

【质量与评价】

新高考选考科目等级赋分 效果的实证分析

——以 H 省新高考方案实施前后的考试数据为例

唐未兵 杨志明

【摘要】新高考方案的突出优点之一是增加了学生选考的自由,但相应的学业水平考试选考科目的等级赋分规则引发过一些争议。对 H 省实施“3+1+2”高考新方案前后两年数据的分析发现,所有选考科目的人数比例都足够大,没有触发等级赋分保障机制;而且各选考科目的得分分布接近正态分布,没有发生低估或高估考生等级分数的问题。这个实证结果不仅肯定了处理问题过程中综合施策的经验,而且为进一步完善选考科目等级赋分规则提供了有益的思路。

【关键词】新高考;学业水平考试;等级赋分

【作者简介】唐未兵,湖南师范大学教育科学学院/马克思主义学院教授,博士生导师,湖南省教育考试院原院长(湖南 长沙 410081);杨志明,湖南师范大学测评研究中心主任,外国语学院教授,博士生导师(湖南 长沙 410081)。

【原文出处】摘自《湖南师范大学教育科学学报》(长沙),2023.1.114~119

【基金项目】国家教育考试科研规划 2019 年度重点课题“高考改革促进基础教育与高等教育衔接贯通的作用研究”[GJK2019010]。

根据 2014 年 9 月颁布的《国务院关于深化考试招生制度改革的实施意见》^[1],我国先后推出了“3+3”和“3+1+2”高考新方案,目前已经覆盖了大部分省市^[2]。新方案的突出特点是赋予了学生一定的选考自由,并实施了高中学业水平选择性考试等级赋分制度等^[3]。有调查显示,选考自由政策深得老百姓认可^[4],但其等级赋分方法引发了一些争议^[5-9],部分省市在高考新方案试点之初曾发生过大量学生弃考物理或化学等问题^[10-13]。为了解决这些问题,相关管理部门先后出台了等级赋分保障机制、物理或历史科目必须而且只能二选其一、大量理工类专业要求必考物理等措施,取得了显著成绩。不过,关于等级赋分效果的实证研究目前并不多。因此,很有必要对高考的真实数据做一些分析,以评估等级赋分方案的效果,确保高考改革的稳妥推进。

一、问题的提出

现代教育测量理论认为,使用科学的等级赋分

规则,可以确保考生总体的等级分数基本服从正态分布^[14-17],为不同选考科目组合总分之间的直接比较、保障各科成绩在总分中的预设分量、有效减轻或消除考试太难或太易可能带来的负面影响等,提供了坚实的理论基础。不过,要获得科学的等级分数,就必须满足一定的前提条件。其中,选考群体背后的考生总体的能力水平基本为正态分布就是实施等级赋分的前提条件之一,也是国际考试行业的惯用做法。若因为某些原因导致了选考群体或样本的能力水平为非正态分布,则无论考试命题的质量如何,选考群体的得分分布极有可能表现为极端的负偏态分布(选考科目太难吓跑了大量低水平考生)、或极端的正偏态分布(选考科目不难吸引了太多低水平考生)。

有研究表明,若根据选考群体极端负偏态分布确立等级赋分规则,则会低估大约 66% 的中等水平考生 1 分~3 分的等级分数;若根据选考群体极端正偏态分布确立等级赋分规则,则会高估大约 84%

的中等水平考生3分~6分的等级分数^[14]。这不仅会影响考试的科学性和公平性,影响高校的人才选拔质量,而且会影响中学的教学工作,导致选考选考中的田忌赛马等问题。

大量研究表明,在假定全省考生能力水平基本为正态分布的条件下,只要能从选考群体中构造出一些全省代表性小样本,然后依据这些小样本确定等级赋分规则,并应用在全省所有选考学生身上,则无论选考群体的能力水平分布如何,都可以减轻甚至杜绝低估或高估考生等级分数的问题^[14-17]。事实上,在全省考生总体能力水平为正态分布的条件下,其代表性小样本的得分分布一般会接近正态分布。于是,只要用于确定等级赋分规则的样本为正态分布,就会自动消除低估或高估考生等级分数的问题。换句话说,正态分布是确保等级赋分质量的前提条件。

本研究依据H省新高考方案实施前后的考试数据对新高考选考科目等级赋分效果进行实证分析,目的在于考察等级赋分保障机制的科学性和实施效果,并评估各选考科目的得分分布在多大程度上满足了等级赋分的前提条件。

二、等级赋分保障机制的分析

1. 等级赋分规则概述

第一和第二批参与高考改革试点的省市采用的是“3+3”新高考方案,即语文、数学和英语3科为必考科目,另外3科从政治、历史、地理、物理、化学和生物(浙江含信息技术)中任选。必考科目一般由教育部统一命题,使用原始分数,每科满分150分。选考科目一般由各省自主命题,使用等级分数,每科满分一般为100分(上海为70分)。第三批及其随后的高考改革试点省市普遍采用了“3+1+2”的新高考方案。^[18]其中,“3”与“3+3”方案中的第一个“3”含义相同;“1”指的是从物理或历史两科中必选其一的科目,由各省自主命题,使用原始分数,每科满分100分;“2”指的是从化学、生物、政治、地理中任选2科,由各省自主命题,使用等级分数。

关于等级赋分规则,教育部《关于普通高中学业水平考试的实施意见》规定:“计入高校招生录取总成绩的学业水平考试3个科目成绩以等级呈现,其他科目一般以‘合格、不合格’呈现。以等级呈现成绩的一般分为五个等级,位次由高到低为A、B、C、D、E。原则上各省(区、市)各等级人数所占比例

依次为:A等级15%,B等级30%,C等级30%,D、E等级共25%。E等级为不合格,具体比例由各省(区、市)根据基本教学质量要求和命题情况等确定。”^[3]

2. 等级赋分保障机制的数理分析

在高考改革试点过程中,由于出现过大量考生弃考部分较难科目的问题^[10-13],因此,部分省市出台了等级赋分保障机制。比如,采用“3+3”模式的浙江省规定:当选考科目(如物理等)的考生人数不足6.5万人时,各个等级人数比例由高到低按母为6.5万人的基数计算。采用“3+1+2”模式的江苏省规定:当选考科目(如较难的化学)的考生人数比例低于全省考生总数的25%时,以全省考生总数25%的人数为基数实施等级赋分。

显然,这种保障机制仅仅缓解了选考人群低水平考生比例太少情况下低估中等水平考生成绩的问题,没有解决选考人群低水平考生比例太大情况下高估中等水平考生成绩的问题,并新添了选考人数极少时水平极低者可以轻松获得高分的问题。以浙江省的等级赋分保障机制为例,由于保障基数为6.5万人,因此当选考人数在5万人以下时,没有任何知识基础的考生都一定可以得到61分以上的分数;若选考人数不足3万人,则问题会更加突出,因为任何考生即使是胡乱作答也一定可得73分以上的成绩。

上述结果表明,在选考人数特别少的情况下,现行的等级赋分保障机制存在着高估中低水平考生等级分数的问题,给考试的科学性和公平性带来了较大隐患。

3. 等级赋分保障机制的实证分析

根据高考数据,可以直接判断H省是否需要使用等级赋分保障机制的问题。

数据分析可知,由于物理和历史科目在“3+1+2”高考模式中属于二选一的必考科目,所以其选考人数与新高考方案实施之前的2020年人数比例比较接近,分别为55.53%和44.45%,理科选考比例略为下降了3.53%,但没有出现大面积弃考物理的情况。在另外4门任意选考科目中,化学比较难,而趋易避难往往是学生的天性,因此需要特别关注是否会出现大量考生弃考化学等某个学科的问题。而真实的数据显示,全省有44.56%的考生选考了化学,应考人数达到了16万5千多人,远远超过了

人们预期的6万5千人的基数。同样的结果也在生物、政治和地理学科上得到了验证。

上述结果说明,该省2021年实施“3+1+2”高考新方案时,全省所有选考科目的人数都足够多而且分布均匀,没有出现大规模弃考某个特定较难科目的问题,所以不需要使用等级赋分保障机制,避开了使用保障机制有可能高估极低水平考生等级分数的陷阱。

值得特别注意的是,并非所有省市的化学选考人数都达到了H省的水平。有研究指出,J省的化学选考人数比例由改革前的45.9%下降到了38.5%(没有触发J省的等级赋分保障机制),而且化学考生中重点中学的学生比例由改革前的68.1%提升到了75%,这种分布极有可能导致低估相当一部分中等水平考生的化学成绩问题^[13]。这个问题需要所有省市密切关注。

三、等级赋分前提条件的检验

1. 选考科目得分的偏态量数分析

调查显示,在实施新高考方案之前的2020年,该省所有考试科目的得分分布都基本接近正态分布(略微负偏态)。其中,语文、理科数学、文科数学、英语、理科综合和文科综合科目考试得分的偏态量数分别为-1.37,-0.54,-0.11,-0.18,-0.33和-0.73,只有语文学科考试得分分布的负偏态情况明显。这说明语文学科考试得高分的人数比正态分布所期望的人数更多一些,而其他学科考试得高分的人数仅比正态分布所期望的人数略微多一些,但总体表现都很正常。

在实施“3+1+2”高考新方案第一年的2021年,所有必考科目和选考科目的得分分布也都十分接近正态分布(部分科目得分略微负偏态,部分科目得分略微正偏态)。其中,语文、数学、英语、物理、历史、化学、生物、政治和地理科目考试得分的偏态量数分别为-1.17,-0.30,-0.25,0.16,-0.06,0.15,-0.25,-0.54和-0.13,也只有语文学科考试得分分布的负偏态情况明显。这说明语文学科考试得高分的人数比正态分布所期望的得分人数要多一些,数学、英语、历史、生物、政治和地理学科考试得高分的人数比正态分布所期望的人数略微多一些,而物理和生物学科考试得高分的人数比正态分布所期待的人数略微少一些,但所有学科的得分表现都非常正常。

此外,各科考试成绩的差异系数表明,除了语文学科差异系数偏小(12.48%和12.75%),其他所有学科差异系数都达到了较高的水平。在2020年的数据之中,差异系数最大的4门科目由高到低依次为文科数学、英语、理科综合和理科数学,分别为39.72%,34.25%,30.55%和30.45%。在2021年新高考中,差异系数最高的4门科目由高到低依次为数学、物理、英语和化学,分别为38.51%,37.64%,36.54%和34.06%,这说明这些科目具有非常好的区分效果。

2. 选考科目得分的分布图分析

分析可知,H省2021年4门选考科目(化学、生物、政治、地理)的得分分布没有出现极端偏态分布情况。因此,按照等级赋分规则对这4门选考科目做等级赋分换算,不会出现严重高估或低估考生等级分数的问题。

四、结论与启示

本次实证研究得到以下三点结论:

第一,该省2021年的等级赋分效果良好,没有触发保障机制。实证数据分析显示,该省2021年所有科目的选考人数都足够大,而且分布均衡,没有触发等级赋分保障机制。

第二,“3+1+2”新高考方案有效遏制了大量学生弃考物理的问题。数据分析显示,H省实施高考新方案前后两年的物理考生人数比例基本一致,人数足够多,均达到了一个理性的水平,弃考物理的现象没有发生。

第三,“3+1+2”新高考方案没有出现大规模弃考化学或某个单科的问题。根据国际经验,在选考自由政策下,避难就易的问题在所难免。比如,美国高考科目的选择完全是自由的,但选择科学、技术、工程和数学(STEM)的学生数就非常少。令人欣慰的是,本研究中H省所有选考科目的人数不仅数量足够大、占比合理,没有触发等级赋分保障机制,而且各科考试分数都服从正态分布,完全满足了等级赋分的前提条件,保障了等级分数的科学性和公平性。

与此同时,本次实证研究也得到了如下一些启示。

第一,等级赋分问题的解决需要综合施策。单纯靠行政管理手段或教育测量学手段并不能有力应对等级赋分过程中所有可能遇到的问题。事实

上,单凭行政手段设置保障机制的做法只能缓解低估中等水平考生等级分数的问题,不能解决高估中低水平考生等级分数的问题。完全照搬国际考试行业处理等级赋分误差的“标准设定”与“测验等值”相结合的测量学方法,也因为在新题试测和考后保密等方面存在极大困难而无法实现。因此,把行政力量与考试科学相结合所研发的“基于全省考生总体的代表性小样本分数分布确立等级赋分规则”的方法^[14]值得参考。

第二,设置正态分布为等级赋分的前提条件,可有效控制高估或低估考生等级分数的风险。本次等级赋分方案的成功实施主要是因为所有选考科目的人数足够大、其原始得分都接近正态分布,满足了等级赋分的前提条件。一旦这个前提条件不具备,则不应当直接根据应考群体的得分分布转换等级分数。

第三,对物理和历史使用等级分数的方案值得考虑。由于物理或历史为必考科目,所以考生的人数普遍较多,其相应的得分分布也会基本接近正态分布,而基于正态分布的等级赋分规则不会高估或低估考生的等级分数,所以对这两科使用等级分数的方案值得考虑。这样做的主要优点是有利于高考总分的合成,可以保障物理或历史在高考总分中的应有地位,避免出现因考试太难或太易而引发的各种负面影响等。

第四,建设高素质命题队伍以确保命题质量是保障等级赋分质量的关键环节。H省2021年首次实施新高考,选考科目的原始得分分布非常接近正态分布,表明其命题质量达到了较高水平,说明其学科命题队伍,包括学科秘书队伍的建设受到了应有的重视和训练,这个经验值得同行借鉴。

总之,H省首次实施“3+1+2”新高考方案的等级赋分效果良好。不仅所有选考科目的报考人数足够多、分布均匀、没有触发等级赋分保障机制,而且所有选考科目的得分都十分接近正态分布,满足了实施等级赋分的前提条件,没有出现高估或低估考生等级分数的问题,说明该省的考试命题及相关的管理工作比较有效。当然,这仅仅是基于一个省份2021年新高考数据分析的结果,未来还需要积累更多更广的证据来评估等级赋分的效果。

参考文献:

- [1]国务院.关于深化考试招生制度改革的实施意见[EB/OL].(2014-09-03).http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_1778/201409/174543.html.
- [2]最新版2023年高考改革新政策解读[EB/OL].(2022-07-29).<https://www.liuxue86.com/a/4331313.html>.
- [3]教育部.关于普通高中学业水平考试的实施意见(教基二[2014]10号)[EB/OL].(2014-12-10).<http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s4559/201412/181664.html>.
- [4]新高考改革研究课题组.沪浙新高考改革4年效果如何?等级赋分技术待改进[EB/OL].(2018-07-13).www.xinhuanet.com/politics/2018-07/13/c_1123119079.htm.
- [5]杨志明.学业水平考试成绩等级化中的风险及其规避办法[J].教育测量与评价(理论版),2015(9):62-64.
- [6]杨志明.做好高中学业水平考试所必须的四项测量学准备[J].中国考试,2017(1):8-13.
- [7]温忠麟.新高考选考科目计分方式探讨[J].中国考试,2017(12):23-29.
- [8]章建石.关于选考科目等级赋分的改进:历史经验、现实限制与可能方向[J].华东师范大学学报(教育科学版),2018(3):79-86,168.
- [9]于涵,韩宁,关丹丹,等.关于改进新高考选考科目赋分方案的若干思考[J].中国高教研究,2018(6):44-49.
- [10]朱邦芬.为什么浙江省高考学生选考物理人数大幅下降值得担忧[J].物理,2017(11):761-763.
- [11]陈爱文,胡银泉.尴尬的物理:浙江新高考下的学科失衡与制度改进[J].中小学管理,2017(9):19-21.
- [12]浙江省教育厅办公室.关于纠正部分普通高中学校违背教育规律和教学要求错误做法的通知(浙教办基[2016]97号)[EB/OL].(2016-10-28).www.51gaoxiao.com/zhejiang/zixun/93053.html.
- [13]李青春.江苏新高考改革背景下学生科学学科态度实证研究——以化学学科为例[J].江苏高教,2022(9):76-80.
- [14]杨志明.基于大数据的学业水平选考科目赋分方案[J].教育测量与评价,2019(1):3-10.
- [15]罗冠中,罗照盛,夏园林.新高考选考科目赋分方案的数据模拟研究[J].中国考试,2020(4):17-21.
- [16]DORANS N J. Recentering and realigning the SAT score distributions:How and why[J]. Journal of Educational Measurement, 2002(1):59-84.
- [17]PETERSEN N, KOLEN M, HOOVER H D. Scaling, normalizing, and equating. In: Lin, R. L., Ed. Educational measurement (3rd ed.) [M]. New York: Macmillan, 1995:221-262.
- [18]唐未兵.全力做好高考改革后的普通高校招生考试与录取工作[J].湖南教育,2019(4C):33-35.