



我国台湾学科能力测验中自然科学选择题的设置及评价*

肖巧玲¹ 叶永谦^{2**}

(1. 福建教育科学研究所 福建福州 350003; 2. 福建省南安市侨光中学 福建南安 362314)

高考改革

摘要 我国台湾学科能力测验自然科学的题型全部为选择题,通过增加选项数目、增加多项选择题比例、改变多项选择题编排模式、改变计分方式、改变题量与总分关系等可以减少选择题存在的强猜现象、无法考查思维过程及无法顾及学生差异性不足,改善了选择题的考查评价功能,使考试更加公平科学,可以为我国大陆即将全面铺开的学业水平考试选择题部分的设计提供参考。

关键词 台湾学科能力测验 考试公平 自然考科 选择题 学业水平考试

DOI: 10.13884/j.1003-3807hxjy.2015120063

我国台湾地区自2005年起实行多元化入学考试制度,其中学科能力测验(以下简称“台湾学测”)是评估高中生是否具备进入台湾各大学学习的重要依据,与我国大陆的学业水平考试具有类似的评价功能。台湾学测包括国文、英文、数学、社会、自然等5个考科,其中自然考科包括物理、化学、生物、地球科学4个学科^[1]。自然考科早在1994年起就开始试行,考试时间为100 min,试卷为典型的标准化试卷类型,题型全部为选择题,包括单选题、多选题。

台湾学测是对台湾接受高中阶段教育的学生进行教育效果评价、学业水平认定的大型考试,具有较高的信度、效度、必要的区分度和适当的难度,保证了考试评价的程序公平。考试程序公平,即要尽量减小考试本身的系统误差,实现一碗水端平,对于所有的考生同等对待。Cleary(教育测量学专家)在1968年定义考试公平时,认为“如果一项考试不会系统地高估或低估某一组人,这个考试就是公平的”。由于大型考试评卷耗时费力,而单项选择题可以避免阅卷人对评分标准的理解不一致而造成差异,评分环节相对来说较为客观准确,并且可用机器阅卷,提高阅卷的速度和准确度,减少阅卷的劳动强度^{[2][14]},已然成为各类标准化考试的主要题型。台湾学测也不例外地以选择题作为试题的主要形式。

但是,选择题存在的局限性也显而易见。首先,选择题难以避免“强猜”现象。如单项选择题考查时,考生排除2个错误选项后,在其余2个选

项中做出猜测选择。“强猜”的结果可能导致对同一认知档次的2位考生评价截然不同,一个零分、一个满分。其次,选择题难以考查概念深刻理解、综合解题和逻辑推理等思维过程。思维过程难以考查,导致评价测量目标与教学目标无法达成一致,降低了考试的有效性。再次,选择题难以顾及考生的差异性。考生学习潜能、兴趣爱好等存在着不同程度的差异,但选择题无法尊重这些差异,影响了考试的公平性。而台湾学测多年来坚持以选择题作为唯一的试题类型,其如何规避这些局限性,值得研究与借鉴。力求挖掘台湾学测选择题设计的方式,从降低“强猜”率、提高思维过程性考查、增加选题机会等方面进行分析,诠释提高选择题命题公平性的策略。

1 降低“强猜”率的设计

选择题的强猜性,在单项选择题上表现尤为显著。在《选择题中的概率问题》中翟铁倪认为,在不同类型的选择题中,猜对的概率问题也有高低之分。甚至有学者认为,考查学生的学业水平并对其进行考核的时候,选择题这样的命题形式不提倡被采用,尤其是像正误判断题这样的题型,更是不提倡采用^{[3][42]}。台湾学测在试题的备选选项数量和选择题类型多样化上颇具特色。

1.1 增加备选选项数目

吴庭光在概率统计基础上提出,如果一定要采用选择题这种命题形式,建议每道选择题的备选选项个数一定要在5个以上,这样才能使考生由于猜

* 全国教育科学“十二五”规划2014年度单位资助教育部规划课题“闽台高校招生学业水平考试设计的比较研究”(课题批准号:FJB140101);福建省教育科学规划科研基地专项课题“闽台高校招生学业水平考试化学学科比较研究”(课题批准号:JAS151453)

** 通信联系人, E-mail: 15880968680@139.com

测而取得及格成绩的几率低于百分之一,这样一来考生考试成绩才能更加接近于考生的真实水平^{[3]13}。从1994年到现在,台湾学测自然考科单项选择题的备选选项数目总体呈现由4个为主过渡到以5个为主的规律(图1),其中1994、1995年备选选项数目5个及以上的题目占当年单项选择题的5%以内,2011年以来比例达到88%以上,有3年达到100%。

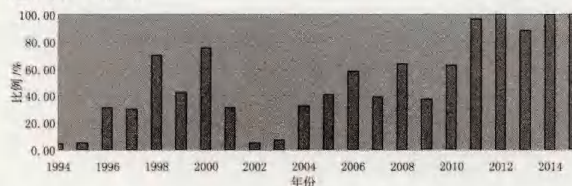


Fig. 1 Percentage of questions with 5 or more options since 1994

图1 1994年以来备选选项数目5个及以上的题目占当年单项选择题比例

备选选项数目为5个选项,使试题正确答案的猜中率降低到1/5。如2015年自然考科第7题,考查的是氧化还原反应,选项有5个(例1)。更有甚者,有个别年份单项选择题中个别题目备选选项数目超过5个,其中2004年备选选项数目为6个的有4道题,如第32题,考查的是氧元素守恒问题,选项有6个(例2)。2009年有一道题目甚至达到7个选项。

【例1】下列反应,哪一项化学变化属于氧化还原反应?

- (A) 铅蓄电池的放电反应
- (B) 加热氯化铵与氢氧化钙混合物产生氨气
- (C) 在高温下,碳酸钙分解产生氧化钙与二氧化碳
- (D) 硝酸银水溶液与溴化钾水溶液混合产生溴化银沉淀
- (E) 将浓硫酸缓慢加入氯化钠粉末,产生的气体经干燥后得氯化氢,及留于溶液中的硫酸钠或硫酸氢钠

【例2】水中的微生物会使水中的有机物(例如 $C_6H_{10}O_5$)转化为 CO_2 与 H_2O ,在这过程中所需 O_2 的量,称为生化需氧量(BOD)。试问要使水中的1个 $C_6H_{10}O_5$ 分子完全变成 CO_2 与 H_2O ,需要几个氧分子?

- (A) 1 (B) 2 (C) 3 (D) 4
- (E) 5 (F) 6

纵观台湾学测自然考科单项选择题选项数目的设置,备选选项数总体呈现4个到5个反复出现的

现象。近5年来,则以5个选项为主,有效减少了强猜的几率,使评价更加公平。

1.2 设置多项选择题

与单项选择题不同,多项选择题的猜中率明显较低。在《多项选择题的编写》中宋锡全、王素英认为:多项选择题的优点除了能够增加知识的覆盖度外,其评分较为客观准确,学生在不选错选项的前提下,有可能会得到中间分数。在5个选项是等可能设置的条件下,其猜测正确的概率仅为 $1/26$ ^{[4]31}。因此,在试卷中设置适当的多项选择题,可以有效地降低选项被猜中的概率。台湾学测自然考科第1部分分为单选题、多选题、综合题3种类型;第2部分为多选题。从整份试卷的角度看,单选题只有第1部分,其余的多选题、综合题及第2部分题目均为多选题。统计近5年台湾学测自然考科多选题所占的分值(表1),可以看出,多项选择题的考查比例总体呈现增加的趋势。特别是近2年,甚至达到64.06%的高比例,可见台湾对于多项选择题考查的重视。多项选择题的分值比例越大,整体试卷的猜测度越低,考查的准确性越高。

表1 近5年台湾自然考科多项选择题分值及所占比例情况

Table 1 Scores and percentages of multiple-choice questions in Taiwan natural science test in the last five years

年份	2011	2012	2013	2014	2015
多项选择题分值	68	72	68	82	82
分值比例/%	53.13	56.25	53.13	64.06	64.06

宋立新等认为,在标准化考试中,单项选择题和多项选择题所规定的给分规则中完全选对所给分数相同的时候,应设计单项选择题的备选选项难度大些,多项选择题的备选答案简单些;同理,如果在单选题下备选答案难度与多选题的相当,那么单选题所规定满分值应该大于多项选择题所规定的满分值^{[4]32}。台湾学测试卷自然考科也表现出当备选答案难度大致相同时,多项选择题的难度明显大于单项选择题的特点。为了保持试卷的难度不会因为多项选择题的存在而出现太大的提高,多项选择题题干后均注明正确的选项数目。如2014年试卷中的第32题,题后注明应选2项(例3)。同时,不同题目的正确选项数目不尽相同。多项选择题总体属于不定项选择题,但具体到某一道题目却属于定项选择题。通过这种设计,既提高了考查的准确度,又可以控制试卷的整体难度。

【例3】将100.0 mL、0.40 mol/L的HCl溶液加于4.24 g的 Na_2CO_3 固体,会产生气泡。下

列关于此反应的叙述,哪些正确?(应选2项)

- (A) 此反应的平衡反应式为: $\text{Na}_2\text{CO}_3 + \text{HCl} \rightarrow 2\text{NaCl} + \text{H}_2\text{O} + \text{CO}_2$
- (B) 若反应完全,则可产生 0.88 g 的 CO_2
- (C) 反应后会剩余 0.01 mol 的 Na_2CO_3
- (D) 此反应的限量试剂为 HCl
- (E) 此反应为沉淀反应

2 提高思维过程性考查的设计

传统测量理论认为,选择题无法考查出思维的过程。但是,从现代测量理论中的认知诊断角度看,如果对多项选择题的干扰项包含的信息进一步诊断,有助于提高估计被试知识状态的准确性。学生的学习过程除了知识和技能外,还有当新知识与旧知识结合后向更高水平发展的过程,这种过程可以形成更深的理解。在设计多项选择题选项时,如果能够结合学生认知发展的有序水平进行设计,将选项与不同的认知水平相结合,再根据学生的答题情况诊断学生的掌握程度,就可以找出学生具体的弱项^{[2]175}。因此,通过适当的编制方法和计分方式,多项选择题可以一定程度考查出思维的过程性。

2.1 多项选择题编制过程中注重思维的过程性

研究表明,高质量的多项选择题不但可以考查考生的知识,还可以考查考生的理解、解释、应用、分析和综合能力^{[2]142}。台湾学测多项选择题是如何实现对思维过程的考查呢?

以2014年第32题(例3)为例,该题以“将含有 0.04 mol 溶质的盐酸滴到 0.04 mol 的碳酸钠固体,会产生气泡”为信息,分别考查反应方程式的书写(A选项)、反应产物的计算(B、C、D选项)、反应类型的判断(E选项)等方面的知识。从考查思维过程性的角度认识这道题目,可以将其改编为一道填空题,具体题目如例4:

【例4】将 100.0 mL、0.40 mol/L 的 HCl 溶液加于 4.24 g 的 Na_2CO_3 固体,会产生气泡。

(1) 请写出此反应的化学方程式: _____;
该反应是否属于沉淀反应? _____ (填“是”或“否”)。

(2) 若完全反应,则可产生 _____ 克的二氧化碳;反应后会剩余 _____ mol 的碳酸钠;在此反应中,决定二氧化碳体积的反应物是 _____。

对比以多项选择题和填空题出现的这2道题目,考查的内容是相同的。从考生填空题的答题情况可以清楚地判断其对所考查知识的掌握情况,并

能从该反应是否属于沉淀反应的判断中考查出学生思维的过程。如果考生认为这是一个沉淀反应,说明其思维还是从溶液中的反应这个常见的知识储备入手解题,便会得到反应不会产生气体,反应后无碳酸钠剩余的结论,从而造成后面几个空格答题错误。但如果真的设计为填空题,考生在二氧化碳填为 0 g 时,会意识到题干明确了“会产生气泡”,从而重新从固体与液体反应角度思考这个题目。此时,便显示出台湾的这种多项选择题甚至优于填空题的优势来了,除了不能考查考生书写化学方程式的能力,其他的思维过程都可以考查到,并且如果考生如上所述误以为这是一个溶液中的反应,就有可能选择 D、E 2 个答案,而不会再深究“会产生气泡”这个现象产生的原因。

通过以上对比,可以看出,台湾学测题在题干和选项的设计上源于填空题或者简答题、计算题等主观题,再进行适当地改编,使之成为一道可以考查出思维过程的多项选择题。

2.2 多项选择题计分过程中注重思维的过程性

我国大陆的多项选择题计分方法一般是只要有一个选项错误,则该题得 0 分。台湾学测自然考科自 1995 年起开始出现的多项选择题,开始也与大陆相同。1996 年起计分方式开始发生改变,试卷中的描述为“每题答对得 2 分,答错不倒扣,未答者不给分。只错 1 个可获 1 分,错 2 个或 2 个以上不给分”。其中“只错 1 个可获 1 分”是最明显的变化,一定程度上顾及了思维的过程性。

我国台湾学测自然考科多项选择题计分方式的最大改革发生在 2011 年,自当年起实行的计分方式有效地体现了思维的过程性。还是以 2014 年第 32 题为例,在大陆的高考或者学业水平考试中,如果以填空题的形式出现(例4),即便其他空格都做错只要有 1 个空格的答案是对的,就该给予相应的得分。因此,以填空题形式出现更合理,计分结果更符合思维的过程性。但是台湾的题型只有选择题,如何合理评价思维过程性的不同结果呢?解决这个问题的方法就是要改变计分方式,使计分的结果更能体现学生的思维过程。台湾的计分方法以多选题为例:

每题有 n 个选项,其中至少有 1 个是正确的选项。各题之选项独立判定,所有选项均答对者,得该题全部的分数;如答错 k 个选项者,得该题 $(n - 2k)/n$ 的分数,但得分低于零分或所有选项均未作答者,该题以零分计算。

例：某多选题题分 2 分，有 5 个选项 A、B、C、D、E，正确答案为 A、C、D，甲、乙、丙、丁、戊 5 位考生之答案分别为 [A. C. D]、[A. B. C. D]、[C. D. E]、[A. B]、[未作答]，则各生在该题之得分如表 2 所示：

表 2 学生得分情况统计
Table 2 Statistics of students' scores

考生	作答答案	各选项对错	答错选项数	计分方式	该题得分
		A B C D E			
甲生	A. C. D	O O O O O	0	2	2
乙生	A. B. C. D	O X O O O	1	$[(5-2 \times 1)/5] \times 2$	1.2
丙生	C. D. E	X O O O X	2	$[(5-2 \times 2)/5] \times 2$	0.4
丁生	A. B	O X X X O	3	0	0
戊生	未作答	- - - - -	-	0	0

同一道多选题，满分是 2 分，如果正确选项是 3 个，可能得到的分数有 2 分、1.2 分、0.4 分、0 分 4 种^{[1]9}。

按照多选题计分方法，2014 年第 32 题如果考生的答案是 B、D，可得 2 分。如果除了 B、D 外，多选了 1 个错误选项，还可得 1.2 分；多选了 2 个错误选项，还可得 0.4 分。如果只选 B 或 D，也意味着选错 1 个答案，仍然可以得 1.2 分。如果 5 个选项全选或全不选，则该题得 0 分。以上这种计分方法，与大陆多项选择题对则满分、错则 0 分计分方法相比，可以更加合理地评价考生的思维过程。可以看出，这种计分方法还是源于主观题分步得分的评价办法。

3 尊重考生差异性的设计

所谓公平，就是从公正的角度出发，平等地对待每一个与之相关的对象。教育所追求的公平，不完全是“均等”和“划一”，因为“均等”和“划一”只是注重数量和程度等方面的一致性，而“公平”更强调受教育者所受到的教育与其社会权利和个体素质之间的匹配^[5]。考试也一样，不同考生的学习潜能、兴趣爱好等方面存在差异性。面对这些差异性，除了保证他们受教育的权利和机会“均等”以外，还应通过适当的评价机制，正视并尊重差异，让考生获得最大可能发展的机会。

在我国大陆目前实施的高考理综试卷中，为了尊重考生的差异性，化学部分由必考题和选考题组成。选考题包含物质结构、有机化学、化学与技术等，考生可以根据自己的学习兴趣和专长自主选择相应的题目答题。我国台湾学测自然考科都是选择题，考查的都是必修的内容，无法从选修的题目提供自主选题的机会。为了解决这个问题，台湾的做法是设计比满分值更多的题量，给予考生更多的

选择机会。

自 2002 年开始，台湾学测自然考科在第 2 部分计分方法中首次出现题量大于满分值的设计。自该年起到现在，台湾学测自然考科满分都是 128 分，共 68 题选择题，每题 2 分。如果 68 道题目全部得分，满分应该是 136 分，为什么试卷满分却只有 128 分呢？这与考卷中第 2 部分的计分方法有关。以 2015 年试卷为例，第 2 部分从 41 题到 68 题，总计 28 题，每题 2 分，总分应为 56 分，但实际满分是 48 分。因该部分得分有个特殊的计分方法，即得分超过 48 分以上，以满分 48 分计，具体计分办法：

第 41 题至第 68 题，每题 2 分。单选题答错、未作答或画记多于 1 个选项者，该题以零分计算；多选题每题有 n 个选项，答错 k 个选项者，得该题 $(n-2k)/n$ 的分数；但得分低于零分或所有选项均未作答者，该题以零分计算。此部分得分超过 48 分以上，以满分 48 分计。

这种计分方法充分考虑了考生的差异性，提高了评价的公平性。由于试卷中提供了比总分多的考题，如 2015 年提供 28 题，如果每题均得分，只需做 24 题即可满分。即便题目个别做错，也有可能得到满分。这种提高题目自主选择权的方法更有利于人才的选拔，不会因考生在某一知识点存在短板而埋没优秀人才。

通过以上分析，可以看出台湾学测自然考科通过增加选项数目与多项选择题比例有效地降低了强猜率；通过基于主观题改编的选择题编制方法、参考主观题分步得分的计分方法设计，使考生思维的过程性得以一定程度的考查和评价；通过设计比满分值更多的题量，充分尊重了考生的差异性，这些设计使选择题考查的公平性得以明显改善。

参 考 文 献

- [1] 财团法人大学入学考试中心. 104 学年度学科能力测试简章. 台湾: 财团法人大学入学考试中心, 2015
- [2] 孙建明. 新课程高考化学学科试题命制研究. 武汉: 华中

师范大学博士学位论文, 2014

- [3] 吴庭光. 上海教育科研, 1986 (6): 42, 13
- [4] 宋立新, 刘影, 王清. 数理统计与管理, 2001 (1): 30—32
- [5] 陆军. 教学月刊: 中学版, 2001 (7/8): 71—73

Setting and Evaluation of Multiple-Choice Questions for Natural Science Test in Chinese Taiwan's Academic Competence Test

XIAO Qiao-Ling¹ YE Yong-Qian^{2**}

(1. Fujian Institute of Education Research, Fuzhou 350003, China; 2. Nan'an Qiaoguang Middle School, Nan'an 362314, China)

Abstract Questions for natural science test in Chinese Taiwan's academic competence test were all choice questions; in order to reduce the tendency to randomly guess answers and to examine the students' thinking process and differences, it increased the number of options and the proportion of multiple choice questions, and changed the arrangement mode of multiple choice questions and the way of scoring, which improved the evaluation function of multiple choice questions and made examination more fair.

Keywords Taiwan's academic competence test; fair of test; natural science test; multiple choice question; academic proficiency test