

# 减负背景下的训练量与结果提升间的关系研究——以上海市嘉定区九年级英语区域质量监测为例

王超杰<sup>1</sup> 沈晨<sup>2</sup> (通讯作者) 梁辰<sup>3</sup>

1. 江苏大学, 江苏 镇江 212013

2. 上海市教育考试院, 上海 200433

3. 朗鹰教育科技, 上海 200040

**摘要:** 在“双减”政策背景下, 不能继续使用原有教学方法, 而要通过科学训练来提升成绩。本文以上海市嘉定区组织的两次九年级英语区域质量监测为例, 通过数据统计分析, 研究了减负背景下的训练量与结果提升间的关系, 以及互联网教学管理平台在减量提质, 减负增效方面的积极作用, 以期实现“双减”政策的有效落地。

**关键词:** 双减; 教育信息化; 训练量; 数据分析

**中图分类号:** G640 **文献标识码:** A

## 0 引言

2021年7月, 中共中央办公厅、国务院办公厅印发《关于进一步减轻义务教育阶段学生作业负担和校外培训负担的意见》(以下简称“双减”《意见》)。

“双减”《意见》指出“全面压减作业总量和时长, 减轻学生过重作业负担”, 同时提出要“提高作业设计质量。发挥作业诊断、巩固、学情分析等功能, 将作业设计纳入教研体系”。~~错误!未找到引用源。~~这给新时期的教学管理带来了挑战, 如何做到“减负不减质”是每个教育管理者必须面对的难题。为了推动“双减”政策精准落地, 笔者经过研究实践发现, 通过互联网智能教学管理平台, 可以从课堂内外或考前、考中、考后等多个环节对学生的学习效果进行训练、跟踪、反馈和评价。在考前模拟训练中, 学生不仅可提前熟悉考试流程, 在正式考试时克服紧张情绪, 同时平台能提供过程性评价数据, 起到反拨教学的正向作用(washback effect)。同时, 通过一定时间的科学训练, 学生能有效提升成绩, 但不宜过度训练, 而要在教学方法突破和创新。

## 1 研究现状

随着“双减”政策的落地, 人们开始研究如何减负增效、如何通过人工智能来减负提质。何俊(2022)<sup>[2]</sup>认为“双减”要达到的目标, 是减轻学生过重的课业负担。这里的“减”, 不是简单的“减少”, 而是有

目的地在减少学生课业负担的同时实现学生“减负不减能”, 主要是减少那些无效的重复作业, 而大数据可精准反馈哪些作业是无效的。杨晓哲和叶露(2022)<sup>[3]</sup>指出线上教学可采取嵌入式、伴随式评价, 有助于及时反馈学生学习情况, 基于学生学习数据的分析是实现大规模个性化教育的必经之路。刘涛等(2022)<sup>[4]</sup>研究发现教师一般只能在周测、月测或期中期末考试后才能发现问题, 而互联网教学管理平台能及时反馈作业情况, 大量的数据统计, 能让教师在“双减”背景下及时调整课堂教学目标和内容、课后作业内容和难度等。杨志明和吴仪(2021)<sup>[5]</sup>研究发现习题训练存在着边际效益递减的规律, 即在学习的初始阶段, 习题训练效益递增明显, 但增长到某个临界点之后, 习题训练效益就会递减。王开和汪基德(2022)<sup>[6]</sup>认为基于大数据的深度挖掘和解析能对现实学业负担状态进行测量评估, 并预判未来学业负担走向, 及时采取有效措施进行调控, 使学业负担稳定在合理区间, 还能实时监测学生的学业负担, 实现科学减负、精准减负。

## 2 数据集

本次分析的数据集来自上海朗鹰教育科技平台在上海市嘉定区组织的两次九年级英语听说区域质量监测的成绩数据, 共包含 5393 个学生样本, 以及 10 个解释变量:

学生 ID、学校、班级、姓名(为保护学校和学生隐私, 下表用星号\*和 XY 代号代替真实信息)。

第一次区域质量监测：AB 卷情况、分数、训练量 首选，选取所有在第一次质量监测和第二次质量

表 1 部分数据集

学生 ID	学校	班级	姓名	AB 卷 1	分数 1	训练量 1	AB 卷 2	分数 2	训练量 2
810***	X 中学	1	田**	A	25.5	0	A	28	0
810***	X 中学	1	杨**	B	22	0	B	25.5	0
810***	X 中学	1	王**	B	20.5	0	A	20	0
810***	X 中学	1	陈**	B	24	0	A	24.5	0
810***	X 中学	1	张**	B	21	0	B	24.5	0
810***	X 中学	1	徐**	A	24.25	0	A	25	0
810***	X 中学	1	潘**	A	23.75	0	B	24	0
810***	X 中学	1	金**	A	20.5	0	A	25	0
810***	X 中学	1	郑**	A	13.5	0	A	10.5	0
810***	X 中学	1	黄**	A	23.5	0	B	29	0
810***	X 中学	1	袁**	B	24.25	0	A	29.5	0
810***	X 中学	1	俞**	A	23	348	A	27	348
810***	X 中学	1	陆**	B	25	0	A	28	0
811***	Y 中学	1	黄**	A	22	236	B	22	294
811***	Y 中学	1	施**	B	18.5	1990	B	22.5	2622
811***	Y 中学	1	蔡**	A	17.5	678	A	22.5	794
811***	Y 中学	1	殷**	A	24.25	378	B	29	378
811***	Y 中学	1	张**	B	20.5	0	A	17	0
811***	Y 中学	1	顾**	A	27.75	974	B	27	1094
811***	Y 中学	1	朱**	B	20.25	966	A	22	1376
811***	Y 中学	1	陈**	A	24.75	1024	B	29.5	1024
811***	Y 中学	1	刘**	B	25.5	0	B	26.5	0
811***	Y 中学	1	李**	B	26	462	B	29	462

第二次区域质量监测：AB 卷情况、分数、训练量

由于两次区域质量监测采用 AB 卷模式，因此无法直接对学生的成绩进行比较，需要对 A 卷和 B 卷分别进行处理。两次质量监测 AB 卷学生数统计如下：

表 2 AB 卷人数统计

测试一	人数	测试二	人数
A	2672	A	2683
B	2721	B	2710
合计	5393	合计	5393

由于两次质量监测的难度、评分标准、总分值等因素不完全相同，甚至在同一次质量监测中 AB 卷也会存在差异，因此作为绝对指标的成绩并不能很好地衡量学生的学习效果。

为了避免这种差异的影响，平台采用作为相对指标的标准分来评价学生的学习效果：

第 i 名学生标准分 = 
$$\frac{\text{第 i 名学生成绩} - \text{试卷平均分}}{\text{试卷标准差}}$$

例如：ID 为 810\*\*\* 的学生，在第一次质量监测中使用 A 卷，成绩为 19.0 分，而第一次质量监测 A 卷的平均分为 23.3 分，标准差为 4.6 分，则其第一次质量监测的成绩换算成标准分为：

$$\frac{19 - 23.3}{4.6} = -0.93$$

3 数据分析

下面将通过统计学数据分析的方法证明，在采用平台的模拟训练后，学生能够显著提高学习成绩。

监测之间的时间段内进行过模拟训练的学生（即训练量 2>训练量 1），共计 843 个学生样本。该组样本标准分统计数据如下：

表 3 843 个进行过模拟训练的学生两次质量监测的标准分统计

	质量监测一	质量监测二
平均值	-0.3046	-0.2182
标准差	1.0984	1.1486

考虑对这 843 个样本两次质量监测的成绩进行配对 t 检验 (Pairwise t-test)：

令  $x_i$  表示第 i 个学生第一次质量监测的百分制分位数， $y_i$  表示第 i 个学生第二次质量监测的百分制分位数，其中  $i = 1, \dots, 843$ 。

假设  $z_i = y_i - x_i$  服从正态分布  $N(\mu, \sigma^2)$ 。此时，有原假设为  $H_0: \mu = 0$ ，即两次质量监测成绩没有明显变化。同时构造检验统计量：

$$t = \frac{\sqrt{n}(\bar{z} - 0)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2}} \sim t(n-1)$$

其中， $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ ， $n = 843$ 。利用观测到的 843 个学生样本代入进行计算，得到  $t = 4.0005$ ，对应的 p 值 (p-value) 为  $6.876 \times 10^{-5}$ ，远小于显著性水平  $\alpha = 0.05$ ，因此拒绝原假设，说明两次质量监测成绩有显著性变化。同时第二次质量监测的平均成绩要高于第一次质量监测，说明在经过模拟训练后，学生的成绩明显提高了。

另外,本次研究还分析了训练量对提高程度的影响。

令  $X_i$  表示第  $i$  个学生在第一次质量监测和第二次质量监测之间的新增训练量,  $Y_i$  表示第  $i$  个学生在第二次质量监测比第一次质量监测标准分的增加值,拟合以下回归方程:

$$Y = \beta_0 + \beta_1 \ln(X) + \epsilon$$

其中  $\epsilon \sim N(0, \sigma^2)$ 。利用 843 个学生样本数据拟合结果为:

$$Y = -0.1747 + 0.0358 \ln(X)$$

拟合的回归方程如下图(横轴单位是分钟):

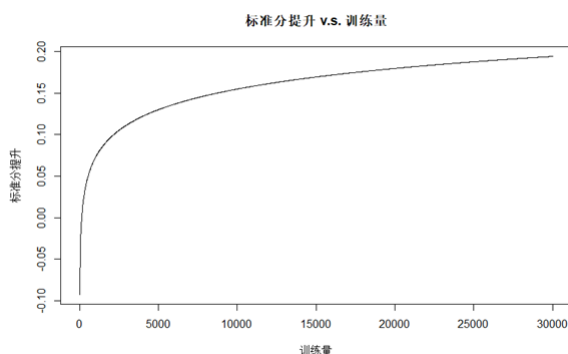


图1 拟合的回归方程

从图中可以清楚地看出,随着模拟训练量的增大,标准分提升的效果越好。

回归方程进行显著性检验:对训练量作为解释变量的显著性进行  $t$  检验,其对应的  $p$  值为 0.0365,小于显著性水平  $\alpha = 0.05$ ,说明解释变量显著。对回归方程整体进行  $F$  检验,其对应的  $p$  值为 0.0365,小于显著性水平  $\alpha = 0.05$ ,说明回归方程整体是显著可靠的。

由此我们可以得出结论:学生模拟训练量越大,对其成绩提升的效果越好。

同时,通过图 1,我们也可以发现,在减负背景下,仍然需要一定时间的积累和巩固所学内容。图 3 中有两个临界值,一个是 2500 分钟,另一个是 5000 分钟。可以看出:在 0 到 2500 分钟的时间段内,每多投入 1 分钟时间成本的训练量,标准分提高得越显著,边际效用递增(increasing marginal utility)明显;

在 2500 到 5000 分钟的时间段内,每多投入 1 分钟时间成本的训练量,仍有进步,但边际效用递增幅度低于 0 到 2500 分钟;

在高于 5000 分钟的时间段,每多投入 1 分钟时间成本的训练量,仍有进步,但边际效用递增幅度低于 2500 到 5000 分钟。虽未至于边际效用递减,但出现类

似高原效应(plateau effect)的问题。

教育部在 2021 年 4 月印发的《关于加强义务教育学校作业管理的通知》中明确规定:初中每天书面作业完成时间平均不超过 90 分钟。<sup>[7]</sup>在一个周期内,比如一个学年或一个学期,按照每天最多 90 分钟的各学科作业总时间计算,2500 分钟大致相当于 28 天的训练量,5000 分钟大致相当于 56 天的训练量。所以,在有限时间内,当训练量达到一定程度后,沿用既有的教学训练方式作用就不大了,单纯的练习解决不了学习认知提升了,需要在教学本身的方法和训练手段、结构上创新和突破,减量提质,减负增效,遵循教育规律,避免机械、无效、重复性的训练,避免给学生过重的负担。这时,省市区级的各类质量监测就是一种及时的诊断、反馈、巩固机制,可提供多维度的过程性客观数据,同时注重综合评价和增值评价,使评价、提升、改进融为一体,为教学反拨和决策提供可信依据。

《义务教育英语课程标准(2022 年版)》里提到要“坚持以评促学、以评促教,将评价贯穿英语课程教与学的全过程。推动‘教-学-评’一体化设计与实施。‘评’主要发挥监控教与学过程和效果的作用,为促教、促学提供参考和依据。要注重三者相互依存、相互影响、相互促进,发挥协同育人功能。”<sup>[8]</sup>

另一方面,作为对比,我们也研究了在两次区域监测之间的时间段内,未进行模拟训练的学生的成绩表现。在数据集中,未进行训练的学生样本数为 4550 个,其标准分统计数据如下:

表 4 4550 个未进行训练的学生两次质量监测的标准分统计

	质量监测一	质量监测二
平均值	0.0564	0.0404
标准差	0.9702	0.9645

同样,考虑对这 4550 个样本两次质量监测的成绩进行配对  $t$  检验,得到  $t = -1.8755$ ,对应的  $p$  值( $p$ -value)为 0.0607,高于显著性水平  $\alpha = 0.05$ 。由此可以得出结论:未进行模拟训练的学生成绩没有得到显著性提高,甚至有所下降。

## 4 评估分析

通过以上实例表明,经过模拟训练后,学生的表现更好。此外,平台对学生学业情况的多维度、全方位分析,以及各层级机构间、不同机构分组间、不同学生群体间的对比分析,便于教育主管部门和学校快

速掌握学生学业情况。

	3/4	8.25
--	-----	------

表 5 区域质量监测的试题难度和区分度

监测内容		题型	分值	考试要求			预估 难度值	实际 难度值	区分度
题号	知识点/功能/话题			a	b	c			
1	连读	朗读	0.5	√			0.90	0.91	0.33
2	不完全爆破	朗读	0.5	√			0.90	0.89	0.31
3	词块朗读	朗读	0.5	√			0.90	0.90	0.33
4	语调	朗读	0.5	√			0.90	0.88	0.29
5	意群和停顿	朗读	0.5	√			0.90	0.87	0.30
6	请求允许	交际应答	0.5		√		0.80	0.85	0.27
7	提议	交际应答	0.5		√		0.80	0.85	0.25
8	道歉	交际应答	0.5		√		0.80	0.87	0.25
9	赞扬	交际应答	0.5		√		0.75	0.78	0.28
10	警告	交际应答	0.5		√		0.75	0.73	0.30
11	学校生活	复述	2.5			√	0.75	0.76	0.29
12	日常活动	表达	2.5			√	0.70	0.69	0.30

表 7 区域质量监测的交际应答考生主要回答样例占比

试题	考生主要回答样例	评语	考生占比数		
			本校	本区	外区 A
Sorry, but I've broken your cup.	That's all right.	对	48.0%	44.6%	53.7%
	Never mind.	对	17.9%	21.5%	19.0%
	It doesn't matter.	对	13.7%	14.3%	15.1%
	Not at all.	错	12.5%	13.1%	10.3%
	Zat's all right.	发音不准	7.9%	6.5%	1.9%

表 8 区域质量监测的不同学生群体成绩比较

性别	学生来源	人数	均分	朗读	交际应答	复述	表达
男		2778	8.9	2.5	2.2	2.1	2.2
女		2615	9.3	2.5	2.2	2.3	2.3
	本市	2562	9.4	2.5	2.3	2.3	2.4
	非本市	1954	8.7	2.4	2.1	2.0	2.2
	居住证 120 分	837	9.5	2.5	2.3	2.4	2.4
	港澳台外	29	9.5	2.5	2.3	2.4	2.4
	随班就读	11	5.6	1.9	1.3	1.2	1.2
总体数据		5393	9.1	2.5	2.2	2.2	2.3

目前，平台已经能够以题目、题型、功能、话题、考试要求等进行多维度分析，并能对区分度、难度、信度、效度、峰度、偏度等指标进行全方位分析。

平台可以对试题的知识点、功能、话题、难度、区分度进行即时统计汇总，便于后效观测和质量分析。同时，通过对质量监测结果的自动描述性统计，可以便捷获取一次质量监测结果的全方位数据，并进行纵向和横向比较分析。如下表。

表 6 区域质量监测的结果描述性统计

有效人数		5393
平均分（满分 10 分）		7.18
最高分		10
最低分		0
优良率		44.7%
合格率		80.2%
标准差		1.91
全距		10
中位数		7.75
众数		8.25
偏度		1.27
峰度		1.44
四分位	1/4	6
	2/4	7.25

另外，学校和教师可通过平台提供的考生主要回答样例占比来观测在本校考生在某道题上的表现，可以和本区均值和外区均值进行对比，看准差距，了解短板，为教学反拨提供精准的客观依据，包括知识技能问题和发音问题等。如下表。

此外，平台还可以对不同学生群体的成绩进行比较，便于促进教育公平和教育均衡化，帮助实现“教、评、管”一体化目标。如下表。

5 结语

通过以上研究发现，学生模拟训练量越大，对其成绩提升的效果越好。但也不能一味盲目训练，要在不同阶段采取不同的对策，因时制宜，科学训练，减量提质。当今社会正处于信息技术、人工智能技术飞速发展并与教育教学不断融合的时期，教育部门需要与时俱进，征集、开发优质的线上教育教学资源，利用好互联网教学管理平台，加强平台对学习过程中的数据收集、数据分析、数据诊断与数据反馈，从而更好地贯彻党的教育方针，确保“双减”政策落到实处。

### 参考文献

- [1]何俊.“双减”政策下中小学学校管理改革:理念转变与难点突破[J].当代教育理论与实践,2022(2):10.
- [2]杨晓哲,叶露.新技术支持下义务教育的学习环境与方式变革[J].全球教育展望,2022(5):67.
- [3]刘涛,张熹,赵顺.“双减”背景下人工智能赋能“有效作业”模式的探究[J].中国现代教育装备,2022(5):12.
- [4]杨志明,吴仪.“双减”背景下从边际效益递减规

律看习题训练[J].教育测量与评价,2021(9):64.

- [5]王开,汪基德.人工智能赋能课堂教学减负提质的机制、风险与应对[J].当代教育科学,2022(2):65.

**作者简介:**王超杰,男,江苏镇江人,博士,江苏大学数学科学学院副教授,研究方向:多元统计分析;梁辰,男,上海人,硕士,朗鹰教育科技,智能测评研究员。通讯作者:沈晨,男,上海人,硕士,上海市教育考试院网络信息中心主管,研究方向:教育考试信息化应用。

基金项目:江苏省“双创”博士基金。