

# 德育评价的新思路——基于具身德育的迫选式量表

董圣鸿<sup>1</sup>, 周李文渊<sup>2</sup>

(1. 江西师范大学 心理学院 江西 南昌 330022; 2. 豫章师范学院 学前教育学院 江西 南昌 330103)

**摘要:**在具身德育理论逐渐应用于德育工作的当下,根据普通德育理论编制的量表已渐渐不能满足实际德育工作的要求。为了解决德育测评中遇到的理论落后实际、被试作答含有社会称许性导致结果不够准确、经典测量理论固有的统计性能缺陷等问题,本研究从 Likert 式初中生具身德育量表的编制入手,自编了迫选式初中生具身德育量表,并在无压力与压力两种情境下施测。结果发现,迫选量表拥有良好的信效度,其实测数据能够很好地拟合 Thurstone IRT 模型,解决了自模式数据的问题。在与 Likert 式量表的估计结果比较可以看出,迫选量表通过 Thurstone IRT 模型拟合后估计出的被试特质能够很好地消除社会称许性。

**关键词:**迫选;瑟斯顿 IRT 模型;具身德育;量表编制

中图分类号:G410 文献标志码:A 文章编号:1000-579(2023)06-0058-07

## A New Idea of Moral Education Evaluation—Forced Selection Scale Based on Embodiment Moral Education

DONG Shenghong<sup>1</sup>, ZHOULI Wenyuan<sup>2</sup>

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

2. School of Preschool Education, Yuzhang Normal University, Nanchang, Jiangxi 330103, China)

**Abstract:** At present, embodiment moral education theory is gradually applied to moral education work, the scale compiled according to ordinary moral education theory can not meet the requirements of practical moral education work. In order to solve the problems encountered in moral education assessment, such as the theory lagging behind practice, the results being inaccurate due to the social approval of the subjects, the inherent statistical defects of the classical measurement theory, etc. Starting with the Likert scale of embodiment moral education for junior middle school students, this study designed the forced selection scale of embodiment moral education for junior middle school students and tested it under two situations: no pressure and pressure. The results show that the forced selection scale has good reliability and validity, and its measured data can fit Thurstone IRT model well, which solves the problem of self-pattern data. Compared with the estimation results of the Likert-style scale, it can be seen that the subject traits estimated by the forced selection scale after fitting the Thurstone IRT model can well eliminate social approval.

**Key words:** forced selection; Thurston IRT model; embodiment moral education; scale compilation

收稿日期:2023-05-17

基金项目:江西省社会科学“十二五”规划项目(15WTZD12);江西省教育厅研究生创新基金项目(编号:YC2020-B084)

作者简介:董圣鸿(1971—),男,江西九江人,江西师范大学心理学院教授、博士生导师。研究方向为教育与心理统计测量、认知心理学、教育心理学。

周李文渊(1991—),男,江西奉新人,心理学博士,豫章师范学院学前教育学院助教。研究方向为心理统计与测量。

## 一、引言

纵观我国教育史,德育都占有举足轻重的地位。2019年,国务院发布《中共中央国务院关于深化教育教学改革全面提高义务教育质量的意见》,明确强调了“坚持‘五育’并举,全面发展素质教育”,并将“坚持立德树人,着力培养担当民族复兴大任的时代新人”放在“五育”中的首要位置,足见德育工作的重要性。而在德育工作中,德育评价是检测德育工作完成质量的重要环节,科学的评价方式能够更加清晰、真实地反映出德育工作合格与否。如何更好地对德育效果作出准确的评价,一直是我国德育工作者长期研究的话题。目前,我国德育评价方法主要存在两大问题,其一是以落后、陈旧的“离身德育”为理论基础;其二是德育评价中含有社会称许性等测量问题,而目前的测验工具依然以Likert式量表为主。

### (一) 德育评价中的“离身”与“具身”

我国早期德育工作经历了长期的探索,虽然取得了一定的成效,但也存在一些问题。这些问题不仅是在我国的德育工作中出现,也是整个世界德育遇到的难题。皮亚杰认为道德推理不仅独立于感知和身体运动,而且不受情感的影响<sup>[1]</sup>。整个教育界在早期都受到皮亚杰德育观深刻的影响,各级学校都以“离身德育”的形式为主。2017年,孟万金指出了离身德育的五大缺陷,包括:脱离学生自身生活实际;脱离学生身心发展规律;脱离学生身体与环境的交互作用;脱离学生身体力行;脱离学生主观能动性。这种德育导致的后果是,学生知行无法合一,具体表现为言行不一、表里不一、华而不实、装腔作势、死记硬背、填鸭式灌输、简单粗暴、机械教条,总结为三个字就是“假”“大”“空”<sup>[2]</sup>。

许多学者也提出了相似的看法,刘慧指出脱离个体的道德教育是不完善的,那是为了德育而德育<sup>[3]</sup>。郭志芹也指出德育不应离开身体的维度进行空洞的说教,必须在身心和谐中,真切地关注人的道德水平的提高<sup>[4]</sup>。我国德育工作的困境,说到底还是德育方法的落后。据此,中国教育科学研究院的孟万金经过多年研究,将德育与认知心理学中的“具身”概念相结合,提出了“具身德育”的教育体系。

“具身”是认知心理学中的一个概念,具身哲学的奠基人法国学者Merleau-Ponty(1908—1961)将其解释为身体嵌入世界之中,它是知觉、身体和世界的统一<sup>[5]</sup>。具身与离身相对,强调情境、文化和身体在认知当中的作用。在具身认知视角下,人的抽象思维由隐喻而得来,因此人总是通过具体事物来理解和思考抽象事物;而人们最熟悉的就是自己的身体,所以有大量的抽象概念是来自我们身体的<sup>[6]</sup>。而道德正是一种十分抽象的概念,因此人们在关于道德的判断和决策中,往往会受到自身身体体验的影响,这已经得到了许多研究的证实,不仅如此,还有一些实验也已经证明,人的身体感觉会影响人的情绪、认知等精神活动<sup>[7]</sup>。这些都为具身德育奠定了理论基础。

“具身德育”在“具身”的基础上应运而生。其教育体系主张德之根在心,人之本在劳,二者合起来即立德树人的根本,在身心一体的教育中,实现价值内化。这与布鲁姆情感领域的教育目标不谋而合。布鲁姆提出,价值内化经历了接受、反应、价值化、组织、价值体系个性化形成这五个过程。Krothwohl等人也将情感目标细分为不同层级,并且用常用的情感用语:兴趣、鉴赏、态度、价值、适应对品格形成的各个阶段进行标注。由此可见,想要内化形成道德的价值观,决不能忽略道德情感所发挥的作用,对学生进行离身德育无法培养出知行合一的道德之士,“化人成德,其必由具身德育”<sup>[8]</sup>。

“具身德育”在提出后,很快就得到了教育界的认同。《中国教育报》文章指出,新时代的德育应向具身化转向,应当要关心人的“情绪体验、认知冲突和身心投入”,这释放出具身德育必将成为未来德育方向的信号<sup>[9]</sup>。在科研方面,具身德育已经成为研究热点<sup>[10-11]</sup>,并且在实践中开始应用<sup>[12]</sup>。然而,已有的测评工具依然以陈旧的“离身德育”为基础开发,并没有开发出新的工具以适应具身德育这一新的德育方式,不能很好地适应当前的具身德育环境<sup>[13-14]</sup>。第一,从量表内容结构上来说,以前的测评工具主要适应离身德育,因此包含大量书本上浮于表面文字的品德测量项目,对于学生真正的内化品德的测量,这样的项目是无效的。第二,以往的测评工具往往只测知和行,很少测量情和意,而道德情感和道德意志都是具身德育理论所强调的道德的重要组成部分。为解决以上问题,开发以“具身德育”为理论基础的德育测评工具是十分必要的。

### (二) 德育评价中的社会称许性问题

分析现存的德育测评工具可以看出,目前的德育量表依然立论于经典测量理论,采用Likert式量表

设计,而这样开发的测评工具也存在许多问题<sup>[15-16]</sup>。第一,Likert 式量表的测验形式容易产生各种反应偏差,即社会称许性问题,如默许反应(acquiescence response)、光环效应(halo effect)、印象管理(impression management)等<sup>[17]</sup>。第二,还存在经典测量理论容易导致的一些问题,例如观察分数等权重累加的不合理性等<sup>[18]</sup>。

为了解决 Likert 量表中存在的社会称许性等问题,有研究者开发了迫选式量表。迫选式量表不同于 Likert 式量表,它不要求被试对每一个陈述评分,而是在配对的多个陈述中选择最符合自己和最不符合自己的一项。这种项目已经被证明可以有效避免反应定势和光环效应<sup>[17]</sup>。不仅如此,配对的陈述是具有相同社会称许性的,这又解决了一般量表的社会称许性问题。

然而,迫选式量表的数据称为自模式数据(ipsative data),这种数据的特点在于,不论被试作出何种反应,其总分都是定值。这一特点导致迫选式量表使用普通的线性模型或 IRT 模型进行分析会带来一系列的问题,如分数解释困难、测验信度估计不准确等。为此,Brown 和 Maydeu-Olivares 于 2011 年专门开发了 Thurstone IRT 模型解决以上问题。目前有许多研究都已证明,使用 Thurstone IRT 模型能很好地拟合迫选式量表数据<sup>[19]</sup>。因此,本文将以具身德育理论为基础,在 Thurstone IRT 框架下开发迫选式的具身德育量表。

小学、初中是我国义务教育德育工作的关键阶段,了解这一时期的学生的道德发展情况能够帮助教育部门更有效地开展德育工作。综上,本研究拟选取初中阶段,在前人德育量表研究的基础上,解决现存测验的关键问题,编制出测验性能良好并且适应于新型德育工作——具身德育环境下的测验。

## 二、对象和方法

### (一) 样本

开放式调查与访谈样本:抽取江西省吉安市 6 所初中,每所学校访谈校领导 1 名、班主任 1 名、德育老师 1 名。此外,每所学校访谈初一至初三每个年级学生各 6 名,根据班主任提供的信息保证抽取的学生中成绩好、中、差比例大致相同。

预测样本:抽取江西省宜春市 3 所初中,发放问卷 502 份,全部有效回收;使用 R 进行异常作答侦测后,保留了 486 份数据,问卷有效率 96.81%。为了提高研究结果的外部效度,本研究对人口学变量进行了控制,有效问卷中的人口学变量情况如下:男生 252 人,女生 234 人;初一 182 人,初二 168 人,初三 136 人;城镇初中 233 人,农村初中 253 人。

Likert 式量表正式施测样本:抽取江西省宜春市 3 所初中,发放问卷 473 份,全部有效回收;使用 R 进行异常作答侦测后,保留了 457 份数据,问卷有效率 96.62%。其人口学变量情况如下:男生 226 人,女生 231 人;初一 189 人,初二 142 人,初三 126 人;城镇初中 232 人,农村初中 225 人。

迫选式量表正式施测样本:抽取江西省宜春市 2 所初中,发放问卷 300 份,全部有效回收;通过 R 进行异常作答侦测后,保留了 292 份数据,问卷有效率 97.33%。其人口学变量情况如下:男生 151 人,女生 141 人;初一 97 人,初二 107 人,初三 88 人;城镇初中 207 人,农村初中 85 人。

以上各研究问卷有效率均在 95% 以上,达到了问卷分析有效率的要求。

### (二) 量表编制

#### 1. 测量维度的确定

本研究首先进行了测量维度的确定<sup>[20]</sup>,主要参考中华人民共和国教育部提出的《中小学德育工作指南》以及孟万金教授对具身德育的详细阐述,此外还参考了国内较为成熟的其他道德相关量表<sup>[21-25]</sup>。在 8 位心理学博士、10 位心理学硕士形成的专家组的集体探讨下,确定了从具身德育的定义出发,以《中小学德育工作指南》为指导而产生的 6 个维度为基本理论构想,并同时设计了包含 23 个方面内容的半开放性问题的访谈提纲。经讨论研究,6 个维度确定为:理想信念、爱党爱国、中华优秀传统文化、团结友善、心理健康、生态文明。

#### 2. 项目收集

以两种方式收集具身德育量表的原始题项:(1)文献分析。通过检索国内相关文献,选取已有德育量表成熟问卷中的典型项目,对一些项目进行修改,使之适合初中学生。(2)深度访谈。以文献分析为

基础,设计了包含23个方面内容的半开放性问题的访谈提纲,对吉安市6所中学的30名初中生进行了线下的半开放式访谈,并在征得访谈对象同意的前提下对访谈内容进行了录音,便于进行后续编码和分析。考虑到初中生的专注力,每名学生的访谈时长控制在20分钟左右。访谈内容包括教育部对中小学生学习德育要求的内容、具身德育对学生德育工作的新要求等,涵盖知、行、意、情四方面,问题描述形如:“长大后你想成为什么样的人?为什么想成为那样的人?你知道的大人中(可以是生活中你认识的、书上学习到或电视、电影里等知道的人),有没有类似于你长大后想成为的人?如果有的话,请详细描述一下。”同时,对6名校长、6名班主任、6名德育教师共18位德育一线工作者进行深度访谈,主要内容为目前初中阶段德育工作的实际情况。

### 3. 项目整理与内容归纳

首先,专家组基于文献分析、深度访谈,经讨论筛选出含义清晰、符合初中生德育要求、契合具身德育要求、符合初中德育工作实际情况的初试项目共307条。接着,由16位心理学专家(其中博士6人,硕士10人)将项目归类至各维度。考虑到初中学生注意力集中水平,为了将测验时长控制在15到20分钟,要求专家同时筛选出更具有代表性的项目。经过5轮筛选,最终保留了50个项目。最后,邀请了10名心理测量方向专家(其中博士3人,硕士7人)对形成的问卷陈述进行评价,并给出了修改建议。并邀请了另外5名专家(其中博士2人,硕士3人)完成了维度匹配问卷,归类一致性结果在90%—100%之间,表明项目与维度匹配合理且有效;同时在讨论后,对80%的专家认为不匹配的1个项目进行了替换。此时已初步形成试测问卷。

为了考察试测问卷的清晰性、可读性及适宜性,邀请2位初一学生及1位初中教师进行试测及访谈,得到了试测问卷大致的完成时间及对试测问卷项目内容的反馈。对于可读性较低的项目,通过受访者的意见进行了仔细推敲和修改。在以上步骤结束后,邀请1位心理测量教授和3位心理测量方向的博士对基本形成的初始问卷进行最后的审定和修改,主要检验问卷内容设计的合理性、题目的代表性与可读性等,最终形成含50个项目的初始问卷。问卷采用Likert式5点计分法,从“1=非常不同意”到“5=非常同意”。

### 4. 迫选式量表组卷

在获得了良好信效度的Likert式量表后,邀请30位心理测量方向专家(其中博士8人,硕士22人)完成社会称许性评价量表,要求专家为每个项目进行社会称许性进行评分,采用5点计分法,从“1=社会称许性很低”到“5=社会称许性很高”。结果表明,各项目社会称许性得分方差都小于2,标准差都小于1.4,表明专家对项目社会称许性的评价具有一致性。

使用R4.0中的autoFC包进行迫选式量表快速组卷,该R包由Li、Sun、Zhang三人设计开发。为了得到比项目组块为2更多的自比数据,提高项目的利用率,同时避免出现缺失比较数据,综合考虑后,设置项目组块为3。将项目难度、项目所在维度、社会称许性得分在组卷中所占权重设置为相等,最终得到了由10个组块、每个组块3个项目构成的初中生具身德育迫选式量表。

## 三、结果与分析

### (一) 项目分析

采用预测样本数据进行项目分析。首先,计算所有被试的总分,并将总分按照从高到低的顺序排列,取总分分值为前27%和后27%的个案组成高分组和低分组,对两组被试在每一个项目上的得分进行独立样本t检验,以检验项目区分能力。结果删除了两个得分差异未达显著的项目( $p > 0.05$ ),其余所有项目上的得分均存在显著差异( $p < 0.001$ ),显示出很好的区分能力。

随后计算剩余项目的题总相关,结果表明,所有项目与总分的相关均达到0.4以上( $p < 0.01$ )。对初始问卷进行内部一致性信度分析,结果表明其Cronbach's系数为0.949,删除任何一道题均不会引起其值的明显提高。

### (二) Likert式量表的信效度检验

#### 1. 探索性因子分析

采用预测样本进行探索性因子分析(EFA)。在EFA分析前求得KMO值为0.905,且Bartlett球形

检验为 2996.585 ( $p < 0.001$ ) ,表明变量间关联度较高 ,适合进行因子分析。EFA 采用主成分分析法 ,以特征大于 1 的标准及正交旋转法抽取因子 ,具体项目的剔除标准为: (1) 共同度小于 0.3; (2) 因子载荷小于 0.4; (3) 交叉载荷大于 0.3; (4) 删除后因子包含的项目大于 3 个; (5) 因子归属不当或不易解释的项目; (6) 考虑到迫选式量表编制需要构造每个组块项目数为 3 ,保证最终项目数为 3 的倍数。经过多轮探索 ,删除 20 个项目后得到收敛效度及区分效度良好的因子结构 ,最终得到含 30 个项目共 6 个显著因子 ,6 个因子解释了总方差的 58.273%。

根据 EFA 结果及各个项目的含义 ,将中学生具身德育需要考察的内容归纳为 6 个方面: (1) 因子 1 为“有理想有目标” ,内容为学生的理想、抱负以及未来视野 ,包含知、行两方面; (2) 因子 2 为“爱党爱国” ,内容为对党和国家的政治认同、情感认同、价值认同 ,包含知、情、行三方面; (3) 因子 3 为“中华优秀传统文化” ,主要内容对中华优秀传统文化的了解、热爱和传承 ,包含知、情、行三方面; (4) 因子 4 为“团结友善” ,主要内容为友善对待他人 ,不欺凌同学 ,包含知、行两方面; (5) 因子 5 为“心理健康” ,内容为健全的人格、积极的心态和良好的个性心理品质 ,包括知、行、意三方面; (6) 因子 6 为“生态文明” ,主要内容为节约资源、环境保护和健康的生活方式 ,包括知、行、意三方面。

## 2. 探索性结构方程建模

探索性结构方程建模(ESEM) 可用于验证因子模型 ,且与传统验证性因子分析(CFA) 相比更容易拟合 ,且更容易接近真实结果 ,并能更有效地支持一些后续的高级统计分析。因此本文主要采用 ESEM 对正式施测样本数据进行分析 ,并将分析结果与 CFA 进行对比。

本研究构建了由 1 到 7 共 7 个因子模型进行探索性结构方程建模 ,并比较各个模型的拟合指数。结果发现 ,仅有六因子( $CFI = 0.951$  , $TLI = 0.921$  , $RMSEA = 0.047$ ) 和七因子( $CFI = 0.966$  , $TLI = 0.940$  , $RMSEA = 0.041$ ) 模型的 CFI、TLI 均大于 0.90 , $RMSEA < 0.05$  ,达到拟合良好的标准。而七因子模型的拟合指标虽然优于六因子模型 ,但  $RMSEA = 0.06$  ,没有达到 Asparouhov 等人提出的提高标准( $RMSEA \geq 0.15$ ) ,这说明六因子模型的设置更为合理。

CFA 的各项拟合指标值为:  $CFI = 0.918$  , $TLI = 0.908$  , $RMSEA = 0.061$ 。从传统 CFA 拟合结果也可以看出 ,六因子模型的各拟合指标都达到了拟合良好的标准。

## 3. 效标关联效度

采用正式施测样本数据进行效标关联效度分析 ,效标采用学生的“道德与法治”科目期末考试成绩以及班主任对学生道德水平的评价。结果发现 ,初中生具身德育量表总分及各维度得分与学生“道德与法治”期末考试成绩、班主任对学生道德水平的评价均呈现显著正相关。

## 4. 信度检验

采用正式施测样本数据进行信度分析。(1) 内部一致性信度与分半信度: 分析结果显示初中生具身德育量表的 Cronbach's 系数为 0.903 ,分半信度值为 0.806。(2) 重测信度: 在正式施测的被试中抽取了 188 人在一个月后进行重复测量 ,并成功回收所有数据。分析前根据异常作答规则删除了 13 个无效数据 ,剩余 175 名被试作答用于计算两次测验的重测信度。结果表明 ,重测信度值为 0.689 ,各维度的重测信度都达到了 0.70 以上。所以 ,初中生具身德育量表具有良好的内部一致性、跨时间的稳定性。

## (三) 迫选式量表的分析

### 1. 整体拟合分析

数据采用迫选式量表正式施测样本 ,利用 Thurstone IRT 模型对数据进行拟合 ,估计方法为适用于分类变量的 WLSMV。结果表明 ,卡方/自由度为 1.234 ,小于 2;  $RMSEA$  为 0.028 ,小于 0.05; CFI 及 TLI 的指标分别为 0.925 和 0.913 ,皆大于 0.90 ,证明模型拟合良好。

### 2. 参数估计

采用 WLSMV 估计项目参数及被试参数 ,结果显示 ,每个项目至少在一个维度上的因子载荷在 0.3 以上 ,项目参数的标准误皆小于 0.6 ,表明参数估计结果良好。

### 3. 抗作假效果检验

本研究在迫选测验的正式施测中将被试分为两组 ,并设置两种不同的测验情境 ,在施测前 ,主试都会详细说明作答方式与作答要求。在说明后 ,第一组被试在无人监督的情形下作答 ,属于无压力情

境;而第二组被试在班主任老师的巡视下进行作答,属于压力情境。两组都要完成一份 Likert 式量表和一份迫选量表,采用抵消平衡法控制顺序效应。

独立样本  $t$  检验结果表明,被试在两种不同情境下作答 Likert 式量表时,在理想信念( $t=2.555, p<0.05$ )、爱党爱国( $t=5.082, p<0.001$ )、心理健康( $t=2.616, p<0.01$ )、生态文明( $t=1.872, p<0.05$ )四个维度及总分( $t=3.448, p<0.001$ )均存在显著差异;而作答迫选量表时,除中华优秀传统文化( $t=2.914, p<0.01$ )维度外,其余各维度上估计出的被试的特质得分均不存在显著差异。

## 四、讨论

### (一) 迫选量表的编制及 Thurstone IRT 模型的适用性检验

迫选量表的编制过程一般包括编写陈述、社会称许性评定及陈述配对等过程。本研究严格按照迫选量表的编制规范,并在迫选测验编制前很好地进行了项目分析及信效度分析,以用于陈述配对。陈述配对采用 Li 等人开发的“autoFC”R 包完成,配对的三个指标分别为项目所在维度、项目难度与社会称许性评分,其中项目难度采用两参数 IRT 模型进行估计。据此,不仅配对陈述的质量得到了保证,配对过程的质量亦得到了保证。

Thurstone IRT 模型是一种验证性因子分析方法,对施测数据进行模型拟合,发现整体拟合指标及估计出题目参数结果都十分理想。一方面,这说明 Thurstone 模型能够很好地描述迫选量表中项目与所测特质的关系,也能很好地描述被试的作答机制。另一方面,也说明迫选式初中生具身德育量表的结构效度很好,能够很好地测量研究者理论构想的心理特质。因此,Thurstone IRT 模型对迫选量表的实测数据是适用的,其估计的特质分数是对被试神经质程度的良好度量。

### (二) 迫选和 Likert 式量表抗作假效果的对比检验

已有的许多文献已经证明, Likert 式量表存在作假问题,这一问题不仅存在于人格量表中<sup>[17]</sup>,也已被证实存在于道德量表中<sup>[26]</sup>。本研究也得到了一致的结果。在压力情境和无压力情境两种作答情境下,两组被试 Likert 式量表的总分及社会称许性较高的理想信念、爱党爱国、心理健康与生态文明四维度的得分呈现显著差异,但在社会称许性较低的中华优秀传统文化及团结友善两维度上的得分未呈现显著差异。可以看出压力情境下,被试有意识地提高了对敏感道德问题的评分。而被试在迫选量表中仅有一个维度上的得分存在显著差异,其余五个维度上的得分均没有明显差异。这一结果有力地说明,迫选量表能够有效阻挡被试作假。尤其在采用 Thurstone IRT 计时,可以看出两组被试的能力参数不易受到压力情境的影响。

### (三) 迫选式初中生具身德育量表的实用性

我国教育部门及义务教育阶段的各级学校每三到四年就会开展大规模地德育监测,这就需要大规模地进行德育量表的施测<sup>[27-28]</sup>。然而,在监测过程中会遇到几个问题:首先,德育监测没有跟上德育理论的进步,尤其是具身德育理论的提出;其次,大规模调查的结果容易受到学生及学校层面社会称许性的影响,其结果的准确性还有待商榷;再次,为了解决以上问题,有时会用到一些高级的统计方法,对于学校一级的教师来说,可能在分析中容易遇到困难。

本研究很好地解决了以上三个问题。首先,迫选式初中生具身德育量表以教育部中小学生德育工作指南及孟万金教授的具身德育理论为理论基础,融入道德的知情意行四方面,很好地紧跟着德育理论的步伐,适用于现今的德育环境;其次,使用迫选式量表,很好地规避了社会称许性带来的影响,结果更具准确性;最后,分析简单,各级学校只需要使用 Mplus 就可以自行得到分析结果,结果也一目了然,易于解读。总之,本研究编制的迫选式初中生具身德育量表具有很强的实践性和实用性。

### (四) 研究的局限性

孟万金教授在具身德育评价的问题上指出“可以断言,最有效的评价是学生的自我评价,尤其是自我形成性评价,其次为教师与学生间的评价,最差的评价是外界人员的评价。”<sup>[25]</sup>本研究虽然采用了自陈式的量表进行德育评价,也力求获取更接近真实的结果,但由于研究的条件限制,难以完成复杂而多次的形成性评价。未来的研究者可以根据本文的思路,进一步编制适合义务教育各阶段的形成性评价量表,以求更好地辅助德育工作。

由于学校工作安排的原因,两次重测之间间隔的时间为两个月,略长于一般的重测信度时间间隔设计,故重测信度结果略低于一般要求的 0.7,但也均达到了 0.6 以上。此外,由于题量和测试安排的原因,无法将压力情境下的测验设计成被试内设计,不利于更深入的探讨结果,未来研究可对此改进。

## 五、结论

本研究根据目前较新的具身德育理论,扎根于我国现实的德育情境,编制完成了易实施测试、易分析数据、易解读结果的迫选式初中生具身德育量表。研究完整实现了迫选式初中生具身德育量表的编制过程,并很好地拟合了 Thurstone IRT 模型;并且通过研究结果可以看出,迫选式量表的抗作假能力优于传统 Likert 式量表,Thurstone IRT 模型估计的特质得分也比传统计分具有更强的抗作假能力。

### 参考文献:

- [1]章振乐. 正心立德 劳动树人:小学“新劳动教育”的实践与思考[J]. 中国特殊教育 2017(5):27-29.
- [2]孟万金. 具身德育:背景、内涵、创新:一论新时代具身德育[J]. 中国特殊教育 2017(11):69-73.
- [3]刘慧. 生命道德教育:基于新生物学的范式建构[D]. 南京师范大学 2022.
- [4]郭志芹. 马克思主义身体观视域中的道德教育[J]. 社科纵横 2012(12):208-209.
- [5]叶浩生. 有关具身认知思潮的理论心理学思考[J]. 心理学报 2011(5):589-598.
- [6]叶浩生. 具身认知:认知心理学的新取向[J]. 心理科学进展 2010(5):705-710.
- [7]叶红燕,张凤华. 从具身视角看道德判断[J]. 心理科学进展 2015(8):1480-1488.
- [8]孟万金,官群. 积极心理健康教育的转型升级[J]. 中小学心理健康教育 2016(16):43-46.
- [9]查亮亮,孙炳海,范丽婷. 推动新时代德育的具身化转向[N]. 中国教育报 2023 2023-05-04.
- [10]谔文莉,王若光,啜静. 寓教于乐:论《丸经》中的具身德育[J]. 体育研究与教育 2019(5):28-32.
- [11]沈啸忱,张卫国,张红,等. 具身德育的四个认知属性探析[J]. 中小学心理健康教育 2020(17):4-6.
- [12]苗小燕. 杭州下城实验区率先启动“具身德育”实践[J]. 中国特殊教育 2017(11):73.
- [13]张梦艳. 对当前德育测评趋势的思考:基于道德发展阶段理论[J]. 教育教学论坛 2018(50):218-219.
- [14]黄晓芬. 基于协同理念的德育测评题库设计与开发[J]. 教育信息技术 2020(3):27-29.
- [15]杨峰. 初中生品德测评量表的研制[D]. 广西师范大学 2000.
- [16]刘玉娟,孟万金. 小学生积极道德品测评量表的编制研究[J]. 中国特殊教育 2011(1):70-74.
- [17]王珊,骆方,刘红云. 迫选式人格测验的传统计分与 IRT 计分模型[J]. 心理科学进展 2014(3):549-557.
- [18]罗照盛. 项目反应理论基础[M]. 北京:北京师范大学出版社 2012.
- [19]洛方,刘红云,张东,等. 神经质人格迫选量表的开发及其抗作假效果研究[J]. 心理学探新 2013(5):460-464.
- [20]郑日昌. 编制心理量表须注意的几个问题[J]. 中国心理卫生杂志 2009(11):761-762.
- [21]中华人民共和国教育部. 教育部关于印发《中小学德育工作指南》的通知:教基[2017]8号[EB/OL]. (2017-08-22) [2022-08-23]. [http://www.moe.gov.cn/srcsite/A06/s3325/201709/t20170904\\_313128.html](http://www.moe.gov.cn/srcsite/A06/s3325/201709/t20170904_313128.html).
- [22]孟万金. 具身德育:源泉、体系、模式:二论新时代具身德育[J]. 中国特殊教育 2018(1):65-69.
- [23]孟万金. 具身德育:机制、精髓、课程:三论新时代具身德育[J]. 中国特殊教育 2018(4):73-96.
- [24]孟万金. 具身德育:原则、课堂、课外:四论新时代具身德育[J]. 中国德育 2019(5):14-18.
- [25]孟万金. 具身德育:纲领、实效、评价:五论新时代具身德育[J]. 中国德育 2019(6):15-19.
- [26]孙丽娟. 瑟斯顿 IRT 模型在迫选式道德价值观量表中的应用[D]. 西南大学 2015.
- [27]中华人民共和国教育部. 2017 国家义务教育质量监测报告[R]. 北京:教育部基础质量监测中心 2018.
- [28]江西省教育厅. 江西省 2019 年义务教育阶段德育质量监测报告[R]. 南昌:江西省教育评估监测研究院基础质量教育研究所,江西师范大学心理学院 2021.

(责任编辑:韩桥生)