

托福改革前后的考试分数量化数据对比

刘烁炀, 姚宇西

新东方国际教育培训事业部总部, 北京

收稿日期: 2024年4月1日; 录用日期: 2024年5月13日; 发布日期: 2024年5月23日

摘要

托福考试于2023年7月改革, 改革后考试题量与题型出现了变化。本文根据新东方某分中心托福学生数据, 针对改革前后的学员出分情况进行定量分析, 对比了考试前后学生分数的各项特征。根据学生数据对比发现, 托福考试改革之后的成绩变化并不明显。本文同时也给出了部分统计数据, 为更加清楚地了解托福考生的整体画像提供了信息。

关键词

托福, 托福考试改革

A Quantitative Comparison of Test Scores Before and After Adjustments to TOEFL iBT

Yuxi Yao, Shuoyang Liu

Global Education and Training Office, New Oriental Education & Technology Group, Beijing

Received: Apr. 1st, 2024; accepted: May 13th, 2024; published: May 23rd, 2024

Abstract

The TOEFL test made adjustments to its format in July, 2023, making changes to the number of test items in various sections, as well as to certain question types. This study, using test scores from students in New Oriental, focuses on a quantitative comparison and analysis of test scores from before and after the said adjustments were made. According to findings of this study, the effect of the adjustments on test score of students was not significant. This study also gives certain descriptive statistics to help give a better description of the TOEFL test takers in China.

Keywords

TOEFL, Enhanced TOEFL iBT

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Open Access

1. 研究背景

托福考试从 2023 年 7 月 26 日开始进行了改革[1], 考试从过去的 3.5 个小时左右缩短到 2 小时以内。为了可以在更短的时间内完成考试, 本次 ETS 选择取消了所有的加试环节, 并且将阅读的文章从 3 篇减少为 2 篇。与此同时, 过去需要 30 分钟完成的写作独立话题也将被新的 10 分钟长度的学术讨论题型所替代。

随着考试时长和题目数减少, 许多疑问也应运而生: 题目减少是否意味着容错率降低? 写作题型的改变会不会影响考生成绩? 考生分数数据分布是否会产生变化? 这些问题都是想要选择托福考试或正在备考托福考试的考生非常关注的。不过目前 ETS 没有给出相关数据, 而从以往经验来看, ETS 在一年一度的 Test and Data Summary 中给出的会是全年数据[2] [3] [4], 也就意味着今年晚些给出的 2023 年数据将无法区分考试前后; 而由于今年全球托福成绩一直在提升, 利用未来 2024 年的数据对比 2022 年的改革前数据会产生一些误差。

因此本研究希望通过新东方考生的数据帮助托福考生和从业者更好地了解改革前后托福考试的分数分布情况, 让考生们对于考试有更好的掌握。

2. 数据来源

本次研究将利用新东方某地大学生业务在过去一年多记录的学生考试数据, 对考试改革前后的考试成绩进行计算, 对比改革前后统计数据数值和统计学特征, 以判断本次托福考试改革对于考生的影响。

由于所有参与统计的学生都来自于单一城市并均为同一年龄段, 因此前后两组数据虽然数量不同(改革前数据较多), 但是其数据特征是有横向比较价值的。

3. 结论与分析

我们首先看一下考试改革前后四个单项的统计数据情况, 见表 1。

Table 1. A sample of modified POST

表 1. 调整后的语篇标签示例

单项	时间	样本量	平均数	标准差	第一四分位数	中位数	第三四分位数	四分位距
阅读	改革后	595	23.8	4.9	21	25	28	7
	改革前	922	23.8	4.9	21	25	28	7
听力	改革后	595	22.0	5.2	19	23	26	7
	改革前	922	21.9	5.5	18	23	26	8
口语	改革后	595	20.3	2.9	18	21	22	4
	改革前	922	20.3	2.9	18	21	22	4
写作	改革后	595	20.9	3.5	19	21	23	4
	改革前	922	21.0	3.6	19	21	23	4

可以看出, 四个单项在平均分数上几乎都没有变化, 均在 0.1~0.2 分上下, 而中位数则是四个单项改

革前后均没有任何的变化。数据离散度方面也相对比较相似, 其中相比于其他四个单项, 听力在改革后数据的离散度稍有降低。这一点在听力改革后四分位距缩小 1 分上也有所体现。

如果将两组数据的箱线图拿出来比较也可以得到比较有趣的结论, 见图 1。其中听力除了第一四分位数上升之外, 其像下图的下边缘也有所提升。由于本次考试听力改变较少, 只是取消了加试, 可以看出加试的取消对于听力分数不高的同学还是非常友好的, 由于加试疲劳度导致的一些问题应该是得到了缓解。

除了听力的离散度降低之外, 通过箱线图可以看出口语单项在改革之后其上边缘有所下降。此部分有可能是由于改革后数据在高分段较少, 使得范围内没有取到理论最大值导致, 与考试改革本身无关。

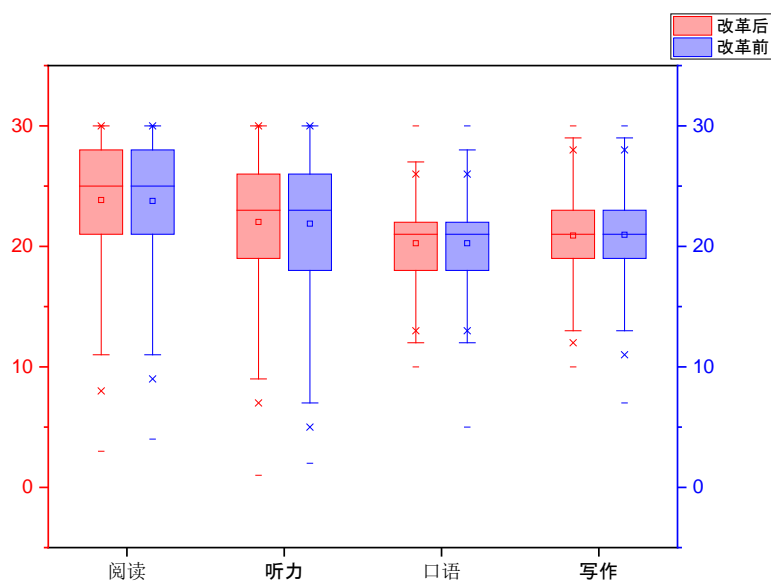


Figure 1. A comparison of student scores of the four sections
图 1. 改革前后学生四个单项成绩对比

在 ETS 刚刚宣布考试改革的初期, 很多同学会担心由于题目减少, 导致高分段容错率降低, 使得高分比较难以获取。本次研究做了考试改革前后数据的累计频率的对比, 并计算了改革后与改革前累计频率的差值, 见表 2。可以看出, 在 28 以上、25 以上分数段, 改革后的累计频率都有所下降, 不过下降幅度较小, 为 1% 左右, 影响不大。不过有趣的是阅读单项在 20 以上的累计频率反而是有所上升的。可以看出目前统出的数据还是支持大部分人“高分更为困难, 中等分数更为容易”的预判的。不过考试改革前后, 四个单项在高分段中最大的差值也仅仅为 2.7%, 因此可以看出考试改革对于最终分数影响并不显著。

Table 2. Changes of accumulative frequency change in different score ranges

表 2. 不同分数段学员占比变化

分数段	阅读	听力	口语	写作
28+	-1.5%	-0.6%	0.2%	-0.2%
25+	-1.3%	-0.8%	-0.8%	-1.6%
20+	1.8%	-0.2%	-0.2%	-1.4%

下面我们再看一下各个单项直方图的对比, 见图 2 和图 3。

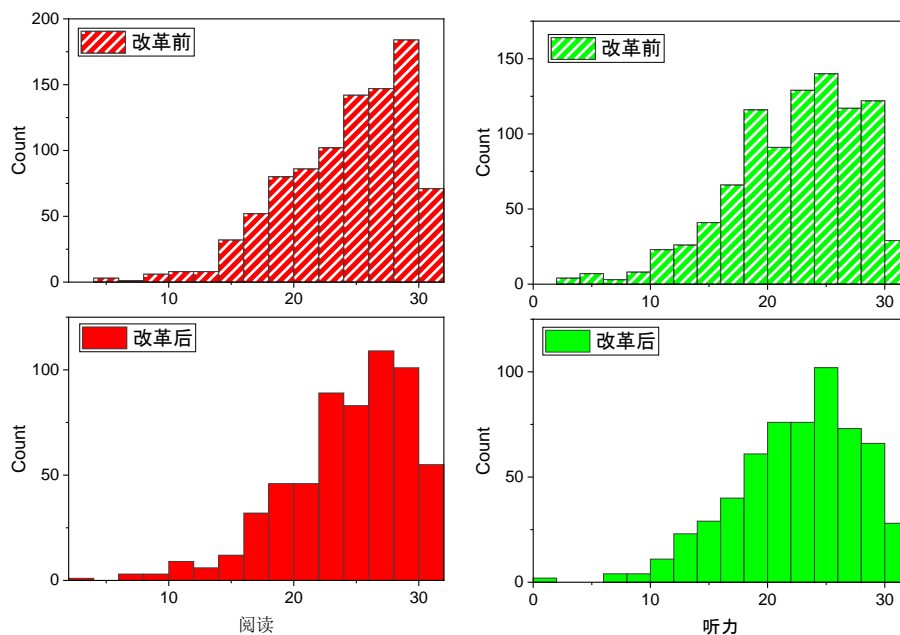


Figure 2. A comparison of frequency in Reading and Listening sections
图 2. 阅读与听力学生分布对比

在阅读和听力两个单项, 可以看出改革对于直方图形状的影响稍有不同。阅读单项在改革之后直方图在高分段相对拉平了一些, 而在听力单项则感觉数据向中位数更加靠拢了一些。不过两者其实都是在 20~30 这个分数段“损有余而补不足”的趋势, 与前面在累计频率上提到的趋势相似。

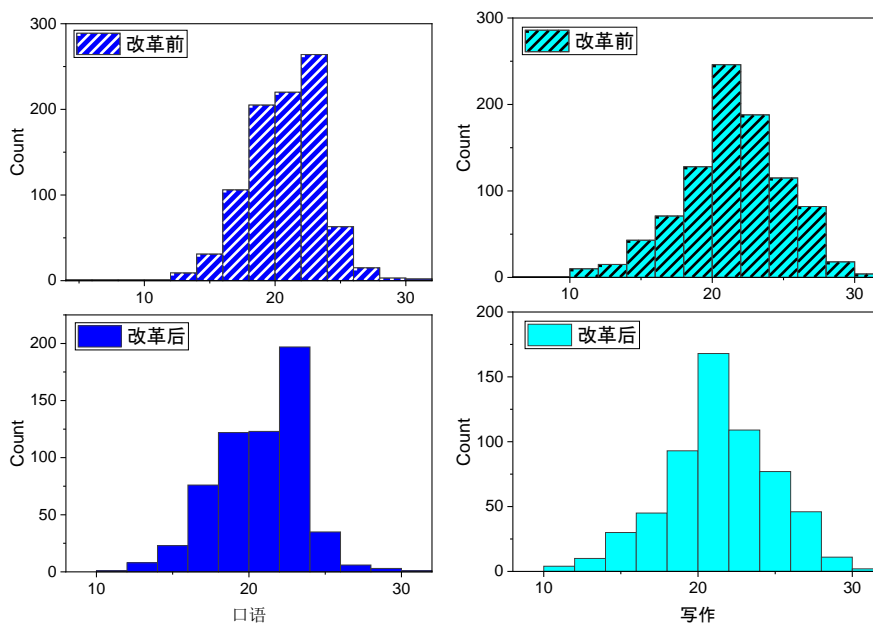


Figure 3. A comparison of frequency in Speaking and Writing sections
图 3. 口语和写作学生分布对比

在口语和写作两个单项, 考试改革前后直方图的变化则并不明显。口语方面, 由于并没有发生变化, 因此与之前直方图相似属于预料之中。而在写作方面直方图和其他数据的稳定则表明新的学术讨论题型

在替代了独立写作的情况下保持了分数的稳定性。

最后我们来看一下各个单项分数与总分之间的关系在考试改革前后的是否有所变化, 见图 4。

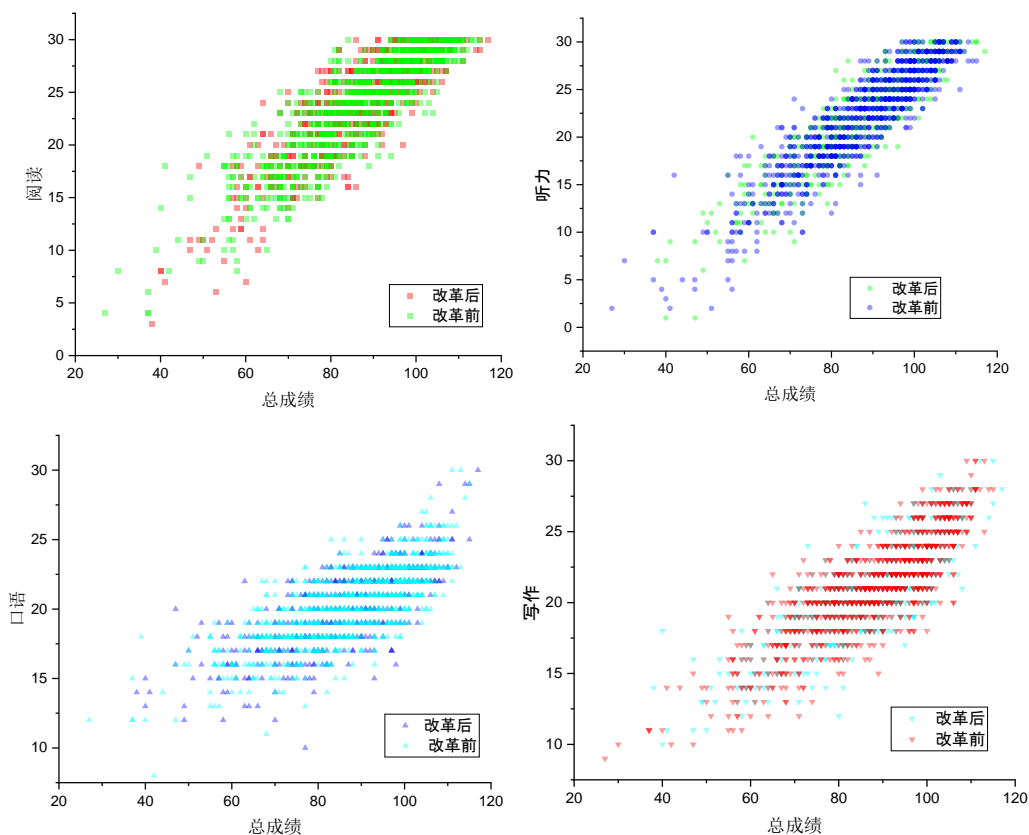


Figure 4. Distribution comparison of four sections
图 4. 各单项改革前后分数分布对比

4. 总结

2023 年托福考试的改革比较 2019 年的改革还是更为大胆的, 不仅减少了题目, 还改变了过去非常稳定的写作题目并大胆地去掉了托福极其特色的“加试”。新的变化虽然在改革初期让很多考生和潜在考生感受到了极大的不确定性, 但是随着后续陆续考生的出分, 考生们所担心的问题也一一得到了解答。

本次研究通过改革前后 1500 多名考生的真实数据, 可以初步给出“托福考试本次改革在考试出分维度影响较小”的结论。托福考试在经历较大的改变的情况下, 依旧维持了很好的一致性, 从学生实际出分的角度来看, 托福考试依旧非常可靠。

不过, 本次研究由于受目标人群所限, 结论更多地集中在单一地区的成年考生方面。不同地区考生的横向对比, 以及改革在低龄考生的影响方面依旧有很大的研究空间。特别是考虑到低龄考生在阅读和听力单项上表现与成年考生相比较弱, 对其影响, 以及不同年龄段考生影响大小的横向对比还是比较值得进一步研究的。

参考文献

- [1] ETS (2023) An Enhanced TOEFL Experience. <https://www.ets.org/toefl/ibt-enhancements.html>
- [2] ETS (2023) TOEFL iBT Test and Score Data Summary 2022.

<https://www.ets.org/pdfs/toefl/toefl-ibt-test-score-data-summary-2022.pdf>

- [3] ETS (2022) TOEFL iBT Test and Score Data Summary 2021.
<https://www.ets.org/content/dam/ets-org/pdfs/toefl/toefl-ibt-test-score-data-summary-2021.pdf>
- [4] ETS (2021) TOEFL iBT Test and Score Data Summary 2020.
<https://www.ets.org/content/dam/ets-org/pdfs/toefl/toefl-ibt-test-score-data-summary-2020.pdf>