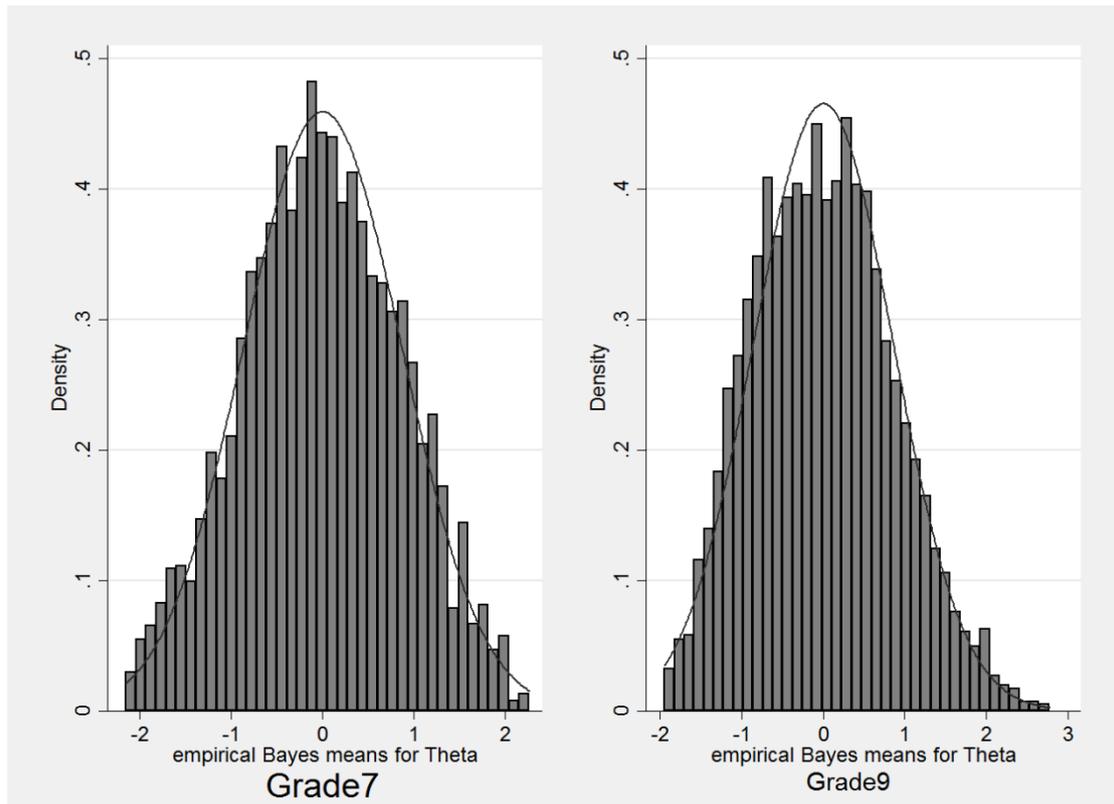


图 3: 3PL 能力得分估计值的分布

参考文献:

[1] Yang, M. L., Tam, T., & Huang, M. H. (2003). Psychometric Report for the Ability Tests of TEPS 2001. Taipei: Center for Survey Research, Academia Sinica.

[2] Baker, F. B., and S. Kim. 2004. Item Response Theory: Parameter Estimation Techniques. 2nd ed. New York: Dekker