

对GRE写作自动化评分器e-rater评分准确性的实证研究

杜 璟¹, 冷 楠²

¹新东方在线, 北京

²武汉新东方学校, 湖北 武汉

Email: jingdu42@gmail.com

收稿日期: 2020年6月1日; 录用日期: 2020年8月1日; 发布日期: 2020年8月27日

摘 要

本研究通过分析ETS推出的ScoreItNow!付费评分服务, 来探讨ETS研发的自动化阅卷程序e-rater在GRE分析性写作评分上的准确性。研究试图回答以下四个问题: (1) ScoreItNow!打分与实际考试的得分有多大差别; (2) ScoreItNow!是否能够对文章实际内容进行准确判断; (3) 文章字数对ScoreItNow!评分有多大影响; (4) ScoreItNow!作者分析工具给出的反馈是否准确。研究发现ScoreItNow!给出给出的分数和实际考试分数接近。但是, ScoreItNow!无法对文章实际内容进行准确判断。对于字数越多的文章, ScoreItNow!也越倾向于给出更高的分数。ScoreItNow!提供的作者分析工具中给出的反例准确性不令人满意。对于GRE写作教师而言, 在教学过程中可以适当强调语言表达的重要性, 但同时也要教授实用的写作方法, 提高学生的批判性思维的能力。

关键词

GRE分析性写作, e-rater, 自动化阅卷

Accuracy of the e-rater Automated Scoring Engine in the GRE Writing Section: An Empirical Study

Jing Du¹, Nan Leng²

¹New Oriental Online, Beijing

²Wuhan New Oriental School, Wuhan Hubei

Email: jingdu42@gmail.com

Received: Jun. 1st, 2020; accepted: Aug. 1st, 2020; published: Aug. 27th, 2020

Abstract

This study aims to evaluate the accuracy of the e-rater automated scoring engine by investigating the scores and feedback provided by ScoreItNow! paid scoring service in the GRE analytical writing section. The study attempts to answer the following four questions: (1) Is there any difference between the ScoreItNow! scores and the scores test-takers gotten from the GRE analytical writing section? (2) Is ScoreItNow! able to accurately evaluate essay content? (3) To what extent are the scores influenced by essay length? (4) Is the Writer's Analysis Tools able to provide accurate feedback? The study found that scores provided by ScoreItNow! generally agree with the scores test-takers received from the GRE analytical writing section. However, ScoreItNow! cannot give accurate evaluation concerning the content of the essays. ScoreItNow! is more likely to give higher scores when essays are longer. The feedback in Writer's Analysis Tools is far from satisfactory. GRE writing teachers are advised to not only emphasize the importance of language use in their teaching, but also impart useful writing skills and improve students' ability to think critically.

Keywords

GRE Analytical Writing, e-rater, Automated Scoring

Copyright © 2020 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. GRE 写作部分介绍

GRE General Test (美国研究生入学考试, 以下简称 GRE)是由 ETS (Educational Testing Service, 美国教育考试服务中心, 以下简称 ETS)出题和举办的学术能力测试。GRE 考试主要考查研究生阶段学习所需要的语言推理、数学推理、和分析性写作能力。

在分析性写作部分, 考生需要完成两篇作文: Analyze an Issue (分析一个问题, 以下简称 Issue)和 Analyze an Argument (分析一段论证, 以下简称 Argument)。Issue 题目要求考生根据具体的写作指令对一个常见的话题进行分析; Argument 题目要求考生根据具体的写作指令对一段论证进行评价。每篇文章均要求在 30 分钟内完成。每篇作文得分均为 0~6 分, 考生最终作文部分的总分为两篇作文各自得分的平均分, 精确到 0.5 分。

考官在评分时, 会针对考生的作文质量做出整体的评价。考官主要会关注考生是否能够针对话题给出深入的分析、用强有力的论证和例子展开观点、文章扣题并且组织合理、运用多样的句式和准确的词汇有效表达观点[1]。

ETS 每年在写作评分上花费的时间成本和资金都十分巨大。所以, 为了节约成本以及提高评分效率, ETS 研发了自动化阅卷程序 e-rater 来辅助打分[2]。考生的每一篇作文会分别由人类评分员和 e-rater 来评分。如果两者分数相近, 则考生的得分为两者的平均分。如果两者分差较大, 则考生的作文需要由第二名人类评分员打分, 最终得分为两名人类评分员分数的平均分。在研发过程中, e-rater 需要学习成千上万篇人类评分员已经给出分数的文章。这些文章从质量很高到水平较低不等, 以便让 e-rater 能够较全面地学习。在评分时, e-rater 会寻找被评分文章和学习过的高质量文章的相似之处。另外, e-rater 打分效率很高, 每篇作文只需要几秒钟时间便能出分。虽然早期版本的 e-rater 打分准确性非常令人怀疑, 但经

经过多年的发展, 如今 e-rater 评分的准确度已经非常高, 其与人类评分员给分的平均分差与两名人类评分员给分的平均分差几乎是相同的[3]。

e-rater 在评分时, 主要根据文章的语言质量打分, 具体打分项如下:

- 词汇复杂程度(lexical complexity)——文章中生僻词的数量以及单词的平均长度;
- 语法错误的占比(proportion of grammar errors)——文章中语法错误(如主谓不一致)出现的比例;
- 用法错误的占比(proportion of grammar errors)——文章中用法错误(如冠词使用错误)出现的比例;
- 错误行文规则的占比(proportion of mechanics errors)——文章中行文规则错误(如单词拼写错误)出现的比例;
- 文章风格(proportion of style comments)——文章中出现的被动语态、过度重复的单词或者过长或过短的句子数量;
- 结构和展开得分(organization and development scores)——对文章中背景信息、总论点、分论点、支持观点、和结论等句子的识别;
- 对文章措辞恰当的奖励(features rewarding idiomatic phraseology)——文章中地道的搭配和正确使用的介词数量。

ETS 推出的 ScoreItNow!在线 GRE 作文打分收费服务也使用的是 e-rater 技术[4]。使用 ScoreItNow!评分能够让考生在备考 GRE 写作期间得到与实际考试相近的评分。用户在购买服务之后可以获得两次作文打分机会, 并且可以选择通过练习模式(Practice Option)或者模考体验(Test Experience)来完成。

练习模式中, 用户可以从规定的六个题目中(Issue 和 Argument 各三题)选择两个题目进行练习。用户可以自由组合所选题目。例如, 可以 Issue 和 Argument 各选择一题, 也可以两题都选择 Issue 或者两题都选择 Argument。在写作过程中, 用户可以随时保存作文并退出系统, 以后再重新登陆完成作文。在模考体验中, 系统会从前面提到的六道题目中随机选择一道 Issue 题和一道 Argument 题给用户进行限时模拟测试。不论用户选择哪种模式, 最终都会得到每篇文章的分数以及相应的反馈。选择了模考体验的用户还会得到一个 GRE 分析性写作的总分。此外, ScoreItNow!反馈部分还包括作者分析工具(Writer's Analysis Tools), 该工具会从语言表达和文章展开以及文章结构角度对文章进行分析反馈。

由于 ScoreItNow!使用的是与 GRE 考试相同的 e-rater 来进行评分, 所以对于 ScoreItNow!评分机制的分析即可以被视作对于 e-rater 的分析。

2. 研究问题

本研究主要试图回答以下问题:

- (1) ScoreItNow!打分与实际考试的得分有多大差别;
- (2) ScoreItNow!是否能够对文章实际内容进行准确判断;
- (3) 文章字数对 ScoreItNow!评分有多大影响;
- (4) ScoreItNow!作者分析工具给出的反馈是否准确。

3. 研究样本

本研究一共选择了来自 3 位 GRE 写作教师和 1 位正在学习 GRE 的学生的作文, 4 位参与者都参加过 GRE 实考。他们分别完成了 ScoreItNow!所给出题目中 Issue 和 Argument 各一题, 并且完成的题目相同。

Issue 题目为:

A nation should require all of its students to study the same national curriculum until they enter college.

Write a response in which you discuss the extent to which you agree or disagree with the recommendation and explain your reasoning for the position you take. In developing and supporting your position, describe specific circumstances in which adopting the recommendation would or would not be advantageous and explain how those examples shape your position.

Argument 题目为:

The following is a memorandum from the business manager of a television station.

“Over the past year, our late-night news program has devoted increased time to national news and less time to weather and local news. During this time period, most of the complaints received from viewers were concerned with our station's coverage of weather and local news. In addition, local businesses that used to advertise during our late-night news program have just canceled their advertising contracts with us. Therefore, in order to attract more viewers to the program and to avoid losing any further advertising revenues, we should restore the time devoted to weather and local news to its former level.”

Write a response in which you discuss what specific evidence is needed to evaluate the argument and explain how the evidence would weaken or strengthen the argument.

4. 结果与讨论

4.1. ScoreItNow!打分与实际考试的得分有多大差别

从 ScoreItNow!给出的分数来看(表 1), 教师 1 和教师 2 ScoreItNow!分数和 GRE 实考写作分数一致。教师 3 和学生 1 ScoreItNow!和 GRE 实考写作分数相差 0.5 分。由于 0.5 的分差属于正常范围, 所以我们可以认为, 对于这 4 位参与者而言, ScoreItNow!能够比较准确地反映实际考试分数。

Table 1. Comparison between participants' ScoreItNow! scores and GRE analytical writing section scores

表 1. 参与者 ScoreItNow!得分和 GRE 实考写作分数对比

参与者	ScoreItNow!得分	GRE 实考写作分数
教师 1	5	5
教师 2	5	5
教师 3	5.5	5
学生 1	4	3.5

4.2. ScoreItNow!是否能够对文章实际内容进行准确判断

通过研究, 我们发现 ScoreItNow!并不能够针对文章的内容进行准确评价, 这一结论和前文提到的 e-rater 各打分项的内容一致(这里所说的文章内容, 主要是指作文观点是否扣题, 给出的论据是否足够有说服力, 句子之间联系是否紧密等)。

在研究这一问题时, 作者进行了如下尝试: 1) 将教师 1 的 Issue 作文放到 Argument 题目的答题框并提交作文; 2) 将教师 1 的 Argument 作文放到 Issue 题目的答题框并提交作文; 3) 对学生 1 Issue 作文内容进行分析。

当我们将教师 1 的 Issue 作文放到 Argument 答题框提交作文时, ScoreItNow!给出的分数为 5 分, 与教师 1 正常提交 Issue 作文的得分一致。将教师 1 的 Argument 作文放到 Issue 的答题框提交作文时, 我们同样得到了 5 分, 这也与正常提交 Argument 作文得分一致。也就是说, 虽然“文不对题”, 但是由于教

师 1 的 Issue 和 Argument 两篇作文语言优秀, 所以仍然得到了 5 分的高分, 而没有因为内容与题目不符而造成扣分。

学生 1 的 Issue 作文在 ScoreItNow! 中得到了 4 分, 但是参与研究的 3 位教师在读过学生 1 的文章后一致认为这篇 Issue 不能够达到 4 分的水平。本文选取学生 1 Issue 作文其中一段进行分析:

What's more, varied situations, such as cultural and resource distinction, restrict the application of the same curriculum. It is possible that if a nation has wild territory and multi-ethnic people populated in, there might be considerable differences in areas no matter on culture or geography. For instance, in southwest China, if soccer class become a national demand, schools in that area will face a big problem for it is karst landform, hard to find large plains. In this case, schools taking as an alternative the sport that is special for local people would be a better choice, not only solve the problem but flourish local culture as well. Additionally, when the unify demand take most developed areas into account but hit the underdeveloped area, it is also impossible for the school in the underdeveloped area to offer computer classes and meet the demand of uniform curriculum.

这一段的观点是文化和资源的差异使得推行同样的课程比较困难。后文分别用足球课和电脑课作为例子来证明这一观点。在对足球课例子进行说明时, 解释的是由于喀斯特地貌导致无法修建场地, 但这一解释是不够合理的。另外, 电脑课的例子只用了一句话带过, 没有具体说明和展开。除此之外, 这一段同样比较严重的问题在于有一些难读的句子, 例如“*For instance, in southwest China, ..., hard to find large plains. In this case, ...but flourish local culture as well.*”所以, 虽然从机器的角度来看, 这篇文章可以拿到 4 分, 但是如果把内容要素考虑进去, 却很难达到 4 分的要求。

4.3. 文章字数对 ScoreItNow! 评分有多大影响

在 e-rater 打分的各项指标中, 不同的指标在评分过程中权重不一。但不论是 Issue 还是 Argument, 各评分项中权重最大的是文章结构(organization)和文章展开(development)。在 Issue 中, 这两项分别占比 31.16% 和 29.63%; 在 Argument 中, 这两项占比分别为 33.17% 和 28.31%。而其它指标在评分时的权重相对低很多, 只占到了 2%~8% 不等[5]。文章结构的计算标准为文中的“话语元素(number of discourse elements)”的数量。所谓话语元素指的是 e-rater 在分析文章结构时会把文中的内容划分为引入内容(introductory material)、总论点(thesis)、分论点(main ideas)、支撑观点(supporting ideas)和文章结论(conclusion), 以上几项即为话语元素。对于文章展开, e-rater 的计算标准则是话语元素的平均长度(length of discourse elements) [6]。不难发现, 按照这样的计算标准, 不论是文章结构还是文章展开都与文章的长度几乎直接成正比。再考虑到两者的打分占比, 理论上文章的最终得分将主要取决于文章的长度。

为了对这一理论进行验证, 本研究将教师 3 的 Issue 作文和学生 1 的 Argument 作文的篇幅进行调整, 再提交到 ScoreItNow! 进行第二次评分。在调整篇幅时, 我们将教师 3 Issue 文章的第三个中间段删掉, 文章的段落数量从 5 段变为 4 段, 总字数从 546 减少到 419。调整后文章 ScoreItNow! 给出的分数由调整前的 5 分变为了 4 分。我们又将学生 1 Argument 作文第二段的部分内容复制, 并分别粘贴到每个中间段后面, 以增加段落长度。调整后的文章字数从 401 增加到 580, ScoreItNow! 给出的分数由调整前的 4 分变为了 5 分。

由此可以看出, 文章长度确实会显著影响 ScoreItNow! 最后对文章的评分。

4.4. ScoreItNow! 作者分析工具给出的反馈是否准确

ScoreItNow! 给出的反馈主要包括对于各个语言要素(例如语法, 用法等)和结构及展开的识别(图 1)。

点击相应项目, 文中会出现相关标记。比如, 点击 **Transitional Words and Phrases** (连接词和词组), 会高亮显示 **ScoreItNow!** 识别出的文章中所使用到的所有连接词和词组(图 2)。本研究想要探讨的是 **ScoreItNow!** 给出的这些标记是否准确。结果见表 2。

Writer's Analysis Tools	
Missing Comma	<p>First of all, schools should provide all students with the core curricula, and these can pave the way for the students' further study no matter which field they chose in the universities. Classes such as language and math, which is fundamental for almost every field should be classified to core units for they will cultivate students' writing, listening, speaking skills, especially logic skills. Such skills are significant for them to express themselves properly and systematically in academic development. What's more, only taught in the unified language can the students communicate with each other fluently without misunderstanding. For example, India has a lot of dialects which are pervasively spoken even more than the official language. Some students can only speak dialects at home, no official language practising, so unified language teaching wipes out the barriers and differences in communication in the complicated environment.</p>
Missing Final Punctuation	
Missing Question Mark	
Spelling	
Style	
Repetition of Words	
Organization & Development	
Conclusion	
Main Ideas	
Supporting Ideas	
Thesis Statement	
Topic Relationship and Technical Quality	
Transitional Words and Phrases	
Missing Comma	
Missing Final Punctuation	<p>First of all, schools should provide all students with the core curricula, and these can pave the way for the students' further study no matter which field they chose in the universities. Classes such as language and math, which is fundamental for almost every field should be classified to core units for they will cultivate students' writing, listening, speaking skills, especially logic skills. Such skills are significant for them to express themselves properly and systematically in academic development. What's more, only taught in the unified language can the students communicate with each other fluently without misunderstanding. For example, India has a lot of dialects which are pervasively spoken even more than the official language. Some students can only speak dialects at home, no official language practising, so unified language teaching wipes out the barriers and differences in communication in the complicated environment.</p>
Missing Question Mark	
Spelling	
Style	
Repetition of Words	
Organization & Development	
Conclusion	
Main Ideas	
Supporting Ideas	
Thesis Statement	
Topic Relationship and Technical Quality	
Transitional Words and Phrases	

Figure 1. A sample of ScoreItNow! feedback page
图 1. ScoreItNow! 反馈样式

Transitional Words and Phrases

However, **when** the basic skill is guaranteed, classes that help students to show their personalities is optional **instead of** applied to all. There are different thoughts between students for **while** some of them take an interest in piano, some others in violin. **so** in this situation, it is not possible to offer all the student the same curriculum on their interest. **But if** the school forced students to learn as much as they can, **for example,** both piano and violin, there is **so** much burden, and it will limit the professionalization on interest. In the modern world, with all its complexity, personalized classes are essential **if** diversity is to be made so the curriculum should not be the same.

What's more, varied situations, **such as** cultural and resource distinction, restrict the application of the same curriculum. It is possible that **if** a nation has wild territory and multi-ethnic people populated in, there might be considerable differences in areas no matter on culturer or geography. **For instance,** in southwest China, **if** soccer class become a national demand, schools **in that** area will face a big problem for it is karst landform, hard to find large plains. **In this case,** schools taking as an alternative the sport **that is** special for local people would be a better choice, not only

Figure 2. Transitional words and phrases given in ScoreItNow! feedback
图 2. ScoreItNow!反馈中对于连接词的标注

Table 2. Accuracy of ScoreItNow! feedback based on the 8 essays from study participants
表 2. 参与研究的 8 篇文章 ScoreItNow!反馈准确性

	总次数	正确次数	错误次数	准确率
Grammar	8	4	4	50%
Usage	20	14	6	70%
Mechanics	21	10	11	47.6%
Style	16	16	0	100%
Organization and Development				
Conclusion	8	6	2	75%
Introductory Material	5	4	1	80%
Main Ideas	20	20	0	100%
Supporting Ideas	20	20	0	100%
Thesis Statement	7	4	3	57.1%
Transitional Words and Phrases	243	227	16	93.4%

通过分析 ScoreItNow!对每篇文章给出的反馈,我们发现 ScoreItNow!对文章的标记准确度不高。在我们分析的十项内容中,语法(grammar),行文规范(mechanics)和总论点(thesis statement)标记的准确度只有 50%左右,其中行文规范的正确率还不到 50%。本文作者猜测,这些指标 ScoreItNow!识别不够准确的原因主要是因为英文写作变化比较多,不能够完全按照一套固定的规则来评判对错。例如,“If this is the case, the local farmers are not making excessive profits and regulating retail milk price would be unnecessary.” ScoreItNow!标注“retail milk price”缺少冠词,但是从整句来看,这里并不需要额外加冠词。

ScoreItNow!在文体(style)、分论点(main ideas)和支持观点(supporting ideas)三项上的识别准确率达到了 100%;连接词和词组(transitional words and phrases)的准确率也较高,达到了 93.4%。但是需要注意的是,通过分析 ScoreItNow!的反馈,我们发现这四项准确率较高的原因之一是这几项比较好识别。比如,在文体这一项中,标记得最多的就是重复的单词,而这一点对于 ScoreItNow!来说十分好操作。其次,在标记连接词和词组时,ScoreItNow!需要做的就是找到与系统语料库相对应的连接词(however, but, when 等)即可,所以正确率也比较高。另一方面,ScoreItNow!的标记十分机械。比如 ScoreItNow!会把所有中间段的第一句话作为分论点,所有除分论点之外的内容作为支持论点。ScoreItNow!的这种标注在多数情况下能够做到准确,因为多数考生的作文结构比较固定:一般都是 4~5 段论,其中第一段为开头,最后一段为结尾,剩下的 2~3 个段落为主体段。在主体段中,多数的文章也都是第一句为主题句,后面的内容为支持观点。但是一旦文章的写法超出 ScoreItNow!预期的结构,这种机械的标记则有失准确。比如,ScoreItNow!会将文章的最后一段都标记为总结段落。为了测试 ScoreItNow!是否能够准确识别出最后一个段落是否为总结段,我们将教师 2 的 Issue 文章删掉结尾段再提交。这时,文章的最后一段实际上是一个主体段,但是 ScoreItNow!还是将最后一段标记为总结段。另外,ScoreItNow!对于总论点的标注也不总是准确的。当开头段中明确出现了“as far as I am concerned”等表示观点的表达时,ScoreItNow!能够准确识别。但如果没有明确的标志词或者词组,ScoreItNow!的判断就会有偏差。比如,“In the argument, the author concludes that the television station should restore the time devoted to weather and local news to its former level. While the conclusion might seem reasonable, the argument is based on a series of unsubstantiated assumptions. In order to better evaluate the argument, the author needs to provide us with more evidence.”这是 argument 作文的开头段。ScoreItNow!将这一整段的内容都标记为总论点,显然是不准确的。

除此之外,本研究还发现,文章中有一些内容 ScoreItNow!并没有标记。例如,“In this case, few complain cannot lead the conclusion that majority of them are happy with the change.”在这一句话中 ScoreItNow!指出了缺乏冠词的问题,但是并未指出句子中其它的语法问题。再比如,“Classes such as language and math, which is fundamental for almost every field...”这句话中的语法错误 ScoreItNow!也未能识别出来。

5. 结论

通过研究发现,ScoreItNow!给出的分数和实际考试得到的写作分数基本一致。但是 ScoreItNow!评分后给出的反馈不够准确。本文作者猜测这主要是由于 ScoreItNow!反馈不够准确的项目,例如语法,行文规范等,在评分时所占权重较低,所以即使准确度不够高,也不会造成最后分数上的过大偏差。

研究还发现 ScoreItNow!主要对文章的语言进行评价,无法对文章的内容进行准确判断,并且评分时受文章长度影响较大。理论上只要作文中间段篇幅足够,并且尽量避免语法错误,即使文章内容答非所问,句子之间没有太强逻辑关系,e-rater 也能够打出高分。

但要注意的是,实际考场上由于还有人类评分员对作文进行评分,因此考生不大可能去写完全与题目无关的文章。从这个意义上来说,e-rater 仍然可以胜任对考试作文打分的工作。

6. 局限

本研究选取的样本数量较少, Issue 和 Argument 文章数量分别是 4 篇, 并且只有一名学生参与到研究中。此外, 几位参与者, 除了学生 1 之外, 写作的分数差别不大, 所以没有很好地研究 ScoreItNow! 对于不同分数段或者不同写作方式在给分上的差异。特别是缺乏 2.5~3.5 分这一区间的文章, 而 2.5~3.5 分是中国考生 GRE 写作部分最常见的分数[7]。

7. 对教学的启示

本研究得出的结论对 GRE 写作教学也有一定的指导意义。GRE 写作教师在教学的过程中可以适当跟学生强调语言表达的重要性。这里所说的语言表达, 主要是指语言表达上的准确性, 即尽量避免语言错误。也可以建议学生在保证作文质量的前提下, 适当加长文章主体段落的篇幅。另外, 对于不同水平的学生, 教师也可以适当补充一些不同难度的逻辑连接词。这些都可以一定程度上帮助学生在 e-rater 评分时获得一些优势。

当然, 除了语言和篇幅之外, 教师应该更关注 GRE 作文的写法, 教授学生真正实用的方法, 提高学生思维的深度, 让学生不仅能够 GRE 考试中取得高分, 更能够在出国之后的学习中得心应手。

参考文献

- [1] Educational Testing Service (2016) The Official Guide to the GRE General Test. 3rd Edition, McGraw Hill Education, 11-41.
- [2] Attali, Y. and Burstein, J. (2005) Automated Essay Scoring with e-rater v.2.0 (ETS RR-04-45). ETS, Princeton, NJ. <https://doi.org/10.1002/j.2333-8504.2004.tb01972.x>
- [3] 新东方教育科技集团国外考试推广管理中心. 揭秘 GRE——从出题到评分[M]. 北京: 海豚出版社, 2018.
- [4] ScoreItNow!™ Online Writing Practice Service for the GRE® General Test. <https://dxrgroup.com/scoreitnow>
- [5] Breyer, F.J., Rupp, A.A. and Bridgeman, B. (2017) Implementing a Contributory Scoring Approach for the GRE® Analytical Writing Section: A Comprehensive Empirical Investigation (ETS Research Report No. RR-17-14). Educational Testing Service, Princeton, NJ. <https://doi.org/10.1002/ets2.12142>
- [6] Ramineni, C., Trapani, C.S., Williamson, D.M., Davey, T. and Bridgeman, B. (2012) Evaluation of e-rater® for the GRE® Issue and Argument Prompts (ETS Research Report No. RR-12-02). Educational Testing Service, Princeton, NJ. <https://doi.org/10.1002/j.2333-8504.2012.tb02284.x>
- [7] Educational Testing Service (2019) A Snapshot of the Individuals Who Took the GRE® General Test July 2014-June 2019. Princeton, NJ.