

# 我国大学生的能力水平与高等教育增值

——基于“2016 全国本科生能力测评”的分析

沈 红, 张青根<sup>①</sup>

(华中科技大学 教育科学研究院, 湖北 武汉 430074)

**摘 要:** 本项目于2016年12月对16省83所高校的1.6万名本科生进行了批判性思维能力、创造力、人际交往能力、问题解决能力的水平测量。结果显示:(1)我国大学生能力水平随院校群体的学术声望而变化,且工科生的创造力和问题解决能力较强。(2)我国高等教育对本科生的能力水平产生了增值效应但其效应量有限;分组后的院校群体的学术声望越高,并不会对其学生能力增值产生的效应越大。(3)高校对文科生的批判性思维和人际交往能力、对理科生的创造力和问题解决能力的增值效应最大。

**关键词:** 高等教育增值; 批判性思维能力; 创造力; 人际交往能力; 问题解决能力

**中图分类号:** G525.5 **文献标志码:** A **文章编号:** 1000-4203(2017)11-0070-09

## The Level of Students' Capacities and Higher Education Value-Added in China

——Based on the “2016 National Assessment of Collegiate Capacity”

SHEN Hong, ZHANG Qing-gen

(School of Education, Huazhong University of Science & Technology, Wuhan 430074, China)

**Abstract:** With the sampling of 16,000 students across 83 higher education institutions (HEIs) in sixteen provinces in China in December 2016, it measured four general capacities, namely critical thinking, creativity, interpersonal communication and problem-solving. It found that, these general capacities were changing along with institution's reputation after grouping and engineering students took on strong capacities in creativity and problem-solving. Chinese higher education exerted certain value-added effect on undergraduates' capacities, though the effect size was limited. After grouping, institution's better reputation didn't guarantee the enhancement of its effect on students' value-added among capacities. Comparatively, HEIs exerted the largest value-added effect on students of liberal arts as to critical thinking and interpersonal communication capacities, and the largest value-added effect on students of sciences as to creativity and problem-solving capacities.

**Key words:** higher education value-added; critical thinking; creativity; interpersonal communication; problem-solving

<sup>①</sup> 收稿日期:2017-10-08

基金项目:国家自然科学基金面上项目(71673097)

作者简介:沈红(1956—),女,湖北武汉人,华中科技大学教育科学研究院教授,博士生导师,从事高等教育经济、财政与管理研究;张青根(1990—),男,江西上饶人,华中科技大学教育科学研究院博士后研究人员,从事教育经济学、高等教育管理研究。

## 一、问题的提出

教育给人以知识,也提高人的能力。大学生的年龄一般在 18—22 岁,其知识与能力的现状水平是通过自然增长、社会影响与教育作用的叠加形成的。在日益强调高等教育质量、回归学生本位、回应社会问责的今天,我国大学生的知识和能力水平如何,高等学校在提高学生的知识和能力水平上发挥了怎样的作用,成为社会各界关注的重要议题。

对知识和能力及其增长的测量主要有 3 种方式。一是“以投入测产出”,即以测量学生的学习投入程度来估计其学习结果。前提假设是,学生在学习上投入的时间和精力越多,其学习成果就越多。代表性的研究有:美国印第安纳大学的全美大学生学习投入调查 NSSE(National Survey of Student Engagement)<sup>[1]</sup>,澳大利亚大学生学习投入调查 AUSSE(Australasian Survey of Student Engagement)<sup>[2]</sup>,以及清华大学的中国本科教育学情调查 NSSE-China<sup>[3]</sup>等。然而,在知识与能力的获取过程中,个人的天赋、各类知识的吸收程度、各种能力习得的方式以及学习者和教学者之间的互动程度等变量,都置于投入与产出的“过程之箱”中,使测量得到的学习成果可能只是一种猜测,且“猜测与真实间的距离”受到这种测量的细致程度、变量选取的科学程度、被测人员的诚实程度等因素的影响。二是自我测量,由被测者“扪心自问”来评判其知识或能力的增长及增长程度,如英国的全国学生调查 NSS(National Student Survey)<sup>[4]</sup>、澳大利亚的学生课程体验问卷 CEQ(Course Experience Questionnaire)<sup>[5]</sup>和北京大学的全国高校学生调查<sup>[6]</sup>,都要求被试对自己的能力提高状况进行自评。这种测量来自个人感性,结果容易被“美化”或“自黑”。三是出题考试,由测试方出题来考查被测者的知识或能力的现实状况及增长幅度。代表性测量有:美国教育考试服务中心的 EPP(ETS Proficiency Profile)<sup>[7]</sup>,美国大学考试中心的学生学业水平测试 CAAP(Collegiate Assessment of Academic Proficiency)<sup>[8]</sup>,美国教育资政理事会的大学学习测试 CLA(Collegiate Learning Assessment)<sup>[9]</sup>以及澳大利亚教育研究委员会的毕业生技能评估 GSA(Graduates Skills Assessment)<sup>[10]</sup>等。这种测量来自客观判断,结论生硬,没有解释机会,结果的科学性和准确度易受试题质量和测试环境的影响。上述方式都是专门对知识和能力的“即期状态”或意欲求出“两次或多次即期

状态之差”所做的测量。尽管它们存在局限,但都能在一定程度上反映大学生的能力水平。若将知识与能力分开测量,大学生的知识增长则以来自教学和“第二课堂”的学习过程为主,且以学业成绩作为测量的替代结果不具备院校可比性。若要比不同院校或不同学科的学生培养质量的差异,可通过学生的能力水平测试来实现。

本研究综合应用上述 3 种测量方法,对我国四年制本科院校中的本科生的能力进行了测评,并试图探究高等教育对增长大学生能力的效用。总体上看,大学生的能力可划分为通用能力和专用能力。通用能力指可在不同工作、组织和情境下转移且不损失有效性的能力,专用能力指只适用于特定场合的,在不同工作、组织或情境下难以转移的能力。本文主要研究以批判性思维能力、创造力、人际交往能力、问题解决能力等为代表的通用能力。本文研究的主要问题有:第一,我国大学生的批判性思维能力、创造力、人际交往能力、问题解决能力的现状水平和校间差异如何?若把四年制高校分组为“985 工程”大学、“211 工程”大学、四年制大学、四年制学院 4 组,把 13 个学科门类“群聚”为文、理、工、医 4 类,学生的 4 项能力水平是否具有高校组间差异和学科偏向?第二,我国高等院校是否对本科生的批判性思维能力、创造力、人际交往能力、问题解决能力的提高做出了贡献?也就是说,学生接受高等教育的经历对其能力增值产生了多大的效应?学生能力增值是否存在院校差异、院校组间差异、学科差异?第三,面对大学生的 4 项能力现状水平及高等教育增值状况,我国高等学校可以采取哪些改进行动?

## 二、研究样本与测评工具

### 1. 研究样本

全国本科生能力测评(National Assessment of Collegiate Capacity, NACC)对全国 812 所四年制本科院校(截至 2016 年 5 月教育部数据)进行了 10% 抽样,实际测评地处东、中、西部 16 个行政省<sup>①</sup>的 83 所高校,包含“985 工程”大学、“211 工程”大学、四年制大学、四年制学院 4 个层次。为降低样本的同质性,要求每所抽样高校仅 200 名本科生被试,其中大一新生和大四学生各占 50%,文、理、工、医科(含生物和生命科学)学生各占 25%。经过在 6 所高校进行的 3 次试测与工具调整后,2016 年 12 月经专门培训的 8 支测评工作队分赴全国 16 省 83 所高校,

分场次集中被试学生,现场发放纸质测评材料。超过16000名本科生接受了测评。经对测评试卷和调查

问卷的匹配整理、标准化判分、数据输入与清洗,有效样本为15336份。具体样本分布如表1所示。

表1 NACC样本高校和学生分布

	高校层次				高校地域			学科				年级*					总数
	“985工程” 大学	“211工程” 大学	四年制 大学	四年制 学院	东部	中部	西部	文科	理科	工科	医科	大一	大二	大三	大四	大五	
高校 (所)	13	15	36	19	31	29	23	81	65	77	43	83	31	37	78	3	83
学生 (人)	2218	2869	6760	3489	5578	5562	4196	5476	3179	5414	1267	8245	495	1118	5457	21	15336

注: \*表示实际测试和原始设计在抽样学生的年级上有差异;测试中的非大一、非大四学生计入样本总数,但不进入学生能力增值计算中。

## 2. 测评工具

NACC测评工具由来自哲学、逻辑学、心理学、教育学、语言学等学科的学者组成的国际化团队研制开发和对经典量表进行删减和调试而成。本研究中的批判性思维测试共33道客观题,集中被试在50分钟内独立闭卷完成;创造力、人际交往能力、问题解决能力测试用的是倾向性量表,给出李克特七等选项,不限定回答时间<sup>②</sup>。在测试4项能力的同时对被试者发放问卷,包括年龄、性别、民族等人口统计学背景数据;生源地、户口类型、家庭收入、父母教育和职业等家庭经济背景数据;高中学校与学业状况、高考分数、高校和专业录取与志愿相符程度、流动与留守求学经历、高校学费额度与学生资助需求等教育背景数据和对毕业后求职、求学、创业意愿等信息;对非大一新生的问卷特别增加了在高校

就读期间的课程学习、科研参与、社团活动、实习经历、创业尝试等学生个体对教育和社会活动投入的相关信息,并要求每一个被试都自评其本人的4项能力状况。

表2呈现了测评工具的信效度检验结果。在信度上,4套测评工具的克伦巴赫系数均大于0.6,说明其信度良好。在多维效度上,若以被试对其4项能力的自评分数为效标,4套测评工具的效标效度均呈高度显著正相关( $P < 0.001$ );因绝大部分项目载荷在0.3以上,说明测试创造力、人际交往能力、问题解决能力的3套倾向性量表的会聚效度良好;若以简约适配指数为结构效度的判断指标,则3套力量表的PGFI和PNFI值均大于0.5,表明其结构效度满足测评需要。

表2 NACC测评工具的信效度

	信度	效标效度	会聚效度	结构效度—简约适配指数	
				PGFI值	PNFI值
批判性思维能力	0.623	0.094***	—	—	—
创造力	0.851	0.412***	绝大部分项目载荷>0.4	0.747	0.674
人际交往能力	0.773	0.429***	绝大部分项目载荷>0.3	0.707	0.661
问题解决能力	0.908	0.303***	绝大部分项目载荷>0.4	0.736	0.735

注:表2中使用克伦巴赫(Cronbach) $\alpha$ 系数进行信度判定,当系数大于0.6时便可判定为信度良好;\*\*\*表示 $P < 0.001$ ;“—”表示该指标未如此检验或不宜如此检验。绝大部分项目载荷高于0.3表明会聚效度良好<sup>[11]</sup>;简约适配指数>0.5,表明模型的拟合度可达到观测水平,所用量表的结构效度满足测量需要。

## 三、我国大学生的能力现状及组间差异

客观试题测评和倾向性量表测评的判分标准是不同的。为可比较,将上述4项能力得分统一进行百分制转换,使各单项能力满分均为100。从总体

样本4项能力得分的描述性统计均值来看,创造力得69.5分,问题解决能力得67.4分,人际交往能力得66.4分,批判性思维能力得54.2分。4项能力的标准差、最低分和最高分见表3。T检验发现4项能力均分中两两能力之间存在差异且都为高度显著( $P$ 值均小于0.001)。

表 3 NACC 测得的学生能力的描述统计分析

	有效样本数	得分均值	标准差	最低分	最高分
创造力	13211	69.5	10.1	18.1	100
问题解决能力	13306	67.4	10.8	17.9	100
人际交往能力	13070	66.4	9.1	30.9	100
批判性思维	15189	54.2	12.7	6.1	93.9

1. 院校差异：学生能力随院校群体的学术声望而变化

若将参评的 83 所院校分成“985 工程”大学、“211 工程”大学、四年制大学和四年制学院 4 组，相应的被测学生数分别是 2218 人、2869 人、6760 人和 3489 人。可从院校个体层面和分组的群体层面来

分析大学生能力测评结果的院校差异。

图 1 显示，被测高校抽样学生在创造力、人际交往能力、问题解决能力的倾向性测量均分上的院校差异不大，3 项能力的各校得分区间都在 63—73 分之间，并且不随高校所在层次的变化而变化。但经客观性测试的批判性思维能力得分均值的院校差异明显，从最低的 40 分（一所四年制大学）到最高的 68 分（一所“985 工程”大学）。尽管方差分析发现，在批判性思维能力、创造力、人际交往能力、问题解决能力上均存在着高度显著的院校差异（F 值分别为 58.25、4.33、2.80 和 5.19，P 值均小于 0.001），但这些院校差异并无规律性。

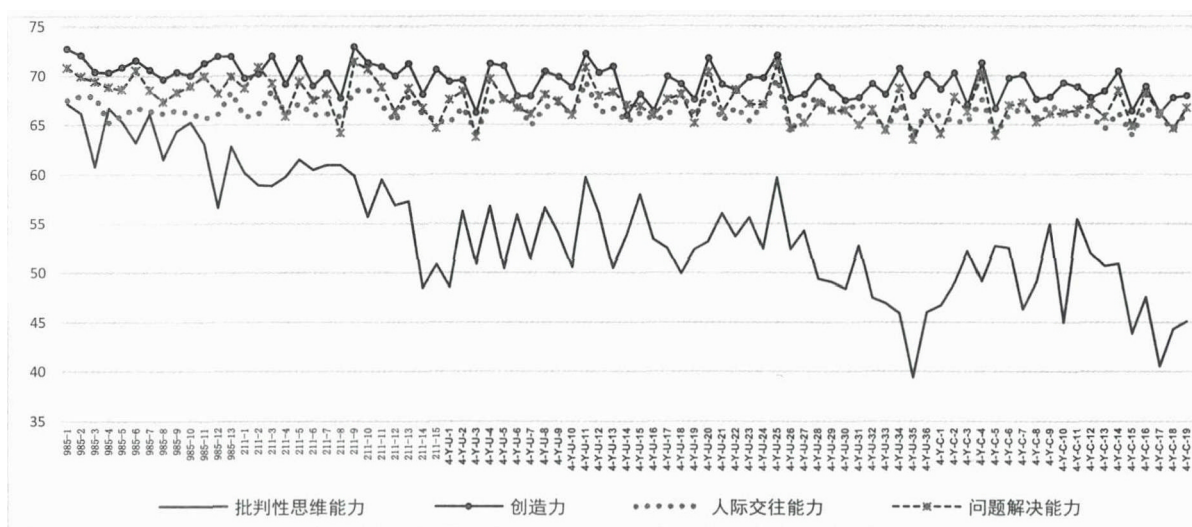


图 1 分院校的大学生 4 项能力测评得分均值

注：横轴为被测高校代号：从 985-1 到 985-13, 211-1 到 211-15, 4-Y-U-1 到 4-Y-U-36; 4-Y-C-1 到 4-Y-C-19; 省份顺序同注释①；全文同。

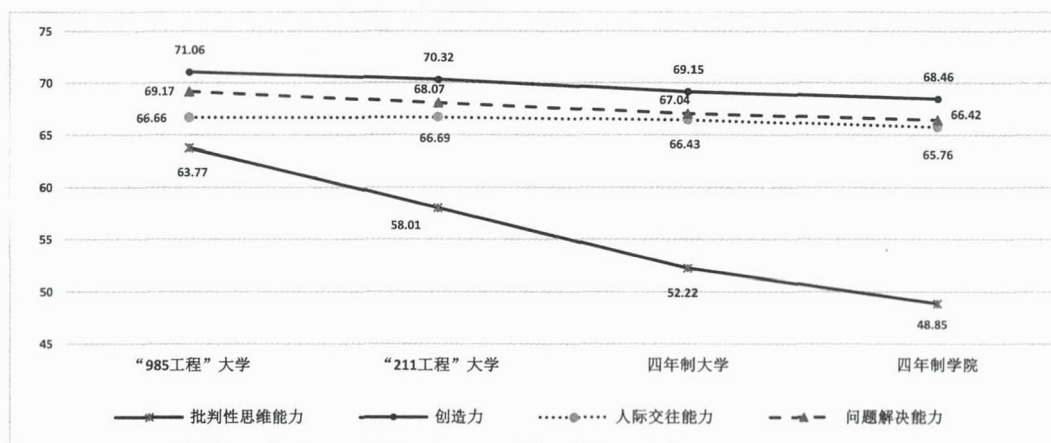


图 2 分院校层次的大学生 4 项能力测评得分均值

我们尝试计算分组后院校群体的学生的 4 项能力得分均值及组间差异。图 2 显示，在批判性思维能力上院校组间差异最大且依学术声望的高低顺序变化：“985 工程”大学、“211 工程”大学、四年制大

学、四年制学院的各组均值分别约为 64 分、58 分、52 分和 49 分，高低分相差 15 分。另外，虽然其他 3 项能力得分的院校组间差异也朝着同一方向——高校组的学术声望越高，学生能力平均得分越高——

但组间差异值没有批判性思维能力的组间差异那么大。用“985工程”大学组均分减去四年制学院组均分时,问题解决能力、创造力和人际交往能力的得分差分别是 2.8 分、2.6 分和 0.9 分。另经方差分析可知,4 项能力均存在高度显著的院校组间差异(F 值分别为 856.83、31.49、5.52 和 27.96,P 值均小于 0.001)。

2.学科差异:工科生的创造力和问题解决能力较强,医科生的批判性思维能力较强

表 4 显示抽样高校被测学生在 NACC 中得到的 4 项能力均分的学科差异。其中,医科生的批判性思维能力较为突出,文、理、工三科生之间的得分差异很小;工科生的创造力和问题解决能力得分均最高,文、理、医三科生的得分相差不大;理科生的人际交往能力得分最低,文、工、医三科生得分相近。

表 4 NACC 测评结果的学科差异 (单位:分)

	批判性思维能力	创造力	人际交往能力	问题解决能力
文科生	54.1	69.1	66.7	67.3
理科生	53.8	69.1	65.8	66.7
工科生	54.1	70.5	66.4	68.2
医科生	55.9	68.7	66.3	66.5

注:抽样高校并不都拥有 4 个学科大类。本研究以某学科或某年级有效样本数 12 作为进入数据分析的下限,如此,拥有文、理、工、医科生的测评院校分别有 80 所、58 所、76 所和 28 所。

方差分析发现,在批判性思维能力、创造力、人际交往能力和问题解决能力上均存在着高度显著的学科差异(F 值分别为 9.33、20.38、6.29 和 15.37,P 值均小于 0.001)。

3.年级差异:大四学生批判性思维和问题解决能力高于大一新生,但创造力的表现恰好相反

NACC 测评中的大一新生有 8245 人,大四学生有 5457 人。若用大四学生的 4 项能力得分分别减去大一新生的相应得分,图 3 的均分差可用来解释大四学生和大一新生能力的年级差异。批判性思维能力和问题解决能力的“大四大一均分差”均为正,说明分层次院校组的大四学生的这两项能力都高于大一新生的相应能力的水平,四年制大学组中大四学生与大一新生的批判性思维能力相差不大;创造力的“大四大一均分差”均为负,说明分层后各院校组中大四学生的创造力水平低于大一新生的相应水平;“211工程”大学、四年制大学和四年制学院组的大四学生的人际交往能力高于大一新生的相应能力。最后在“985工程”大学组中,大四学生与大一新生在能力得分上的均分差,与其他 3 组高校相比,在问题解决能力上,“985工程”大学组中大四学生比大一新生高出的程度最低;在创造力上,“985工程”大学组的大四学生与大一新生的得分相比,低得最多;在人际交往能力上,只有“985工程”大学组的大四学生比大一新生的得分要低。

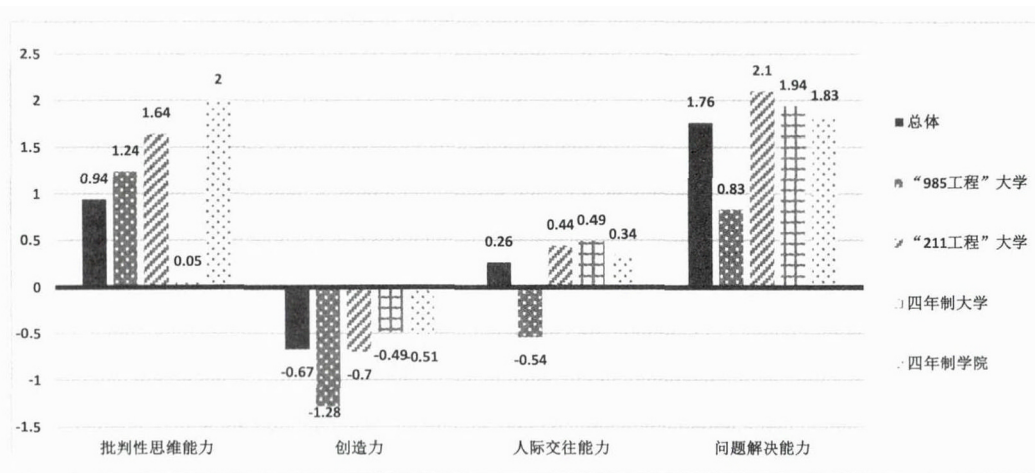


图 3 分院校层次的大学生 4 项能力“大四—大一”均分差

再基于全样本的方差分析发现,除人际交往能力的年级差异不显著外(F 值为 0.38,P 值大于 0.1),批判性思维能力、创造力、问题解决能力均存在显著的年级差异(F 值为 3.59、18.02 和 52.89,P 值均小于 0.1)。

本研究对学生的 4 项能力采用了不同的测评方

法,不宜对其得分结果进行横向比较,但可对测评方式相同的 3 项能力的测评结果作出观察性评价:就我国本科生群体而言,其创造力水平要高于问题解决能力水平,且这两种能力水平又都高于人际交往能力的表现。总体上看,学生能力水平在院校个体差异上不具有规律性,但学科差异和年级差异明显。

## 四、我国大学生的能力增值及组间差异

增值研究需要用同一人或同一批人在高年级时的测评结果减去其在低年级时的测评结果。但在NACC的测评对象中,大四学生与大一新生并非同一批人。两批人的成长环境、教育经历、生理和心理成熟度都不相同,不可对非同一批人的测评结果进行简单相减来计算增值。本文尝试借用量化研究常用的残差差异分析法进行增值研究。

### 1. 残差差异分析与计量步骤

残差(residual)是指实际观测值与回归估计值之间的差。在具体方法上,我们借鉴他人已有的相关研究成果<sup>[12-14]</sup>,在数据分析上,我们分5个步骤测算高校对大学生4项能力增值产生的效应及其院校差异。

第一步,求模型的拟合系数。基于大一新生(freshman)和大四学生(senior)个体样本数据拟合模型(1)和(2),分别得出两个模型中相应变量的拟合系数 $\beta$ 值,

$$y_{ij, fre} = \beta_{0, fre} + \beta_{1, fre} \vec{X}_{ij, fre} + \epsilon_{ij, fre} \quad (1)$$

$$y_{kj, sen} = \beta_{0, sen} + \beta_{1, sen} \vec{X}_{kj, sen} + \epsilon_{kj, sen} \quad (2)$$

式中, $j$ 为院校代号; $y_{ij, fre}$ 和 $y_{kj, sen}$ 分别为院校 $j$ 的大一新生 $i$ 和大四学生 $k$ 的某项能力的实际得分; $\vec{X}_{ij, fre}$ 和 $\vec{X}_{kj, sen}$ 分别为院校 $j$ 的大一新生 $i$ 和大四学生 $k$ 的个人特征变量(包含已测得的人口统计学背景、家庭经济背景、进高校前的教育背景等)。

第二步,求样本学生个体的回归估计得分。根据从模型(1)和(2)中求出的拟合系数 $\beta$ 值,分别估计院校 $j$ 的大一新生 $i$ 和大四学生 $k$ 的某项能力得分的预测值: $y_{ij, fre}^{\wedge} = \beta_{0, fre}^{\wedge} + \beta_{1, fre}^{\wedge} \vec{X}_{ij, fre}^{\wedge}$ ,  $y_{kj, sen}^{\wedge} = \beta_{0, sen}^{\wedge} + \beta_{1, sen}^{\wedge} \vec{X}_{kj, sen}^{\wedge}$ 。这里, $\beta_{0, fre}^{\wedge}$ 、 $\beta_{1, fre}^{\wedge}$ 、 $\beta_{0, sen}^{\wedge}$ 和 $\beta_{1, sen}^{\wedge}$ 分别为模型(1)和(2)中 $\beta_{0, fre}$ 、 $\beta_{1, fre}$ 、 $\beta_{0, sen}$ 和 $\beta_{1, sen}$ 的估计值。

第三步,求残差,即实际观测值与回归估计值之间的差。从模型(1)和(2)看,高校 $j$ 的大一新生 $i$ 和大四学生 $k$ 的该项能力残差 $\epsilon_{ij, fre} = y_{ij, fre} - y_{ij, fre}^{\wedge}$ ,  $\epsilon_{kj, sen} = y_{kj, sen} - y_{kj, sen}^{\wedge}$ 。残差 $\epsilon_{ij, fre}$ 和 $\epsilon_{kj, sen}$ 的技术含义是指未纳入模型(1)和(2)中的其他变量对该项能力带来的影响以及可能存在的系统性偏差。

第四步,求某高校对其学生的某项能力所产生的增值(Value-Added)效应。用院校 $j$ 的大四学生

的某项能力的残差均值减去其新生的残差均值,得到的是消除其他未进入模型的变量所产生的影响和系统性偏差后的某高校对其学生的某项能力产生影响的净值,净值的大小和方向表示院校 $j$ 提供的教育对该校学生样本群体产生的平均增值效应( $\overline{VA}_j$ ),即

$$\overline{VA}_j = \overline{\epsilon_{j, sen}} - \overline{\epsilon_{j, fre}} = \frac{1}{m} \sum_{k=1}^m \epsilon_{kj, sen} - \frac{1}{n} \sum_{i=1}^n \epsilon_{ij, fre}$$

式中的 $m$ 和 $n$ 分别为高校 $j$ 的大四学生和大一新生样本数。

第五步,求经标准化处理的某高校对其学生的某项能力所产生的增值效应量(effect size)大小。由于高校被试在样本容量上的差异会影响增值计算结果的院校可比性,因此需要进行样本容量的标准化处理,其转换公式为:

$$D_j = \frac{\overline{VA}_j}{\sigma_j} = \frac{\overline{VA}_j}{\sqrt{((m-1)\sigma_{sen}^2 + (n-1)\sigma_{fre}^2)/(m+n-2)}}$$

其中, $\sigma_j$ 为院校 $j$ 大一新生和大四学生混合样本的某项能力残差的标准差, $\sigma_{sen}^2$ 和 $\sigma_{fre}^2$ 分别为院校 $j$ 的大四学生和大一新生在该项能力上的残差的方差。

### 2. 绝大多数院校对学生的能力增值具有正向效应

基于以上步骤,计算满足最小样本数要求(见表4后注释)的72所被测院校<sup>③</sup>对其抽样学生4项能力所产生的增值效应量。结果显示,高校对本科生批判性思维能力产生的增值效应量最大,对学生问题解决能力、创造力、人际交往能力产生的增值效应量依次变小。在这些院校中,90%的院校提高了其抽样学生的批判性思维能力,82%的院校提高了其抽样学生的问题解决能力,81%的院校提高了其抽样学生的创造力以及54%的院校提高了其抽样学生的人际交往能力。院校对学生能力增值产生效应的具体数据是:批判性思维能力增值平均效应量为0.3,最高与最低增值效应量分别为0.84(一所“985工程”大学)和-0.79(一所四年制大学);问题解决能力增值平均效应量是0.15,最高与最低增值效应量分别为0.66和-0.71(均在“985工程”大学组中);创造力增值平均效应量为0.09,最高与最低增值效应量分别为0.53和-0.65(均在四年制大学组中);人际交往能力增值平均效应量仅0.05,最高与最低增值效应量分别为0.54(一所四年制学院)和-0.47(一所“985工程”大学),如图4所示。

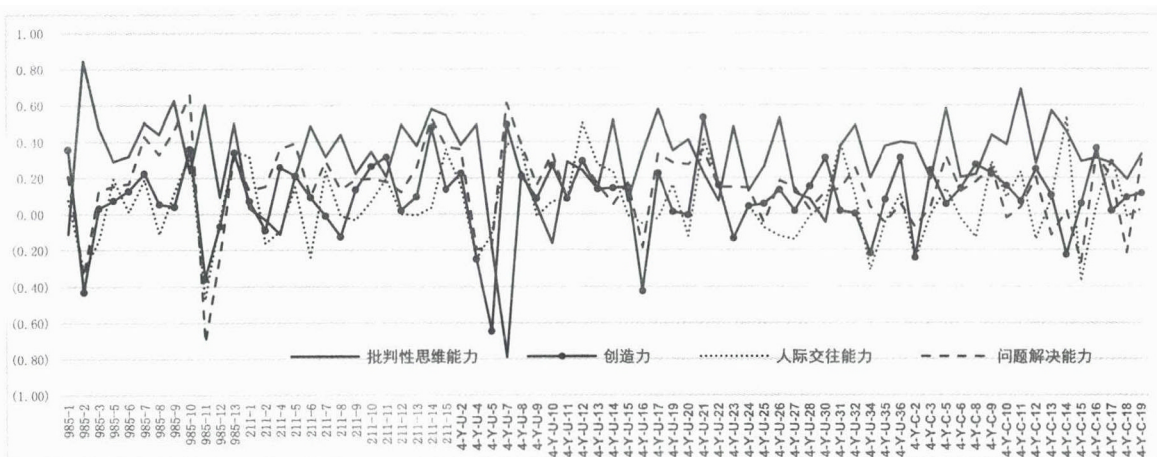


图4 高校对其学生4项能力增值产生的效应量

残差差异分析的结果表明：在被测高校对学生4项能力增值效应量计算的288个数据量中，77%的数据量为正向增值，23%为负向增值；单项能力增值主体为正向，80%以上的被测高校对其学生3项能力增值产生正效应，过半数的高校对其学生的人际交往能力增值有正向作用；单项能力的平均增值效应量由大到小依次是：批判性思维能力、问题解决能力、创造力和人际交往能力。方差检验显示，整体上，被测高校对学生批判性思维能力、创造能力、人际交往能力、问题解决能力的增值效应量存在高度显著差异(F值分别为229.07、23.42、11.95、58.28，P值均小于0.001)。

### 3. 增值效应量具有组间差异

已知大多数被测院校对其学生的4项能力增值都具有正向作用，那么，这种高校对学生能力增值产生的效应量是否具有院校个体差异、分层后的院校群体差异以及学科差异呢？计算结果显示：“985工程”大学对学生的批判性思维能力增值最大但对其创造力和人际交往能力的增值最小，四年制学院对学生的批判性思维和创造能力的增值效应值得肯定；一般来说，高校对文科生的批判性思维能力和人际交往能力的增值效应量最大，对理科生的创造力和问题解决能力的增值效应量最大。

(1) 高校对学生能力增值效应量的院校个体差异。对图1和图4进行综合比较后发现，被测学生的4项能力现状水平、被测院校对其学生4项能力增值所产生的效应量，都不具有院校个体在差异表现上的规律性。也就是说，从学生能力增值的角度来判断，并不是学术声望越高、公共投入越多、教师和学生筛选越严的高校，对其学生本科期间的能力增值产生的效应就越大。

(2) 高校对学生能力增值效应量的院校组差异。

图5显示，高校对学生批判性思维能力增值效应量的院校组均值由高到低依次是：“985工程”大学组、四年制学院组、“211工程”大学组、四年制大学组。

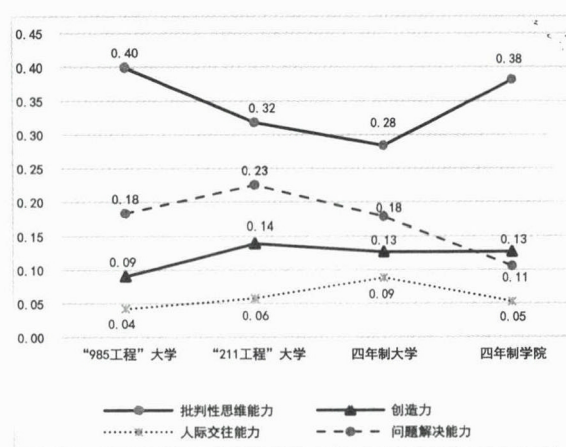


图5 高校对学生能力增值效应的院校组差异

在高校对学生创造力增值的效应量上，“211工程”大学组最大，“985工程”大学组最小，四年制大学组和学院组居于中间且两组产生的效应量相等；高校对学生人际交往能力增值的效应量，四年制大学组最高，“985工程”大学组最低，“211工程”大学组和四年制学院组居于中间；对学生问题解决能力增值的效应量，“211工程”大学组最大，四年制学院组最低，“985工程”大学组和四年制大学组居于中间且两组的效应量相等。图5清楚地说明，并不是所处学术层次越高的院校，对其学生能力的增值效应量就越大。该结果的形成具有多种原因，如不同层次高校的学生努力程度存在差异等，学生能力提高的“边际效应递减”也会产生作用，即能力初始分相对较高和以高分升学的学生，在后续提高中的增值空间相对较小。<sup>[15]</sup>

(3) 高校对学生能力增值效应量的学科差异。

图6显示,高校对学生批判性思维能力和人际交往能力产生效应的学科变化趋势相似,大小顺序是:文科生>理科生>工科生>医科生。其中,高校对医科生人际交往能力的增值效应量为负值(-0.05)。

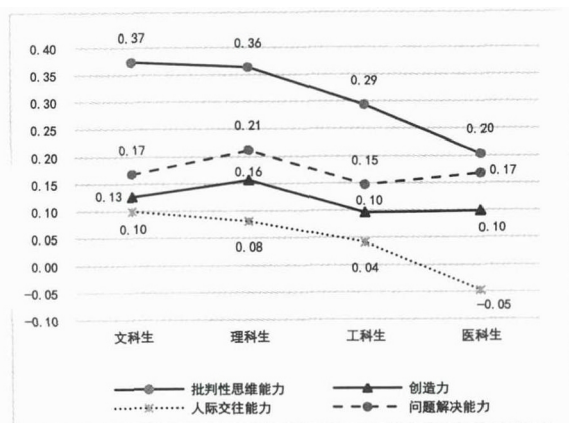


图6 高校对学生能力增值效应的学科差异

高校对不同学科的学生创造力增值产生的效应量存在差异,大小顺序是:理科生>文科生>工科生=医科生。高校对学生问题解决能力增值产生的效应量也存在学科差异,大小顺序是:理科生>文科生=医科生>工科生。

## 五、结论与建议

### 1. 结论

第一,NACC的测评结果可信。来自东、中、西部16省4个层次院校组的83所高校的16000名本科生参加了测评,院校样本和学生样本具有中国本科院校的代表性;经检测,4项能力测评工具均具有较高的信效度;用描述统计分析法研究大学生的能力水平和用残差差异分析法研究高校对大学生的能力增值效应,在研究方法上具有科学性和合理性。

第二,我国大学生4项能力水平测试得分在54—70分之间且存在组间差异。方差分析表明,4项能力水平的院校差异显著。院校组的学术声望越高,该院校组的学生4项能力的得分均值越高。

第三,我国绝大多数高校对其学生4项能力增值具有正向效应且存在组间差异。高校在对本科生的批判性思维能力、问题解决能力、创造力、人际交往能力上产生的增值效应量依次变小。值得关注的是,“985工程”大学组对学生的批判性思维能力增值产生的效应是最大的,但对其创造力和人际交往能力产生的增值效应却最小;四年制学院组对学生的批判性思维能力和创造能力的增值效应都值得肯定。也就是说,与学生能力的现状水平不同的是,并

不是学术声望越高的高校组对其学生4项能力增值所起的作用越大。分学科来看,被测院校整体上对文科生的批判性思维能力和人际交往能力的增值效应量最大,对理科生的创造力和问题解决能力的增值效应量最大。

NACC用测评数据证明我国高校对在校生的4项通用能力增值具有促进作用,增值效应量由大到小依次为:批判性思维能力、问题解决能力、创造力、人际交往能力。

### 2. 建议

本研究发现,若分“上、中上、中下、下”四档计,我国大学生的4项能力水平都居中上段(54—70分之间);54%、81%、82%和90%的被测高校分别对其学生的人际交往能力、创造力、问题解决能力、批判性思维能力产生正向增值效应;平均而言,4项能力的高等教育增值效应量在0.05—0.3之间;高校对学生能力增值产生的效应量大小与其学术声望、资源投入、新生筛选等要素并无明确关联。所以说,我国高校对学生能力增值产生了效应,但效应量有限,部分学术声望高的高校比部分学术声望低的学院更具有改进的空间。

高校的主要职能是教学、科研和服务,高校人的主体构成为学生、教师和行政人员。如何使“三职能”和“三群人”在提高本科生能力增值上做到“无缝对接”?在履行科研和服务职能的过程中,教师为绝对的主导,本科生能在教师指导下参与,行政永远是提供管理、协调与服务。本科生参加科研首先是得到近距离观察科研活动的机会,并对什么是科研、怎么做科研、自己是否合适从事科研和合适于什么样的科研等问题得出自己的答案;然后才是在真正的科研实践中提高自己的能力。高校的社会服务职能建立在高等教育产出的基础上,如思想观点的产出、各类人才的产出、科技成果的产出等。这些产出的社会转化结果是大众创业、万众创新,形成新的生产力,推动社会进步。18—22岁的本科生在高校的社会服务中能够做的事情是尽早地了解社会,了解自己,了解书本和课堂知识与社会实践的结合点,进行反思性思维,得到创造的动力和解决实际问题的经历,在与各种人的交往中磨炼成长。在高校课堂内外教学中,虽然教师是教方,学生是学方,但在知识的分化、综合与更新速度快捷,提供知识和信息的数据网络遍地,学生的能动性、自主性和个性化得以彰显的今天,高校教师与学生的主导与参与角色需要适时互换。教师主导在于课程的计划性、科学的前沿性、知识的结构性、为人的经验性上,学生主导在

于信息的适时性、工具的便捷性、年龄的可塑性、发展的需要性上;教师参与在于学习如何聆听、如何引导、如何服务得更好,学生参与在于学习为学之道、做事之道、为人之道。教学行政管理则体现出更为浓厚的计划、协调、服务等方面的职能。这样的角色关系使师生在互动中成长,从认识学术、认识社会中来反思学术、反思社会,从而激发创造潜力,在与人与社会的交往中提高解决问题的能力。

高等学校是分层次分类型的,大学和学院的定位和目标有所不同,这就带来了“三职能”之间的不同任务比重和不同目标层次的分配。无论是为科学发现还是为地区繁荣,也无论是研究型大学还是教学型学院,只要教师与行政管理人员时刻保持着一颗为学生成长服务的心,就能利用高校的资源,通过学校的各项活动来促进学生能力的增值。要知道,不是所有高校都能成为“单一流”或“双一流”的,但所有高校都可以在提高学生能力上有所作为。

(截至目前,本项目团队成员包括:沈红、张青根、李文平、罗靳雯、林桢栋、胡仲勋、汪洋、梁春晓、张冰冰、于海琴、余荔、张和平、马騫、王小菲、刘之远、贾永堂、易元祥、刘燕、马萍、徐志平、李爱萍、杨小雨、周珪、黎梦菲、刘欢、罗蕴丰、吴永源、俞光祥、熊俊峰、李玉栋、商守卫、王鹏。)

#### 注释:

- ① 东部6省份包括北京、河北、上海、江苏、山东、广东;中部5省包括黑龙江、江西、河南、湖北、湖南;西部5省份包括四川、云南、陕西、甘肃、新疆。
- ② “批判性思维测试题”由加籍华裔学者董毓博士领衔的批判性思维研究专家团队开发与调试而成;“创造力倾向量表”是对《威廉斯创造力倾向量表》汉化版的删减和调试(参见林幸台,王木荣.《威廉斯创造力测验》,台北:心理出版社有限公司,1996);“人际交往能力量表”是对加利福尼亚《人际交往能力问卷》的删减和调试(参见BUHRMESTER D. et al. *Five domains of interpersonal competence in peer relationships*. *Journal of Personality & Social Psychology*, 1988, 55(6): 991-1008);“问题解决力量表”是在借鉴 PISA 中的相关定义及结构划分方法的基础上,综合参考多种已有研究的相关量表后研制和调试而成(参见陆璟. *PISA 测评的理论和实践*. 上海:华东师范大学出版社,2013)。
- ③ 由于部分学生样本在某些变量上存在缺失,导致计量分析时部分学校的学生样本流失,使得满足最小样本数要求的院校数减少。

#### 参考文献:

- [1] NSSE (National Survey of Student Engagement) [EB/OL]. (2017-08-20). <http://www.nsse.indiana.edu/>.
- [2] COATES H. Australasian Survey of Student Engagement: Institution Report 2008 [R]. (2012-06-13). <http://research.acer.edu.au/ausse/17>.
- [3] 史静寰,等. 基于学习过程的本科教育学情调查报告 2009 [J]. *清华大学教育研究*, 2011, 32(4): 9-23.
- [4] HEFCE (Higher Education Funding Council for England) [EB/OL]. (2017-08-02). <http://www.thestudentsurvey.com/>.
- [5] MCLNNIS C, GRIFFIN P, JAMES R, COATES H, DEPARTMENT of EDUCATION, TRAINING and YOUTH AFFAIRS. Development of the course experience questionnaire [R]. 2001:1.
- [6] 马莉萍,管清天. 院校层次与学生能力增值评价——基于全国 85 所高校学生调查的实证研究 [J]. *教育发展研究*, 2016, (1): 56-61.
- [7] ETS Proficiency Profile [EB/OL]. (2017-05-15). <http://www.ets.org/proficiencyprofile/about>.
- [8] ACT CAAP [EB/OL]. (2017-05-15). <http://www.act.org/content/act/en/products-and-services/act-collegiate-assessment-of-academic-proficiency.html>.
- [9] CLA+: Measuring Critical Thinking for Higher Education [EB/OL]. (2017-05-15). <http://cae.org/flagship-assessments-cla-cwra/cla/>.
- [10] HAMBUR S, ROWE K, LUC L T. Graduate Skills Assessment: Stage One Validity Study [R]. 2002.
- [11] 白利刚,凌文壮,方俐洛. 霍式中国职业兴趣量表构想效度的验证性因素分析(I):量表会聚效度和区分效度的验证性因素分析 [J]. *心理学报*, 1996, 28(1): 64-69.
- [12] STEEDLE J T. Selecting Value-added Models for Postsecondary Institutional Assessment [J]. *Assessment & Evaluation in Higher Education*, 2012, 37(6): 637-652.
- [13] KLEIN S, BENJAMIN R, SHAVELSON R, BOLUS R. The Collegiate Learning Assessment: Facts and Fantasies [J]. *Evaluation Review*, 2007, 31(5): 415-439.
- [14] LIU O L. Value-added Assessment in Higher Education: a Comparison of Two Methods [J]. *Higher Education*, 2011, 61(4): 445-461.
- [15] TEREZINI P T. Assessment with Open Eyes: Pitfalls in Studying Student Outcome [J]. *The Journal of Higher Education*, 1989, 60(6): 644-664.

(本文责任编辑 许宏)