

Research on the Effects of Question Formulation on the Test Score

Zhenzhu Wu^{1,2}

¹The Education Examinations Authority of Guangdong Province, Guangzhou Guangdong

²Nyingchi Municipal Bureau of Education, Nyingchi Tibet

Email: 58239962@qq.com

Received: Jun. 15th, 2020; accepted: Jun. 30th, 2020; published: Jul. 7th, 2020

Abstract

By two different tests, we want to find the effect of question formulation on the investigate ability, test score, item difficulty and degree of differentiation, to explore the strategy of how to create high-quality question in biology paper for university entrance examination, and to provide reference to create high-quality question.

Keywords

Question Formulation, Ability, University Entrance Examination, Biology Paper

设问方式对试题得分影响的研究

吴珍珠^{1,2}

¹广东省教育考试院, 广东 广州

²林芝市教育局, 西藏 林芝

Email: 58239962@qq.com

收稿日期: 2020年6月15日; 录用日期: 2020年6月30日; 发布日期: 2020年7月7日

摘要

通过实测两道测试题, 了解不同设问方式对能力考查、试题平均分、试题难度和区分度等方面的影响, 探索生物科优质高考试题编制策略, 为生物科优质高考试题的编制提供参考。

关键词

设问, 能力, 高考, 生物试题

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. 目的和意义

新课改后的高考试题以能力立意, 重视对能力的考查, 突出对人才选拔的作用, 尤其是非选择题, 更加重视对考生学科能力和科学素养的测量。进行基于能力维度的生物科优质高考非选择题编制策略的探究, 对于进一步完善我国优质生物学高考试题的编制方法和策略具有理论意义, 对于提高我国生物学试题质量和命题水平具有应用价值。本文笔者设计了两道测试题进行实测, 以更好地了解非选择题中不同设问方式对能力考查力度、试题平均分、试题难度和区分度等方面的影响, 为生物科优质高考试题的编制提供参考。

2. 研究内容与方法

2.1. 测试题

试题分为测试题 I 和测试题 II, 均为单选题, 测试题 I 即为 2010 年广东理综高考第 28 题[1]原题。与测试题 I 相比, 测试题 II 有两处不同:

一是将测试题 I 的第(2)题由填空题变为简答题, 即将“导致上述男孩患克氏综合征的原因是: 他的_____ (填“父亲”或“母亲”)的生殖细胞在进行_____分裂形成配子时发生了染色体不分离。”改为“导致上述男孩患克氏综合征的原因是什么?”

二是将测试题 I 的第(3) ①题由判断题变为计算题, 即“他们的孩子中是否会出现克氏综合征患者?”改为“他们的孩子患克氏综合征的几率是多大?”

2.2. 测试对象

测试对象共 502 人, 分别来自广州、市、云浮、韶关、肇庆、四所学校的高三学生群体。参与测试题 I 的学生共有 253 人, 参与测试题 II 的学生共有 249 人。

2.3. 研究方法

闭卷, 测试 10 分钟。每个班以小组为单位, 将 1 个班分成两个不同群体, 例如: 1 个班有 8 小组, 第 1、3、5、7 组做测试题 I, 第 2、4、6、8 组做测试题 II。数据统计: 采用 EXCEL2003 软件输入所有学生的数据, 对试题的平均分、标准差、难度、区分度进行统计, 进行差异显著性检验, 进行比较分析。

3. 测试结果

从表 1 可知, 与测试题 I 相比, 测试题 II 中的各小题及该大题总分都低, 尤其是第(2)小题的平均分低很多, 此小问变为偏难题; 各题的区分度值都高于 0.4, 区分效果都优; 试题难度提高, 没有影响到试题区分度下降。本研究的 T 检验结果显示, 第(1)小题、第(4)小题在两套测试题中设问均相同, 平均得分差异均不显著, P 值分别为 0.137 和 0.318 ($P > 0.05$), 意味着测试题 I 和测试题 II 的学生群体可视为等值样本, 对不同设问形式对得分影响的研究提供了平等的前提。第(2)小题、第(3) ①问在两套测试题中设问方式不同, 平均得分差异非常显著($P < 0.001$), 说明设问方式改变对试题的得分影响大; 测试题 I 和测试题 II 的整体平均得分差异也非常显著($P < 0.001$), 表明某些小题的设问方式不同, 会直接显著影响整

道大题的平均得分；测试题 I 和测试题 II 的第(3) ②问虽然相同，但两套卷的平均得分差异很显著($P = 0.001$)，分析可能与第(3) ①问的改变相关。

Table 1. The average score, standard deviation, difficulty, discrimination and T-test results of test I and Test II
表 1. 测试题 I 和测试题 II 的平均分、标准差、难度、区分度以及 T 检验结果

| 题号 | 测试卷 | 满分(分) | 平均分 | 标准差 Sn | 难度 | 区分度 | T 检验(P) |
|--------|-----|-------|-------|--------|------|------|---------|
| (1)题 | I | 6 | 4.90 | 1.66 | 0.82 | 0.81 | 0.137 |
| | II | 6 | 4.66 | 1.74 | 0.78 | 0.66 | |
| (2)题 | I | 4 | 3.47 | 0.95 | 0.87 | 0.65 | 0.000 |
| | II | 4 | 1.76 | 1.52 | 0.44 | 0.66 | |
| (3) ①问 | I | 2 | 1.83 | 0.55 | 0.92 | 0.40 | 0.001 |
| | II | 2 | 1.60 | 0.82 | 0.80 | 0.47 | |
| (3) ②问 | I | 2 | 1.72 | 0.70 | 0.86 | 0.45 | 0.011 |
| | II | 2 | 1.54 | 0.84 | 0.77 | 0.57 | |
| (4)题 | I | 2 | 1.45 | 0.89 | 0.72 | 0.68 | 0.318 |
| | II | 2 | 1.37 | 0.93 | 0.69 | 0.58 | |
| 本大题 | I | 16 | 13.37 | 3.10 | 0.84 | / | 0.000 |
| | II | 16 | 10.93 | 3.54 | 0.68 | / | |

$P < 0.05$ ，表示差异显著； $P < 0.01$ ，表示差异很显著； $P < 0.001$ ，表示差异非常显著。

4. 讨论

测试题 I 的第(2)小题：“导致上述男孩患克氏综合征的原因是：他的_____ (填“父亲”或“母亲”) 的生殖细胞在进行_____ 分裂形成配子时发生了染色体不分离。”；测试题 II 的第(2)小题：“导致上述男孩患克氏综合征的原因是什么？”。不同测试题中，第(2)小题都是要求学生回答“导致患病的原因”，设问方式不同，对能力考查的侧重点也有所不同。测试题 I 采用填空形式，要求学生简单判断疾病何时、由谁传递到子代，考查学生运用所学知识进行简单分析、判断的能力。测试题 II 采用简答形式，没有任何提示，要求考生自己独立思考、分析原因后再组织语言阐述原因，考查了学生运用知识分析问题、文字组织表达的能力，能力层次要求高，一部分学生会因为没有提示无从下手而丢分，另一部分会因为答不全原因而丢失部分分值，还有少部分学生因为不能组织好语言进行作答而丢分，导致测试题 II 中的第(2)题得分比测试题 I 中对应小题的得分低了一半。但是，测试题 II 中第(2)小题的标准差明显高于测试题 I 中对应的小题，说明其更能将考生得分差距拉大，考生得分更离散，有利于区分学生群体；从能力考查力度来看，测试题 II 的第(2)题对能力考查力度更大；从整道题来看，只有测试题 II 的第(2)题，难度偏难，其他各小题的难度偏易，从难度分布来看，测试题 II 的第(2)题可使整道题的难度分布更呈现梯度，难度分布更为合理；从题型分布来看，测试题 II 设问方式更为丰富，包括作图题、简答题、判断题、计算题和填空题五种类型。由此可见，不论从能力考查力度、难度分布、分值离散程度还是设问方式丰富度等方面，将第(2)小题的简单填空改为简答题，更有利于试题对综合人才的选拔。

测试题的第(3)题包括了两问，测试题 I 和 II 这两套题中，只有①问对能力考查的侧重点不同：即测试题 I 第(3) ①问为判断题，主要考查学生的预判、推理和分析能力，学生只要回答是与否，不需要考生进行计算，作答时间相对小些，而且对于一些对知识掌握不牢、生物素养弱的学生而言，即便不知原理但猜对率可高达 50%，平均分相对较高。而测试题 II 第(3) ①问为计算题，主要考查计算推理能力，要

求考生计算出患病的概率，所需作答时间要更长些，计算能力弱的考生可能会放弃作答此题，导致平均分下降。

虽然第(3) ②问在测试题 I 和测试题 II 中相同，但答题结果受到了第(3) ①问设问不同影响，得分差异很显著，可见同一道小题中问与问之间息息相关，前一小问的设问方式改变，会影响到后一小问的得分率。测试题 I 中的第(3) ①问采用判断题形式，学生通过预估、猜测等方式就可以轻松地解答，为第②问提供了轻松的作答环境，而且通过简单的预判得知后代中不会出现克氏综合征患者后，在后面的计算过程中，学生不再将两种疾病搅在一起思考问题，只要考虑单一疾病的计算概率就可，这样使第②问的解答简单化。而测试题 II 中的第(3) ①问采用计算题形式，部分计算能力弱的考生望而却步，更不会去尝试去计算解答第②问了，另外其第①问的设问形式和考查能力与第②问重复，使能力考查的广度变窄。从考查能力更广、题与题之间的牵连度小、节省考生答题时间、考查能力不重复的命题角度来看，在两套题的区分度均好的前提下，测试题 I 第①问的设问方式更符合优质试题的标准。而且，本人认为，能否在短时间内快速对问题进行正确预判、预估能力，是人类生活中重要的能力之一，在高考试卷中值得提倡，本人更倾向采用测试题 I 第①问的设问方式来强化对能力的考查。

5. 结论

从本次测试结果来看，简答题比简单填空题的能力要求高，试题难度大，学生的计算能力相对较弱。以能力立意的高考优质非选择题命制过程中，建议要避免对同一种能力重复考查，能力考查覆盖率尽量要广，以更有利于选拔综合性人才；要避免前后问题相牵连，确保各小问既有联系，又相对独立。

基金项目

本文系全国教育科学规划教育部重点课题《基于能力维度的生物科优质高考试题编制策略探究》(课题批准号：GFA111017)的研究成果。

参考文献

- [1] 2010年广东省高考试卷(理科综合)[Z].