

托福考试成绩的变与不变

——对于2017~2020年全球新托福成绩报告的横向研究

姚宇西

新东方国际教育培训事业部, 北京
Email: nalunacz@hotmail.com

收稿日期: 2021年7月27日; 录用日期: 2021年8月23日; 发布日期: 2021年8月30日

摘要

本文研究了2017年到2020年之间的《全球托福成绩报告》数据, 对于其数据进行了可视化处理, 并对于几年间的数据变化进行了适当分析。尤其对于SpeechRater的引入和今年的报告中分数上涨的情况进行了讨论, 并对于其未来对于备考的影响做了探究。

关键词

新托福, 考试成绩, SpeechRater

TOEFL Scores: What Changed and What Stayed

—A Lateral Study of 2017~2020 TOEFL iBT Test and Score Data Summary

Yuxi Yao

Global Education and Training Division, New Oriental Edu & Tech Group, Beijing
Email: nalunacz@hotmail.com

Received: Jul. 27th, 2021; accepted: Aug. 23rd, 2021; published: Aug. 30th, 2021

Abstract

This paper studies the data provided by *TOEFL iBT Test and Score Data Summary* from 2017 to 2020, through visualization; and analyzes its changes during these years. The study also focuses on the effects brought about by the introduction of SpeechRater, discusses the obvious rise in average test score in this year's report, and explores its potential implications on future test preparation.

Keywords

TOEFL iBT, Test Score, SpeechRater

Copyright © 2021 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. 写作背景

每年, 美国教育考试服务中心(Educational Testing Service, 下简称 ETS)都会公布前一自然年的全球托福成绩。在今年 7 月发布的《2020 年全球托福成绩报告》中, 全球平均成绩和中国学生的成绩均有大幅上升, 其中全球托福平均分上升 4 分, 中国大陆考生平均成绩上升 7 分[1]。这一变化在过去十几年内是绝无仅有的, 其背后的原因, 以及托福考试成绩的趋势十分值得分析。本文将通过对于 2017~2020 年间 ETS 官方公布的考试数据, 对于这 4 年的托福成绩进行全面的分析。

2. 2017~2020 年间托福总分与单项分数变化

2.1. 4 年间成绩变化总览

从 2017~2020 年的成绩报告中可以查得过去四年间全球托福总分与阅读、听力、口语、写作四个单项的平均分[1] [2] [3] [4], 整理如下表 1:

Table 1. Average subscores and the total scores of TOEFL iBT between 2017 and 2020

表 1. 2017~2020 年间新托福四项与总分的平均分

平均分	2017 年	2018 年	2019 年	2020 年
阅读	20.7	20.8	21.2	22.2
听力	20.3	20.6	20.9	22.3
口语	20.4	20.5	20.6	21.2
写作	20.8	20.7	20.5	21.5
总分	82	83	83	87

我们进一步处理以上数据, 计算出各个单项与总分每一年的增长情况, 如下表格:

Table 2. Increase in average subscores and the total scores of TOEFL iBT between 2017 and 2020

表 2. 2017~2020 年间新托福四项与总分的平均分增长情况

平均分增长	2017 至 2018 年	2018 至 2019 年	2019 至 2020 年
阅读	0.1	0.4	1
听力	0.3	0.3	1.4
口语	0.1	0.1	0.6
写作	-0.1	-0.2	1
总分	1	0	4

我们可以看出, 2020 年与 19 年相比, 各个单项的上浮都是巨大的(表 2), 最大的是听力, 涨幅高达 1.4 分, 而写作和口语也一改往年稳定的状态, 分别上涨了 1 分与 0.6 分。考虑到目前写作与口语是有人工智能引擎加持, 并且本身的数据修正就非常成熟, 这么大的涨幅还是比较出乎预料的。

如果我们把四个单项的四年变化的分开来看(图 1), 其实可以发现虽然各个单项的涨幅都比往年要高, 但是阅读、听力和口语单项的增长是与过去几年的趋势相同的, 反而是写作, 一改过去每年下滑一点点的趋势, 开始上升, 甚至上升的涨幅超过了口语。这背后的原因还是很值得未来进行针对性研究的。

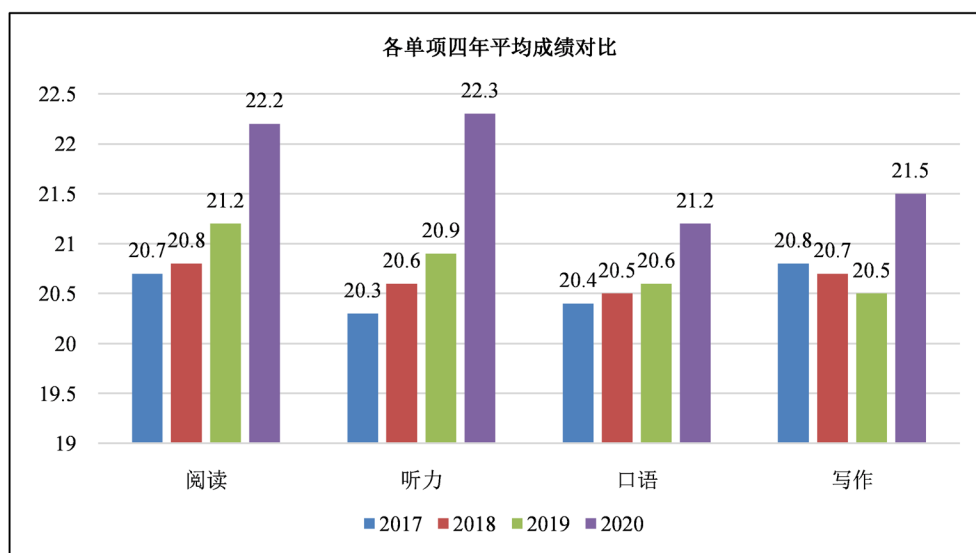


Figure 1. Annual changes in average subscores and the total scores of TOEFL iBT between 2017 and 2020
图 1. 2017~2020 年间托福四项与总分的平均分变化情况

我们再来看一下一个我们一直非常不关注的的数据——标准差。我们把四年的标准差数据进行比较(表 3), 发现 2020 年的标准差不升反降, 而这一下滑其实并非今年特有, 而是和一直以来的数据趋势相同的。这其实表明了, 虽然成绩上升了, 但是学生成绩的差异其实有略微缩小, 也就是说, 考到更高的分数的人更多了, 但是整体数据的差异性在减小, 极端分数(有可能是极端低分)其实在下降。

Table 3. Standard deviation for the subscores and the total scores of TOEFL iBT between 2017 and 2020
表 3. 2017~2020 年间新托福四项与总分的标准差

标准差	2017 年	2018 年	2019 年	2020 年
阅读	6.7	6.7	6.6	6.3
听力	6.7	6.5	6.4	6
口语	4.5	4.5	4.4	4.3
写作	4.8	4.7	4.7	4.6
总分	20	20	19	19

2.2. 4 年间总分的百分位数变化情况

我们将 2017 年到 2020 年在总分上的百分位数的数据进行了可视化处理(图 2), 可以清晰地看出, 今

年的数据出现了整体下移，而过去三年的数据几乎在同一个线上。这个与整体分数上升的趋势相同，意味着同样的分数，“战胜”的竞争对手减少了。考虑到很多时候学生在使用托福成绩的时候是相互竞争而非“达标”，某种程度上来说托福考试的分数出现了一定程度的通货膨胀。反之，对于很多用托福成绩来对标学校或公司要求的考生，这一趋势则是大大的利好。

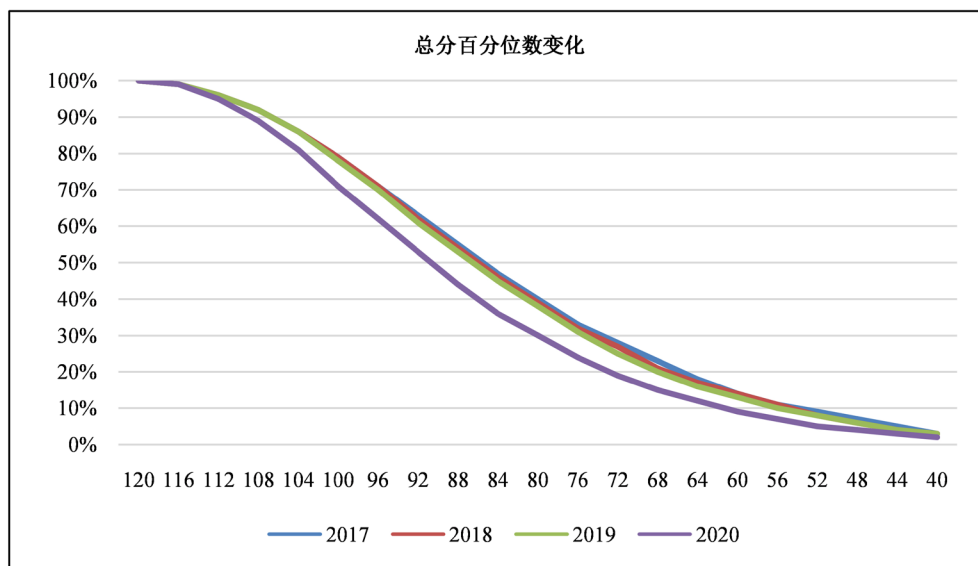


Figure 2. Percentile comparison of total scores of TOEFL iBT between 2017 and 2020

图 2. 2017~2020 年间新托福总分的各分数段百分位数对比

我们对于 2020 年的百分位数和 2019 年进行了对比，同分数端差值的极值为 9%。从 76 分考试到 100 分之间，这一差值一直处于 7% 及以上，也就是说，这一段的的成绩是影响最大的。

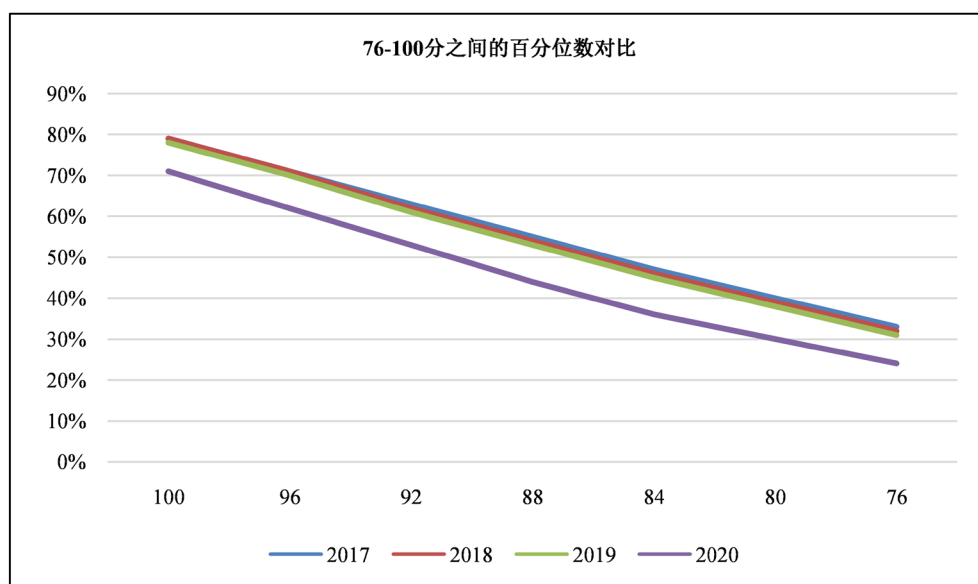


Figure 3. Percentile comparison of total scores of TOEFL iBT between 2017 and 2020 (76-100 range)

图 3. 2017~2020 年间托福总分的各分数段百分位数对比(76~100 分区间)

有趣的是，将这一段的数据放大(如 图 3所示)，可以看到这段的变化几乎为线性的，而且斜率也思路很相似，仿佛有一批考生成绩是直接平移上去一般。这一点其实也能体现出，虽然这次总成绩有所变化，但是托福考试本身的稳定性还是很高的。

2.3.4 年间各单项分数的百分位数变化情况

我们将各个单项过去四年的百分位数变化进行了同样的处理，如 图 4~7所示，可以看出这一百分位数向下移动的趋势在各个单项都是相同的。而与去年差值的对比上可以看出，阅读和口语的差值影响较小(极值为 6%)，而听力和写作变化较大(极值为 9%)，这一对应关系可能与考试本身的算分机制有一定的关系。

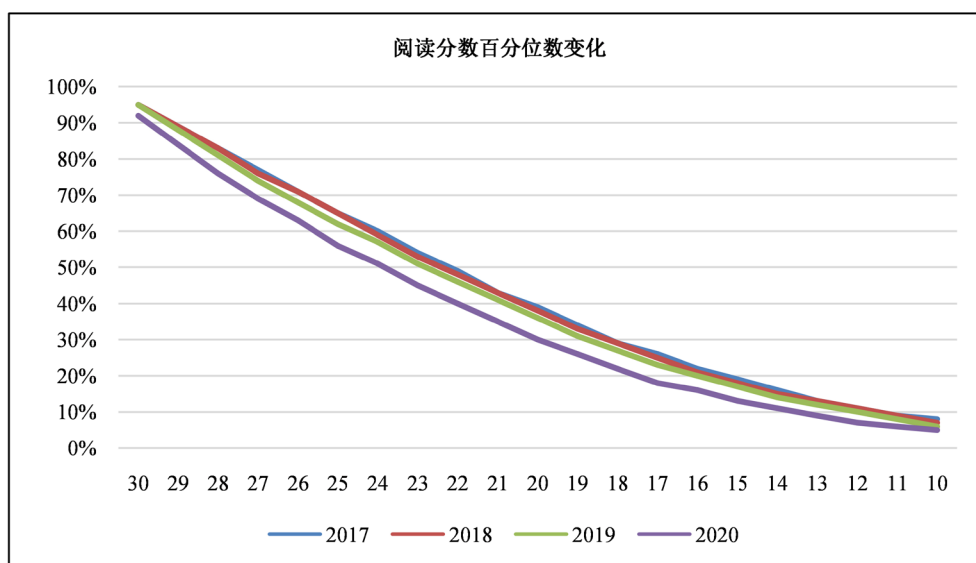


Figure 4. Percentile comparison of TOEFL iBT reading scores between 2017 and 2020

图 4. 2017~2020 年间新托福阅读的各分数段百分位数对比

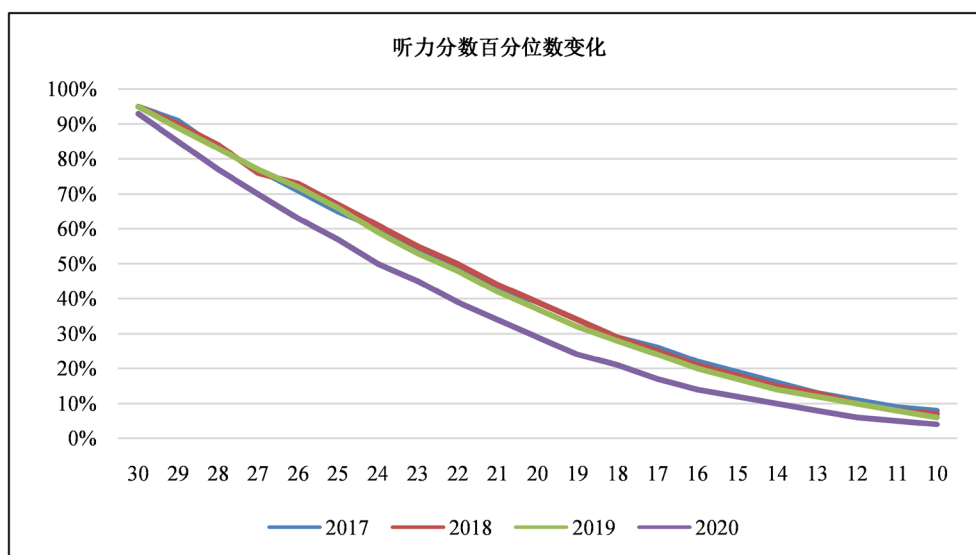


Figure 5. Percentile comparison of TOEFL iBT listening scores between 2017 and 2020

图 5. 2017~2020 年间新托福听力的各分数段百分位数对比

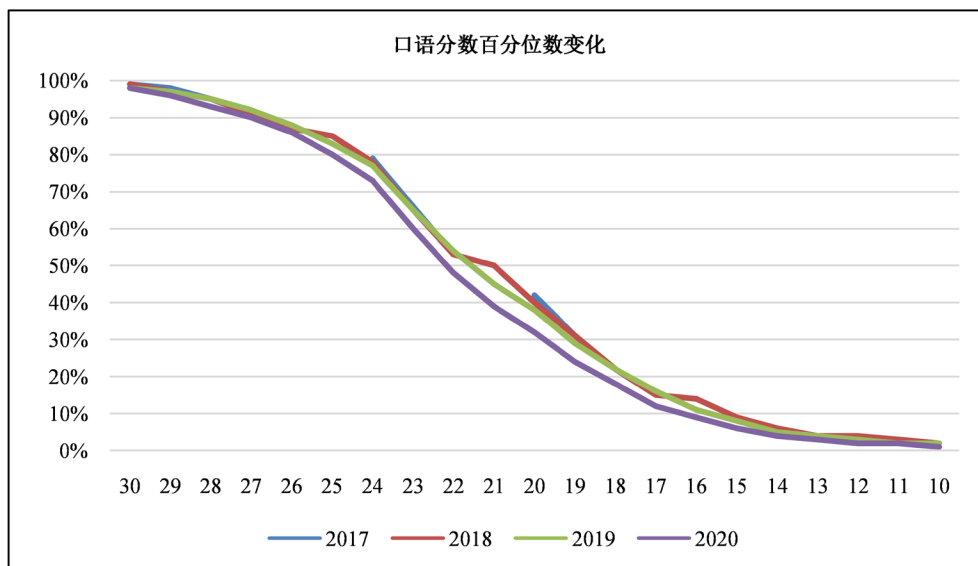


Figure 6. Percentile comparison of TOEFL iBT speaking scores between 2017 and 2020

图 6. 2017~2020 年间新托福口语的各分数段百分位数对比

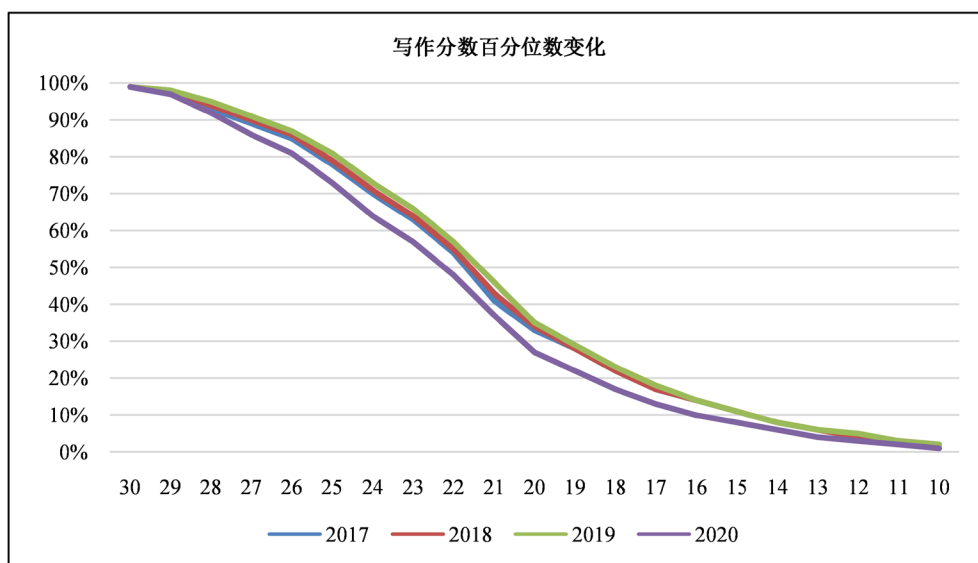


Figure 7. Percentile comparison of TOEFL iBT writing scores between 2017 and 2020

图 7. 2017~2020 年间新托福写作的各分数段百分位数对比

而上面四幅图片中口语分数的四年对比是最为有趣的，可以明显的看出数据从 2017 的断点状态，到 2018 年的折现状态，再到 2019、2020 年的光滑状态的变化。从中也可以看出 SpeechRater 的引入对于宏观数据的影响还是非常明显的。

3. SpeechRater 引入对于托福口语分数的影响

3.1. 口语评分在托福考试中的问题

托福的四个单项中，口语和写作与阅读和听力不同，都是有人工评分存在的单项。在 2019 年之前，口语评分与写作不同，只有人工评分，没有机器评分。如果参考 2018 年之前的全球托福考试成绩报告，

可以发现在一些特殊分数，百分位数数据是不连续的[4]。在 7、12、16、21 和 25 分的分数上，口语单项是没有相应的百分位数数据的，这些分数也成为了托福考生圈内的“都市传说”。然而到了 2019 年开始，在当年的《2018 年全球托福考试成绩报告》中，我们发现这些分数的百分位数出现了，虽然考试中依旧没有这些分数，但是这些分数的百分位数是被计算出来了的。到了 2019 年，我们发现很多学生开始反馈出现了 25 分的口语分数，随后不久，ETS 宣布将 SpeechRater 引入正式考试中，而在随后一年的《2019 年全球托福考试成绩报告》中托福口语的各个分数也都是有了相应的百分位数数据。

3.2. 2017~2020 托福口语数据百分位数对比

我们将 2017~2020 年的口语百分位数数据做可视化处理，以 2020 年为参考，将每一年的分数与之对比，如图 8。

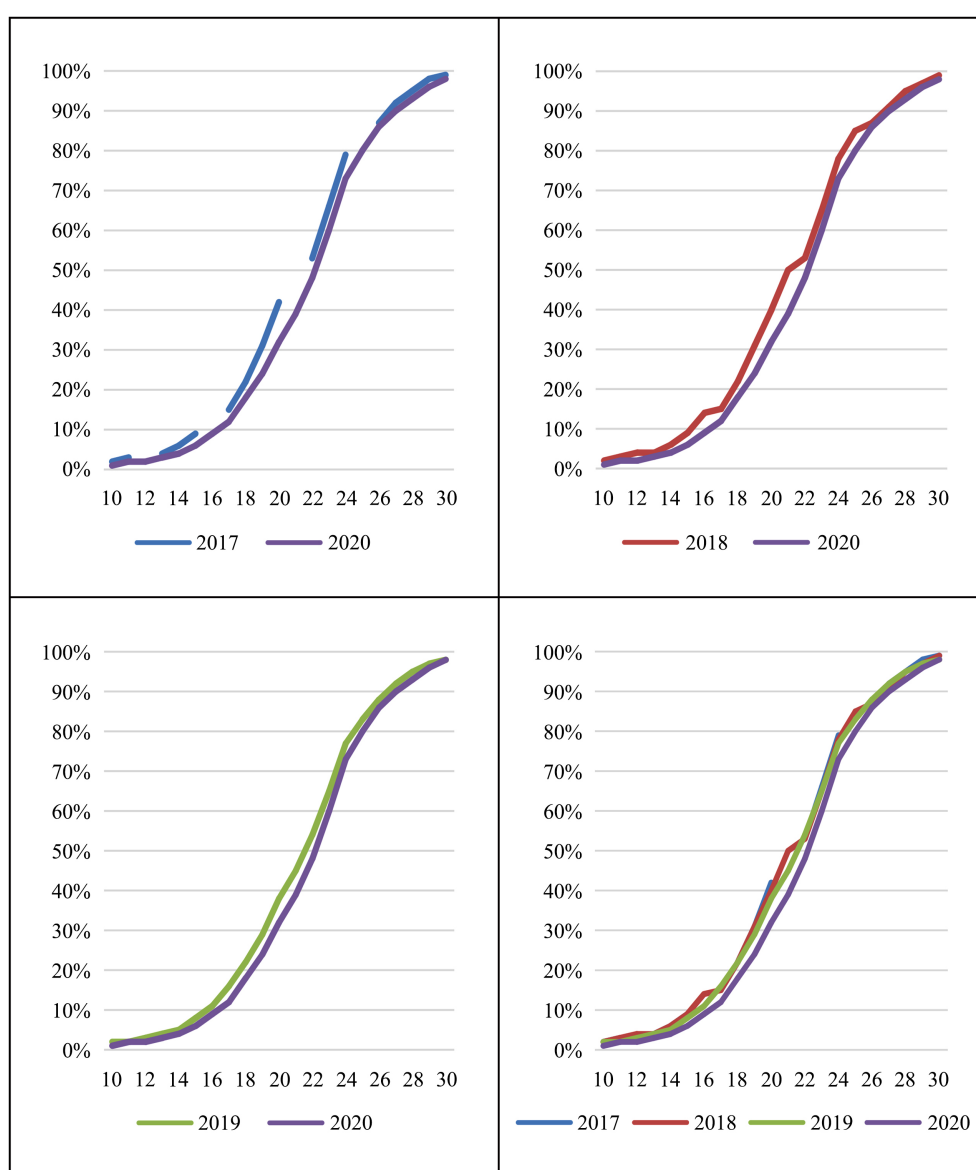


Figure 8. Percentile comparison of TOEFL iBT speaking scores between 2017 and 2020

图 8. 2017~2020 年间新托福口语的各分数段百分位数对比

横向对比发现, 整个数据的光滑程度从 2017 年的断续, 到了 2018 年的锯齿状, 再到 2019、2020 年的光滑, 可以明显看到在 2019 年引入 SpeechRater 之后托福口语成绩的百分位数更加趋于光滑。智能引擎对于人工评分工作的影响还是显而易见的。

4. 成绩上涨对于托福考生未来的启示

4.1. 考试成绩上涨原因分析

2019 年 ETS 宣布托福考试进行“缩编”, 对于除写作考试之外的所有单项题数进行删减。笔者认为这一改变代表着 ETS 对于托福考试的信心的加强, 因为体量是保证信度的重要手段, 为考试“瘦身”代表着 ETS 对于托福考试的稳定性和可靠性的极大信任。

而事实上来看, 在 2019 年的数据(受疫情影响, 该数据的公布日期相较往年的有所推迟)中的确体现了这种稳定性, 在考试体量减少的情况下, 全球成绩在总分上没有变化, 各个单项变化最大的是阅读上升 0.4 分, 其改变幅度虽然与往年相比变大了(2017 年到 2018 年四个单项变化最大的是听力单项上涨的 0.3 分), 但是并没有跳跃性的增长。

2020 年, 在过去几十年人类历史中是非常特殊的一年, 由于新冠疫情, 全球大部分国家和地区都采取了大规模的隔离措施。而 ETS 为了避免学生由于隔离无法参加托福考试, 推出了“托福在家考”的考试形式。这一举措, 让学生们可以自由选择考试时间和地点, 减少了学生在考试时候的不适和紧张感, 让学生们获得了全新的考试体验。有可能正是这一原因, 在本措施推出的初期, 很多学生的托福成绩还是出现了比较大的正向震荡的。客观的来说, 这一现象和托福在全球和中国的成绩上涨有一定的吻合度, 不失为这次成绩上涨的可能性之一。

4.2. 对于考生的启示

虽然这次全球和中国学生平均分的上涨比较出乎大家的预料, 但是如果我们认真研究过去四年的成绩变化, 这一上涨的趋势其实是一直存在的, 并非一次完全的异常。总体来说, 由于大陆考生的上涨幅度领先全球, 使得我们的平均分近十年第一次达到了全球平均数, 去年的成绩报告背后的变化其实是有利于国内考生的。特别是很多以固定达标线为目标的托福考生, 这次的报告给此类考生一个很强的强心剂。

当然, 如果个别考生并未赶上去年的成绩上涨的机会, 也不用担心。新东方根据往年的学生分数大数据, 也明确地给出过各个分数段托福考生应该如何规划自己的四个单项的目标分(如下表 4)。若考生能利用好这些数据化的结论, 应该会为其托福学习和备考带来积极影响。

Table 4. Recommended goals for each subscore of TOEFL iBT based on New Oriental big data (2018)

表 4. 基于新东方大数据的新托福单项分数目标(2018)

目标总分	阅读	听力	口语	写作
60	14	13	17	16
80	21	19	20	20
100	27	26	23	25
105	29	27	23	26
110	30	29	25	27

对于留学语培行业中的教师和从业者来说，这次报告的结论还是非常有趣的，我们也很期待明年经过进一步的修订，包括 TOEFL Essential 的诞生，会不会对托福 iBT 本身带来影响。毕竟托福考试是一个非常科学，非常严谨，非常稳定的考试，其自我修复性也是很强的。明年的全球成绩会有怎样的变化，相信无论是考生还是从业者都会非常期待。

参考文献

- [1] ETS TOEFL (2021) Test and Score Data Summary for TOEFL iBT Tests, January 2020-December 2020 TestData. https://www.ets.org/s/toefl/pdf/toefl_tsd_data_2020.pdf
- [2] ETS TOEFL (2020) Test and Score Data Summary for TOEFL iBT Tests, January 2019-December 2019 TestData. https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf
- [3] ETS TOEFL (2019) Test and Score Data Summary for TOEFL iBT Tests, January 2018-December 2018 TestData. https://www.ets.org/s/toefl/pdf/toefl_tsd_data_2018.pdf
- [4] ETS TOEFL (2018) Test and Score Data Summary for TOEFL iBT Tests, January 2017-December 2017 TestData. https://www.ets.org/s/toefl/pdf/toefl_tsd_data_2017.pdf