

文章编号:1000-5862(2020)02-0142-06

基于计算机自适应的大学生社会责任感测验

戴步云¹, 谢宇欣¹, 简小珠²

(1. 江西师范大学心理学院, 江西省心理与认知科学重点实验室, 江西 南昌 330022; 2. 井冈山大学教育学院, 江西 吉安 343009)

摘要:构建基于计算机自适应的大学生社会责任感测验的题库,并基于真实数据模拟 CAT 过程,检测 CAT 的测验效率并探讨用 CAT 测量社会责任感的可行性. 使用探索性因素分析对测验进行单维性检验,根据项目反应理论选择高质量的项目,构建基于计算机自适应的大学生社会责任感测验(CAT-SRS). 然后,基于纸笔测验的实测数据进行 CAT 模拟,验证 CAT-SRS 的有效性. 研究表明;CAT-SRS 符合项目反应理论等级反应模型的假设,各个参数比较理想,测量结果与 5 个原量表纸笔测验结果的相关性较高. CAT-SRS 可在测验长度较短时实现对大学生社会责任感的准确测量.

关键词:项目反应理论;计算机自适应测验;社会责任感;大学生

中图分类号:B 841 **文献标志码:**A **DOI:**10. 16357/j. cnki. issn1000-5862. 2020. 02. 05

0 引言

社会责任感(social responsibility)是指个体积极承担社会责任或积极帮助他人的比较稳定的心理品质^[1]. 如中国古语有云“达则兼济天下”“天下兴亡匹夫有责”,这些都昭示着个体对社会的责任感. 国家已经提出将“着力提高学生服务国家人民的社会责任感”作为未来教育的战略重点^[2],可见国家认为社会责任感是一种十分重要的人格特质. 因此十分有必要对大学生开展社会责任感的教育,而开发高效化的测量工具能够节省测验时间,并有助于高效准确地估计出学生的社会责任感水平.

国内外不少学者对于社会责任感的测量展开了研究. 国内外学者已经编制了许多社会责任感量表^[3-7]. 但是,这些测验都是基于经典测量理论(classical test theory, CTT)而编制的传统纸笔测验. CTT 对被试特质水平的估计依赖于具体测验项目,对项目参数的估计依赖于具体被试. 为了分数的可比性,它们只能是“千人一面”的,这使得测验对每个被试有冗余的题目,这使得测验效率较低,需要耗费较多时间.

为了解决 CTT 的这些不足,学者们提出了项目反应理论(item response theory, IRT). IRT 模型描述

被试的潜在特质与其作答之间的关系^[8],通过被试在具有一定难度和区分度的项目上的作答就可以确定被试的潜在特质(Latent Trait). 与 CTT 相比,IRT 具有以下优势:(i)对被试的特质水平估计不依赖于具体测验项目;(ii)对测验的项目参数估计不依赖于具体被试样本;(iii)对测量误差的估计更精确等. 这使得 IRT 近年来被广泛应用^[9]. 计算机自适应测验(Computerized adaptive testing, CAT)是基于计算机技术与 IRT 发展而来的一种新的心理与教育测验方式,它允许每个被试所做题目不同,而估计的特质水平最终仍能相互比较,从而能使得针对每个被试不同的特质水平进行测验,达到用较少的项目对被试的特质水平(θ)做出有效的测量. 已经有学者将基于传统纸笔形式的焦虑测验改造成了 CAT 形式并获得了成功^[10-11].

本研究尝试用 IRT 方法对来自 5 个常用的社会责任感问卷的 147 个项目进行分析,检验其是否满足 IRT 的单维性,并选择高质量项目构建一个基于计算机自适应的大学生社会责任感测验(Computerized adaptive testing-Social Responsible Scale, CAT-SRS)的题库,并基于真实数据模拟 CAT 过程,检测 CAT 的测验效率,并将 CAT 模拟测验所得分数与 5 个原量表纸笔测验分数的一致性,进一步探讨用 CAT 测量社会责任感的可行性.

收稿日期:2019-05-16

基金项目:国家自然科学基金(31660278),江西师范大学青年英才培育资助计划,江西省教育厅科技项目青年课题(GJJ170230)和横向课题“科学学科监测数据分析”资助项目.

作者简介:戴步云(1984-),男,上海人,讲师,博士,主要从事计量心理学研究. E-mail:biweijianpsy@qq.com

1 对象与方法

1.1 测验工具

本文将组合使用国内外测量社会责任感的 5 个量表(或分量表),其中 2 个国外量表(学生责任感量表和大五人格量表中的尽责性分量表)已经在国内得到了实证研究,具有良好的信效度。

1.1.1 青少年责任心问卷 采用程岭红^[3]编制的青少年学生责任心问卷,共 54 个条目,采用 5 级计分,得分越高说明青少年社会责任感越强。本研究中测量的 α 系数为 0.923。

1.1.2 大学生社会责任心问卷 采用赵兴奎^[4]编制的大学生社会责任心问卷,共 36 个条目,采用 5 级计分,得分越高说明大学生社会责任感越强。本研究中该问卷的 α 系数为 0.824。

1.1.3 中学生社会责任心问卷 采用李雪^[5]编制的中学生社会责任心问卷,共 35 个条目,其中包含了 9 个因素成分:评价性、自控性、效能性、敏感性、灵活性、反映性、主动性、独立性和坚持性。采用 5 级计分,得分越高说明中学生社会责任感越强。本研究中该问卷的 α 系数为 0.920。

1.1.4 学生责任感量表(Student Personal Responsibility Scale-10) 采用 S. Singg 等^[6]编制的学生责任感量表进行评定,共 10 个条目,采用 4 级计分,得分越高说明大学生社会责任感越强。本研究中测量的 α 系数为 0.750。

1.1.5 大五人格问卷(The Big Five Inventory,BFI)尽责性分量表 采用 O. P. John 等^[7]编制、张建新修订的大五人格问卷中的“尽责性”分量表测量社会责任感。该分量表共 12 个条目,采用 5 级计分,得分越高说明公民社会责任感越高。该量表是目前国内外应用最普遍的量表,是测量责任感的常用工具,本研究中尽责性分量表的 α 系数为 0.786。

1.2 被试

从南昌市四所高校中随机选择 1 000 名大学生被试,每个参与测验的被试均需要完成 5 个社会责任感量表(或分量表)共计 147 题的组合形式的纸笔测验(其中还包括 3 组测谎题)。在删除不完整及不合格的测验作答(连续作答、比例作答超过 10 题,测谎题绝对等级差之和大于 5)后,共得到有效答卷 653 份。其中包括女生 405 人,男生 248 人,平均年龄为 19.54 岁($S_D=1.28$)。

1.3 分析工具

使用的数据分析处理工具包括 R 语言和 SPSS

23.0 软件。

2 结果

首先让被试对包含这 147 题的社会责任感组合测验进行纸笔测验施测,然后根据纸笔测验的分析结果选择高质量的项目构建 CAT-SRS 的题库。

2.1 共同方法偏差与各原测验的内部一致性信度

为了探查是否存在共同方法偏差,采用 Harman 单因素分析^[12]对纸笔测验的实测数据进行检验。结果表明,在未旋转时共生成 35 个因子,它们共解释了 64.52% 的方差,其中第 1 个因子解释了 20.42% 的方差,远低于 D. A. Harrison 等^[13]推荐的 50% 的判断标准,这意味着共同方法偏差不存在。

5 个原测验在这次施测中得到的 α 系数分别为 0.923、0.824、0.920、0.750 和 0.786,这表明 5 个原测验的内部一致性信度均可接受。

2.2 CAT-SRS 题库建构

对纸笔测验的结果使用单维 IRT 模型进行数据分析,并构建题库,具体过程如下。

2.2.1 单维性检验 为了保证最终题库里所有的题目测到的是相同的心理特质,故要对题目进行单维性检验,并剔除不满足需要的题目。

使用 SPSS 23.0 统计软件包对所有题目进行探索性因素分析。将因子数量固定为 1 个,删除因子负荷小于 0.3 的项目^[14],直至符合单维性,最终剩余 121 题。根据 D. Andrich^[15]提出的单维性检验标准,第 1 特征根与第 2 特征根之比超过 3,第 1 特征根所解释的方差大于 20%,即可认为满足单维性假设。在本研究第 1 个因子解释的总方差为 24.67%,第 1 特征根为 29.86,第 2 特征根为 9.06,2 者之比为 3.29,故可认为这个题库是单维的。

2.2.2 模型比较及选择 使用 AIC、BIC 和 -2LL 这 3 个模型拟合度指标来比较 2 个常见的多级评分 IRT 模型:一个是等级反应模型(Graded Response Model,GRM)^[16],另一个是拓广分部评分模型(Generalized Partial Credit Model,GPCM)。其中 AIC、BIC 和 -2LL 的值越小,则说明模型拟合效果越佳。

模型拟合度指标如表 1 所示。由表 1 可以看出,GRM 的 AIC、BIC 和 -2LL 值均低于 GPCM 模型,这表明 GRM 模型更能够拟合数据,故选择 GRM 模型来进行分析。

表1 2个IRT模型拟合度指标

模型	拟合度指标		
	-2Log-Likelihood	AIC	BIC
GRM	144 529.1	145 547.1	147 830.6
GPCM	147 210.7	148 228.7	150 512.2

2.2.2.3 局部独立性检验 进行局部独立性检验,根据 W. M. Yen^[17] 提出的 Q_3 统计量,其原理是计算由原始分数减去预测分数所形成的项目残差之间的相关.当 Q_3 值介于 0.26 ~ 0.36 时,偏差处于合理状态,当 Q_3 值超过 0.36 时,项目间存在较大的偏差,此时应删除两两比较的 2 个项目中的 Q_3 累加量较大的一个项目^[18].在本文中,题库剩余的 121 题中有 18 题的 Q_3 值大于 0.36,所以删除了 18 题,剩余 103 题.

2.2.2.4 区分度 题库项目质量越高,对被试特质水平的测量越准确,P. K. Hambleton 等^[19] 提出:为了对项目参数及被试的特质水平参数进行有效估计,项目区分度一般要求位于 0.5 ~ 2.0 之间.本文以区分度 0.8 作为删题标准.分析结果发现在题库剩余的 103 题中有 12 题的区分度小于 0.8,所以删除这 12 题,题库剩余 91 题.

2.2.2.5 项目拟合检验 采用 $S-\chi^2$ 指标来做项目拟合检验.当项目的 $S-\chi^2$ 值小于临界值 0.01 时,项目拟合较差,应删除.题库剩余的 91 题中有 11 题的 $S-\chi^2$ 值小于 0.01,所以删除这 11 题,题库剩余 80 题.

经过以上步骤对测验结果的检验分析与项目筛选,CAT-SRS 的题库构建完成.对最终构建后的题库进行单维性检验,发现题库里的 80 题覆盖了原有的中学生社会责任心问卷 9 个维度里的 6 个,以及另外 4 个测验的全部维度.由此可以认为,题库中的 80 题所测量的是社会责任感,而不是社会责任感的一部分.

2.2.2.6 测验质量分析 模拟 1 000 个特质水平居于 -3 ~ 3 之间均匀分布的被试,估算整个题库对被试的信息量,得出的结果如图 1 所示.

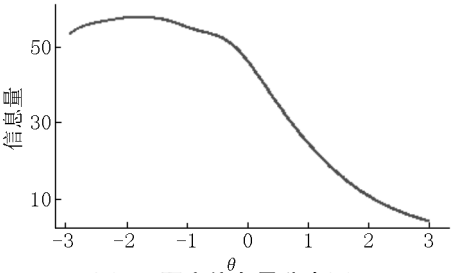


图1 题库信息量分布图

从图 1 的信息曲线可以看出,当特质水平为 -1.5 时,CAT-SRS 题库总信息量最大值为 55.题库的信息量比较充足,特质水平处于 -3 ~ 2 的被试能

够获得的信息量均达到了 10 以上;而在估计特质水平处于 -1.5 左右的被试时,题库提供的信息量是最高的.

用等级反应模型对整个题库的项目参数进行估计,发现题库中的 68 道题的区分度参数均值为 1.46 ($S_D = 0.51$),各个项目的难度都位于 -3 ~ 2 之间,这表明整个题库具有较高的质量.

2.3 CAT-SRS 的实现及效率与效度验证

2.3.1 基于真实作答数据在 CAT-SRS 中模拟测试 在建立了题库的基础之上,基于纸笔测验中实测的被试作答数据,将估计出的被试特质水平 θ_{80} 作为被试特质水平真值进行 CAT 模拟,具体的模拟过程如下:

(i) 将每个被试在选择试题时特质水平初始值都设定为 0.

(ii) 选取一道试题给被试作答.若是初始作答试题,则采用随机选题方式;若不是初始作答试题,则用最大 Fisher 信息 (Maximum Information, MI) 法,根据估计出的被试特质水平,从剩余题库里选择对当前被试能够提供最大 Fisher 信息量的项目.

(iii) 根据被试的作答情况,用贝叶斯期望后验概率估计法 (Expected a posteriori, EAP) 来估计被试的社会责任感特质水平.

$$\hat{\theta} = \sum_{h=1}^q Z_h L_i(Z_h) W(Z_h) / (\sum_{h=1}^q L_i(Z_h) W(Z_h)), \quad (1)$$
其中 Z_h 是节点特质水平值, $L_i(Z_h)$ 是节点 Z_h 对应项目 i 的作答的似然函数, q 是节点数, $W(Z_h)$ 是相应节点的权重,它代表概率密度曲线在相应节点处的高度.它将先验分成许多积分点而不是将其作为连续分布来进行评估.

(iv) 终止策略.尝试 3 种不同的终止策略:规则 1 为固定测验精度、规则 2 为固定测验长度、规则 3 是“精度为主,定长为辅”.当被试作答达到终止标准时,即终止作答;若未达到终止标准,则回到第 (ii) 步.

2.3.2 CAT-SRS 的测验效率分析 在 CAT-SRS 模拟之后,可获得被试在 CAT-SRS 测验中估计的特质水平 $\theta_{\text{CAT-SRS}}$,将 $\theta_{\text{CAT-SRS}}$ 与基于之前被试在纸笔测试中的作答估计出的结果 θ_{80} ,对 CAT-SRS 的测验效率进行分析.

本文通过以下指标对 CAT-SRS 的效率进行评价:每个项目的测量标准误 S_e 的平均值,题量节省比例,纸笔测验估计结果 θ_{80} 与 CAT-SRS 估计结果之间的偏差 (B_{bias})、平均绝对离差 (A_{BSE})、近似误差的平方根 (R_{MSEA}) 以及相关系数 (r).

用能力估计标准差 $S_E(\theta)$ 作为衡量选题策略是否稳定的指标;采用引入了潜在特质估计值到计算中的边际信度(Marginal Reliability) 的平均值作为测验成绩可靠性的指标^[20].

这几个指标的计算公式如下:

$$M_R = 1 - 1/I_i(\theta), \tag{2}$$

题量节省比例 = (题库大小 - 测验平均题数)/
题库大小 × 100%, \tag{3}

$$A_{BSE} = \sum_{j=1}^N | \theta_{80,j} - \theta_{CAT-SRS} | / N, \tag{4}$$

$$B_{ias} = \sum_{j=1}^N (\theta_{80,j} - \theta_{CAT-SRS}) / N, \tag{5}$$

其中 θ 表示被试 j 的特质水平估计值; $I_i(\theta)$ 表示被试 j 在项目 i 上获得的 Fisher 信息量; $\theta_{80,j}$ 表示纸笔测验估计的特质水平; $\theta_{CAT-SRS}$ 表示 CAT-SRS 估计的结果.
在 3 种终止规则下的测试结果如表 2 所示.

表 2 被试在 3 种终止规则下的测试结果

N = 656

终止规则	终止条件	平均用 题数	边际信度 M_R	相关系数 r	A_{BSE}	R_{MSEA}	B_{ias}
固定测验精度 (规则 1)	$S_E(\theta) < 0.2$	30.03	0.96	0.99	0.11	0.16	-0.00
	$S_E(\theta) < 0.3$	16.43	0.91	0.97	0.20	0.25	-0.02
	$S_E(\theta) < 0.4$	9.13	0.86	0.95	0.26	0.33	-0.03
	$S_E(\theta) < 0.5$	5.84	0.79	0.92	0.30	0.39	-0.02
固定测验长度(规则 2)	10 题		0.87	0.95	0.24	0.31	-0.02
	15 题		0.91	0.96	0.21	0.27	-0.03
	20 题		0.93	0.97	0.18	0.24	-0.02
	25 题		0.94	0.98	0.16	0.21	-0.03
	30 题		0.95	0.98	0.14	0.19	-0.02
“精度为主,定长为 辅”(规则 3)	$S_E(\theta) < 0.3$ 或测 验项目数量为 20	12.21	0.90	0.96	0.22	0.20	-0.03

研究表明:在规则 1 中,在设置 4 种不同精度的终止条件下边际信度之间存在明显差异.尤其是在设置终止规则为 $S_E(\theta) < 0.5$ 时,边际信度低于 0.8,这说明此结果与纸笔测验的被试特质真值之间可靠性不够高,与此同时,测验项目最少的只有 2 个,最多达到了 80 个,并且个体之间也存在显著差异($S_D=6.12$).当 $S_E(\theta) < 0.2$ 时,纸笔测验的被试特质真值与 CAT-SRS 估计结果的相关性是最高的($r=0.99$).值得一提的是,在 4 种终止条件下均存在被试作答完所有项目的情况.进一步分析发现,在这 4 种终止条件下,被试特质水平 (θ) 的最终估计值大于 2.4 的被试,在模拟测验过程中做完了题库中所有试题.这种情况的出现是因为这些被试属于特质水平偏高的被试,在进行模拟测验过程时,模拟测验对他们提供的信息量一直偏少,测验精度一直达不到终止标准,于是被试只有将题库做完才能终止测验.

采用规则 2,当测验长度分别固定为 10、15、20、25、30 题时,CAT-SRS 估计结果与纸笔测验结果的相关系数都达到了 0.94 以上,属于高相关.这说明只要测验长度达到了 10 题以上,CAT-SRS 估计结果与纸笔测验结果就高度一致了.

对于测验终止规则 3,其设置依据是:测验边际信度和平均测量标准误差反映得都是测量的信度和

可靠性,当 $S_E(\theta) < 0.3$ 时,边际信度大于 0.9,这说明此时测验的可靠性高,而在固定项目长度为 20 时,标准误也低于 0.3,所以,可以设置当 $S_E(\theta) < 0.3$ 或者答题长度达到 20 时终止测验.这样既可以达到精度的要求,又可以节约项目,由此减少测验时间并降低曝光率,进而节约成本.图 2 是在规则 3 测验模拟时,被试特质估计值 (θ) 与使用题量和信息量的关系.

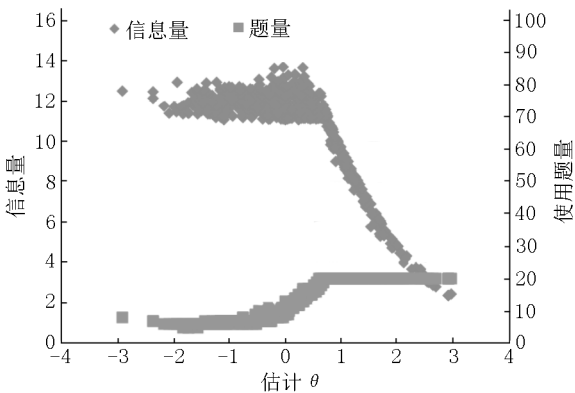


图 2 在 $S_E(\theta) < 0.3$ 或测验题数为 20 下估计 (θ) 与使用题量和信息量的关系

由图 2 可知,在规则 3 下,当特质水平处于 $-3.0 \sim 0.5$ 时,被试的答题量都低于 20 题,且当特质水平处于 $0 \sim 0.5$ 时,被试的答题量成持续上升趋势,当特质水平位于 $0.5 \sim 3.0$ 时,被试的答题量都是 20 题.另外,题量节省比例越大越好,数值越大表明被

试需要完成的项目越少. 本研究中该比例已经达到 84.74%, 这说明被试在题库中的答题量偏少. ABSE 反映地是 CAT-SRS 估计的结果与纸笔测验的被试特质真值之间的差异大小,其数值是越小越好,数值越小表示 CAT 估计的结果与纸笔测验的被试特质真值之间越接近^[21]. 在表 2 的规则 3 下, A_{BSE} 的值小于 0.3, 这说明 CAT-SRS 有较好的测验模拟返真性 (Recovery); 纸笔测验的被试特质真值与 CAT-SRS 的偏差 (B_{ias}) 接近于 0, 且 R_{MSEA} 小于 0.5, 这说明 CAT-SRS 的估计结果与被试特质真值之间的误差很小; 2 者之间的相关系数达到 0.90 ($P < 0.01$), 进一步证实了 CAT-SRS 估计结果的有效性和可靠性.

表 3 规则 3 下, CAT-SRS 模拟测验结果与效标之间的相关关系

效标关联度	青少年责任心问卷	中学生社会责任心问卷	大学生社会责任心问卷	学生责任感问卷	大五人格“尽责性”分量表
r	0.80**	0.79**	0.79**	0.64**	0.54**

注: ** 表示在 0.01 的显著性水平下显著.

由表 3 可知: 在规则 3 下 CAT-SRS 模拟测验的被试特质估计值与效标之间存在着显著的正相关, 且相关系数都大于 0.5, 其为大效应量. 这表明 CAT-SRS 的效标关联度较高.

2.3.4 对 CAT-SRS 的进一步讨论 从 CAT-SRS 模拟的结果来看, CAT-SRS 能够达到既有准确性又有较高的测验效率, 特别是对于社会责任感特质水平位于偏低水平区间的被试来说. 在对 CAT-SRS 的有效性进行验证时, 考虑到极端人格人群需要作答所有题目方能终止测验的情况, 本研究中尝试了“精度为主, 定长为辅”的终止规则, 并发现它为最佳终止规则, 结果表明: CAT-SRS 可以从题库中选择能提供大部分信息的试题来测试, CAT-SRS 的平均作答项目数减少到了 12.21, 比 5 个原有量表中的 80 要少很多, 这说明 CAT-SRS 可以大大减少被试的作答负担, 提高被试的作答效率. 模拟结果还表明, CAT-SRS 估计的结果与原有 5 个量表的纸笔测验总分均达到了较高的相关性, 这说明该测验的与 5 个原有量表测量结果具有较高的一致性.

在研究过程中, 为了保证良好的项目拟合, 本研究将重点放在了单维性检验上, 删除的项目较多 (删除了 67 题), 未来可以考虑在 Bifactor 模型的基础上进行 CAT 化分析. 同时题库数量只有 80 题, 部分项目的曝光度可能会过高; 以及 CAT 测验信度曲线显示在估计特质水平偏低的被试时的估计结果会更加稳定和可靠, 这与研究目的——研究者想要实现的是在估计特质水平居于中间的被试时的精度和信度更高——有所出入, 这都需要在未来研究中进

综合来说, 在人格测验当中, 由于极端人格的存在, 在设置测验终止规则时, 可以采用“以精度为主, 同时兼顾定长”的做法.

总之, 在 CAT-SRS 估计结果与纸笔测验的被试特质真值之间比较分析, 发现模拟返真性能良好, 测验误差较小, 所需题量相对较少, 这说明 CAT-SRS 的测验效率比较理想.

2.3.3 CAT-SRS 的测试效度验证 分别以纸笔测验下 5 个量表的总分作为效标, 以终止规则 3 下的 CAT-SRS 模拟结果代表 CAT-SRS 模拟测验结果, 计算 CAT-SRS 模拟估计结果与效标之间的相关系数, 即估算 CAT-SRS 的效标关联度 (见表 3).

一步增加项目、扩充题库.

3 讨论

本文根据项目反应理论, 构建了基于计算机自适应测验方式的社会责任感测验 (CAT-SRS). 根据纸笔测验结果显示: CAT-SRS 符合项目反应理论等级反应模型的假设, 符合单维性假设, 测验参数比较理想.

基于纸笔测验实测数据进行 CAT-SRS 模拟, 结果显示: 该测验的长度比多数同类纸笔测验更短, 即测验的效率较高; 而且, 测验模拟返真性能较好, 测验误差小. 其中以“精度为主, 定长为辅”的终止规则为最佳终止策略: 在 CAT 模拟过程中, 当满足 $S_e < 0.3$ 或测验长度达到 20 题时即终止测验. 而且, CAT-SRS 估计的结果与纸笔测验的总分之间存在着较高的相关性, 这表明 CAT-SRS 测验具有较高的测量效度.

由测验模拟结果可知, CAT-SRS 可以凸显社会责任感测量在采取 CAT 测验方式中的优势: 在相近的测验准确度下, CAT-SRS 能够大大缩短测验所需的长度, 节省施测时间, 对实现高效准确地大学生社会责任感测量具有实测参考价值. 将本研究应用于实践中, 可以更有效和准确地测量社会责任感, 这使得该测验可以适用于广泛的大学生人群, 这为今后应用社会责任感的 CAT 版式提供了理论基础.

如果读者对本文的题库感兴趣, 可向本文笔者索取.

4 参考文献

[1] 黄四林,韩明跃,张梅. 人际关系对社会责任感的影响[J]. 心理学报,2016,48(5):578-587.

[2] 黄四林,韩明跃,宁彩芳,等. 大学生学校认同对责任感的影响:自尊的中介作用[J]. 心理学报,2016,48(6):684-692.

[3] 程岭红. 青少年学生责任心问卷的初步编制[D]. 重庆:西南师范大学,2002.

[4] 赵兴奎. 大学生社会责任心结构及发展特点[D]. 重庆:西南大学,2007.

[5] 李雪. 中学生社会责任心结构及其发展特点研究[D]. 重庆:西南师范大学,2004.

[6] Singg S, Ader J A. Development of the student personal responsibility scale-10 [J]. Social Behavior and Personality, 2001, 29(4):331-335.

[7] John O P, Donahue E M, Kentle R L. The big-five inventory [J]. Journal of Personality and Social Psychology, 1991, 18(5):367-385.

[8] Lord F M. Applications of item response theory to practical testing problems [M]. Hillsdale, NJ: Lawrence Erlbaum, 1980.

[9] 刘慧,简小珠,张敏强,等. 多水平IRT的发展与应用述评[J]. 心理科学进展,2012,20(4):627-632.

[10] Weiss D J. Improving measurement quality and efficiency with adaptive testing [J]. Applied Psychological Measurement, 1982, 6(4):473-492.

[11] Gibbons R D, Weiss D J, Pilkonis P A, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety [J]. The American Journal of Psychiatry, 2014, 171(2):187-194.

[12] 熊红星,张璟,叶宝娟,等. 共同方法变异的影响及其统计控制途径的模型分析[J]. 心理科学进展,2012,20(5):757-769.

[13] Harrison D A, Mclaughlin M E, Coalter T M. Context, cognition, and common method variance: psychometric and verbal protocol evidence [J]. Organizational Behavior and Human Decision Processes, 1996, 68(3):246-261.

[14] Reeve B B, Hays R D, Bjorner J B, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS) [J]. Medical Care, 2007, 45(5):22-31.

[15] Andrich D. A hyperbolic cosine latent trait model for unfolding polytomous responses: reconciling thurstone and likert methodologies [J]. British Journal of Mathematical and Statistical Psychology, 1996, 49(2):347-365.

[16] Samejima F. Estimation of latent ability using a response pattern of graded scores [J]. Psychometrika, 1969, 34(1):1-97.

[17] Yen Wendy M. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model [J]. Applied Psychological Measurement, 1984, 8(2):125-145.

[18] Cohen J. Statistical power analysis [J]. Current Directions in Psychological Science, 1992, 1(3):98-101.

[19] Hambleton R K, Swaminathan H. Item response theory principles and applications [M]. Leiden: Kluwer-Nijhoff Pub, 1985.

[20] Kim S. A note on the reliability coefficients for item response model-based ability estimates [J]. Psychometrika, 2012, 77(1):153-162.

[21] 刘馨婷. 基于Bifactor-model的抑郁症计算机自适应测验开发[D]. 南昌:江西师范大学,2017.

The College Students' Social Responsibility Testing Based on Computerized Adaptation

DAI Buyun¹, XIE Yuxin¹, JIAN Xiaozhu²

(1. School of Psychology, Jiangxi Key Laboratory of Psychology and Cognitive Science, Jiangxi Normal University, Nanchang Jiangxi 330022, China; 2. School of Education, Jinggangshan University, Ji'an Jiangxi 343009, China)

Abstract: Using exploratory factor analysis, a single-dimensional test is carried out in this paper. The high-quality project is selected according to the item response theory(IRT), and the college students' social responsibility test (CAT-SRS) is built based on computerized adaptation. Then, based on the measured data of the paper and pencil test, the cat simulation is carried out to verify the effectiveness of CAT-SRS. The results show that CAT-SRS meets the assumption of the graded response model of IRT, and parameters of items are ideal, and the correlations between the results of CAT-SRS and the five original paper and pencil tests are relatively high. CAT-SRS can exactly measure social responsibility of college students when the test length is relatively short.

Key words: item response theory; computerized adaptive test; social responsibility; college student

(责任编辑:冉小晓)