

文章编号: 1000-5862(2019)05-0448-06

一种新的自尊测量技术: 计算机化自适应测量

郑泽宁, 蔡 艳*

(江西师范大学心理学院, 江西 南昌 330022)

摘要: 以项目反应理论(IRT)为基础, 开发了一种新的自尊测量技术——计算机自适应测验。通过单维性检验、局部独立性检验、区分度和项目功能差异检验进行试题筛选, 建立了由高质量的110题组成自尊题库。研究结果显示: 在CAT下, 被试平均只需作答不到10题就能达到以往110题的效果(2者的相关达0.94以上), 而且此时IRT的边际信度为0.91。即该自尊的CAT测量不仅有较高的测量精度, 还可以大大减轻被试的测试负担; 同时, 与传统的纸笔测验(P&P)不同, 该文对国内外多个不同的自尊量表进行整合, 从而避免了因量表不同而产生的测量偏差, 并真正实现了对具有不同自尊水平被试的因人施测(个性化测试), 从而达到自尊的高效、快速、准确的测量效果。因此, 该研究为自尊的测量提供了新的技术及方法支持。

关键词: 自尊; 计算机自适应测验; 项目反应理论

中图分类号: B 841.7 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2019.05.02

0 引言

计算机自适应测验(Computerized Adaptive Testing, CAT)是近年来发展起来的一种新的测量形式, 该测量以项目反应理论(Item Response Theory, IRT)为基础, 以计算机技术为手段, 在题库建设、选题策略等方面形成了一套理论和方法。与传统的纸笔测验相比, CAT可以实现因人而异地选题、施测。CAT依据每位被试当前的实际作答情况来调整出题策略、选取下一个合适的题目, 最终给予被试一个恰当的评价。其次, CAT能够精确地测量被试的能力, 每次挑选出匹配估计能力的信息量最大的题目, 与传统的纸笔测验相比, 可以用更少的题目达到更高的测量精度, 并且有效减少天花板和地板效应^[1]。CAT的测验时间和程序较为灵活、快速, 由于每人所测试题目不同, 所以并不需要固定时间。

目前, CAT在很多领域得到了广泛应用, 教育类考试如TOEFL、GRE等都是采用这种形式; 美国医生护士资格考试以及军事服役职业能力倾向成套测验也相继推出了CAT版本。近年来, CAT技术已逐渐应用于人格测验、临床评定量表等方面, 如MMPI-2、MMPI-A、MASS、EORTC QLQC、ICQ等成熟

量表已发展出相应的CAT测验。

自尊(self-esteem, SE)是自我概念中一个极为重要的成分, 它是个体有关自我的一系列评价。关于自尊的研究由来已久, 它也是心理学领域研究最多的课题之一^[2]。大量研究表明, 自尊作为最重要的保护因素之一, 与个体社交障碍、心理障碍的发展方面息息相关^[3-6]。目前对于自尊水平的测量主要采取自我报告的方法, 并且存在超过200种的自尊量表, 其中国外常用的有M. Rosenberg自尊量表(SES)^[7]、S. Coopersmith自尊问卷(CSEI)^[8]、I. L. Janis等^[9]缺陷感量表(FIS)、德克萨斯社会行为问卷(TSBI)^[10]、2维自尊量表(SLCS-R)^[11]以及身体自尊量表(BES)^[12]等; 而国内主要有黄希庭等^[13]编制的自我价值感量表、魏运华^[14]的儿童自尊量表等的自我描述问卷等。

不同学者采用不同量表对自尊水平进行测量, 可能会造成自尊的测量结果出现一定程度的偏差, 从而对实验结论产生影响; 其次, 这些量表都是基于经典测量理论(classical test theory, CTT)下固定测验长度、固定测验顺序的测验形式, 被试需要作答全部题目才能呈现结果, 耗费大量人力物力, 且不同量表之间无法等值, 不好进行比较。而将CAT技术运用于自尊的测量, 可对几个知名的自尊量表进行整

收稿日期: 2019-04-16

基金项目: 国家自然科学基金(31760288)资助项目。

通信作者: 蔡 艳(1979-) 江西宜春人, 教授, 博士, 博士生导师, 主要从事心理统计与测量研究。E-mail: cy1979123@aliyun.com

合,有效避免传统纸笔测验下产生的种种弊端,因此,本实验基于项目反映理论,为自尊的测量提供一种全新的高效、快速、准确且因人而异的测量技术。

1 研究过程

1.1 研究对象

在浙江与江西2省的2所中学发放问卷2 251份,回收有效问卷1 754份,有效率为77.9%;被试年龄为10~15岁,平均年龄为(12.98±0.89)岁;其中来自城镇的被试有842人,来自农村的被试有912人;男性被试837人,女性被试917人;初一年级748人,初二年级664人,初三年级342人。

1.2 测量工具

通过大量文献资料显示,Rosenberg自尊量表、Coopersmith自尊问卷、Janis和Field缺陷感量表和2维自尊量表等4个量表为国内外最为常用的测量自尊的量表,并被经常作为效标、研究手段进行使用,具有较高的测量信度和效度^[15-19]。自我价值感量表是我国本土化的自尊量表,是评估青少年自我价值感使用最广泛的工具之一^[20]。鉴于此,该文主要采用以上5个国内外相对较知名的自尊量表,并根据这些量表来构建自尊计算机化自适应测验(computerized adaptive testing for self-esteem, CAT-SE)的题库。

以上5个量表共152题,不存在重复项目,所有被试均被要求在规定时间内完成作答。5个量表具体介绍如下。

1.2.1 Rosenberg自尊量表(SES) 采用M. Rosenberg^[7]编制,王孟成等^[21]修订的中文版自尊量表。该量表由10个项目组成,采用4级记分,由1(很不符合)~4(非常符合),总分在10~40之间,得分越高,自尊感越强。本研究中该量表的内部一致性系数 α 为0.865。

1.2.2 Coopersmith自尊问卷(CSEI) 该量表由50个项目组成,其中8题为测谎题,采用2级记分,由1(像我)~2(不像我),总分在50~100之间,得分越高,自尊感越低。在本研究中,为了与其他量表统一记分方法,对该量表得分进行了反转,即得分越高,自尊感越高。量表的内部一致性系数 α 为0.898。

1.2.3 Janis和Field缺陷感量表(FIS) 该量表由36个项目组成,采用7级记分,由0(从来没有)~6(总是如此),总分在0~216之间,得分越高,自尊感越强。本研究中该量表的内部一致性系数 α 为0.947。

1.2.4 自我价值感量表(ASES) 采用黄希庭等^[13]编

制的《青年学生自我价值感量表》,去除其中道德价值感与生理价值感16题,剩余40题,采用5级记分,由1(非常不符合)~5(非常符合),总分在40~250之间,得分越高,自尊感越强。本研究中该量表的内部一致性系数 α 为0.945。

1.2.5 2维自尊量表(SLCS-R) 采用R. W. Tafarodi等^[11]编制,张进辅等^[22]修订的中文版量表。该量表共由16个项目组成,采用5级记分,由1(非常不符合)~5(非常符合),总分在16~80之间,得分越高,自尊感越强。本研究中该量表的内部一致性系数 α 为0.915。

2 研究方法

2.1 CAT-SE的题库建设

本研究在现代测量理论之项目反应理论(item response theory, IRT)框架下,开发自尊计算机化自适应测验(CAT-SE)的题库。IRT采用数据数学建模的思路,描述被试的潜在特质与项目特征是如何影响被试的得分概率。而这些数学模型是建立在数据单维性、局部独立性等假设之下,因此对于数据是否符合这些假设需要进一步检验,以保证IRT分析的科学性和可靠性。

2.1.1 单维性检验 虽然单维性并不是做CAT的必要条件,但大多数IRT模型的假设为单维。这表示在项目反映中单个潜变量解释了所有的项目^[23]。研究采用探索性因素分析(EFA)考察量表的单维性^[24],采用里斯卡标准进行判别,即第1特征根的解释方差超过20%,且第1特征根与第2特征根的比值大于5,则认为模型符合单维性假设^[25]。此外,为了保证题库项目的质量,删除因子负荷小于0.4的项目^[26]。

2.1.2 模型比较与选择 研究采用资料模型拟合检验的方法,从适用于多级评分的拓广分部评分模型(GPCM)、称名反应模型(RSM)和等级反应模型(GRM)3个IRT模型中选择最恰当的模型进行参数估计和数据分析。资料模型拟合检验采用相对拟合指标AIC、BIC、-2Log Likelihood进行评估,拟合指数越小,表明模型拟合越好。

2.1.3 项目局部独立性检验 局部独立性是IRT模型的重要假设,它表示被试在每个项目上的反应都是独立的,不受其他项目的影响。本研究采用S. L. Franzoi等^[12]提出的 Q_3 统计量作为指标进行局部独立性检验。根据Cohen's提出的规则,当 Q_3 值介于0.26~0.36时,表明偏差合理,处于合理范

围;而当 Q_3 值超过 0.36 时,表示项目间存在较大的偏差,此时应在相互比较的 2 个项目中删除 Q_3 累加量较大的一个项目。

2.1.4 项目区分度分析 具有良好区分度的项目质量高,能将不同水平的被试区分开来,对于被试能力的测量也更准确。因此,在 IRT 中,项目区分度是衡量题目质量非常重要的指标。为了保证 CAT-SE 题库项目的高质量,删除区分度小于 0.7 的项目^[27]。

2.1.5 项目功能差异(DIF) 项目功能差异检验用于探究是否出现由于群体不同而导致的系统误差^[28]。本文使用 McFadden's pseudo R^2 对性别、城乡 2 个变量进行检验,当 R^2 改变量大于 0.02 时,表明此项目存在 DIF,需要删除^[29]。

2.2 CAT-SE 算法、效果验证

CAT 算法包括初始能力值、选题策略、能力估计方法以及终止规则等方面。本文将被试初始能力值定为 0,使用最大信息量作为选题策略,使用期望后验法(Expected A Posteriori, EAP)作为能力估计的方法。根据公式 $\rho = 1 - S_E^2$,当信度选 0.80、0.90、0.95 和 0.96 时,所对应的 S_E 分别为 0.45、0.32、0.22 和 0.20。将对这 5 个终止规则下的答题情况进行比较,选出最佳的终止规则。

对 CAT 过程的模拟,分别采用真实被试和虚拟被试 2 种方式进行模拟研究。其中真实被试是指使用本次参与问卷作答的被试的真实数据进行 CAT 的过程模拟,称为“真实数据模拟”^[30-34],将用此模拟选出最佳终止规则以及探究 CAT 作答的效果;关于虚拟被试,使用选定的终止规则在能力区间为 $-3.5 \sim 3.5$ 之间每隔 0.25 模拟 100 个被试,共模拟出 $29 \times 100 = 2900$ 个被试的作答情况,再进行 CAT 效果的检验,详细过程参照 Wang Shudong^[35] 的模拟过程。

2.3 研究工具

研究采用 SPSS 24.0、R 3.31 软件包 2 个软件进行数据分析和处理。

3 研究结果

3.1 题库建设

3.1.1 单维性检验 通过 SPSS 24.0 进行探索性因素分析(EFA),共删除主因子上因子负荷小于 0.4 的项目 15 题,剩下 137 题。对余下的题目再次进行探索性因素分析,结果显示第 1 特征根的值为 38.51,方差解释贡献率比例为 28.11% > 20%;第 2

特征根的值为 6.43,第 1 与第 2 特征根值的比值为 $5.98 > 5$ 。根据里斯卡标准^[25],可知 137 题的作答数据符合单维性的假设,即这 137 题主要测量了一个自尊维度。

3.1.2 模型比较与选择 由表 1 可知,在 3 个多级评分的 IRT 模型中,等级反应模型(GRM)在 -2Log Likelihood 、AIC、BIC 3 个相对拟合指标上的数值都是最低的,即拟合效果最佳。因此,在后续 IRT 分析研究中均采用 GRM 模型进行项目的参数估计和 CAT 模拟与评估。

表 1 不同 IRT 模型与数据拟合检验结果

IRT 模型	-2Log Likelihood	AIC	BIC
拓广分部评分模型(GPCM)	543 693.4	544 951.4	548 391.8
称名反应模型(RSM)	541 150.8	543 118.7	548 500.9
等级反应模型(GRM)	538 676.4	539 934.4	543 374.9

3.1.3 项目质量分析 项目质量分析主要从项目是否满足局部独立性假设、项目区分度高低以及项目是否存在功能差异 3 个方面进行。

局部独立性是 IRT 理论的前提,通过局部独立性检验及 Q_3 指标值可知:137 题中有 17 题不符合局部独立性的假设,给予删题,剩余 120 题。

项目区分度反映了项目对被试的区分能力,区分度越高说明区分能力越强。根据参数估计结果,结合区分度筛选指标发现:120 题中有 10 个项目的区分度低于 0.7,删除后剩余 110 题,且项目区分度在 0.81 ~ 3.68 之间。

为保证测验的公平性,分别对被试关于性别、城乡 2 个变量进行了项目功能差异检验,检验发现所有项目的 R^2 改变量均小于 0.02,即 110 个项目均不存在功能差异,故没有删除题目。

综上,本小节共删除 27 题,最终题库由剩余的 110 题组成,题目筛选的具体情况详见表 2。

表 2 项目质量分析过程中题目筛选的具体情况

指标	题目	题数
单维性	$b_4, b_8, b_9, b_{23}, b_{25}, b_{26}, b_{28}, b_{37}, b_{46},$	15
	$b_{50}, b_{55}, b_{58}, c_{26}, c_{31}, e_{15}$	
局部独立	$a_2, a_{10}, b_5, b_{19}, b_{24}, c_{10}, c_{15}, c_{18},$	17
	$c_{21}, c_{23}, c_{37}, c_{33}, c_{34}, d_3, e_{11}, e_{20}, e_{31}$	
区分度	$b_3, b_{10}, b_{16}, b_{21}, b_{32}, b_{45}, b_{49}, b_{52},$	10
	c_{29}, e_{13}	
项目功能差异	无	0

注: $a \sim e$ 分别代表 Rosenberg 自尊量表、Coopersmith 自尊问卷、Janis 和 Field 缺陷感量表、2 维自尊量表和自我价值感量表。

3.2 CAT 研究结果

3.2.1 基于真实数据的 CAT 结果 表 3 为不同终止规则下基于真实数据的 CAT 结果,其中平均答题量是指所有被试平均作答的题目数量,反映了 CAT 的测试效率,平均题量越少效率越高;边际信度是指各误差水平下估计出潜在特质的平均信度,反映了测试的可靠程度,信度值越高测试成绩越可靠;相关系数 r 反映了 CAT 能力估计结果与被试作答 110 题

(题库所有项目)的相关程度,相关系数越大 CAT 的测试精度越高.分析比较不同终止规则下的相关指标,发现在误差 S_E 为 0.32 时,所需的题目较少,与误差为 0.45 情况下几乎一致,且测量出的被试能力与完成整个题库所测出的能力呈现高度相关($r = 0.94, P < 0.001$),当相关系数 $r \geq 0.9$ 时,模型被认为是良好的^[23,32].因此,本文终止规则定为 $S_E \leq 0.32$.

表 3 在真实被试不同终止规则下基于真实数据的 CAT 结果

终止规则(S_E)	答题数量		平均误差 $S_E(\theta)$	边际信度	相关系数 r
	均值	标准差			
完成整个题库	110.00	0.00	0.17	0.97	1.00
$S_E \leq 0.45$	8.89	9.23	0.37	0.86	0.91
$S_E \leq 0.32$	8.88	9.92	0.30	0.91	0.94
$S_E \leq 0.22$	18.27	16.25	0.22	0.95	0.97
$S_E \leq 0.20$	23.44	18.40	0.20	0.96	0.98

对真实被试的数据进行模拟 CAT 过程,结果显示能力值在 2 个标准差内的被试,平均只需作答 7.40 题($S_D = 2.99$)就能满足 $S_E \leq 0.32$,即测量的边际信度 $r \geq 0.9$ 的标准.超过 2 个标准差的被试共 86 名,他们平均需作答(37.59 ± 31.28)题就能达到该指标.

为了进一步考察 CAT 估计的被试能力值与其它自尊量表间的相关一致性程度,绘制了相关散点图(见图 1)并计算了其相关系数.在终止规则 $S_E \leq 0.32$ 条件下测得的被试能力值与被试做完所有题

目所测得的能力值呈高度相关($r = 0.94, P < 0.001$);在与测量自尊最为广泛使用的 Rosenberg 自尊量表的比较中,也发现 2 者呈显著正相关($r = 0.76, P < 0.001$);与在国内广泛使用的自我价值感量表中的总体自我价值维度(共 6 题)进行比较,被试需要作答的题量相似,2 者也呈高度正相关($r = 0.85, P < 0.001$);在与总题量相似的 2 维自尊量表(共 16 题)进行比较发现,2 者存在高度相关($r = 0.88, P < 0.001$).

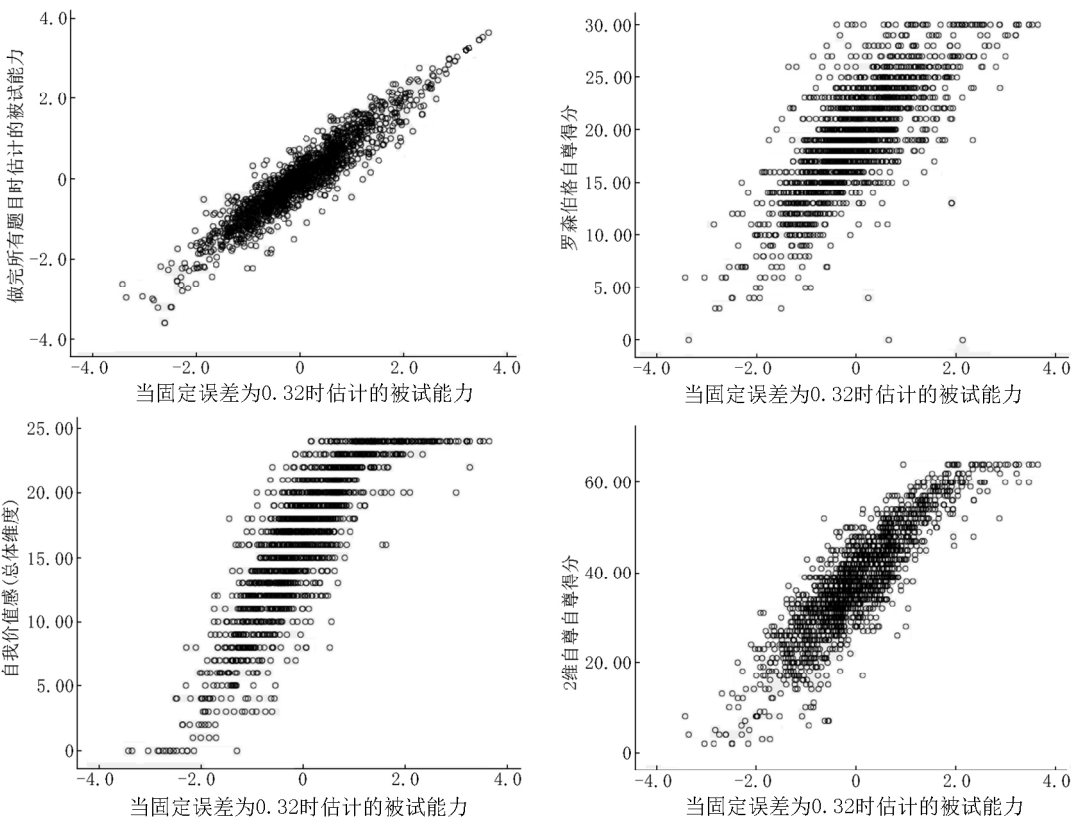


图 1 不同终止条件下被试能力估计值与其它自尊量表分数的相关散点图

3.2.2 基于虚拟被试的 CAT 结果 图 2 显示了在 $S_E = 0.32$ 的终止规则下,能力值为 $-3.5 \sim 3.5$ 的被试平均需要做题的数量.对于潜在能力值在 $-2.0 \sim 2.0$ 范围内的被试,平均需作答 8.70 题($S_D = 4.31$)就能满足 $S_E \leq 0.32$,即测量的边际信度 $r \geq 0.9$ 的标准;对于能力水平超过 2 个标准差的被试而言,平均作答题量会变多($M = 34.20$, $S_D = 30.37$);对于所有被试而言,平均需做 19.25 题($S_D = 23.5$).

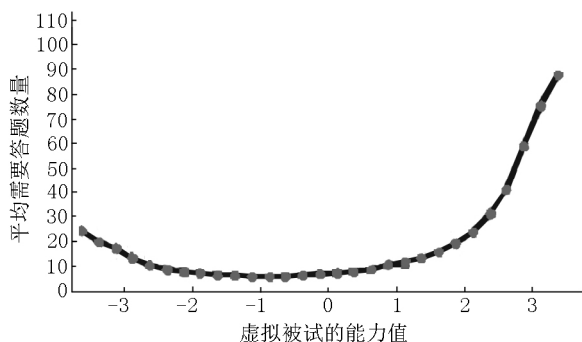


图 2 不同能力的虚拟被试平均所需答题量

4 讨论

自尊对个体的发展具有重要作用,大量研究表明:自尊水平与抑郁、焦虑、人际关系、情绪等都息息相关^[34-37],用于测量自尊水平的量表较多,且大多是基于经典测量理论下的纸笔测验,受测者需完成全部问卷才能估计其能力,费时费力,且使用不同的量表其测量结果可能存在偏差;而本文依据项目反应理论建立的 CAT-SE 题库,对国内外知名的自尊量表进行了整合,通过“因人施测”的手段,在保证精确度的同时,大大提高了测试的效率与灵活性,节省了测试所需的人力、物力和时间.同时, CAT-SE 的实现,使得来自不同自尊量表的题目建立了统一的量尺,评价标准统一,避免了不同量表间评估结果的不可比性.

本研究最终选取的终止规则为 $S_E \leq 0.32$,在此精度下,基于真实被试的 CAT 研究中,能力值在 2 个标准差内的被试平均只需作答 7.4 题就与做完全部 110 题所估计出的能力值几乎一致;同时,所有被试平均只需作答 8.88 题就能达到做 110 题的效果(2 者的相关高达 0.94 以上),且此时 IRT 下的边际信度也高达 0.91.即自尊的 CAT 测量不仅具有较高的测量精度,还可以大大减轻被试的测试负担;同时,与传统的纸笔测验(P&P)不同,本研究对国内外多个不同的自尊量表进行整合,从而避免了因量表

不同而产生的测量偏差,并真正实现了对具有不同自尊水平被试的因人施测(个性化测试),从而达到自尊的高效、快速、准确地测量.

当 $S_E \leq 0.32$ 时, CAT 的估计精度和信度将更高,但平均使用的题量将增加,效率下降.因此,在不同情况下,使用者可以根据精度和效率要求,合理设置 S_E 的取值,更好地服务于实际需求.另外,图 2 结果表明:低到中等能力的被试使用的题量相对较少,显示了题库对于低自尊水平的被试具有更加精确的估计,这与目前自尊相关研究的诊断目标相一致,原始的自尊量表的开发,本身也有筛选出低自尊水平的个体并对其进行干预的目的;而近年来,越来越多的研究开始发现过高的自尊也会对个体产生不利影响^[38-42],本题库对于高能力即高自尊水平的被试而言所需题量更多一些,这可能与题库中缺少测量高能力的题目、选用的量表本身就是为了更好地区分出低自尊水平的被试有关,后续需要扩建题库.

综上所述,本研究将计算机自适应技术应用于自尊的题库建设中是合适的,它实现了对自尊的高效、快速、准确地测量,为今后开发自尊量表的 CAT 版式提供了理论基础,同时,本研究也为自尊的测量提供了新的技术及方法支持.

5 参考文献

- [1] Embretson S E, Reise S P. Item response theory [M]. London: Psychology Press 2013.
- [2] Tesser A. On the confluence of self-esteem maintenance mechanisms [J]. Personality and Social Psychology Review 2000 4(4): 290-299.
- [3] Lee A, Hankin B L. Insecure attachment, dysfunctional attitudes and low self-esteem predicting prospective symptoms of depression and anxiety during adolescence [J]. Journal of Clinical Child and Adolescent Psychology, 2009 38(2): 219-221.
- [4] Martyn-Nemeth P, Penckofer S M, Velsor-Friedrich B, et al. The relationships among self-esteem, stress, coping, eating behavior and depressive mood in adolescents [J]. Research in Nursing and Health 2009 32(1): 96-109.
- [5] 肖崇好, 黄希庭. 社交焦虑个体外显与内隐自尊的研究 [J]. 心理科学 2011 34(2): 289-292.
- [6] 钟佑洁, 张进辅. 大学生评价恐惧在自尊与社交焦虑间的中介效应分析 [J]. 心理发展与教育 2011(5): 506-512.
- [7] Rosenberg M. Society and the adolescent image [EB/

- OL]. [2019-01-12]. https://www.researchgate.net/publication/238322868_Society_and_the_Adolescent_Image_Maker.
- [8] Coopersmith S. The antecedents of self-esteem [EB/OL]. [2019-01-16]. https://www.researchgate.net/publication/200008764_The_Antecedent_of_Self-Esteem.
- [9] Zheng Lijun, Richard A L, Zheng Yong. Sex and sexual orientation differences in personality in China [J]. Archives of Sexual Behavior 2011 40(3): 533-541.
- [10] Helmreich R, Stapp J. Short forms of the texas social behavior inventory(TSBI): an objective measure of self-esteem [J]. Bulletin of the Psychonomic Society, 1974, 4(5): 473-475.
- [11] Tafarodi R W, Jr W B S. Two-dimensional self-esteem: theory and measurement [J]. Personality and Individual Differences 2001 31(5): 653-673.
- [12] Franzoi S L, Shields S A. The body esteem scale: multidimensional structure and sex differences in a college population [J]. J Pers Assess, 1984 48(2): 173-178.
- [13] 黄希庭. 青年学生自我价值感量表的编制 [J]. 心理科学, 1998 21(4): 289-292.
- [14] 魏运华. 自尊的结构模型及儿童自尊量表的编制 [J]. 心理发展与教育, 1997(3): 31-38.
- [15] Watkins D, Dong Qi. Assessing the self-esteem of Chinese school children [J]. Educational Psychology, 1994, 14(1): 129-137.
- [16] 蔡华俭. 内隐自尊效应及内隐自尊与外显自尊的关系 [J]. 心理学报 2003 35(6): 796-801.
- [17] 田录梅. Rosenberg(1965) 自尊量表中文版的美中不足 [J]. 心理学探新 2006 26(2): 88-91.
- [18] 杨福义, 梁宁建. 内隐自尊与外显自尊的关系: 多重内隐测量的视角 [J]. 心理科学 2007 30(4): 785-790.
- [19] 张丽华, 李娜. 自尊研究范式的发展 [J]. 苏州大学学报: 教育科学版 2015(4): 33-41.
- [20] 毕重增, 肖影影, 许欢欢. 国内青少年自我价值感量表研究结果的元分析 [J]. 心理科学 2014 37(3): 625-632.
- [21] 戴晓阳, 张进辅, 程灶火. 常用心理评估量表手册 [M]. 北京: 人民军医出版社 2010.
- [22] 韦嘉, 张春雨, 赵清清, 等. 2 维自尊量表修订版在中学生群体中的信效度检验 [J]. 中国心理卫生杂志, 2012 26(9): 715-720.
- [23] Wainer H. Computerized adaptive testing: a primer: L. Erlbaum Associates [EB/OL]. [2019-01-12]. <http://dx.doi.org/10.1037/10244-000>.
- [24] 田建全, 苗丹民, 杨业兵, 等. 应征公民计算机自适应化拼图测验的编制 [J]. 心理学报 2009 41(2): 167-174.
- [25] Hambleton R K. Principles and selected applications of item response theory [M]. 台北: 心理出版社, 1989.
- [26] Nunally J C. Psychometric theory [M]. 2nd ed. New York: McGraw-Hill, 1978.
- [27] Fliege H, Becker J, Walter O B, et al. Development of a computer-adaptive test for depression(D-CAT) [J]. Quality of Life Research 2005 14(10): 2277-2279.
- [28] Holland P W, Wainer H. Differential item functioning [J]. International Encyclopedia of Education, 1995, 7(11): 36-44.
- [29] Choi S W, Grady M W, Dodd B G. A new stopping rule for computerized adaptive testing [J]. Educational and Psychological Measurement 2011 71(1): 37-53.
- [30] Oommen M, Pajer K A, Kelleher K J, et al. Computerized adaptive measurement of depression: a simulation study [J]. BMC Psychiatry 2004 4(1): 13-15.
- [31] Sands W A, Waters B K, McBride J R. (1997). Computerized adaptive testing: from inquiry to operation [EB/OL]. [2019-01-12]. http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_082118.pdf.
- [32] Nunnally J C, Bernstein I. Psychometric theory [M]. 3rd ed. New York: McGraw-Hill, 1994.
- [33] Demeyer I, Romero N, Raedt R D. Assessment of implicit self-esteem in older adults: the role of actual and ideal self-esteem in negative mood [J]. Assessment, 2018, 25(3): 302-309.
- [34] Orth U, Robins R W. Understanding the link between low self-esteem and depression [J]. Current Directions in Psychological Science 2013 22(6): 455-460.
- [35] Wang Shudong. The accuracy of ability estimation methods for computerized adaptive testing using the generalized partial credit model [EB/OL]. [2019-01-12]. <https://dl.acm.org/citation.cfm?id=930298>.
- [36] 高爽, 张向葵, 徐晓林. 大学生自尊与心理健康的元分析: 以中国大学生为样本 [J]. 心理科学进展 2015 23(9): 1499-1507.
- [37] 李海江, 杨娟, 贾磊, 等. 不同自尊水平者的注意偏向 [J]. 心理学报 2011 43(8): 907-916.
- [38] 倪凤琨. 自尊与攻击行为的关系述评 [J]. 心理科学进展 2005 13(1): 66-71.
- [39] 潘益中, 许燕. 脆弱高自尊在自我威胁后的归因与情绪转换 [J]. 心理科学 2011 34(1): 166-171.
- [40] 田录梅, 袁竞驰, 李永梅. 同伴在场和自尊水平对青少年冒险行为的影响: 来自 ERPs 的证据 [J]. 心理学报, 2018 50(1): 47-57.
- [41] 田录梅, 张向葵. 高自尊的异质性研究述评 [J]. 心理科学进展 2006 14(5): 704-709.
- [42] 王曼, 陶嵘, 胡姝婧, 等. 新的视角: 从脆弱高自尊看人格障碍症状 [J]. 心理科学进展 2009 17(7): 1141-1146.

The Investigation and Bryoflora Analysis in Miyagi Relics of Shangjing Longquanfu Balhae ,China

CONG Mingyang^{1,2}, TANG Luyan³, LI Jinjiang², XU Yueyue⁴, JIAN Minfei^{1*}

(1. Jiangxi Provincial Key Lab of Protection and Utilization of Subtropical Plant Resources ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China; 2. Analytical & Testing Center ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China; 3. College of Life Science ,Guizhou Normal University ,Guiyang Guizhou 550025 ,China; 4. School of Economics and Management ,Nanchang University ,Nanchang Jiangxi 330031 ,China)

Abstract: The miyagi relics of Shangjing Longquanfu in Balhae have great scientific value ,because it is the centre exhibited primitive style of prosperous Tang Dynasty. Bryophytes distribute widely ,and often attach on surface of the relics. Based on field investigation ,specimen collection and taxonomic identification of bryophytes in miyagi relics , the species composition and floristic characteristics are analyzed. Results show that in the field of historical relics protection ,the species richness of bryophytes in miyagi relics is higher relatively. A total of 35 species are recorded , including 1 species in Hepatiaceae and 34 species in Musci (17 genera of 7 families) . Dominant families are Brachytheciaceae ,Pottiaceae and Bryaceae ,with proportion 68.57% in total. Dominant genus are *Brachythecium* B. S. G. ,*Entodon* Müll. Hall. ,*Bryum* Hedw. ,*Eurhynchium* B. S. G. ,*Brachymenium* Schwaegr. and *Schwetschkea* Müll. Hall. ,altogether with proportion 65.71% . All of the species belong to 7 geographical components at the species level ,in which East Asian elements are predominant (36.36%) . North temperate elements and Endemic to China types account for a considerable proportion (24.24% each) . Compared with other 5 sites ,species similarity coefficient of miyagi relics is closet to that of Wudalianchi new-period volcano. The results of this study will provide a scientific basis for the protection of miyagi relics of Shangjing Longquanfu.

Key words: bryophyte; flora; geographical types; similarity coefficient

(责任编辑: 刘显亮)

(上接第 453 页)

The Development of a Computer-Adaptive Test for Self-Esteem

ZHENG Zening ,CAI Yan*

(College of Psychology ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China)

Abstract: Based on item response theory(IRT) ,the study develops a new self-esteem measurement by computerized adaptive testing(called as CAT-SE) . After a series of analyses of one-dimensional test ,local independence test ,discrimination test and differential item function ,the CAT-SE item bank composed of 110 high-quality questions is finally established. Results show that the proposed CAT-SE can achieve the similar precision of 110 items only by answering less than 10 questions on average(the correlation between them is as high as 0.94) ,and the marginal reliability under IRT is as high as 0.91. That is to say ,the CAT-SE not only has a high measurement accuracy ,but also can greatly reduce the test burden of the subjects. At the same time ,unlike the traditional paper-pencil test(P&P) , the study integrates several different widely-used self-esteem scales ,thus it can avoid the measurement deviations caused by the different scales. All in all ,the proposed CAT-SE in this study provides a new technical and methodological support.

Key words: self-esteem; computer adaptive testing; item response theory

(责任编辑: 冉小晓)