

# 古诗知识经验测评问卷的初步编制

陈庆荣  窦  芬

**〔摘 要〕** 本研究采用理论与经验相结合的方法,初步编制古代诗歌知识测评问卷检验个体古代诗歌知识经验水平。选取 217 名大学生进行施测,并对测评项目的质量和测评问卷的信度和效度进行统计检验。结果表明,测评问卷的平均难度为 0.56,各部分的难度在 0.31—0.60 之间;测评问卷的整体鉴别力指数为 0.51,其中 89.13% 项目的鉴别力指数在 0.3 以上;测评问卷整体的内部一致性系数和重测信度系数均在 0.80 以上。进一步分析被试的古诗知识背景发现,诗歌经验组被试加入过与古诗相关的团体组织进行经验交流(64% vs. 21%),古诗创作经历(71% vs. 38%),古诗作品发表(44% vs. 11%)的比例均显著高于非经验组被试。与非经验组被试相比,诗歌经验组被试会花费更多的时间阅读古诗,且阅读量也显著增加。上述测评结果表明,古代诗歌知识测评问卷题目的难度、区分度以及信度和效度均良好,可初步用于测评个体的古代诗歌知识经验水平。

**〔关键词〕** 古代诗歌知识测评问卷;项目分析;信度;效度;难度

## 一、引言

“中华优秀传统文化源远流长、博大精深,中华民族形成和发展过程中产生的各种思想文化,记载了中华民族在长期奋斗中开展的精神活动、进行的理性思维、创造的文化成果,反映了中华民族的精神追求,其最核心的内容已经成为中华民族最基本的文化基因”<sup>①</sup>。诗歌是音律的纯文学(朱光潜,2001),是对语言材料的艺术化萃取。诺贝尔文学奖获得者 Eliot 指出“诗比任何艺术都更顽固地具有民族性”(Eliot, 1957)。较之于以英语为代表的印欧语诗歌,中国古代诗歌具有平仄、押韵、对仗等典型韵律特征(王力, 1978, 2002, 2003),已经成为中华民族最基本的文化基因之一(陈庆荣、杨亦鸣,2017),其历史悠久、题材丰富、内容广博、思想深邃,是历经千百年时间流水的冲刷而愈发闪烁着夺目光彩的民族文化的瑰宝。

那么,人们如何阅读诗歌?其内在的心理和神经机制是什么?诗歌阅读对学生知情意的形成和发展有何影响?如何测评?关于这些重要问题,已有的理论假设和实验研究主要围绕印欧语诗歌的典型韵律特征展开(如重音、节奏等)。在理论假设层面,Roman Jakobson(1960)提出了诗意性功能理论,指出欣赏诗歌时读者会有意识地知觉诗歌语言结构所具有的文本语言模式。基于此,Hanauer(1998)进一步提出了“特殊体裁假设”(genre-specific hypotheses),认为文本的加工模式受到其体裁特征的影响,从而使得具有不同外显特征的文本体裁呈现出显著的加工模式差异(Zwaan, 1991)。2011 年, Jacobs 教授

陈庆荣,心理学博士,南京师范大学心理学院教授、博士生导师(南京 210097);窦芬,南京师范大学心理学院博士生(南京 210097)。本文系国家自然科学基金重点项目“中国古代诗歌阅读中韵律和诗意理解的认知和神经机制”(18AYY010)的研究成果。

①引自习近平总书记在中共中央政治局第十八次集体学习时的讲话,2014 年 10 月 13 日。

首次提出“神经认知诗学”的概念,并综合修辞学、美学、诗学、心理学、情感神经科学的研究成果构建了神经认知诗学模型(Neurocognitive poetics model,NCPM,2015),从而尝试将关于神经元、认知、情感和行为过程的假设与语言学和诗学的假设联系起来,预测和测评文本元素的认知或审美过程(Jacobs,2018,2019;Jacobs & Kinder,2020;Jacobs,Lüdtke,Aryani,Meyer-Sickendieck & Conrad,2016)。在实验研究层面,研究者发现,阅读诗歌时人们会用更多的认知资源加工具有明显诗歌特征的前景要素(Yaron,2002);诗歌的节奏和韵律特征会显著减少阅读时间,增加人们对诗歌信息的记忆(Lea,Rapp,Elfenbein,Mitchel & Romine,2008;Tillmann & Dowling,2007),生成更强烈的情感体验和审美评价(Obermeier,Menninghaus,Von Koppenfels,Raettig,Schmidt-Kassow,Otterbein & Kotz,2014),且这种审美评价也受到读者对诗歌的喜爱度调节(Kraxenberger & Menninghaus,2017);词汇与声音的相似性会促进诗意性文本的词汇识别和语义通达;诗歌阅读激活了与韵律及复杂音乐模式加工相关的脑区,如右侧颞前回、左侧颞上回等(Zeman,Milton,Smith & Rylance,2013)。在诗词的测评和认知计算层面,研究者构建了古腾堡英语诗歌语料库(Gutenberg English Poetry Corpus,GEPC),且运用机器学习的方法构建了SentiArt工具,定量分析和计算诗歌文本的情感和审美潜力(Jacobs,2019;Jacobs & Kinder,2020)。在诗歌阅读对学生认知能力的影响层面,Kuppen等(2017)发现,有节奏的韵律训练可以提升5—6岁低社会经济阶层儿童的语音意识。国内研究者主要采用眼动、EEG和fMRI等实验技术考察了中国古代诗歌的认知和神经机制。Chen等(2016)采用EEG技术考查了人们阅读律诗时押韵和语义加工的神经认知机制。研究发现,阅读古代诗歌时押韵有典型的神经电生理成分P200,且押韵规则性在晚期时间窗口会调节诗句语义理解。进一步的眼动追踪研究发现,中国古代诗歌具有自上而下的、全时程的典型押韵效应,使得中国人在阅读过程中期待声律和谐、形意相合的文句(陈庆荣、杨亦鸣,2017)。针对汉语诗歌平仄节奏,Chen等(2016)采用眼球追踪技术分析了不同文本分割条件对阅读汉语格律诗的影响,发现节奏切分与字符切分并不利于诗歌阅读,而句法切分则促进了阅读过程。此外,Teng等(2020)运用MEG技术揭示了诗歌结构以及听众对诗歌结构的先验知识影响语音分段的脑机制。

综上所述,较之于印欧语诗歌的理论、实验和测评研究,国内研究者围绕押韵、节奏等典型古诗韵律特征开展了一些实验研究,但是缺乏古代诗歌阅读与个体若干认知能力发展的互动机制研究,以及建筑在系统科学研究基础上的理论构建。造成这种现象的主要原因之一是缺乏测量个体古诗知识经验水平的科学工具。鉴于此,研究者在已有诗歌认知机制实验研究的基础上,根据心理测量学的理论和方法,尝试初步编制《古代诗歌知识经验测评问卷》,并对问卷条目的难度、区分度以及信度和效度进行初步检验,以此测查个体的古代诗歌知识经验水平,为后续中国古代诗歌相关的实证和计算研究提供有参考价值的测评工具。

## 二、研究工具

### (一) 确定测验内容

中国古代诗歌,一般称作旧诗,广义的古代诗歌,是指用文言文和传统格律创作的诗,包括古代的韵文,如赋、词、曲等;狭义上则仅包括古体诗和近体诗。本问卷的测评题目是测量个体古代诗歌知识经验水平的一种“量表”,聚焦于狭义的中国古代诗歌。依据心理测验的理论框架,结合已有研究文献、调查获得的实际情况以及已有相关研究(卢莉君,2019),确定了测评问卷的结构。<sup>①</sup>研究者

<sup>①</sup>“在此感谢本人指导的研究团队成员卢莉君、马晓雅、陈雨绮、梁婧涵、高攀科等在部分初始题目编写和部分数据收集过程中付出的辛勤劳动。”

将测评问卷划分为三部分,第一部分设置成填空题,属于客观题;第二部分设置为主观题;第三部分设置为诗句补全题,属于客观题(诗句难度分为难、中、易,按 3:5:3 的比例分配)。

(二) 收集与编写测评问卷项目

本研究运用经典测验理论来编制和分析测评问卷。首先,查阅相关教材、文献中与中国古代诗歌相关的知识,对这些知识点进行细致筛选,编入测评问卷,形成测评问卷条目。在项目编写的过程中,试题中各个题目彼此独立,不含暗示其它题目或参考答案的线索。为保证项目经过筛选后仍有足够的正式项目量,研究者汇总了尽可能多的各类题目。经过小范围的预测后,剔除不符合条件的条目,对个别条目进行修订,并请教古代诗歌经验丰富的专家进行评选,最终形成测试题本。

(三) 抽取实测样本

实测样本抽取了 217 名被试,其中,诗歌经验者 107 人,非诗歌经验者 110 人;男生 103 人,女生 114 人;被试年龄为 18—28 岁,平均年龄 21.82 岁( $SD = 2.32$ )。测试由经过训练的心理学专业研究生主持,指导语和测试程序一致。

(四) 分析测验结果

将所有施测所得数据录入 SPSS 21.0 软件进行统计分析。具体如下:

1. 难度和区分度分析

古诗知识测评问卷中除第二部分为主观题,第一部分的填空题和第三部分的诗句补全题均为客观题,填空题答对记 1 分,答错记 0 分;诗句补全题答对记 2 分,答错记 0 分;第二部分主观题按照对应知识点给分,共包括三道题目。测评问卷项目的难度使用通过率来表示,区分度则采用项目鉴别力指数法和项目与总分的点二列相关来考察。

2. 信度分析

使用 Cronbach's  $\alpha$  系数分析测评问卷条目的同质性信度。使用古诗知识测评问卷对同一组被试前后施测两次,以两次测评问卷分数的稳定性程度表示测评问卷的重测信度。

3. 效度分析

采用专家评定法和经验法检验测评问卷的内容效度。从考察测评问卷的同质性方面进行结构效度检验。

三、研究结果

(一) 测评结果

两组被试在古诗知识测评问卷各部分得分以及总分的平均值和标准差见表 1。结果表明,两组被试在古诗知识测评问卷上的得分存在显著差异,诗歌经验组被试在测评问卷上的得分更高( $t(215) = 60.92, p < 0.001, \text{Cohen's } d = 6.35$ )。此外,被试在古诗知识测评问卷总分和三个部分上的得分均不存在性别差异( $ts < 1.42, ps > 0.16$ )。

表 1 古代诗歌知识经验测评问卷的原始分 ( $M \pm SD$ )

测评问卷	诗歌经验组	非诗歌经验组	总体
第一部分(N=32)	25.83 $\pm$ 2.54	12.79 $\pm$ 4.11	19.22 $\pm$ 7.38
第二部分(N=3)	22.78 $\pm$ 5.50	5.64 $\pm$ 2.32	14.09 $\pm$ 9.56
第三部分(N=11)	17.10 $\pm$ 3.25	6.05 $\pm$ 3.42	11.50 $\pm$ 6.46
总分(N=46)	65.71 $\pm$ 7.53	24.48 $\pm$ 6.35	44.81 $\pm$ 21.25

(二) 项目分析

1. 难度分析

本研究将被试在测试条目上的通过率作为难度的指标,所得整体和不同部分的难度分布见表2。总体而言,古代诗歌知识测评问卷各部分的难度在0.31-0.60之间,平均难度为0.56。与客观题(第一、三部分)相比,主观题部分(第二部分)的通过率更低,因此对被试来说该部分题目难度更高。

表2 古代诗歌知识测评问卷各部分以及总体的通过率(M±SD, %)

测评问卷	诗歌经验组	非诗歌经验组	总体
第一部分	80.72 ± 12.62	39.97 ± 12.43	60.35 ± 10.29
第二部分	49.48 ± 3.55	12.17 ± 3.49	30.82 ± 2.92
第三部分	77.74 ± 14.58	27.52 ± 21.91	52.63 ± 17.56
总分	77.97 ± 14.73	35.18 ± 16.72	56.57 ± 14.14

研究者进一步分析了不同组被试在诗歌知识测评问卷上的通过率。结果发现,诗歌经验组和非诗歌经验组被试存在显著差异( $t(90) = 13.02, p < 0.001, \text{Cohen's } d = 2.69$ )。具体而言,诗歌经验组被试在测评问卷条目上的通过率高,因此对他们而言题目难度偏小;非诗歌经验组在测评问卷条目上的通过率偏低,意味着题目偏难。此外,本研究进一步分析了每个项目在不同组别以及总体中的难度,其通过率分布如图1所示。

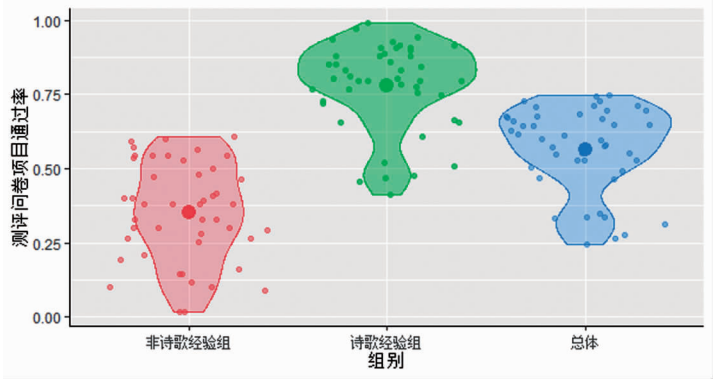


图1 古代诗歌知识经验测评问卷项目的通过率

2. 区分度分析

本研究从两个方面来考察项目的区分度,一方面计算鉴别力指数,另一方面计算各条目与问卷总分的相关。

首先,将诗歌知识测评问卷的三个部分以及总体得分按照从高到低的顺序进行排列;然后,从分布的两端各抽取27%的被试得出高分组和低分组,将两组在项目上的通过率相减即为鉴别力指数。每个项目的鉴别力指数分布如图2所示。其中,鉴别力指数在0.3以上的项目占总项目89.13%,其中,测评问卷的第一、二、三部分的平均鉴别力指数分别为0.48、0.45、0.59,总鉴别力指数为0.51。根据美国测验专家伊贝尔(L. Ebel)提出的鉴别力指数评价项目性能的标准(郑日昌,2002),即鉴别力指数 $\leq 0.19$ 差,0.20-0.29尚可,0.30-0.39良好,0.40以上很好,本研究中测评问卷各部分及整体的鉴别力指数很好。

其次,采用相关法计算测评问卷中每个项目与测评问卷总分的相关系数进一步考察项目的区分度。古代诗歌测评问卷各项目与问卷总分的相关分布如表3所示。此外,古诗知识测评问卷的第一、二、三部分的得分与问卷整体得分的平均相关系数在0.88-0.93之间,均达到优良水平(具体见表6)。

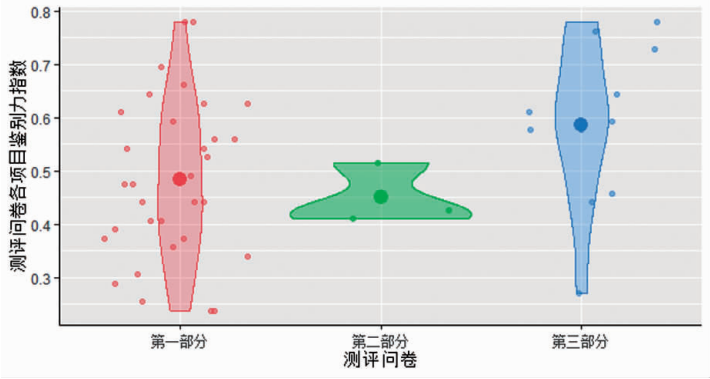


图2 古代诗歌知识经验测评问卷各项目的鉴别力指数分布图

表3 古代诗歌知识经验测评问卷各项目与问卷总分的相关

r 等级	第一部分	第二部分	第三部分
≤0.29	8 (25.0%)	0 (0)	0 (0)
0.30~0.39	8 (25.0%)	0 (0)	1 (9.1%)
0.40~0.49	3 (9.4%)	0 (0)	2 (18.2%)
≥0.50	13 (40.6%)	3 (100%)	8 (72.7%)

注:表中括号内数据为测评问卷题数占总数的百分比。

(三) 信度分析

1. 同质性信度

本研究使用 Cronbach’s  $\alpha$  系数对各部分以及总分的内部一致性信度进行检验。结果分析,测评问卷的一、二、三部分的  $\alpha$  系数分别为 0.89、0.86、0.84,总问卷的  $\alpha$  系数为 0.90,说明其内部一致性较高。

2. 重测信度

对 217 名被试首次施测后,选取其中 102 名被试相隔 20 天进行重新施测,根据前后两次施测的相关来计算信度系数,以考察测评问卷的稳定性和等值性。从表 4 可知,诗歌知识测评问卷三个部分的重测信度系数在 0.74~0.85 之间,总问卷的重测信度系数为 0.88,均达到显著水平。这表明古代诗歌知识测评问卷具有较好的跨时距稳定性。

表4 古代诗歌知识经验测评问卷的重测相关系数

	初测 (M ± SD)	重测 (M ± SD)	相关系数 (r)
第一部分	21.97 ± 8.14	22.07 ± 8.04	0.82 **
第二部分	16.11 ± 10.30	19.23 ± 10.93	0.85 **
第三部分	13.61 ± 6.57	14.09 ± 5.95	0.74 **
总量表	51.69 ± 22.07	55.04 ± 20.02	0.88 **

注: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ,下同。

(四) 效度分析

1. 内容效度

本研究使用专家判断法和经验法来考察测评问卷的内容效度。首先,邀请 6 位古代诗歌经验丰富的专家对所编制问卷的各项进行符合性评定,采用五点计分(1 表示“较差匹配”,2 表示“一般匹配”,3 表示“较好匹配”,4 表示“很好匹配”,5 表示“完美匹配”)。结果发现,问卷项目的平均符合度为 3.84 ( $SD = 0.52$ ),各项的平均符合度见表 5,这说明该问卷能够较好区分出诗歌经验者和非经验者的诗歌知识水平。

表 5 专家对各项目的评定等级

项目	等级	项目	等级	项目	等级	项目	等级	项目	等级	项目	等级
1	4.33	9	2.50	17	4.17	25	4.00	33	4.50	41	4.33
2	4.33	10	3.67	18	4.33	26	4.33	34	3.67	42	4.33
3	3.67	11	2.83	19	4.00	27	3.67	35	4.00	43	2.67
4	3.33	12	3.50	20	4.00	28	3.67	36	3.00	44	3.67
5	3.33	13	3.67	21	2.67	29	4.33	37	4.00	45	4.33
6	4.33	14	4.33	22	3.33	30	4.33	38	4.00	46	4.00
7	3.5	15	4.50	23	4.33	31	4.33	39	4.33		
8	3.5	16	3.50	24	4.17	32	4.00	40	3.67		

其次,根据经验法,诗歌经验组被试和非经验组被试的古诗知识测评得分应该存在较大差异。统计结果表明,诗歌经验组和非诗歌经验组在测评问卷三个部分的得分( $t(215) = 28.17, p < 0.001$ ,Cohen's  $d = 3.93$ ; $t(215) = 29.77, p < 0.001$ ,Cohen's  $d = 4.10$ ; $t(215) = 24.37, p < 0.001$ ,Cohen's  $d = 3.33$ )和总体得分( $t(215) = 60.91, p < 0.001$ ,Cohen's  $d = 6.35$ )均存在显著差异。具体而言,与非诗歌经验组被试相比,诗歌经验组被试在测评问卷各部分以及总量表的原始得分都明显增高(具体见表1)。此外,总分和题目的通过率也随着诗歌经验而升高(见表2和图1)。这些结果说明该测评问卷具有较好的内容效度。

2. 结构效度分析

结构效度是指一个测验实际测到所要测量的理论结构和特质程度。本研究从同质性方面考察测评问卷的结构效度,计算问卷各部分的相关及总分的相关。古代诗歌知识测评问卷不同部分的题目虽然在内容和认知水平上要求不同,但都属于古代诗歌知识范畴。因此,测评问卷各部分之间应该既有差别也有联系。统计结果表明(具体见表6),各部分之间具有较强的相关,但均低于各部分与总分的相关。

表 6 中国古代诗歌知识经验测评问卷各部分及总量表间的相关系数

测评问卷	第一部分	第二部分	第三部分	总量表
第一部分	1			
第二部分	0.76 **	1		
第三部分	0.71 **	0.73 **	1	
总量表	0.90 **	0.93 **	0.88 **	1

(五) 古诗相关的背景信息调查

研究者在施测时还收集了被试具有的古体诗歌相关的背景信息。统计分析表明,诗歌经验组中大约有 64% 的被试曾经加入过与古诗相关的团体组织进行经验交流,71% 的被试创作过古诗,44% 的被试创作的古诗作品发表在校级及以上刊物上;非经验组被试对应的比例分别为 21%、38% 和 11%。此外,研究者进一步使用 7 点计分评定被试对诗歌的喜爱度,发现诗歌经验组对诗歌的喜爱度( $5.90 \pm 0.76$ )显著高于非经验组被试( $4.36 \pm 1.00$ ), $t(215) = 12.73, p < 0.001$ ,Cohen's  $d = 1.73$ (具体见图3)。除此之外,与非经验组被试相比,诗歌经验组被试会花费更多的时间进行与古诗相关的阅读( $t(215) = 12.61, p < 0.001$ ,Cohen's  $d = 1.72$ ),阅读量也显著增加, $t(215) = 8.18, p < 0.001$ ,Cohen's  $d = 1.12$ (具体见图4和图5)。上述结果也在一定程度上表明该问卷具有良好的效度。

四、讨论

先前的文献综述表明,目前缺乏测量个体古代诗歌知识经验水平的相对科学的测评工具,从而在很大程度上阻碍了研究者基于科学工具深入开展古代诗歌学习及其对我国青少年学生若干认知

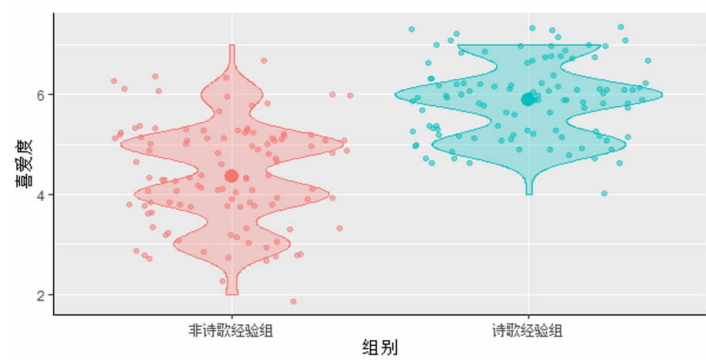


图3 诗歌经验组和非经验组被试对中国古代诗歌的喜爱度

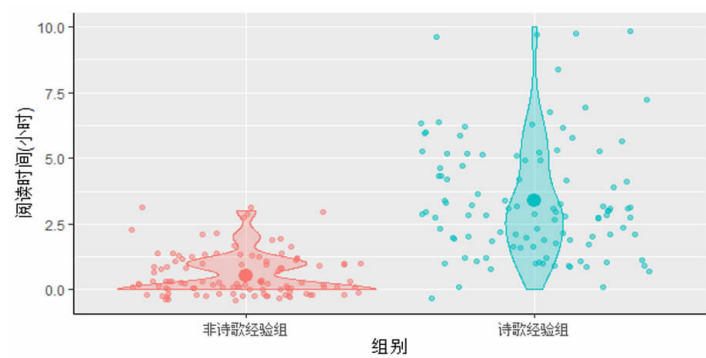


图4 诗歌经验组和非经验组被试阅读古诗及相关信息的时间分布

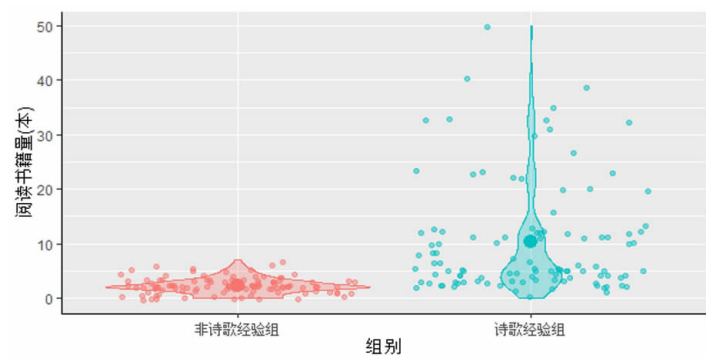


图5 诗歌经验组和非经验组被试对古诗及相关书籍的阅读量

能力影响的基础和应用研究。因此,本研究遵循心理测验编制的基本原理和程序,初步编制了中国古代诗歌知识经验的测评问卷以弥补该领域的不足。该问卷的编制经历了测评结构和内容的确定、测评项目的收集与编写、实测样本的抽取以及测评项目质量的分析四个环节。

基于已有的研究成果以及国内外相关文献,本研究确定了测评内容并对项目进行收集与编写。问卷中的部分题目来自于本人带领的研究团队在古诗阅读神经机制研究中已初步编制的古诗知识测评问卷(卢莉君,2019)。最终抽取实测样本进行检验的测验项目均经过筛选和专家评估。

### (一) 项目分析

#### 1. 测评问卷的难度

理论上一般认为,项目的难度在0.5左右时最具有鉴别能力,同时在实际编制测评问卷时应充

分考虑到不同水平的被试。因此一般来说,特别难和特别容易的项目应尽量少一些,中等难度的项目占多数,所有项目的平均难度应在 0.5 左右(金瑜,2001;王汉澜,1987)。本研究的结果表明,整体而言,中国古代诗歌知识测评问卷各部分的难度在 0.31 - 0.60 之间,平均难度为 0.56。因此,大部分项目的难度分布比较理想。进一步数据分析显示,项目难度分布在不同诗歌经验组被试中表现出明显差异。具体而言,对于诗歌经验组被试,测评问卷项目难度偏低;对于非诗歌经验组被试,测评问卷项目难度偏高。这种难度系数的分布差异是由于被试的诗歌经验水平导致的,符合测评问卷的效度指标。

## 2. 测评问卷的区分度

“测评问卷的区分度是检验测验项目的另一重要指标,本研究从项目的鉴别力指数,以及各条目与总分的相关两个方面对其进行考察。”从鉴别力指数看,89.13% 的项目具有较好的区分度,达到了优良水平(戴海崎、张锋、陈雪枫,2002);测评问卷各部分的平均鉴别力指数在 0.45 - 0.59 之间,整体的鉴别力指数为 0.51,达到了很好水平(郑日昌,1987;漆书青,1998)。从测评问卷项目与总分的相关来看,所有测评项目与总分均呈正相关。诗歌知识经验测评问卷三个部分的得分与整体得分之间的相关系数均在 0.8 以上。以上结果表明,该问卷的大部分项目区分度良好。

## (二) 信度

本研究采用同质性信度和重测信度对测评问卷的信度进行了考察。一般来说,信度水平在 0.80 以上为好。本研究使用 Cronbach's  $\alpha$  系数对测评问卷各部分以及总分的内部一致性信度进行检验。结果表明,三部分的  $\alpha$  系数分别为 0.89、0.86、0.84,总问卷的  $\alpha$  系数为 0.90,说明内部一致性较高。同时,重测信度结果表明,古代诗歌知识测评问卷三个部分的重测信度系数在 0.74 - 0.85 之间,总量表的信度系数为 0.88,均达到显著水平,表明测评问卷分数的稳定程度较高。因此,整体而言,测评题本的信度水平良好,作为测量个体古代诗歌知识经验水平具有一定的可靠性。

## (三) 效度

测验效度的好坏反映该测验能否真实测量出想要测量的心理特质。本研究分别从内容效度和结构效度方面考察中国古代诗歌知识测评问卷的效度。首先,测评问卷的内容效度反映了测评问卷项目对整个测验内容的代表性程度(Leftly & Pennington,2000)。研究者分别使用专家判断法和经验法考察问卷的内部效度。6 位古代诗歌经验丰富的专家的评定结果表明,测评项目的平均符合度为 3.84,说明该问卷比较符合古代诗歌知识的内容,具有较高的一致性。其次,理论上,诗歌经验组被试和非经验组被试在诗歌知识测评问卷上的得分存在显著差异。测评数据表明,与非诗歌经验组被试相比,诗歌经验组被试在测评问卷各部分以及总量表的原始得分都明显增加。该结果验证了我们的猜想,说明该古代诗歌知识问卷的测评项目具有较好的内容效度。

在测评问卷的结构效度方面,本研究从测验的同质性方面对其进行考察。通过统计分析发现,测评问卷各部分之间具有较强的相关,但均低于各部分与总分的相关,说明测评问卷符合预想的结构,即中国古代诗歌知识测评问卷不同部分的题目虽然在内容和认知水平上要求不同,但都属于古代诗歌知识范畴。因此,测评问卷各部分之间既有差别也有联系。这也进一步表明,中国古代诗歌测评问卷具有良好的结构效度。

值得关注的是,本研究进一步分析了被试具有的古代诗歌相关的背景信息。分析结果显示,诗歌经验组被试无论是在诗歌创作和创作的质量还是对诗歌的喜爱以及对与古代诗歌相关的书籍资料的阅读时间、阅读量等方面都显著优于非诗歌经验组。这些结果也进一步说明本测评问卷具有良好的效度。



本研究所编制的中国古代诗歌知识测评问卷无论是在项目的难度和区分度上,还是在信度和效度上均良好,能够在一定程度上科学测评个体的古代诗歌知识经验水平,因此在一定程度上弥补了该领域的不足,为后续在该领域的研究提供了借鉴思路。

#### 参考文献:

- 陈庆荣、杨亦鸣,2017:《古诗阅读的认知机制:来自眼动的证据》,《中国社会科学》第3期。
- 戴海崎、张锋、陈雪枫,1999:《心理与教育测量》,广州:暨南大学出版社。
- 戴海崎、张锋、陈雪枫,2002:《心理与教育测量》,广州:暨南大学出版社。
- 金瑜,2001:《心理测量》,上海:华东师范大学出版社。
- 卢莉君,2019:《中国古代哲理和想象诗歌欣赏的大脑成像研究》,硕士学位论文,南京师范大学发展与教育心理学。
- 罗良功,2002:《英诗概论》,武汉:武汉大学出版社。
- 漆书青、戴海崎、丁树良,1998:《现代教育与心理测量学原理》,南昌:江西教育出版社。
- 王汉澜,1987:《教育测量学》,开封:河南大学出版社。
- 王力,1978:《诗词格律十讲》,北京:北京出版社。
- 王力,2002:《诗词格律概要》,北京:北京出版社。
- 王力,2003:《王力近体诗格律学》,太原:山西古籍出版社。
- 郑日昌,1987:《心理测量》,长沙:湖南教育出版社。
- 郑日昌、蔡永红、周益群,2002:《心理测量学》,北京:人民教育出版社。
- 朱光潜,2001:《诗论》,上海:上海古籍出版社。
- Chen, Q. R., W. T. Gu & C. Scheepers, 2016, “Effects of text segmentation on silent reading of Chinese regulated poems: Evidence from eye movements”, *Journal of Chinese Linguistics*, Vol. 44, pp. 265 – 286.
- Chen, Q. R., J. J. Zhang, X. D. Xu, C. Scheepers, Y. M. Yang & M. K. Tanenhaus, 2016, “Prosodic expectations in silent reading: ERP evidence from rhyme scheme and semantic congruence in classic Chinese poems”, *Cognition*, Vol. 154, pp. 11 – 21.
- Eliot, T. S., 1957, *The Social Function of Poetry*, London: Faber and Faber Inc.
- Hanauer, D., 1998, “The genre-specific hypothesis of reading: Reading poetry and encyclopedic items”, *Poetics*, Vol. 26, No. 2, pp. 63 – 80.
- Jacobs, A. M., 2018, “(Neuro-) cognitive poetics and computational stylistics”, *Scientific Study of Literature*, Vol. 8, No. 1, pp. 165 – 208.
- Jacobs, A. M., 2019, “Sentiment Analysis for Words and Fiction Characters from the Perspective of Computational (Neuro-) Poetics”, *Frontiers in Robotics and AI*, Vol. 6, pp. 1 – 13.
- Jacobs, A. M. & A. Kinder, 2020, “Computing the affective-aesthetic potential of literary texts”, *Artificial Intelligence*, Vol. 1, No. 1, pp. 11 – 27.
- Jacobs, A. M., J. Lüdtke, A. Aryani, B. Meyer-Sickendieck & M. Conrad, 2016, “Mood-empathic and aesthetic responses in poetry reception: A model-guided, multilevel, multimethod approach”, *Scientific Study of Literature*, Vol. 6, No. 1, 87 – 130.
- Jakobson, R., 1960, “Linguistics and poetics”, in T. Sebeok (ed.), *Style in Language*, Cambridge, MA: MIT Press, pp. 350 – 377.
- Kraxenberger, M. & W. Menninghaus, 2017, “Affinity for poetry and aesthetic appreciation of joyful and sad poems”, *Frontiers in Psychology*, Vol. 7, pp. 1 – 7.
- Leftly, D. L. & B. F. Pennington, 2000, “Reliability and validity of the adult reading history questionnaire”, *Journal*

*of Learning Disabilities*, Vol. 33, No. 3, pp. 286 – 296.

Obermeier, C., W. Menninghaus, M. Von Koppenfels, T. Raettig, M. Schmidt-Kassow, S. Otterbein & S. A. Kotz, 2014, “Aesthetic and emotional effects of meter and rhyme in poetry”, *Frontiers in Psychology*, Vol. 4, pp. 1 – 10.

Teng, X. B., M. Ma, J. B. Yang, S. Blohm, Q. Cai & X. Tian, 2020, “Constrained structure of ancient Chinese poetry facilitates speech content grouping”, *Current Biology*, Vol. 30, No. 7, pp. 1 – 7.

Tillmann, B. & W. J. Dowling, 2007, “Memory decreases for prose, but not for poetry”, *Memory & Cognition*, Vol. 35, No. 4, pp. 628 – 639.

Yaron, I., 2002, “Processing of obscure poetic texts: Mechanisms of selection”, *Journal of Literary Semantics*, Vol. 31, No. 2, pp. 133 – 170.

Zeman, F. Milton, A. Smith & R. Rylance, 2013, “By heart: An fMRI study of brain activation by poetry and prose”, *Journal of Consciousness Studies*, Vol. 20, No. 9/10, pp. 132 – 158.

Zwaan, R., 1991, “Some parameters of literary and news comprehension: Effects of discourse type perspective on reading rate and surface structure representation”, *Poetics*, Vol. 20, No. 2, pp. 139 – 156.

(责任编辑: 蒋永华)

## Constructing a Questionnaire to Test the Knowledge of Ancient Chinese Poetry: A Preliminary Result

CHEN Qingrong, DOU Fen

**Abstract:** In this study, we adopted the method of combining theories and empirical evidence to construct a questionnaire to examine readers' knowledge of ancient Chinese poetry. We recruited 217 college students to test the quality of items and the reliability and validity of this questionnaire. The results showed that the difficulty of each part of the questionnaire ranged from 0.35 to 0.68, and the mean difficulty was 0.56. The discrimination of 89.13% items was over 0.3 and the discrimination index of the questionnaire was 0.51. The coefficient and the test-retest reliability coefficient were both above 0.80. Furthermore, we analyzed the participants' related experience and knowledge in ancient Chinese poetry, such as the membership of the communities concerning poetry, and the creation and publication of ancient Chinese poems. The results showed that in contrast to the inexperienced group, more experienced participants had experience in joining the related ancient poetry community (64% vs. 21%), creating (71% vs. 38%) and publishing ancient poems (44% vs. 11%) than the inexperienced participants. Additionally, the group with high poetic experience has spent a lot of time reading ancient Chinese poems. These findings suggest that the knowledge questionnaire of Chinese ancient poetry is satisfactory in terms of the difficulty, discrimination, reliability and validity, and it could be an effective and reliable tool to evaluate one's knowledge of Chinese ancient poetry.

**Keywords:** knowledge of ancient Chinese poetry; item analysis; reliability; validity; difficulty

**About the authors:** CHEN Qingrong, PhD in Psychology, is Professor at School of Psychology, Nanjing Normal University (Nanjing 210097); DOU Fen is PhD Candidate in Psychology at School of Psychology, Nanjing Normal University (Nanjing 210097).