

基于评价量规的实验报告新评估模式

杨鲜梅^{1,2}, 陆红^{1,2}, 何正平^{1,2}, 娄慧玲^{1,2}, 吴燕华^{1,2}(✉), 乔守怡^{1,2}(✉)

1. 生物科学国家级实验教学示范中心(复旦大学), 上海, 200433
2. 复旦大学生命科学学院, 上海, 200433

摘要:传统的生物学实验教学重视实验技能与操作的训练,在数据处理与论文撰写方面训练不足。本文结合高级生化技术与基因工程实验两门课程的教学改革实践,总结出针对实验报告中摘要撰写及图表制作的评价量规,在利用该量规指导学生的同时,也将其用于实验报告的成绩评估。初步实践结果显示量规的使用有利于学生的主动学习,显著提高了学习成效和评价的一致性信度:实验报告的摘要撰写和图表制作更符合科学规范,教师与助教对报告的评分更客观更稳定。笔者认为评价量规是一个在实验课程改革中值得推广的好方法。

关键词:生物学实验教学,评价量规,实验报告,摘要撰写,图表制作

Application of Rubric in Lab Report Evaluation

YANG Xian-mei^{1,2}, LU Hong^{1,2}, HE Zheng-ping^{1,2}, LOU Hui-ling^{1,2}, WU Yan-hua^{1,2}(✉), QIAO Shou-yi^{1,2}(✉)

1. National Demonstration Center for Experimental Biology Education (Fudan University), Shanghai 200433, China
2. School of Life Sciences, Fudan University, Shanghai 200433, China

生物学是一门实验科学,生物学教学应注重培养学生提出并解决科学问题的能力。传统的生物学实验教学重视实验技能与操作的训练,在数据处理与论文撰写方面训练不足,与人才培养的需求之间存在不平衡,急需通过有效的教学方法加以改进和提升。在实验课程中传授和训练数据处理与论文撰写方法存在以下困难:首先,生物学实验类型多样,

实验数据种类繁多,实验数据的处理及展示很难有一个非常具体又清晰的标准。其次,如何利用标准指导学生的实践,针对性地提高数据处理、论文撰写的能力,缺乏科学的、可衡量的方法。

过去一年,笔者所在团队在面向高年级本科生的高级生化技术和基因工程实验两门实验课程中,初步建立了一套针对实验报告中摘要部分写作及实验数据图表展示的评价量规,并利用该量规指导学生的数据处理和论文撰写。在用评价量规来指导规范实验报告中相关内容后,学生的实验报告质量有了明显的提高,对摘要的撰写和实验数据的展示等更加符合科学规范。虽然量规仍需进一步完善,但初步的实践效果显示这是一种在实验课程中值得借

收稿日期:2019-02-18;修回日期:2019-03-22

基金项目:复旦大学2019年度教学研究与改革实践项目(C类:学习评测设计与实施)“基于评价量规(Rubric)的实验报告新评估模式建设”(2019C001)

通讯作者:吴燕华, E-mail: yanhuawu@fudan.edu.cn;
乔守怡, E-mail: syqiao@fudan.edu.cn

鉴的好方法。

1 实验课程的培养目标与存在问题

科学研究需要科学交流，这其中就包含了科学地理解数据、理性地归纳结果、对这两项进行科学且清晰地表达。在这些方面对本科生进行严谨的训练对他们将来从事科学研究会非常有益^[1, 2]。这也是本科生实验课程的教学目标之一。通过实验课程，学生不仅应该掌握基本的实验技能，还应该通过实验报告的撰写锻炼实验数据的科学整理与归纳。

目前的实验课程中，实验报告普遍存在下面几个问题。第一，生物学实验材料多样、方法条件多变、实验数据的处理通常较难有一个清晰具体的标准，在传统的实验报告撰写要求中，针对实验数据的分析和展示尚缺乏相对详细的评价规范；第二，由于缺少评价规范，在实验报告撰写的过程中，大部分学生无章可循，实验数据的处理粗糙，描述结果的文字不规范，一些同学甚至不知道要标识图名与图注。第三，在多数实验课程的成绩评定中实验报告会占一定比重。但由于缺乏规范实验报告的客观评价标准，教师和助教难以给出相对客观合理的评价，评价成绩常与学生的预期之间存在分歧。

和科研实践相比，实验课程涵盖的实验内容一般较为确定，学生获得的实验数据也较单一。针对实验课程的这些特点，我们认为，如果根据课程所获得的具体实验数据类型制定相对统一的标准，将有助于指导和规范学生实验数据的处理方式和实验报告的科学表述。与此同时，这一标准还可以用于实验报告的成绩评定，解决实验报告评分主观性强的问题。

2 在实验报告撰写及评分中引入评价量规

评价量规 (rubric) 是一种评分指南，用于评价学生在建构学习过程中的响应质量。量规通常包括评价标准、在特定水平对标准的质量定义及与之对应的评分策略^[3, 4]，一般以表格形式呈现^[5]。评价量规最初用于上世纪七十年代中期的写作评估^[6]，之后（特别是上世纪九十年代中期之后）在国际上越来越流行。目前，评价量规主要用于对不太适合用客观考试形式来衡量的学习过程的主观性评价，

常用领域包括写作、课堂讨论、课堂报告等^[7, 8]。制定量规，首先必须明确教学目标，并细化目标（也即对学生的要求）到具体的评分项。其次，对每一个评分项的不同完成程度，量规需给出清晰详细的评价标准以及具体得分。

作为主观性评价标准的评价量规，其真正的价值在于对教师的教学行为及学生的学习行为进行科学的指引。分析学生在量规不同评分项的得分情况，教师可以更清晰地掌握学生的学习行为，从而及时给出针对性的指导。同时，在学习活动开始前向学生公开量规，学生可以全面了解教师对学习的具体要求，并用其来规范自身的学习行为。课程后，教师及时反馈学生的评价得分，学生在了解自己表现的同时，可以对照量规查找具体的差距，并在后期的学习过程中有针对性地调整并提高 [学习行为初期的评分可以不计入学生最终成绩，称为“形成性评价 (formative assessment)"]^[9]。因此，评价量规既是一个科学的评价标准，也是一个有效的教学工具，它将教学活动和成绩评价紧密且有效地连接起来。

目前，将评价量规用于实验报告的撰写与评分，在国内的实验课程中还接近空白。笔者近年来在面向生物科学 / 生物技术专业的高年级学生的基因工程和高级生化技术两门实验课程中，对实验报告引入了评价量规，改革实践初显成效。

3 评价量规的设计及初步应用

实验报告包含的内容较广。在最初对其引入评价量规时，选择侧重点来培养学生的科研素养在实施上更具有可行性。笔者团队开设的基因工程实验课程，实验设置系统性非常强，类似于开展一个相对完整的科研课题，所以对于实验报告的评价侧重于对摘要撰写。而团队开设的高级生化技术课程由三个不同的实验组成，故而选择对数据处理 / 图表制作引入量规。同时，由于两门课程量规的侧重点不同，对于这两门实验课程均选修的高年级学生来说得到了训练互补。

3.1 科学论文摘要撰写的评价量规的设计与应用

摘要是科学论文的重要组成部分。摘要利用精

炼准确的语言概要性地描述研究项目的选题背景、研究目标、实验方法、关键数据、核心结论和科学意义，具有独立性、科学性、完整性和概括性的特点。在文献检索中，摘要阅读是读者简要了解论文内容的主要途径，摘要的撰写质量直接关系到研究论文的学术影响力。因此，摘要撰写也应纳入生物学实验课程的科学研究训练内容之中，且评价量规的使用将有助于摘要撰写的学习指导与衡量。

笔者以基因工程实验课程的综合实验项目“转基因哺乳动物细胞基因表达水平分析”为例，介绍摘要撰写的评价量规的设计与应用情况。如表 1 所示，根据该实验内容的侧重点以及摘要撰写的要求，评分点包括语法、学术规范、研究目的、研究材料与方法、研究结果与结论五个部分（本实验内容是教师规定的基因表达分析实验，不涉及选题，因此此评价量规中没有设立对选题背景和科学意义的考察指标）。此外，学术规范是本课程的学习目标之一，也纳入到摘要的评分点中。每个评分点根据完成度从差到好分为 3 个等级，0 分，1~2 分和 3~4 分，每个等级均有相应的详细描述，总分为 20 分。

在学生完成全部实验内容后，将评分点发布给学生并进行考察重点的讲解说明，以便于学生在摘

要撰写的过程中利用评价量规促进主动学习和自我改进。在评价量规的帮助下，摘要撰写的总体水平较往届学生有明显提高，通过与部分学生进行口头交流获知评价量规在摘要撰写过程中发挥了重要的指导作用，能够帮助明确改进的方向。课程评阅阶段，笔者设立一组实验进行评估：两位助教在没有评价量规的指导下，按照他们的科学研究经验给予独立评分；另外两位助教在有评价量规的指导下，按照评分量规表进行独立评分。对两组评分结果进行相关性分析的结果如图 1 所示，全部 19 份报告的 4 份成绩的最低分为 10 分，最高分为 19 分，平均成绩为 13.9。在没有评价量规的帮助下，两位助教凭借自己的学术积累，给出的分数具有一定的相关性，R 值达到 0.47，提示这种经验性的主观评阅虽有一定的可靠性，但存在明显的个体差异和主观随机性。在有评价量规的帮助下，两位助教的评分达到了 0.95 的相关系数，很好地证明了评价量规可以有效改善主观评阅中存在的个体差异和随机性问题。更重要的是，在评价量规的帮助下，助教直接指出学生普遍存在的不足存在于哪些方面，如研究目的表述不准确、研究结果与研究结论混淆等问题，教师和学生均获得了重要的反馈信息，利于制定提升学习成效的具体措施。

表 1 “转基因哺乳动物细胞基因表达水平分析”实验项目摘要撰写评价量规

得分点	0 分	1~2 分	3~4 分
语法	<ul style="list-style-type: none"> 语言生涩 有明显错别字或错误标点符号 	<ul style="list-style-type: none"> 语言较流畅 有明显错别字或错误标点符号 	<ul style="list-style-type: none"> 语言流畅 无明显错别字或错误标点符号
学术规范	<ul style="list-style-type: none"> 基因、蛋白质等符号书写格式不准确；且第一次出现时无英文全称 计量单位使用不准确且数字和符号之间无空格 	<ul style="list-style-type: none"> 基因、蛋白质等符号书写格式不准确或第一次出现时有英文全称 计量单位使用不准确或数字和符号之间无空格 	<ul style="list-style-type: none"> 基因、蛋白质等符号书写格式准确；第一次出现时有英文全称 计量单位使用准确；且数字和符号之间有空格
研究目的	<ul style="list-style-type: none"> 没有研究目的 	<ul style="list-style-type: none"> 有研究目的，但表述不准确 	<ul style="list-style-type: none"> 有研究目的，且表述准确（核心：哺乳动物细胞中外源基因表达分析）
研究材料与 方法	<ul style="list-style-type: none"> 没有研究材料 没有研究方法 	<ul style="list-style-type: none"> 有研究材料或研究方法，但撰写不完整 	<ul style="list-style-type: none"> 有研究材料（哺乳动物细胞 U2OS） 有研究方法，且撰写完整（包括：RNA 抽提、反转录 PCR、常规 PCR、定量 PCR）
研究结果与 结论	<ul style="list-style-type: none"> 没有研究结果 没有研究结论 	<ul style="list-style-type: none"> 有研究结果或研究结论，但撰写不完整或不准确 	<ul style="list-style-type: none"> 有研究结果，且撰写完整（RNA 浓度、常规 PCR 结果、定量 PCR 结果） 有研究结论，且撰写准确（区分高/低表达外源基因的细胞株、野生型）

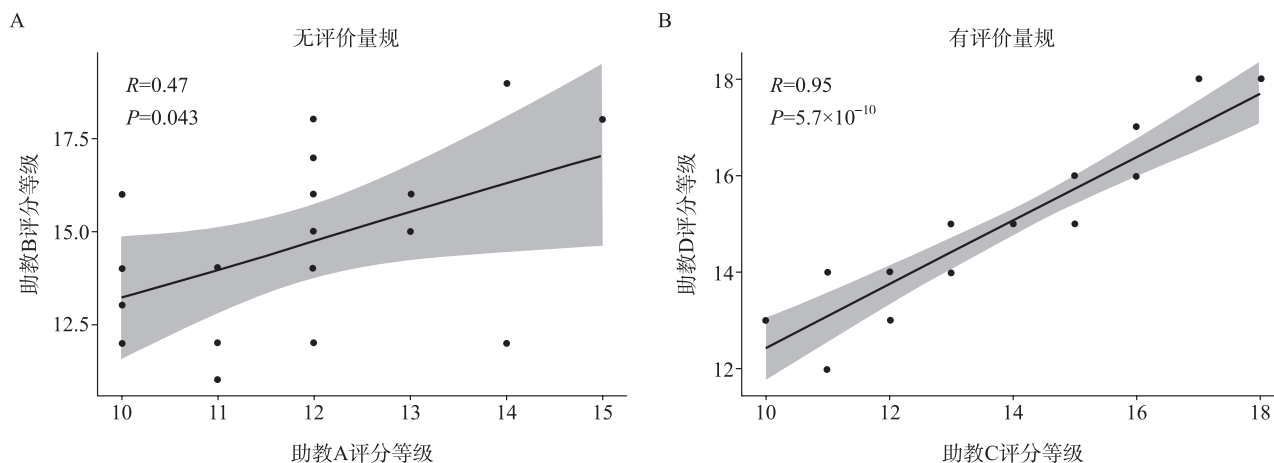


图1 “转基因哺乳动物细胞基因表达水平分析”实验项目摘要的评分结果 ($n=19$)

(A) 在没有评估量规情况下两位助教独立评分的 spearman 相关性分析;

(B) 在有评估量规情况下两位助教独立评分的 spearman 相关性分析

3.2 数据图表制作的评价量规的设计与初步实施效果

数据图表是科学论文的重要组成部分。数据图表

的制作应符合国际学术规范、逻辑清晰且图注文字简明扼要，便于读者准确地、便捷地、直观地从图表中获得论文研究信息。为初步训练本科生的数据整理展示能力，笔者在“高级生化技术”课程中初步设计了针对实验报告中数据图表的评价量规（表2）。

表2 “高级生化技术”课程实验报告数据图评价量规

评分项	评分项说明	不扣分	扣1分
板块标示	<ul style="list-style-type: none"> 若数据图包括多于一个板块，则每个板块用大写英文字母标示 大写英文字母标示于板块的左上方 若数据图只包含一个板块，则无需标示，此项也无需评分 	<ul style="list-style-type: none"> 标示规范 	<ul style="list-style-type: none"> 标示不规范或没有标示
图身标示	<ul style="list-style-type: none"> 若数据图有 x 轴和 y 轴，则注明各轴所示内容及单位 数据图中需要特别进行说明的数据，应以适当的方式，如数字、字母、箭头、符号等在图中相应处标示 	<ul style="list-style-type: none"> 标示规范、完全 	<ul style="list-style-type: none"> 标示不规范或不完整
图题	<ul style="list-style-type: none"> 数据图须有图号（如图1、图2） 数据图需要图题，接于图号之后 	<ul style="list-style-type: none"> 图号图题完整 	<ul style="list-style-type: none"> 图号图题不完整
图注	<ul style="list-style-type: none"> 对数据图进行注释与说明（若数据图包含多于一个板块，则每一板块均给出题目） 	<ul style="list-style-type: none"> 有图注 	<ul style="list-style-type: none"> 没有给出任何图注
图题图注位置	<ul style="list-style-type: none"> 图注接于图题之后 图题和图注一般著于数据图的下方或右侧（左侧亦可） 	<ul style="list-style-type: none"> 图注接于图题之后 图题图注位置规范 	<ul style="list-style-type: none"> 图注未接于图题之后 图题图注位置不规范
图尺寸	<ul style="list-style-type: none"> 黑白线条图分辨率不低于 600 dpi，照片图分辨率不低于 350 dpi 数据图宽度为半栏或通栏 若图题图注著于数据图右侧（或左侧），则数据图尺寸可不遵循上条，但其与图题图注的总宽度应为通栏 	<ul style="list-style-type: none"> 数据图分辨率及尺寸规范 	<ul style="list-style-type: none"> 数据图分辨率及尺寸不规范
图注具体内容	<ul style="list-style-type: none"> 对应图身各标示（数字/字母/箭头/符号等）给出必要说明 列出实验的关键条件 表述清晰且简要 	<ul style="list-style-type: none"> 图注内容达到左侧要求，奖励2分；每份报告此项奖励最多不超过两次 	
数据排版逻辑	<ul style="list-style-type: none"> 数据排版逻辑清晰（如对照组与实验组数据于一张图中一起展示以使结果对比更一目了然） 	<ul style="list-style-type: none"> 排版逻辑清晰奖励2分；每份报告此项奖励最多不超过两次 	

如表2所示,笔者根据往年实验报告中数据图表制作的常见问题,在评价量规设计时着重数据展示的科学规范:量规的评分点主要包括图板块标示、图身标示、图题、图注、图题图注位置规范、图(图题图注)尺寸六项。批改实验报告数据图部分的每幅图时,上述评分项如若不合格,则从总分20分中扣除1分。另外,对于图注具体内容的书写以及数据排版逻辑等更高的要求,则采取奖分制,每次奖励2分,每份报告同一项奖励不超过两次。课程总共包括三份实验报告。

评价量规公开透明,在实验课程正式开始前下发给每一位学生。为了便于学生理解量规的各项评分点,笔者选取了往届学生实验报告中的代表性数据图,结合量规说明其中出现的问题,以及如何修改。在实验报告提交后,教师及时对课程前期的报告进行批改,反馈评分情况,并与学生进行面对面的交流。“评阅—反馈—交流”的成效在后期的实验报告数据图中反映了出来(图2)。以反馈前后改善最明显的数据图图身标示与图注具体内容书写为例,23位选课学生中,第一份实验报告的图身标示合格者(未被扣分)只有1人,而在第三份实验报告中合格者达到了20人;对于图注具体内容,书写完全又简明的获奖分人数,也从反馈前的4人提高到了反馈后的8人。这两个评分项,特别是后者,所涉及的具体内容不如其他评分项直观,尽管是学生们在展示数据时较难掌握的地方,但是均有显著的成效提升。

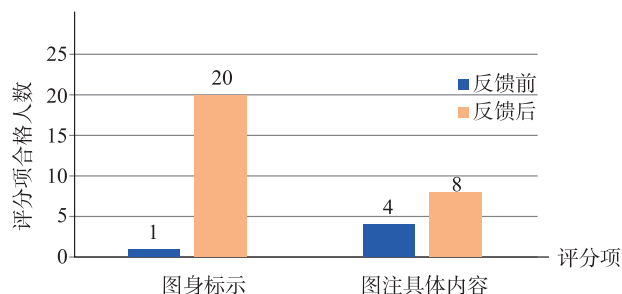


图2 评分结果反馈前后学习行为比较

课程结束后的调查问卷结果显示(回收问卷19份,占全班人数的82.6%),所有人均在第一次撰写报告时就阅读了提前公布的评价量规及说明,且其中17人(占回收问卷的89.5%)自认为进行了“仔细”的阅读,说明学生对于课程的评价要求普遍非

常关心。在阅读评价量规后,有16人表示“很清楚”或“比较清楚”实验报告的撰写要求,另外3人对此表示“仍然很不清楚”,其中包括2位“仔细”阅读了评价量规的同学。在对第一次报告评分结果进行反馈,并回答学生的咨询后,全部同学表示对于实验报告的撰写要求“很清楚”或“比较清楚”,充分证明了评价量规的指导性作用。

4 评价量规的改进方向

评价量规作为一种教学测量手段,与其他测量手段(如试卷)一样,应具有一定的一致性信度、效度和区分度。特别是一致性信度,它反应用量规测量所获结果的稳定性,是衡量量规质量高低的一个重要指标。一致性信度系数常用spearman相关系数来估计。

摘要撰写量规的实施结果显示,两位助教运用量规独立评分,其spearman相关系数为0.95,反应该评价量规的信度很高(图1)。但数据图量规的结果显示,教师和助教独立评分结果的spearman相关系数为0.59(图3),虽具有了一定的可靠性,但仍有待进一步提高。具体的数据分析显示部分评分项,特别是对于图尺寸以及图注内容书写质量的认定,评阅人对达到何种质量程度进行扣分/加分存在主观差异,需要在评阅前进行交流讨论以便于统一。

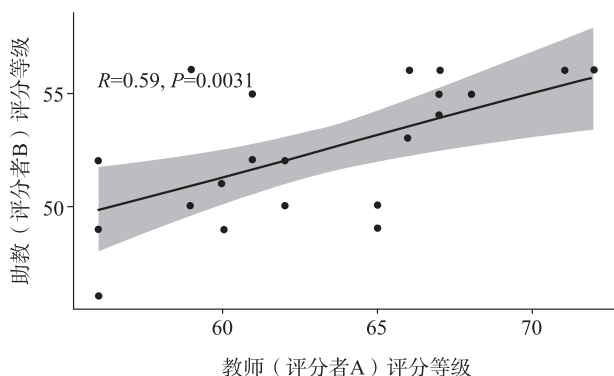


图3 “高级生化技术实验”课程实验报告数据图评分结果的spearman相关性分析($n=23$)。

另外,高级生化技术课程调查结果显示,尽管84%的受调查者在阅读量规后认为“比较清楚”和/或“很清楚”评分标准,且这个比例在反馈后上升到了100%,但数据图的实际评分结果显示仍有相当

学生没有达到评分项要求, 这充分说明: 学生与教师对评价量规的认知会存在一定的分歧。在后续的改进中, 笔者将进一步对数据图量规各个评分项的标准做更具体明确的描述, 同时, 各个评分项的合理性及其评分权重也会纳入改进的范畴。

综上, 笔者在基因工程实验及高级生化技术的教学中初步尝试用评价量规来对实验报告进行规范并评分, 期望能够提升学生科学地处理、展示并表述实验数据的科研能力。教学改革的初步结果显示: 学生根据评价量规进行报告撰写和数据整理, 主动学习效果好, 报告质量有明显提高; 其次, 将评分结果及时反馈学生, 有利于对照评分细则进行个性化的深入学习, 全面提升学习效果; 第三, 虽然现阶段的评价量规的一致性信度仍有待进一步完善, 但利用评价量规对实验报告进行评分显著提高了实验报告审阅的客观性、准确度和便捷性。笔者认为, 评价量规适合在生物学实验课程教学改革中进行学习和推广。

致谢

复旦大学生命科学学院研究生方宜潇、陈茜、刘王杰、陶成秋、尚凌月和李国庆同学对评价量规提出了宝贵意见, 并帮助批改了实验报告。在此一并表示衷心感谢!

参考文献

[1] SHAH P, MAYER R E, HEGARTY M. Graphs as aids to

knowledgeconstruction [J]. Journal of Educational Psychology, 1999, 91: 690-702.

[2] BENGTTSSON L A, OTTOSSON T. What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement [J]. Journal of Research in Science Teaching, 2006, 43 (1): 43-62.

[3] POPHAM J. What's wrong—and what's right—with Rubrics [J]. Educational Leadership, 1997, 55 (2): 72-75.

[4] 史蒂文斯, 利维. 评价量表快捷有效的教学评价工具 [M]. 陈定刚, 译. 广州: 华南理工大学出版社, 2014: 1-173.

[5] DAWSON P. Assessment rubrics: towards clearer and more replicable design, research and practice [J]. Assessment & Evaluation in Higher Education, 2015, 42 (3): 347-360.

[6] DIRLAM D, BYRNE M. Standardized developmental ratings [J]. Developmental Stages, 1978.

[7] 王芳. 《评价量表——快捷有效的教学评价工具》评介 [J]. 上海教育评估研究, 2017, 6 (1): 77-79.

[8] ANGRA A, GARDNER S M. The graph rubric: development of a teaching, learning, and research tool [J]. CBE—Life Sciences Education, 2018, 17: ar65, 1-18.

[9] PANADERO E, JONSSON A. The use of scoring rubrics for formative assessment purposes revisited: A review [J]. Educational Research Review, 2013, 9: 129-144.

(责编 张磊 田红)