

梅尔美术判断测验在中国应用的信、效度检验*

董圣鸿^{1,2}, 黎旋^{1,2}, 郭璐^{1,3}, 李小雨^{1,2}

(1. 江西师范大学心理学院, 南昌 330022; 2. 江西省心理与认知科学重点实验室, 南昌 330022;

3. 中国移动通信集团江西有限公司景德镇分公司, 景德镇 333000)

摘要:目的:修订梅尔美术判断测验(Meier Art Judgment Test)并对其信度、效度进行检验。方法:通过对来自6所大学、中专共2270人施测梅尔美术判断测验,采用CTT区分度和IRT的模型拟合检验、区分度筛选项目,以霍兰德艺术分测验、学生艺术创作水平自评与艺术过往经历分量表为效标,以及采用效标组法(美术与非美术专业)检验效标关联效度。结果:保留的61题都拟合IRT的2参数logistic模型,量表得分与各效标得分相关显著,美术与非美术专业学生得分存在显著差异;但测验信息量分析表明,对高能力被试的测量误差相对较大。结论:修订的量表能测量个体的美术判断能力;今后改进方向应该是增加更难的试题。

关键词:美术才能;梅尔美术鉴赏测验;信度;效度

中图分类号:B848

文献标识码:A

文章编号:1003-5184(2019)06-0549-07

1 引言

能力倾向测验是对个体在不同能力因素上潜在优劣倾向的心理测验。特殊能力倾向测验是专门测验个体某一特种能力倾向而设计的测验,比如,美术能力倾向测验,它所测量的不是个体是否具有足够的美术水平,而是测量个体在未来有没有潜在的美术能力。由于能力倾向测验可以从不同能力因素上来评估个体,因而被广泛应用于专业、职业指导与人事管理工作之中(戴海琦,2010)。

由于文化差异、年代差异等多种因素的影响,艺术品味、风格偏好会存在个体差异,但总体上,比较一致的看法是美术能力倾向可以从审美能力(辨别和感知)和艺术创造技能的角度进行测量。学者们从这两个角度分别开发出了一些测验,如从审美能力的角度开发的测验有桑代克审美测验(Thorndike tests of esthetic appreciation)、梅尔美术鉴赏测验(Meier Art Judgment Test)、格雷夫斯图形鉴赏测验(Graves Design Judgment Test)、戈登艺术判断测验(Gordon tests of esthetic judgments),从艺术创造技能的角度进行测量的有洪恩美术能力倾向量表(The Horn Art Aptitude Inventory)。另外,我国台湾学者邱维城(1969)结合梅尔美术鉴赏测验和格雷夫斯图形鉴赏测验开发了一项测验,有学者后来对该测验又进行了修订,用于测试小学三年级到中学三年级学生的艺术能力倾向(邱维城,1970;陈荣

华,邱维城,王秀雄,卢钦铭,范德鑫,1981)。这些测验当中,前三个测验被心理测量学类的图书广泛介绍(丁秀峰,2001;Lewis & Gray,2005;戴海琦,2010;顾海根,2010;郑日昌,2011;郑日昌,孙大强,2012)。

在测量美术能力倾向的两个角度中,从审美能力的角度进行测试更具有普遍意义。因为艺术创造技能不仅要求个体具有较高的艺术鉴赏能力,还需要一定的专业训练;没有很强绘画能力的人,可能会具有较高的艺术鉴赏能力;对于画家来说,这两种能力是缺一不可的。所以,对个体进行艺术鉴赏能力的评价,对其专业发展指导具有重要意义。但是,以往从评价审美能力角度开发的各种测验中,有些测验存在一定问题。一是有些测验没有区分偏好与艺术判断。例如,威尔士图形偏好测试(The Welsh Figure Preference Test;Welsh & Barron,1949)和格雷夫斯图形鉴赏测验要求的是偏好而不是判断;桑代克审美测验要求的本质上也是偏好,它要求将五种设计从最令人愉悦的到最不令人愉悦的排序,但对“愉悦”并没有进一步定义,模糊地指审美和非审美两种愉悦(Bamossey, Johnston, & Parsons, 1985)。一个人在拥有偏好方面是没有好坏之分的,偏好也不能反映一个人鉴赏能力的高低,偏好是比判断更广泛、更主观的决定。人们喜欢一幅画可能有多种原因,有些原因与它的审美特征无关。例如,偏好可能

* 基金项目:江西省文化艺术科学规划重点项目:美术才能自适应测验开发及其在美术人才早期甄别中的应用(YG20163111)。

会因为这幅画让人想起一些令人愉快的东西,或者因为作品中有自己喜欢的颜色,或者因为对作品主题感兴趣而产生。而审美判断是需要更明显的认知和客观决定的过程。史密斯认为,审美判断的基本过程包括描述(识别和分类)、分析(检查作品各部分之间的关系)、解释(赋予作品整体意义)和评价(基于之前的活动)(Smith, 1971)。另一个问题是,有一些测验没有考虑到艺术品和简化材料之间的区别。艺术品是复杂的“刺激对象”,很难说其中最能够影响个体反应的是是什么。这使得一些经验主义的美学家简化了刺激因素,以便控制对结果的解释。虽然这种实验室方法可以更准确地测量自变量,但此类研究的结果缺乏外部效度。例如,格雷夫斯图形鉴赏测验要求一个人在两个简单的抽象设计中做出选择(Graves, 1939),桑代克(Thorndike, 1926)和戈登(Gordon, 1923)把他们的量表建立在规范的设计元素上。这种关注设计元素的做法限制了评估鉴赏艺术品能力的有效性。而梅尔美术判断测验采用真实的作品,只是对这些杰出作品中的元素作了重新的安排,同时要求被试判断哪个更好(更有艺术才能的、更令人满意的),较好地克服了这两个不足,成为美术判断测验中较为成功的一个测验。梅尔美术判断测验也是当前各类心理测量教材中重点介绍的审美能力测验,但国内翻译的名称各异,有“梅尔艺术判断测验”、“梅耶尔审美感测验”等等。由于英语中的“art”一词既作“艺术”解释,又作“美术”解释,它既可以指音乐、舞蹈、文学、戏剧和电影等其他各种艺术门类,有时也用来专指包括绘画、雕塑、工艺和建筑在内的视觉艺术(贾炜, 2009)。因此,建议统译为“梅尔美术判断测验”。

虽然人们开发出许多审美能力的测验,但由于审美能力是一种特有的高级能力(李静, 2017),美术能力包括哪些心理成分,目前还没有统一的认识。因此,对于美术能力的判断比较困难(丁秀峰, 2001)。艺术测验的争议和复杂使得心理学家不敢进一步尝试更精确的测试(Crannell, 1953)。但是现在我国发展的现实对开发审美能力测验提出了实践需求。《国家中长期教育改革和发展规划纲要》中明确指出,要“提高义务教育质量”,“配齐美术等薄弱学科教师”;2014 年国家颁布《国家艺术基金章程》,指出“国家艺术基金是由国家设立,旨在繁荣艺术创作、打造和推广原创精品力作、培养艺术创作人才、推进国家艺术事业健康发展的公益性基金。”

无论是“配齐美术等薄弱学科教师”,还是要实现《国家艺术基金章程》设立的宗旨,都依赖于艺术人才,所以如何识别及选拔真正具有创新能力的美术人才就成为了关键。而我国目前没有性能优良的美术才能的相关测验,选择学习艺术的学生多数也不是因为擅长而选择艺术专业。有研究表明,超过 91% 的报考美术专业的学生是进入高中后由于文化课学习不佳等各种原因而转入美术班学习的(王芳,饶俊, 2013),这种现象十分不利于真正有天赋的艺术人才的发现与成长。张旭东(2019)也指出,视觉艺术素养是核心素养的分支,而识别人才是因材施教的前提,因此,有必要开发出能鉴别个体美术能力倾向的测验,而近期内快速有效的方式就是修订已有的较好量表。所以,本研究拟对梅尔美术判断测验进行修订,检验其在中国背景下的信、效度。

2 研究方法

2.1 被试

本研究采用线上问卷形式,共对 2270 名本科生、研究生和中专生进行施测,删除作答时间小于 500 秒的被试 692 人(美术生 10 人,非美术生 682 人),有效被试 1578 人,其中男 536,女 1042;美术专业 229,非美术专业 1349;年龄范围 14 - 43 岁,中专生(高中生)356 人,大一学生 754 人,大二学生 296 人,大三学生 107 人,研究生及以上 65 人。

2.2 研究材料

线上测试问卷包含三个部分的内容:

(1)梅尔美术判断测验。该测验包括 100 对图片,施测时要求被试根据图片下方的差异说明,从左右两幅图片中选出自己认为最好(更使人愉悦、更具有艺术特点、更令人满意)的那幅。选对计 1 分,选错计 0 分。另有 25 道题,梅尔认为区分度较大,加权计分,选对计 2 分,选错计 0 分(Meier & Seashore, 1929; Meier, 1942)。因加权计分的也是二分计分,且目前没有证据这些题在中国应用时也有更大的区分度,所以在本研究中一律为 0、1 计分。

(2)被试基本信息。包括姓名、性别、出生年月、学校、专业,以及居住地的情况等信息。

(3)效标量表。采用了一些理论上与审美能力相关的一些问题,包括了四大部分:第一,与审美能力相关的一些过往经历的调查。弗朗西斯·罗杰斯(Rodgers, 1993)通过调查美术天分高的孩子和天分低的孩子的生活环境发现:那些天分高的孩子大多数都来自视觉刺激丰富的环境,并且从小阅读大量

的书籍和图片。此外天分高的孩子会更善于利用他们过去的经验、情感经验和视觉记忆。本研究将这些经历的调查作为效标,一共 10 题。第二,以霍兰德职业兴趣问卷(Holland,1997)中“你喜欢从事下列活动吗”、“你所擅长或胜任的活动”、“你所喜欢的职业”等与美术有关内容作为效标,共 9 题。第三,以被试自我评价“与同龄人相比,你觉得自己的美术创作水平处于什么位置”、“与同龄人相比,你觉得自己的音乐艺术技能水平处于什么位置”作为效标。第四,以美术专业与非美术专业学生为效标组,通过比较专业与非专业学生得分的差异来考察效标。

2.3 施测程序

被试在电脑上识别二维码,或者用手机扫描二维码进行作答。被试作答题目的顺序依次为基本信息量表、效标量表、梅尔美术判断测验;前一部分的量表未回答完毕,不进入下一部分量表的作答;每个部分内的作答没有顺序要求,允许被试跳题作答,但是若有题目未回答,被试将无法提交自己的作答,同时系统会自动返回到该被试第一道未完成的题目位置,如此反复直到被试完成所有项目方可提

交结果。

2.4 数据分析

基于经典测量理论的数据分析和一般的统计分析均采用 SPSS18.0 进行;项目反应理论的数据分析采用 IRTPRO4.1 进行。

3 结果

3.1 项目分析

3.1.1 基于经典测量理论区分度的项目筛选

以各题得分与总分的相关系数,以及采用极端分组法,将被试在 100 个项目的总得分由高到低进行排序,总分前 27% 的为高分组,后 27% 的为低分组,计算项目鉴别指数。表 1 列出了各题得分的题总相关与项目鉴别指数。根据项目鉴别指数在 0.2 以下必须淘汰的标准,应删除第 3、4、5、7、8、16、19、21、25、29、30、34、36、38、53、56、58、59、60、61、63、64、67、71、72、76、77、82、83、92、94、95、100 题,共 33 个项目。由于题总相关全部显著,相关较低的项目在根据项目鉴别指数删除时基本上被删除,且还要进行项目反应理论的分析,所以不再依据题总相关进行删题。

表 1 各题得分的题总相关与项目鉴别指数

项目编号	与总分的相关	CR	项目编号	与总分的相关	CR	项目编号	与总分的相关	CR
1	0.201**	0.207	34	0.263**	0.185	67	0.086**	0.129
2	0.293**	0.261	35	0.236**	0.277	68	0.297**	0.296
3	0.246**	0.167	36	0.173**	0.174	69	0.309**	0.289
4	0.174**	0.183	37	0.209**	0.230	70	0.227**	0.265
5	0.182**	0.174	38	0.118**	0.160	71	0.019	0.033
6	0.236**	0.221	39	0.229**	0.225	72	0.158**	0.185
7	0.192**	0.141	40	0.187**	0.230	73	0.266**	0.265
8	0.170**	0.124	41	0.187**	0.216	74	0.259**	0.308
9	0.234**	0.263	42	0.239**	0.254	75	0.285**	0.317
10	0.189**	0.204	43	0.257**	0.251	76	0.154**	0.195
11	0.216**	0.209	44	0.192**	0.207	77	0.059*	0.080
12	0.220**	0.230	45	0.177**	0.209	78	0.324**	0.275
13	0.216**	0.204	46	0.232**	0.246	79	0.243**	0.279
14	0.162**	0.211	47	0.281**	0.273	80	0.309**	0.263
15	0.168**	0.213	48	0.269**	0.305	81	0.332**	0.357
16	0.186**	0.171	49	0.299**	0.559	82	0.150**	0.167
17	0.235**	0.254	50	0.214**	0.242	83	0.142**	0.020
18	0.241**	0.282	51	0.188**	0.240	84	0.264**	0.301
19	0.102**	0.096	52	0.082**	0.221	85	0.297**	0.312
20	0.237**	0.251	53	0.149**	0.188	86	0.218**	0.256
21	0.150**	0.155	54	0.185**	0.437	87	0.196**	0.228
22	0.228**	0.268	55	0.241**	0.202	88	0.216**	0.258
23	0.216**	0.268	56	0.152**	0.171	89	0.348**	0.303

续表 1

项目 编号	与总分 的相关	CR	项目 编号	与总分 的相关	CR	项目 编号	与总分 的相关	CR
24	0.231**	0.244	57	0.165**	0.216	90	0.359**	0.338
25	0.133**	0.146	58	0.070**	0.089	91	0.235**	0.242
26	0.256**	0.232	59	0.102**	0.108	92	0.066**	0.106
27	0.256**	0.279	60	0.087**	0.101	93	0.178**	0.223
28	0.265**	0.261	61	0.079**	0.178	94	0.135**	0.148
29	0.134**	0.148	62	0.180**	0.204	95	-0.061*	-0.082
30	0.112**	0.131	63	0.060*	0.085	96	0.263**	0.284
31	0.277**	0.298	64	0.087**	0.124	97	0.282**	0.345
32	0.317**	0.310	65	0.200**	0.256	98	0.154**	0.207
33	0.297**	0.245	66	0.249**	0.268	99	0.190**	0.226
						100	0.062*	0.064

注：* 表示在 0.05 水平上（双尾）相关性显著，** 表示在 0.01 水平上（双尾）相关性显著；下同。

3.1.2 基于项目反应理论区分度和模型资料拟合检验进行项目筛选

使用项目反应理论的双参数 logistic 模型,对剩余的 67 题进行项目分析和模型资料拟合检验。结果发现所有项目都拟合模型;依据 0.3 为标准将低于该标准的项目进行剔除(Reeve et al,2007),删除

了第 14、15、52、54、57、98 共 6 个项目。对余下的 61 个题目再次进行项目反映理论中双参数 logistic 模型分析,所得结果表明,各题的模型资料均拟合,区分度均在 0.3 以上。另外,模型的整体检验表明所有题目整体上用双参数 logistic 模型进行拟合是恰当的($\chi^2/df=2.35$,RMSEA=0.03)。

表 2 IRT 双参数 logistic 模型资料拟合检验及项目区分度(α 值)

项目编号	α	χ^2	d.f.	P 值	项目编号	α	χ^2	d.f.	P 值
1	0.30	27.13	35	0.8269	48	0.53	36.05	35	0.4206
2	0.84	34.24	32	0.3598	49	0.88	25.82	32	0.7721
6	0.68	21.56	34	0.9518	50	0.46	27.28	35	0.8215
9	0.56	28.19	33	0.7063	51	0.41	29.86	35	0.7153
10	0.44	18.98	33	0.9758	52	-0.06	21.54	36	0.9730
11	0.44	17.45	33	0.9879	54	0.28	26.95	35	0.8333
12	0.42	14.92	35	0.9988	55	0.79	33.24	33	0.4568
13	0.54	22.06	34	0.9431	57	0.27	30.02	35	0.7081
14	0.27	28.84	35	0.7595	62	0.42	29.18	35	0.7451
15	0.22	30.08	35	0.7050	65	0.34	31.06	34	0.6136
17	0.49	20.66	35	0.9741	66	0.68	30.94	33	0.5710
18	0.49	22.38	35	0.9515	68	0.85	31.25	32	0.5054
20	0.34	28.11	34	0.7516	69	0.97	26.31	31	0.7072
22	0.53	27.25	34	0.7881	70	0.64	31.40	34	0.5967
23	0.32	37.79	34	0.2994	73	0.87	33.22	33	0.4577
24	0.65	35.36	34	0.4053	74	0.59	31.85	35	0.6221
26	0.67	34.44	33	0.4002	75	0.67	37.03	33	0.2874
27	0.61	32.33	34	0.5508	78	1.45	30.27	29	0.4020
28	0.70	32.79	32	0.4295	79	0.52	18.58	33	0.9796
31	0.60	27.85	34	0.7631	80	1.12	34.58	31	0.2998
32	0.96	16.88	32	0.9870	81	0.84	29.90	32	0.5742
33	0.73	18.78	33	0.9777	84	0.46	29.05	35	0.7507
35	0.56	27.46	34	0.7795	85	0.75	31.54	33	0.5412
37	0.51	25.99	34	0.8361	86	0.50	28.56	34	0.7317
39	0.55	27.04	34	0.7963	87	0.38	26.23	35	0.8578

续表 2

项目编号	α	χ^2	$d.f.$	P 值	项目编号	α	χ^2	$d.f.$	P 值
40	0.40	20.45	34	0.9678	88	0.47	41.71	35	0.2014
41	0.44	30.73	34	0.6297	89	1.08	42.07	30	0.0704
42	0.57	33.94	34	0.4719	90	1.24	23.87	31	0.8161
43	0.75	21.79	33	0.9323	91	0.62	27.60	34	0.7734
44	0.42	34.58	33	0.3938	93	0.45	30.82	35	0.6708
45	0.33	18.94	35	0.9877	96	0.38	35.00	34	0.4218
46	0.53	32.65	33	0.4857	97	0.74	27.82	33	0.7234
47	0.73	30.23	33	0.6066	98	0.19	29.67	34	0.6806
					99	0.30	33.84	34	0.4767

3.2 信度分析

从表 3 可见,梅尔美术判断测验的信度基本达到了测量学的要求。

表 3 梅尔美术判断测验信度分析

信度类型	指数
Spearman – Brown 分半信度	0.701
Guttman 分半信度	0.700
Cronbach’s α 系数	0.774

3.3 测验信息量分析

从图 1 可见,从能力值 -3 到 -0.5 左右,测验总信息量较大,信息量最大处大约在能力值为 -2 左右;而从能力值 0.5 左右往右,测验信息量趋于下降,测量标准误逐渐增大。

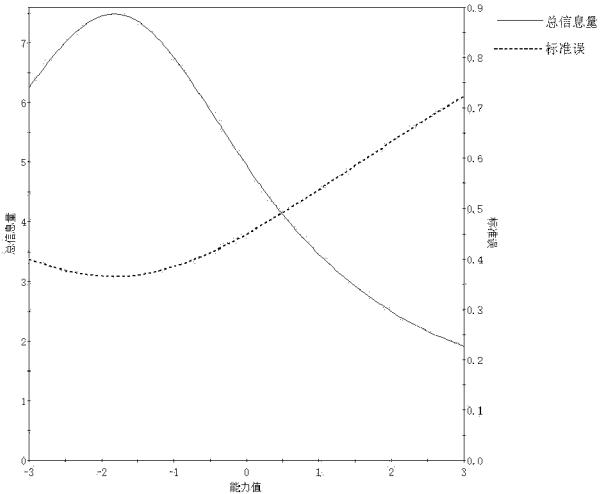


图 1 测验总信息量与标准误分布曲线图

3.4 效度分析

梅尔美术判断测验总分与各效标测验得分的相关见表 4。从中可见,除与自评音乐能力无关之外,与其他效标量表的得分均存在显著的正相关。另外,艺术类学生测验得分 ($\bar{X} = 57.12, S = 7.354$) 高于非艺术类学生测验得分 ($\bar{X} = 50.96, S = 7.925$), t 值为 11.589 且在 .01 水平上显著。由此可见该测

验得分可以反映被试审美水平的高低。

表 4 梅尔美术判断测验总分与各效标测验得分的相关

效标测验	相关系数
喜欢美术活动	0.298**
擅长美术活动	0.368**
喜欢职业(美术类)	0.270**
霍兰德美术总得分	0.387**
自评美术能力	0.248**
自评音乐能力	-0.060
过往经历	0.403**

4 讨论

4.1 测验的信、效度

通过经典测量理论的区分度,以及项目反应理论的模型资料拟合检验和区分度筛选,得到的 61 个项目的内部一致性信度系数 α 系数达到 0.774,说明测验具有较好的信度,符合心理测量学的要求。

效度验证是一个不断累积证据的过程,本研究设法从多个角度收集了证据。一是对与美术才能有关的个人过往经历进行了调查,这些调查是梅尔美术判断测验开发时作为效标的量表,结果表明与测验得分相关达到 0.403,相关显著。这与以往的研究结论“环境刺激会影响个体的美术才能”是一致的(Rodgers,1933)。同时,也表明美术判断能力通过一定的训练是有可能得到提高的,因为以往经历中就包括了是否接受过训练或学习、是否经常参观艺术展览等活动。在以被试自评美术能力水平、音乐能力水平为效标的分析中发现,与自评美术能力相关显著,但与音乐自评能力相关不显著,这符合人们的认识与逻辑。在以霍兰德职业兴趣测验中与音乐相关测题得分为效标的相关分析,也得出了不相关的结论,与个体自评的相关研究是一致的。以霍兰德职业兴趣测验中与美术相关测题得分为效标的分析发现是相关显著的,其中与是否擅长美术活动的相关最高,而与喜欢美术活动和喜欢美术类职业

相关要低一些。这也说明,对美术审美能力测验所测量的是判断还是偏好作出区分是有必要的。另外,从测验信息量的分析来看,整体上还有提高的空间,特别是对高能力水平的被试的测量需要增加难度更大的题目。

4.2 美术判断的维度

梅尔美术判断测验并没有对测验是否测量了多个维度进行分析,但其只提供一个测验分数,隐含了测验只测量了一个维度的假设。但是,在本研究进行因素分析时发现,其内部一致性信度系数为 0.774,KMO 值为 0.796,巴特利球形检验显著,这都表明可适合于进行探索性因素分析,但实际上 61 个题目得出了特征根大于 1 的因子达到 22 个。这一方面可能是 0、1 计分的题目不适合采用一般的线性因素分析,也可能是这些题目与因子之间的相关是一种非线性的相关(吴瑞林,涂冬波,2013),应采用基于 IRT 因子分析等有关方法作进一步研究。但是,也可能是因为美术判断并非单维的。在台湾学者陈荣华等(1981)研究的台湾师大式综合美术性向测验中,将美术特殊才能构成因素归结为审美的智力、敏锐的知觉、手指灵巧、美的判断力、丰富的想象力和创造力等。在梅尔美感测验(也叫做梅尔审美知觉测验,Meier Aesthetic Perception Test;Meier, 1967)提出美术才能取决于六个要素:手工艺技巧、坚定的意志、一般和美术智力、敏锐的知觉、创造性想象、审美判断。其中,审美判断被定义为识别组织内各种元素关系美学品质的能力。这个能力与很多因素有关,比如空间知觉,但单一的空间知觉并不等于审美判断能力,正如前文所说,审美判断包括描述、分析、解释和评价等多个基本过程(Smith, 1971)。因此,美术判断能力可能是多维的,这需要作进一步的探索。

4.3 值得进一步思考与研究的方向

尽管梅尔美术判断测验在同类测验中有效地避免了一些不足,但是其本身也存在一个问题:未能区别审美判断和审美判断原因。也就是说,当一个人做出审美判断得出某作品更好时,其背后的原因可能是不同的。例如,一个人可能会因为一幅画的形式和表现而认为它是好的,另一个人可能会因为曾经在他的教室挂了这幅画的复制品产生熟悉感而认为它是好的。显然,前者的判断才能说明个体具有较好的审美判断能力。因此,将来可考虑从审美判断原因的角度来测量个体的审美判断能力。这样做

有两个好处,一是如刚才所说,可以知道被试做出判断的原因,更利于推断其是否具有良好的审美判断能力。另一就是,在艺术领域专家之间并没有像在数学等其他领域那样容易达成一致,对于一幅特定的画,有些专家总是有可能与其他人意见相左,关于什么是好的艺术,我们有很多共识,但并不总是一致的;而专家对于判断的理由达成一致比就判断本身达成一致要容易得多(Bamosy, Johnston, & Parsons, 1985)。

另一个值得思考与研究的方向就是以中国人的作品开发出相应的审美判断测验。因为艺术中的文化因素是在人类社会实践形成的,和历史、地域、政治、道德、宗教等有密切的关系,因此带有民族的差异性。比如,中国艺术重主观,追求“意”,而西方艺术重客观,讲究“形”。这种文化差异,需要从更底层的数据资料做起,开发出更加适合于中国人的审美能力判断测验。

5 结论

修订的量表能测量个体的美术判断能力,但对高能力水平的个体测查不足,今后改进方向应该是增加更难的试题。

参考文献

- 陈荣华,邱维城,王秀雄,卢钦铭,范德鑫.(1981).我国国民中小学生对美术性向之研究.《师大学报》,26,279-303.
- 戴海琦(主编).(2010).《心理测量学》.北京:高等教育出版社.
- 丁秀峰(主编).(2001).《心理测量学》.开封:河南大学出版社.
- 顾海根.(2010).《应用心理测量学》.北京:北京大学出版社.
- 贾炜.(2009).试论美术与艺术的关系.《文学教育》,3(8),102-103.
- 李静.(2017).新时期中职美术教育与审美能力培养探讨.《艺术评鉴》,7(24),154-156.
- 邱维城.(1969).美术性向测验在我国中学之应用.《测验年刊》,16,16-19.
- 邱维城.(1970).美术性向测验在我国小学之应用与修订.《测验年刊》,17,61-77.
- 王芳,饶俊.(2013).高校艺术学院美术专业学生学习动机调查——以天津财经大学为例.《艺术教育》,18(2),28-29.
- 吴瑞林,涂冬波.(2013).题目因素分析:基于 SEM 和基于 IRT 的两类方法.《心理与行为研究》,11(1),124-131.
- 张旭东.(2019).《基于 21 世纪视觉艺术素养的中小学美术学习评价研究》(博士学位论文).华东师范大学.
- 郑日昌(主编).(2011).《心理与教育测量》.北京:人民教育出

- 版社.
- 郑日昌,孙大强(主编).(2012).*实用心理测验*.北京:开明出版社.
- Bamossy, G. , Johnston, M. , & Parsons, M. (1985). The assessment of aesthetic judgment ability. *Empirical Studies of the Arts*, 3(1), 63 – 79.
- Crannell, C. W. (1953). The validity of certain measures of art appreciation in relation to a drawing task. *Journal of Psychology*, 35(4), 131 – 142.
- Graves, M. (1939). What is your I. Q. in design? *Art Instructor*, 4(1), 11 – 14.
- Gordon, K. (1923). A study of aesthetic judgments. *Journal of Experimental Psychology*, 6(1), 36 – 43.
- Heyner, K. (1932). A method of correcting for guessing in True – False Tests and Empirical Evidence in Support of It. *Journal of Social Psychology*, 3(3), 359 – 362.
- Holland, J. L. (1997). Making vocational choices? *A theory of vocational personalities and work environments* (3rd ed.). Odessa: Psychological Assessment Resources.
- Lewis, R. A. , & Gray, G. M. (2005). *艾肯心理测量与评估* (张厚粲,赵守盈 译).北京:中国人民大学出版社.
- Meier, N. C. , & Seashore, C. E. (1929). *The Meier – Seashore Art Judgment Test*. Oxford, England: Bureau of Educational Research.
- Meier, N. C. (1942). *The Meier Art Tests I Art Judgment; Examiner's Manual*. Oxford, England: Bureau of Educational Research.
- Meier, N. C. (1967). *Meier Art Tests II. Aesthetic Perception*. Iowa city, Iowa: Bureau of Educational Research and Service.
- Reeve, B. B. , Hays, R. D. , Bjorner, J. B. , Cook, K. F. , Crane, P. K. , Teresi, J. A. , & Cella, D. (2007). Psychometric evaluation and calibration of health related quality of life item banks. *Medical Care*, 45, S22 – S31.
- Rodgers, F. (1933). Variation in the aesthetic environment of artistic and non – artistic children. *Psychological Monographs*, 45(1), 95 – 107.
- Smith, R. A. (1971). The philosophical literature of aesthetic education. In B. Reiman (Ed.), *Toward an aesthetic education* (pp. 46 – 73). Washington, D. C. : Music Educators Conference.
- Thorndike, E. L. (1926). Tests for aesthetic appreciation. *Journal of Education Psychology*, 7(10), 509 – 523.
- Welsh, G. S. , & Barron, F. (1949). *Welsh Figure Preference Test*. Palo Alto, Calif: Consulting Psychologists Press.

Validity and Reliability of Meier Art Judgment Test in China

Dong Shenghong^{1,2}, Li Xuan^{1,2}, Guo Lu^{1,3}, Li Xiaoyu^{1,2}

(1. School of Psychology, Jiangxi Normal University, Nanchang 330022;

2. Jiangxi Key Laboratory of Psychology and Cognitive Science, Nanchang 330022;

3. Jingdezhen Branch of China Mobile Jiangxi Co. , Ltd, Jingdezhen 333000)

Abstract: Aims: This study revised the Meier Art Judgment Test (MAJT) and tested its reliability and validity. Methods: Artistic type of Holland Code Career Test, Self – evaluation of students' artistic creation level, individual art experience subscale and the MAJT were administered among 2,270 students from 6 college and community college. Selecting items by Classical Test Theory (CTT) and Item Response Theory (IRT) approaches. Testing calibration correlation validity by calibration group method (art and non – art major). Results: All the remained 61 items were fitted with the 2 – Parameter Logistic model of IRT. The scale score was significantly correlated with each criterion scale. The scores between art and non – art majored students were significantly different. Nevertheless, the test information showed that the measurement error was relatively large for the high – ability subjects. Conclusion: The revised scale provides a secure base to measure the art judgment ability of individuals. The direction of future improvement should be to add more difficult items.

Key words: art talents; Meier Art Judgment Test; reliability; validity