

【评价研究】

促进学习的评价：“整本书阅读”表现性评价工具分析及优化研究

赵舒妮

【摘要】基于31张“整本书阅读”的表现性评价表的分析,梳理评价表存在的混淆学习任务和目标、评估指标与目标不匹配、分级无法涵盖所有水平、量规不易被观察、包含模糊性术语等9个常见问题,以“促进学习的评价”的立场,进一步提出确立指标四步骤、优化量规五策略、试用修订六自查等具体改进策略,以期提升表现性评价工具的科学性、诊断性和实效性。

【关键词】表现性评价;整本书阅读;评价量规

新的人才观、学习观推动着教育改革,教育评价被视为教育改革的杠杆,成为当下教育研究的一大焦点。《义务教育语文课程标准(2022年版)》(以下简称“2022年版课标”)首次出现了学业质量板块,其中对评价的要求引起了中小学语文教师的广泛关注和高度重视,可以说,语文教育评价尤其是学生学业成就评价,在实践领域正发生着空前的变化。众多教师在2022年版课标的指导下,设计评价内容、创新评价形式,表现性评价、备择评价、档案袋评价等方式在当下的语文教学中得到较为广泛的应用。多元评价为语文教学带来新的活力,也让教师面临新的挑战。评价工具开发科学性不足、评价工具使用实效性差等问题普遍存在。在许多公开课上,教师耗费大量时间让学生自评、互评却收效甚微。可见,不少教师已然陷入“为了评价而评价”的新误区。

在很长的一段时间内,我们关注“关于学习的评价”,或称之为总结性评价,为高一级教育选拔学生、认证学业成就、评价教育政策、监测变化趋势等服务。但是,当前教育评价领域更关注“促进学习的评价”,也称之为形成性评价。美国教育评估专家简·查普伊斯认为,促进学习的评价的主要作用是为学生提供改进学业成就的建议、帮助教师诊断并回应学生需求、帮助家长了解学生的学业进展并支持学生的学习。由此可见,促进学习的评价不单纯是工具,也不仅仅是教学活动的一个环节,它应是一个促进学生学习的支持系统。

表现性评价是一种基于观察和判断的评价,华东师范大学副教授周文叶将其定义为“在尽量合乎真实的情境中,运用评分规则对学生完成复杂任务

的过程表现或/与结果做出判断”。本文主要以语文学科的整本书阅读教学为切入点,从“促进学习的评价”视角重新审视当下的表现性评价实践,以期找到关键问题,提出切实可行的优化策略。这里需要强调的一点是“促进学习的评价”并不与“关于学习的评价”对立,我们讨论的目的是要构建一个多元且平衡的评价体系。

一、“整本书阅读”表现性评价工具的现状分析

以主题词“整本书阅读”“表现性评价”,在中国知网中文资源库进行检索,先剔除非语文学科的文献,再进一步筛选出同时符合两个条件的文献,可整理出截至2024年2月1日的有效文献17篇。在2019年之前无二者相关的研究,在2019年至2021年三年间相关文献共有6篇。2022年有3篇相关文献,2023年有8篇相关文献,因为2022年版课标首次将“整本书阅读”列为学习任务群,因而上述11篇文献无一不提到课程标准,可见研究的上升趋势与课标制订与实施关系密切。

上述文献反映了教师在整本书阅读教学中选择表现性评价的原因:①传统纸笔测试的不足。“整本书阅读”需要学生综合运用理解、分析、鉴赏等能力,但传统纸笔测试包含大量选择性反应题目(选择、判断、连线)和填空、简答,在评价学生高阶能力的发展情况方面存在困难。②表现性评价与2022年版课标的评价要求贴合度高。2022年版课标在“评价建议”中提出“注意考察阅读整本书的全过程,围绕读书的主要环节编制评价量表,制作阅读反思单,引导学生从阅读方法、阅读习惯等方面进行自我反思、自我改进”。表现性评价贯穿全程、贴合真实情境、评

价主体多元化等特点都很好地回应了课标要求。
③表现性评价与教师设计的相关活动适配度高。2022年版课标建议整本书阅读教学要综合运用多种方法阅读、多种方式分享。表现性评价不是使用常模参照(评价时以学生所在团体的平均成绩为标准),而是使用标准参照,因而更适用于丰富多彩的阅读活动,如朗诵会、故事会等。

文献中相关的实践研究主要涉及四个方面:
①定位表现性评价目标。②设计“整本书阅读”的表现性任务。③编制表现性评价量表。④构建评价实施的一般路径。这些研究集中体现在了评价表中。剔除任务安排表、教学安排表、汇总表,以及引用的表格,17篇文献中有13篇含有有效的评价表,共计31张。

评价表的内容涵盖了听说读写各个方面,包括整本书的阅读方法、人物赏析评价、批注运用、阅读速度、阅读分享、讨论倾听,也有不少专门针对某一项任务的量表。

评价表的类型包含了表现性评价常用的三种评分记录工具:核查表、等级量表和评价量规。其中核查表1张,帮学生自查是否按进度完成任务;等级量表8张,将评价的目标分解成几个核心指标,由教师按等级评价学生表现对应的具体程度;评价量规也被称作评分规则,共17张,其中14张是在等级量表的基础上对每个层级的特征和表现添加具体描述,属于解析型量规,另外3张是只分等级不分维度的整体型量规。核查表和等级量表具有开发简单、使用便捷的特点,但教师似乎更青睐评价量规,大概是因为量规可以清晰地描述出各等级的标准,具有诊断性,有助于教师和学生获得更加具体的反馈信息,并为下一步教学或学习提供方向。

从评价表的特征来看,若以“促进学习的评价”为出发点,借鉴量规编制的相关研究,从合理性、可信度、使用者体验度等角度分析这31张量表,可以总结出9个较为常见的问题:

①评估时混淆学习任务和学习目标。情境任务只是达成目标的途径,评估应当指向学习目标的特征,不应该指向任务自身的特征。

②评估指标与目标不匹配,或者缺失重要指标。比如,要评价的是“整体感知”,所列指标却是“了解书信形式”“了解书信内容”。

③对无关的外部特征评分。比如,评估目标是为一本书写书评,却对“书写工整”进行评分。

④分级无法涵盖所有水平。分级的数量太多或太少,存在超出水平的情况,或存在隐含层级。

⑤分级标准是数量而不是质量。比如,制作昆

虫卡任务分级描述为“梳理昆虫习性,制作一张卡片”,而后逐级递增为“三张”和“一组”。当然,这里并不是否定数量增加的价值,但只追求数量显然是不够的。

⑥描述性说明对指标的解释不足。描述性说明应该较完整地反映学生在该特征的表现,否则可能意味着预期的目标没有被评估。

⑦包含模糊性术语,如“粗略地”“彻底地”“有见地的”“高于平均水平”等,这些词汇具有模糊性和主观性,实际使用中不同的使用者可能会对同一作品做出截然不同的评价。

⑧量规的描述不易被观察。比如,有的教师对量表情感维度的描述为“能通过阅读本书获得成长的力量,提高独立思考的能力”。这样的描述几乎无法被观察,评价也就缺失了证据。

⑨不符合基本规范。评价基本要素不完整或掺杂大量非评价元素等。

整体而言这些问题可以归纳为评价指标确立和评价量规描述两个方面的问题。

二、“整本书阅读”表现性评价工具的优化探索

“整本书阅读”时间跨度大、涉及关键能力多、阅读水平表征难,“促进学习的评价”要求“整本书阅读”的评价要具有科学性、诊断性和实效性。能用纸笔测试的就不一定要使用评价表,能用核查表和等级量表解决的简单问题就不需要使用复杂、耗时的评价量规。需要评价量规的任务,则要求评价工具的开发要科学、合理、可靠,使用要关注学生的体验和反馈,要能真正服务于学生的学习。

1. 确立指标四步骤

任务情境一:请你从图书馆新到的一批儿童小说中挑选一本,为其撰写200字以内的简介,优秀作品将贴在“新书推荐”公示栏,评价工具见表1。

表1 “新书推荐”评价表(修改前)

评价维度	自评	互评	师评
简要介绍	☆☆☆	☆☆☆	☆☆☆
吸引读者	☆☆☆	☆☆☆	☆☆☆
书写美观	☆☆☆	☆☆☆	☆☆☆
交流讨论	☆☆☆	☆☆☆	☆☆☆

任务情境二:请你选择《水浒传》中的一个人物,分析人物形象,并制作一张“水浒人物介绍卡”,评价工具见下页表2。

对此,我们可以依据评价指标确立流程图(如下页图1),分四步进行优化:

第一步,明晰目标。“整本书阅读”的目标是培养“优秀的阅读者”,包含理解、分析、鉴赏、创作等阅



表2 “水浒人物介绍卡”评价表(修改前)

评价维度	自评	互评	师评
卡片美观	☆☆☆	☆☆☆	☆☆☆
人物分析	☆☆☆	☆☆☆	☆☆☆
内容具体	☆☆☆	☆☆☆	☆☆☆
资料引用	☆☆☆	☆☆☆	☆☆☆

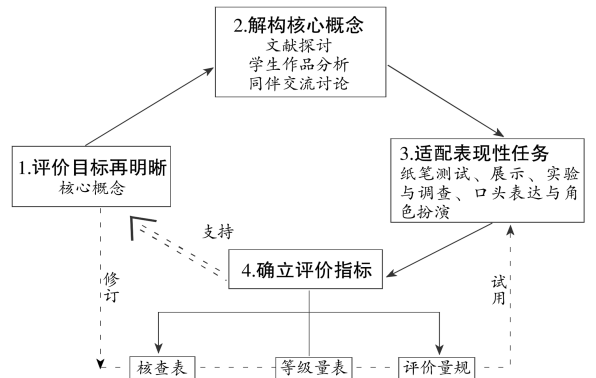


图1 评价指标确立流程图

读思维,也包含良好的阅读方法、习惯等。评价的核心应抓住“整本书阅读能力”中的某一关键概念,如任务情境一的关键概念是把握整本书的内容并加以概括,任务情境二的关键概念是分析小说中的人物形象。

第二步,解构概念。概括的要求是简明、准确、全面、有条理,应以此作为任务情境一指标拟定的基础。任务情境二可以让学生思考“优秀的读者”会如何分析小说人物形象——抓住人物外貌言行等具体细节做出判断、了解作者对某一人物的基本立场、注意在情节中把握人物形象、综合所处环境和身份地位等信息分析人物、关注侧面描写、认识到人物的多面性……教师在解构关键概念时,应力求全面、完整,对于较为复杂的概念,往往还需要辅以文献探讨、同伴讨论、学生作业样本分析等手段。

第三步,适配任务。根据要评估的目标特性,为其选择合适的表现性任务,一般包括纸笔测试、展示、实验与调查、口头表达与角色扮演等形式,必要时,教师还可以撰写任务指导语进一步帮助学生清晰任务要求。同时,教师还要充分考虑任务的数量、完成时长等要素。

第四步,确立指标。基于上述思考,我们可对表1和表2进行如下修改:①剔除表1中与把握整本书内容无关的“书写美观”和“交流讨论”两个指标,将“简要介绍”分解为“内容全面”“简明准确”“条理清晰”三个维度,保留任务情境一的关键特征“吸引

读者”。②修改表2原有指标,并将等级量表改为评价量规。考虑到评价目标相对复杂,可将“评价表”变为“助学表”,具体的描述可以为学生的学习提供更多的支持。同时,教师还需要充分考虑学生的年级特征,结合学生的先备知识及这一阶段的教学重点来确定具体有哪些内容要纳入评价。

2. 优化量规五策略

以《昆虫记》的阅读评价为例,在学生完成阅读后,教师布置了绘制昆虫变态发育图解的任务,表3为本次任务的评价量规初稿。这张量规符合基本规范,也有很多可借鉴之处,但其中存在的问题较为隐蔽。

表3 “好‘变态’的虫虫发育图”
评价量表(修改前)

	3分	2分	1分
变态顺序	各阶段顺序正确。	有一定顺序,但有错误。	没有顺序。
阶段图示	每个阶段都有配图。	有1个以上阶段缺少配图。	没有配图。
发育描述	每个阶段都有准确描述。	每个阶段都有描述,水平一般。	没有描述。
查阅资料	检索到相关信息,并对任务有帮助。	检索到相关信息。	没有检索到信息。

策略一,建议用水平等级代替评分。评价时使用分值看起来似乎更清晰,但实际上会让学生误以为打完分评价就结束了,从而削弱对学习的观察和判断。而且,“入门”“新手”等在一定程度上代表了学生的成长阶段,可减少评分排名带来的压力。

策略二,区分通用和特定情况下的量规。表3实际包含了一般量规和具体任务量规,前3条指标是针对发育图这一特殊任务拟定的,目的是考查学生对文本内容的理解、概括及表达的能力;而“查阅资料”却可用于不同的任务,类似的还有合作、交流、倾听等量规。通用就意味着可以反复使用,教师可以侧重选择一两张通用量规,贯穿于“整本书阅读”的系列任务中,这样既减轻教师开发量规的负担,又有利于学生在反复运用通用标准中理解内涵,迁移能力。

策略三,借助作业样本确认层级数量。教师可以收集10~12份学生作品,将其分成若干质量等级,并尝试对学生各层级的学业表现进行表述,再检验是否存在隐含层级或超越层级的情况。例如,“阶段图示”一项,学生做到“每阶段有配图,配图符合该阶段昆虫发育情况”即达到要求。但实际上,每个班

级总会有学生超出预期,因而教师还可以添加“典范层级”。又如,在实际教学中,“没有配图”“没有描述”的情况几乎没有,所以教师可以删去这一特征。特别值得注意的是,“变态顺序”没有对应设置“典范层级”,因为梳理顺序是相对简单的信息提取能力,作业样品中无法分离出四个层级。

策略四,减少以简单计数为标准的情况。在复杂的表现性任务中,数值统计对质量高低的判断意义不大。例如,在完成这项任务时,制作一组图并不一定意味着比制作三张图的水平更高。因此,教师可以在评价中增加“配图准确、清晰”“符合该阶段昆虫发育的情况”“描述内容完整,且有细节”等实质性描述,帮助判断。

策略五,调整不清晰的描述性语言。31 张量表中多次出现模糊、难以测量的描述语:第一类是评判性质的词语,如“优秀”“有见地”“平庸”,这些词更像是评价等级,而非指标描述;第二类是直接引用价值观的关键词,如“良好的习惯”“端正的态度”“成长的力量”,这些描述难以被观察和评估;第三类与“平均”相关,如“在平均水平之下”“不低于一般水平”,量规是标准参照,与以团体平均值为准的常模参照混淆,会导致难以做出评判;第四类是没有澄清的词语,如“完全”“极”“充分”,描述的表征很模糊,理解自然就因人而异了。这几类词汇都是需要教师在设计量表时尽量避免的。

结合以上五项策略,借鉴美国教育咨询师苏珊·布鲁克哈特的“生命周期项目量规”,我对表 3 进行了优化(如表 4)。

表 4 “好‘变态’的虫虫发育图”
评价量表(修改后)

	典范	熟练	新手	入门
变态顺序		各阶段顺序正确。	有 1~2 处顺序错误。	顺序错误,或没有体现顺序。
阶段图示	每个阶段的配图准确、清晰,很好地辅助说明该阶段昆虫的发育情况。	每个阶段都有配图,配图符合该阶段昆虫的发育情况。	每个阶段都有配图,有些图不符合该阶段昆虫的发育情况。	某些阶段缺少配图,或不符合该阶段昆虫的发育情况。
发育描述	每个阶段都有准确描述,描述内容完整且有细节。	每个阶段都有准确描述。	分阶段使用了一些不准确或不完整的信息。	没有分阶段描述,或阶段描述有错误。

3. 试用修订六自查

要编制一份科学、合宜的评价量规,往往还需要通过试用和修订来进一步完善。教师在试用量规的过程中,可以尝试回答以下六个问题,获得对量规进行评估的证据,从而进一步修正量规。

- 这份量规对于学生来说是否易于理解?
- 这份量规是否能够区分学生的表现水平?
- 这份量规的描述是否易观察、可测量?
- 学生是否可以很容易地通过这份量规修改自己的作品?
- 不同评价者能否借助这份量规做出一致评价?
- 这份量规对于每一个学生来说是否公平?

以上问题是以学生视角对量规进行评判,从评价工具使用的结果逆推量规的质量。教师可以通过观察、访谈、问卷等方式进行了解,完成自查。其中,“公平”问题最容易被忽视,教师应充分考虑学生个体的差异性,如性别、地域、家庭、先备知识等,不要让量规带有某个群体的倾向性。当班级中存在特殊儿童时,教师还要考虑是否需要为其另设一份量规。

本文涉及的 17 篇文献的研究者,均做出了极有意义、极富价值的探索。样本中丰富的描述性语言使表现性评价工具比传统测试更具交互性,让教学评价不再是“冷冰冰”的。此外,一些教师已经开始邀请学生一起制订量规,使评价成为连接教师、学生和课堂的纽带。这些变化都推动着教学评价的转型与深化。应当说,分项指标与质性描述相结合的表现性评价量规,为评估学生的高阶能力提供了有力工具。但是,无论如何,任何工具都只有被使用才能彰显价值,表现性评价工具亦然——其最大的价值就在于被教师运用于教学评价实践中,反馈真实学情,为教师和学生做出下一步决策提供支持,最终促进学生的学习。因此,当我们明确如何设计科学有效的评价工具时,自然就会进一步思考如何合理地、有效地使用它,而这是另一个富有挑战且相当艰难的话题,值得我们继续探索。

【作者简介】赵舒妮,福建师范大学教师教育学院副教授。

【原文出处】《福建教育》:小学版(福州),2024. 3. 29~33

【基金项目】本文系 2023 年度全国教育科学规划教育部重点课题“基于多模态数据教师画像的精准培训课程体系研究”(编号:DKA230449)的阶段成果。