

# 中国青少年社会与情感能力 测评之技术报告<sup>\*</sup>

张 静<sup>1</sup> 唐一鹏<sup>2</sup> 郭家俊<sup>3</sup> 邵志芳<sup>3</sup>

(1. 华东师范大学教育学部教育心理学系, 上海 200062; 2. 华东师范大学教育学部教育学系, 上海 200062; 3. 华东师范大学心理与认知科学学院, 上海 200062)

**摘 要:** 本技术报告基于中国苏州的测评数据, 对 OECD 社会与情感能力测评工具在中国文化下的心理测量学特征进行分析。具体计算 Alpha 系数和 Omega 系数来检验这些量表的信度; 进行了一系列验证性因子分析来检验这些量表的结构效度; 开展多组验证性因子分析来评估不同组(年龄组和性别组)的测量等值性, 以及使用项目反应理论广义分步计分模型来衡量项目并为所有参与测评的学生生成社会与情感能力值。

**关键词:** 社会与情感能力; 信度; 验证性因子分析; 测量等值; 项目反应理论

2019 年, 经济合作与发展组织(Organisation for Economic Co-operation and Development, 简称 OECD)在世界范围内开展了青少年社会与情感能力研究(Study of Social and Emotional Skills), 共有 9 个国家的 10 个城市参与了此次大规模国际测评项目。该研究的目标是测评各个国家(或城市)青少年社会与情感能力的发展水平, 分析影响青少年社会与情感能力发展的家庭、学校和社区因素, 预测社会与情感能力对学生健康、幸福感等生活结果的重要作用, 从而为决策者、管理者与广大中小学教师提供启示和改革建议。

苏州市下辖的 6 个区和 4 个县级市共 151 所中小学校的 7268 名学生参与此次测评项目, 包括 3647 名 10 岁和 3621 名 15 岁学生。作为中国经济、文化和教育的发达地区之一, 苏州的测试结果及其揭示的问题带有指向意义, 对中国进一步深化素质教育也具有借鉴价值。

## 一、OECD 社会与情感能力测评工具

### (一) OECD 社会与情感能力测评框架

OECD 社会与情感能力研究借鉴“大五人格”模型(Big Five Model), 建构社会与情感能力的测评框架。此框架主要分为五大维度: 任务能力(尽责性)、情绪调节(情绪稳定性)、协作能力(宜人性)、开放能力(开放性)和交往能力(外向性)。虽然这五个维度源自大五人格理论, 但是又有发展, 它们体现了人对社会情境的适应能力。每个维度又确立了三项子能力, 任务能力对应大五人格的尽责性, 包括自控力、责任感和毅力; 情绪调节对应大五人格的情绪稳定性, 包括抗压力、乐观和情绪控制; 协作能力对应大五人格的宜人性, 包括共情、合作与信任; 开放能力对应大五人格的开放性, 包括好奇心、创造性和包容度; 交往能力对应大五人格的外向性, 包括活力、果敢和乐群。表 1 呈现了正式测评中包含的五大维度和 15 项子能力。

<sup>\*</sup> 基金项目: 华东师范大学“幸福之花”基金先导项目(2019ECNU-XFZH011)。

## (二) OECD 社会与情感能力测评工具概览

社会与情感能力测评项目主要包括四类测评问卷:(1)学生调查问卷主要涵盖学生对其社会与情感能力的自我测评,但也包含有关学生学校和家庭环境的题项。(2)家长调查问卷包括有关孩子的社会与情感能力、孩子的成长背景、家庭环境以及家长的社会和情感能力等问题。(3)教师调查问卷收集教师对学生社会与情感能力的测评以及教师关于学校学习环境的报告。(4)校长调查问卷涵盖学校及其学生的更广泛的社会背景,以及旨在改善学习环境的学校现有的资源和计划。具体而言,2019 OECD 社会与情感能力正式测评为苏州测评现场提供了7份在线测评问卷。有两份学生问卷调查,一份针对10岁组,另一份针对15岁组。两个年龄组的父母问卷调查是相同的。教师问卷包括两个部分。第一部分由两个年龄组的教师完成,包括教师问卷调查和辅测定锚题。每位参与的教师都完成这一部分。第二部分包括两个版本,一个针对10岁组,另外一个针对15岁组。教师被要求为每个“他/她最了解”的学生完成这一部分。两个年龄组的主要问卷调查是相同的。表2列出了 OECD 社会与情感能力研究所使用的问卷调查信息。

表1 社会与情感能力的测评框架(5个维度和15个子能力)

| 5个维度(大五人格)  | 15个子能力                     |
|-------------|----------------------------|
| 任务能力(尽责性)   | 毅力(PER)、责任感(RES)、自控力(SEL)  |
| 情绪调节(情绪稳定性) | 抗压力(STR)、乐观(OPT)、情绪控制(EMO) |
| 协作能力(宜人性)   | 共情(EMP)、合作(COO)、信任(TRU)    |
| 开放能力(开放性)   | 包容度(TOL)、好奇心(CUR)、创造性(CRE) |
| 交往能力(外向性)   | 乐群(SOC)、果敢(ASS)、活力(ENE)    |

注:毅力=Persistence; 责任感=Responsibility; 自控力=Self-control; 抗压力=Stress resilience; 乐观=Optimism; 情绪控制=Emotional Control; 共情=Empathy; 合作=Cooperation; 信任=Trust; 包容度=Tolerance; 好奇心=Curiosity; 创造性=Creativity; 乐群=Sociability; 果敢=Assertiveness; 活力=Energy。15个子能力的英文缩写以其英文单词的前三个字母表示。

表2 OECD 社会与情感能力研究所使用的问卷调查

| 学生问卷调查        | 家长问卷调查        |
|---------------|---------------|
| 10岁组/15岁组测评量表 | 10岁组/15岁组测评量表 |
| 1. 辅测定锚题      | 1. 辅测定锚题      |
| 2. 直接评估工具     | 2. 间接评估工具     |
| 3. 行为指标       | 3. 行为指标       |
| 4. 学生问卷调查     | 4. 父母问卷调查     |
| <b>教师问卷调查</b> | <b>校长问卷调查</b> |
| 第一部分          | 1. 校长问卷调查     |
| 1. 辅测定锚题      | /             |
| 2. 教师问卷调查     | /             |
| 10岁组/15岁组测评量表 | /             |
| 1. 间接评估工具     | /             |
| 2. 行为指标       | /             |

## 二、OECD 社会与情感能力测评工具的测量学分析

在经过2017年的题项筛选、2018年的现场测试后,结合2019年的正式测试结果,OECD的社会与情感能力的正式测评最终确定学生问卷包含65个问题(15岁组有68个问题);家长问卷包含53个问题;教师问卷包含34个问题;校长问卷包括32个问题。其中学生直接和家长间接评估社会与情感量表中,每项社会与情感能力各有8个题项,共90个题项。教师间接评估量表中,每项社会与情感能力各有3个题项,共45个题项。值得注意的是,OECD综合所有参与国家或城市的数据分析后删减有些不太符合测量学标准的题项,表3呈现了学生直接评估和家长间接评估量表中各项社会与情感能力的

删减题项。为了保持国际标准,本部分沿用 OECD 2019 社会与情感能力研究最终筛选后的各子能力测评题项分别进行测量学分析。

**表 3 学生和家長量表中 15 项社会与情感能力测评题项的删减**

| 能力   | 学生量表 |       | 家長量表 |       |
|------|------|-------|------|-------|
|      | 保留数  | 删除题项  | 保留数  | 删除题项  |
| 果敢   | 7    | ASS07 | 7    | ASS07 |
| 合作   | 7    | COO05 | 7    | COO05 |
| 创造性  | 6    | CRE03 | 6    | CRE03 |
|      |      | CRE08 |      | CRE08 |
| 好奇心  | 6    | CUR01 | 8    | /     |
|      |      | CUR03 |      | /     |
| 情绪控制 | 7    | EMO05 | 6    | EMO01 |
|      |      |       |      | EMO06 |
| 共情   | 6    | EMP05 | 6    | EMP02 |
|      |      | EMP08 |      | EMP08 |
| 活力   | 7    | ENE07 | 7    | ENE04 |
| 乐观   | 7    | OPT04 | 8    | /     |
| 毅力   | 7    | PER06 | 8    | /     |
| 责任感  | 6    | RES02 | 6    | RES02 |
|      |      | RES07 |      | RES07 |
| 自控力  | 6    | SEL05 | 6    | SEL05 |
|      |      | SEL06 |      | SEL06 |
| 乐群   | 6    | SOC06 | 6    | SOC06 |
|      |      | SOC08 |      | SOC08 |
| 抗压力  | 6    | STR06 | 7    | STR07 |
|      |      | STR08 |      |       |
| 包容度  | 7    | TOL08 | 7    | TOL03 |
| 信任   | 6    | TRU03 | 7    | TRU03 |

表 4 列出使用的主要分析类型及其相应的测量学指标。在适用的情况下,这些分析是在分组水平上进行的,包括按年龄进行分析。由于进行了大量的分析,本报告这一部分的结果仅包含部分发现,以突出相关问题。总体来看,15 项社会与情感能力测评工具在中国文化背景下的信效度符合测量学标准。

**表 4 本报告中使用的测量学指标**

| 统计指标                   | 取值范围        | 解释说明  |
|------------------------|-------------|-------|
| 信度系数 (Alpha系数和Omega系数) | <0.60       | 信度低   |
|                        | 0.60 – 0.70 | 可接受信度 |
|                        | >0.70       | 信度高   |
| 验证性因子分析                | /           | /     |
| 比较拟合指数 (CFI)           | <0.90       | 拟合较差  |
|                        | >0.90       | 拟合较好  |
|                        | >0.95       | 拟合合理  |
|                        | >0.10       | 拟合较差  |
| 近似误差均方根 (RMSEA)        | 0.08 – 0.10 | 拟合一般  |
|                        | <0.08       | 拟合合理  |
| 标准化残差均方根 (SRMR)        | >0.08       | 拟合较差  |
|                        | 0.06 – 0.08 | 拟合一般  |

## (一) 信度分析

表5和表6分别列出了根据学生自评和家长间接评估数据估计的15项社会与情感能力量表的信度,主要包括Alpha系数 $\alpha$ (Cronbach, 1951)和Omega系数 $\Omega$ (McDonald, 2013)。可以看出,无论是总体数据,还是基于10岁组或15岁组数据,学生和家长问卷调查中所有15项社会与情感能力各子量表的信度都相对不错(Alpha和Omega系数均高于0.70)。而且相对于10岁组学生,根据15岁组学生自评和其家长间接评估数据估计的量表信度要好。

表5 15项社会与情感能力的信度系数(学生自评量表)

|      | 总体       |          | 10岁组     |          | 15岁组     |          |
|------|----------|----------|----------|----------|----------|----------|
|      | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ |
| 果敢   | 0.80     | 0.85     | 0.77     | 0.84     | 0.84     | 0.88     |
| 合作   | 0.83     | 0.88     | 0.82     | 0.87     | 0.81     | 0.86     |
| 创造性  | 0.81     | 0.86     | 0.77     | 0.80     | 0.82     | 0.84     |
| 好奇心  | 0.81     | 0.87     | 0.77     | 0.84     | 0.81     | 0.87     |
| 情绪控制 | 0.81     | 0.85     | 0.78     | 0.82     | 0.81     | 0.85     |
| 共情   | 0.78     | 0.85     | 0.77     | 0.83     | 0.76     | 0.83     |
| 活力   | 0.79     | 0.85     | 0.73     | 0.81     | 0.81     | 0.84     |
| 乐观   | 0.84     | 0.86     | 0.79     | 0.82     | 0.86     | 0.89     |
| 毅力   | 0.85     | 0.88     | 0.81     | 0.85     | 0.85     | 0.90     |
| 责任感  | 0.79     | 0.85     | 0.75     | 0.82     | 0.80     | 0.85     |
| 自控力  | 0.74     | 0.82     | 0.74     | 0.83     | 0.74     | 0.84     |
| 乐群   | 0.75     | 0.87     | 0.70     | 0.79     | 0.76     | 0.87     |
| 抗压力  | 0.83     | 0.88     | 0.80     | 0.84     | 0.83     | 0.89     |
| 包容度  | 0.73     | 0.76     | 0.70     | 0.76     | 0.77     | 0.80     |
| 信任   | 0.85     | 0.89     | 0.86     | 0.90     | 0.87     | 0.90     |

表6 15项社会与情感能力的信度系数(家长间接评估量表)

| 能力   | 总体       |          | 10岁组     |          | 15岁组     |          |
|------|----------|----------|----------|----------|----------|----------|
|      | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ |
| 果敢   | 0.83     | 0.87     | 0.82     | 0.87     | 0.85     | 0.88     |
| 合作   | 0.83     | 0.87     | 0.82     | 0.87     | 0.84     | 0.86     |
| 创造性  | 0.81     | 0.86     | 0.81     | 0.86     | 0.82     | 0.87     |
| 好奇心  | 0.84     | 0.87     | 0.84     | 0.87     | 0.85     | 0.88     |
| 情绪控制 | 0.80     | 0.84     | 0.80     | 0.84     | 0.79     | 0.84     |
| 共情   | 0.79     | 0.84     | 0.79     | 0.84     | 0.80     | 0.85     |
| 活力   | 0.70     | 0.76     | 0.68     | 0.78     | 0.71     | 0.77     |
| 乐观   | 0.79     | 0.83     | 0.79     | 0.83     | 0.80     | 0.83     |
| 毅力   | 0.86     | 0.89     | 0.86     | 0.89     | 0.85     | 0.88     |
| 责任感  | 0.79     | 0.83     | 0.77     | 0.82     | 0.79     | 0.84     |
| 自控力  | 0.76     | 0.82     | 0.76     | 0.83     | 0.74     | 0.81     |
| 乐群   | 0.77     | 0.86     | 0.75     | 0.85     | 0.77     | 0.84     |
| 抗压力  | 0.83     | 0.86     | 0.83     | 0.86     | 0.83     | 0.86     |
| 包容度  | 0.78     | 0.81     | 0.77     | 0.81     | 0.79     | 0.82     |
| 信任   | 0.79     | 0.81     | 0.79     | 0.81     | 0.79     | 0.83     |

表7显示了15项社会与情感能力量表中,教师对学生间接评估数据的量表信度。由于每个量表只有三个项目,因此预计信度低于根据学生和家长评估数据估计的信度,尤其是活力与自控力量表的

信度比较低( $\alpha$ 和 $\Omega$ 系数均低于0.70)。然而,有一些量表的信度非常高,如毅力量表约为0.90。根据教师数据估计的量表信度在10岁和15岁组中差别不大。

表7 15项社会与情感能力的信度系数(教师间接评估量表)

| 能力   | 总体       |          | 10岁组     |          | 15岁组     |          |
|------|----------|----------|----------|----------|----------|----------|
|      | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ | $\alpha$ | $\Omega$ |
| 果敢   | 0.82     | 0.83     | 0.84     | 0.85     | 0.79     | 0.80     |
| 合作   | 0.67     | 0.73     | 0.73     | 0.77     | 0.59     | 0.68     |
| 创造性  | 0.84     | 0.85     | 0.87     | 0.87     | 0.81     | 0.81     |
| 好奇心  | 0.78     | 0.81     | 0.83     | 0.85     | 0.69     | 0.75     |
| 情绪控制 | 0.67     | 0.70     | 0.71     | 0.73     | 0.62     | 0.65     |
| 共情   | 0.84     | 0.84     | 0.86     | 0.86     | 0.80     | 0.81     |
| 活力   | 0.64     | 0.66     | 0.68     | 0.69     | 0.58     | 0.62     |
| 乐观   | 0.76     | 0.78     | 0.79     | 0.80     | 0.72     | 0.74     |
| 毅力   | 0.89     | 0.89     | 0.92     | 0.92     | 0.85     | 0.85     |
| 责任感  | 0.74     | 0.76     | 0.80     | 0.81     | 0.65     | 0.69     |
| 自控力  | 0.57     | 0.67     | 0.64     | 0.71     | 0.46     | 0.59     |
| 乐群   | 0.83     | 0.83     | 0.85     | 0.85     | 0.80     | 0.80     |
| 抗压力  | 0.69     | 0.71     | 0.67     | 0.70     | 0.71     | 0.72     |
| 包容度  | 0.81     | 0.81     | 0.83     | 0.84     | 0.78     | 0.78     |
| 信任   | 0.75     | 0.77     | 0.78     | 0.79     | 0.78     | 0.79     |

## (二) 结构效度: 验证性因子分析

为了进一步检验学生直接评估、家长和教师间接评估15项社会与情感能力测评工具的结构效度,本部分采用Mplus7.4(Muthén & Muthén, 1998—2017)分别基于总体学生样本、10岁和15岁组学生子样本对这些测评工具进行验证性因子分析(Confirmatory Factor Analysis, 简称CFA)分析。CFA的基本思想在于将所有的测量项(即所有因子对应的测量量表题项)放在一个因子里面,然后进行分析。为了保持国际标准,本部分沿用OECD组织基于所有参与2019社会与情感能力测评项目的国家或城市数据分析,最终筛选后的各子能力测评题目分别进行CFA分析(参见表3)。

判断各测量模型是否达到心理测量学可接受标准,主要依据卡方自由度比、RMSEA、SRMR、CFI等不太受样本大小影响的指标。如表4所示,若CFI(比较拟合指数)大于0.90, RMSEA(近似误差均方根)小于0.08, SRMR(标准化残差均方根)小于0.06,则模型拟合良好(Beauducel & Wittmann, 2005; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Hu & Bentler, 1999)。如果测量模型拟合指标无法达标,即说明所有的测量项并不应该同属于一个因子。此时,可考虑根据模型修正指数(Model Indices, MI)修正模型直到上述各个指数达到可接受水平。如果因子与测量题项间的载荷系数低于0.40,也可考虑删除某测量项。

### (1) 学生直接评估量表

对于学生直接评估数据,表8、表9和表10分别呈现基于总体学生样本、10岁和15岁组学生子样本所有15项社会与情感能力的CFA模型各拟合指标信息。结果显示,最终量表的测量模型几乎所有指标都达到了测量学可接受水平,表明模型拟合良好,而且标准化因子载荷系数(各因子与测量项之间的系数)均大于0.40,说明每项社会与情感能力与其相对应的所有测量项之间有着良好的对应关系,结构效度好。对于10岁组的学生数据,活力这一能力的测量模型指标 $CFI = 0.890$ ,略小于0.90,模型拟合度一般。而对于15岁组的学生数据,好奇心和情绪控制两个能力的测量模型拟合度一般。根据模型修正指数显示,自由估计好奇心的第2题和第6题的误差方差相关系数,以及情绪控制的第3题和第8题的误差方差相关系数之后,这两种能力的测量模型拟合度良好。

表8 15项社会与情感能力测量模型拟合指标信息(总样本)

| 测量模型 | $\chi^2(df)$ | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|------|--------------|----------------------|-------|------------|-------|
| 果敢   | 528.277 (14) | 0.071 [0.066, 0.076] | 0.939 | 142177.448 | 0.034 |
| 合作   | 640.098 (14) | 0.079 [0.073, 0.084] | 0.928 | 110079.783 | 0.037 |
| 创造性  | 227.168 (9)  | 0.058 [0.051, 0.064] | 0.971 | 106223.130 | 0.025 |
| 好奇心  | 705.264 (9)  | 0.103 [0.097, 0.110] | 0.906 | 104600.913 | 0.046 |
| 情绪控制 | 871.614 (14) | 0.092 [0.087, 0.097] | 0.910 | 138205.822 | 0.042 |
| 共情   | 146.424 (9)  | 0.046 [0.040, 0.053] | 0.977 | 101235.364 | 0.020 |
| 活力   | 719.422 (14) | 0.083 [0.078, 0.089] | 0.918 | 140259.194 | 0.040 |
| 乐观   | 537.650 (14) | 0.072 [0.067, 0.077] | 0.950 | 134472.908 | 0.034 |
| 毅力   | 634.254 (14) | 0.078 [0.073, 0.083] | 0.941 | 119039.550 | 0.033 |
| 责任感  | 414.971 (9)  | 0.079 [0.072, 0.085] | 0.936 | 106880.247 | 0.034 |
| 自控力  | 254.878 (9)  | 0.061 [0.055, 0.068] | 0.961 | 109679.055 | 0.032 |
| 乐群   | 304.809 (9)  | 0.067 [0.061, 0.074] | 0.955 | 116169.452 | 0.032 |
| 抗压力  | 561.520 (9)  | 0.092 [0.086, 0.099] | 0.947 | 121556.099 | 0.032 |
| 包容度  | 313.786 (14) | 0.054 [0.049, 0.060] | 0.943 | 131370.499 | 0.030 |
| 信任   | 131.339 (9)  | 0.043 [0.037, 0.050] | 0.986 | 108560.586 | 0.016 |

注:数据来源于苏州市10岁和15岁年龄组学生直接评估社会与情感能力数据。

表9 15项社会与情感能力测量模型拟合指标信息(10岁组)

| 测量模型 | $\chi^2(df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|--------------|----------------------|-------|-----------|-------|
| 果敢   | 189.161 (14) | 0.059 [0.051, 0.066] | 0.955 | 76127.993 | 0.028 |
| 合作   | 269.59 (14)  | 0.071 [0.064, 0.078] | 0.936 | 57710.082 | 0.034 |
| 创造性  | 120.206 (9)  | 0.058 [0.049, 0.068] | 0.963 | 55872.204 | 0.028 |
| 好奇心  | 206.675 (9)  | 0.078 [0.069, 0.087] | 0.934 | 54527.111 | 0.037 |
| 情绪控制 | 333.821 (14) | 0.079 [0.072, 0.087] | 0.918 | 72205.682 | 0.039 |
| 共情   | 44.415 (9)   | 0.033 [0.024, 0.043] | 0.988 | 52201.855 | 0.016 |
| 活力   | 347.688 (14) | 0.081 [0.074, 0.088] | 0.890 | 73334.459 | 0.042 |
| 乐观   | 187.434 (14) | 0.058 [0.051, 0.066] | 0.955 | 69810.756 | 0.031 |
| 毅力   | 282.232 (14) | 0.073 [0.065, 0.080] | 0.936 | 62284.111 | 0.036 |
| 责任感  | 185.009 (9)  | 0.073 [0.064, 0.083] | 0.931 | 55849.160 | 0.034 |
| 自控力  | 72.67 (9)    | 0.044 [0.035, 0.054] | 0.979 | 57657.546 | 0.023 |
| 乐群   | 63.416 (9)   | 0.041 [0.032, 0.050] | 0.978 | 60029.197 | 0.021 |
| 抗压力  | 304.265 (9)  | 0.095 [0.086, 0.104] | 0.934 | 64348.768 | 0.036 |
| 包容度  | 132.896 (14) | 0.048 [0.041, 0.056] | 0.949 | 70016.596 | 0.028 |
| 信任   | 70.203 (9)   | 0.043 [0.034, 0.053] | 0.985 | 57509.657 | 0.017 |

注:数据来源于苏州市10岁年龄组学生直接评估社会与情感能力数据。

表10 15项社会与情感能力测量模型拟合指标信息(15岁组)

| 测量模型 | $\chi^2(df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|--------------|----------------------|-------|-----------|-------|
| 果敢   | 263.73 (14)  | 0.070 [0.063, 0.078] | 0.956 | 62332.493 | 0.028 |
| 合作   | 444.74 (14)  | 0.092 [0.085, 0.100] | 0.905 | 51207.295 | 0.046 |
| 创造性  | 111.787 (9)  | 0.056 [0.047, 0.066] | 0.977 | 48646.385 | 0.023 |
| 好奇心  | 301.76 (8)   | 0.101 [0.091, 0.111] | 0.929 | 48283.742 | 0.041 |
| 情绪控制 | 420.52 (13)  | 0.093 [0.086, 0.101] | 0.926 | 64381.430 | 0.044 |



续表 10

| 测量模型 | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|---------------|----------------------|-------|-----------|-------|
| 共情   | 166.548 (9)   | 0.070 [0.061, 0.079] | 0.944 | 47401.354 | 0.032 |
| 活力   | 440.021 (14)  | 0.092 [0.084, 0.099] | 0.917 | 65059.972 | 0.043 |
| 乐观   | 404.959 (14)  | 0.088 [0.081, 0.095] | 0.939 | 62519.071 | 0.038 |
| 毅力   | 378.971 (14)  | 0.085 [0.078, 0.092] | 0.931 | 54681.336 | 0.037 |
| 责任感  | 223.096 (9)   | 0.081 [0.072, 0.090] | 0.942 | 49523.042 | 0.035 |
| 自控力  | 215.756 (9)   | 0.080 [0.071, 0.089] | 0.938 | 50767.042 | 0.042 |
| 乐群   | 226.686 (9)   | 0.082 [0.073, 0.091] | 0.946 | 53883.923 | 0.038 |
| 抗压力  | 278.213 (9)   | 0.091 [0.082, 0.100] | 0.950 | 55472.028 | 0.031 |
| 包容度  | 219.013 (14)  | 0.064 [0.056, 0.071] | 0.939 | 58963.497 | 0.033 |
| 信任   | 68.963 (9)    | 0.043 [0.034, 0.053] | 0.988 | 49515.226 | 0.016 |

注: 数据来源于苏州市15岁年龄组学生直接评估社会与情感能力数据。

## (2) 家长间接评估量表

对于家长间接评估数据, 表 11、表 12 和表 13 分别列出了基于总体学生样本、10 岁和 15 岁组学生子样本 15 种社会与情感能力的 CFA 模型各拟合指标信息。最终量表的测量模型几乎所有指标都达到了测量学可接受水平, 表明模型拟合良好, 而且标准化因子载荷系数(各因子与测量项之间的系数)均大于 0.40, 说明每种社会与情感能力与其相对应的所有测量项之间有着良好的对应关系, 结构效度好。与学生直接评估数据一致, 对于家长间接评估 10 岁年龄组的数据, 活力这一能力的测量模型指标 CFI = 0.890, 略小于 0.90, 模型拟合度一般。对于家长间接评估 10 岁和 15 岁年龄组的数据, 情绪控制 and 责任感能力的测量模型拟合度一般。根据模型修正指数显示, 自由估计情绪控制的第 2 题和第 4 题的, 以及责任感的第 5 题和第 6 题的误差方差相关系数之后, 其测量模型的拟合度明显提高, 均达到可接受水平。

表 11 15 项社会与情感能力测量模型拟合指标信息 (总样本)

| 测量模型 | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|------|---------------|----------------------|-------|------------|-------|
| 果敢   | 379.37 (14)   | 0.061 [0.055, 0.066] | 0.966 | 108918.963 | 0.024 |
| 合作   | 673.98 (14)   | 0.081 [0.076, 0.087] | 0.932 | 92384.364  | 0.035 |
| 创造性  | 228.826 (9)   | 0.059 [0.052, 0.065] | 0.974 | 87861.480  | 0.021 |
| 好奇心  | 665.74 (19)   | 0.069 [0.065, 0.074] | 0.947 | 114567.305 | 0.036 |
| 情绪控制 | 625.728 (8)   | 0.104 [0.097, 0.111] | 0.922 | 97087.130  | 0.056 |
| 共情   | 577.561 (9)   | 0.094 [0.088, 0.101] | 0.926 | 86440.783  | 0.041 |
| 活力   | 715.635 (13)  | 0.087 [0.082, 0.093] | 0.878 | 122562.346 | 0.050 |
| 乐观   | 471.085 (20)  | 0.056 [0.052, 0.061] | 0.951 | 119153.871 | 0.035 |
| 毅力   | 762.121 (19)  | 0.074 [0.070, 0.079] | 0.944 | 120587.771 | 0.043 |
| 责任感  | 105.657 (8)   | 0.041 [0.035, 0.049] | 0.985 | 99000.089  | 0.017 |
| 自控力  | 484.029 (9)   | 0.086 [0.080, 0.093] | 0.935 | 96787.145  | 0.042 |
| 乐群   | 269.256 (9)   | 0.064 [0.057, 0.070] | 0.964 | 93379.739  | 0.025 |
| 抗压力  | 206.333 (14)  | 0.044 [0.039, 0.049] | 0.983 | 112335.448 | 0.021 |
| 包容度  | 161.757 (14)  | 0.039 [0.033, 0.044] | 0.978 | 94138.826  | 0.019 |
| 信任   | 83.887 (14)   | 0.037 [0.030, 0.045] | 0.980 | 48361.847  | 0.020 |

注: 数据来源于苏州市家长间接评估学生社会与情感能力总体数据。

表 12 15 项社会与情感能力测量模型拟合指标信息 (10 岁组)

| 测量模型 | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|---------------|----------------------|-------|-----------|-------|
| 果敢   | 218.16 (14)   | 0.064 [0.056, 0.071] | 0.959 | 56415.426 | 0.027 |

续表 12

| 测量模型 | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|---------------|----------------------|-------|-----------|-------|
| 合作   | 362.221 (14)  | 0.083 [0.076, 0.091] | 0.928 | 47573.777 | 0.036 |
| 创造性  | 106.541 (9)   | 0.055 [0.046, 0.065] | 0.976 | 45338.326 | 0.021 |
| 好奇心  | 601.61 (20)   | 0.090 [0.084, 0.096] | 0.901 | 58742.994 | 0.049 |
| 情绪控制 | 329.007 (8)   | 0.106 [0.096, 0.116] | 0.921 | 50266.181 | 0.057 |
| 共情   | 295.182 (9)   | 0.094 [0.085, 0.103] | 0.925 | 44799.481 | 0.041 |
| 活力   | 286.00 (13)   | 0.076 [0.069, 0.084] | 0.890 | 62201.313 | 0.045 |
| 乐观   | 241.385 (20)  | 0.056 [0.049, 0.062] | 0.948 | 61325.328 | 0.035 |
| 毅力   | 381.788 (19)  | 0.073 [0.067, 0.079] | 0.947 | 62523.162 | 0.040 |
| 责任感  | 52.116 (8)    | 0.039 [0.029, 0.050] | 0.987 | 51097.030 | 0.017 |
| 自控力  | 254.674 (9)   | 0.087 [0.078, 0.097] | 0.934 | 49966.338 | 0.043 |
| 乐群   | 191.642 (9)   | 0.075 [0.066, 0.085] | 0.944 | 47661.555 | 0.031 |
| 抗压力  | 113.883 (14)  | 0.045 [0.037, 0.052] | 0.982 | 58565.666 | 0.022 |
| 包容度  | 184.82 (14)   | 0.058 [0.051, 0.066] | 0.954 | 55939.370 | 0.032 |
| 信任   | 83.887 (14)   | 0.037 [0.030, 0.045] | 0.980 | 48361.847 | 0.020 |

注:数据来源于苏州市家长间接评估10岁年龄组学生社会与情感能力数据。

表 13 15 项社会与情感能力测量模型拟合指标信息 (15 岁组)

| 测量模型 | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|---------------|----------------------|-------|-----------|-------|
| 果敢   | 179.118 (14)  | 0.058 [0.051, 0.066] | 0.972 | 52168.590 | 0.023 |
| 合作   | 287.060 (14)  | 0.074 [0.067, 0.082] | 0.943 | 44651.209 | 0.032 |
| 创造性  | 97.748 (9)    | 0.053 [0.044, 0.063] | 0.980 | 42209.154 | 0.019 |
| 好奇心  | 538.916 (20)  | 0.086 [0.080, 0.092] | 0.917 | 55481.871 | 0.043 |
| 情绪控制 | 287.062 (8)   | 0.100 [0.090, 0.110] | 0.926 | 46667.130 | 0.056 |
| 共情   | 293.102 (9)   | 0.095 [0.086, 0.104] | 0.928 | 41602.916 | 0.042 |
| 活力   | 313.618 (12)  | 0.085 [0.077, 0.093] | 0.900 | 59720.010 | 0.044 |
| 乐观   | 242.286 (20)  | 0.056 [0.050, 0.063] | 0.955 | 57606.937 | 0.034 |
| 毅力   | 386.041 (19)  | 0.074 [0.068, 0.081] | 0.942 | 57945.769 | 0.046 |
| 责任感  | 64.250 (8)    | 0.045 [0.035, 0.055] | 0.982 | 47723.865 | 0.019 |
| 自控力  | 236.004 (9)   | 0.085 [0.076, 0.094] | 0.932 | 46560.129 | 0.042 |
| 乐群   | 98.409 (9)    | 0.053 [0.044, 0.063] | 0.977 | 45403.232 | 0.021 |
| 抗压力  | 114.113 (14)  | 0.045 [0.038, 0.053] | 0.983 | 53567.129 | 0.022 |
| 包容度  | 203.590 (14)  | 0.062 [0.055, 0.070] | 0.954 | 52400.636 | 0.032 |
| 信任   | 93.179 (14)   | 0.040 [0.033, 0.048] | 0.977 | 45855.524 | 0.020 |

注:数据来源于苏州市家长间接评估15岁年龄组学生社会与情感能力数据。

### (3)教师间接评估量表

不同于学生直接评估和家长间接评估量表,教师间接评估量表中每一项社会与情感能力各有3个测量题项,共45个题目,由最熟悉学生的教师来完成。验证性因子分析理论上要求每个因子至少有三个测量题项,因此,教师间接评估量表中每一项能力的测量模型都有且只有三个题项,是一个饱和模型。这些模型的自由度为零,CFI = 1.00, RMSEA = 0.00, SRMR = 0.00(Brown, 2006)。

#### (三)测量等值性检验:多组CFA

如前所述,OECD社会与情感能力测评项目主要抽取10岁和15岁年龄组学生参加,要比较不同年龄组或同一年龄组不同性别学生社会与情感能力的发展水平差异,首先要确保15项社会与情感能力



的组间年龄和性别比较具有真实性和有效性。因此, 社会与情感能力的测评工具必须具备跨年龄和跨性别测量等值性, 这是组间比较的前提条件(French & Finch, 2006)。

具体而言, 我们使用 Mplus7.4 对每项社会与情感力量表进行了三种测量等值检验: 形态等值(无约束)、弱等值(因子载荷等值)和强等值(因子载荷和截距等值)。至少在满足强等值之后, 组间比较才真实有效(Chen, 2007; Cheung & Rensvold, 2002)。一般来说, 模型拟合度会随着模型约束的增加而降低, 因为模型中约束越多, 说明组间差异的空间就越小。为了判断测量等值的程度, 除了每个模型的整体拟合外, 还可以查看模型拟合的相对变化。

表 14 列出了模型拟合指数以及这些统计数据在不同模型约束条件下的变化, 这些指数变化用于评估测量模型等值水平。一般认为, 当  $RMSEA \leq 0.08$ ,  $CFI \geq 0.90$  以及  $SRMR \leq 0.06$  时, 模型达到较好的拟合水平(Satorra & Bentler, 2001; Hu & Bentler, 1999)。值得注意的是, 卡方检验受样本量影响明显, 随着样本量增大, 即使很小的差异也可能会得到差异显著的结果, 因此采用模型拟合指数 CFI 和 RMSEA 的差异(即  $\Delta CFI$  和  $\Delta RMSEA$ )来评估测量等值性(Chen, 2007; OECD, 2019)。当  $\Delta CFI < 0.01$  或  $\Delta RMSEA < 0.015$  时, 表明两个嵌套模型不存在显著差异, 即认为模型等值可接受(王孟成, 2014)。值得注意的是, 用静态临界水平确定测量等值水平是不合适的, 因为拟合指数只是指示性的, 没有科学确定的接受或不接受标准。然而, 当进行跨组差异比较(不同性别或不同年龄组)得出结论时, 应记住量表测量等值性水平, 以避免错误解读研究结果。

表 14 测量等值模型拟合指数信息

| 等值水平 | 拟合指数           | 数值           |
|------|----------------|--------------|
| 形态等值 | RMSEA          | $\leq 0.08$  |
|      | SRMR           | $\leq 0.06$  |
|      | CFI            | $\geq 0.90$  |
| 弱等值  | $\Delta RMSEA$ | $\leq 0.015$ |
|      | $\Delta SRMR$  | $\leq 0.03$  |
|      | $\Delta CFI$   | $\leq 0.01$  |
|      | $\Delta BIC$   | $\leq 0$     |
| 强等值  | $\Delta RMSEA$ | $\leq 0.015$ |
|      | $\Delta SRMR$  | $\leq 0.03$  |
|      | $\Delta CFI$   | $\leq 0.01$  |
|      | $\Delta BIC$   | $\leq 0$     |

### (1) 跨年龄组测量等值检验

运用多组验证性因子分析分别检验 15 项社会与情感力量表在 10 岁和 15 岁年龄组的测量等值性, 结果如表 15 所示。首先检验形态等值, 即检验潜变量的构成形态是否相同, 同时也为下一步检验设定基线模型。在形态等值检验中, 允许各种参数自由估计, 得到的拟合指数见表 15 的 M1。各拟合指数均达到心理测量学要求, 形态等值成立, 形态等值模型可以作为下一步检验的基线模型。在基线模型 M1 的基础上进行弱等值, 即检验同一题项在不同年龄组载荷等值。年龄弱等值检验的拟合结果(见表 15 的 M2)显示  $\Delta RMSEA \leq 0.015$ ,  $\Delta SRMR \leq 0.030$ , 或  $\Delta CFI \leq 0.01$ , 这一结果支持各量表对应题项的因子载荷跨年龄组等值, 即弱等值成立。在此基础上检验强等值, 分别设定每个指标在 10 岁和 15 岁两个年龄组别的截距等值。跨年龄组强等值检验的拟合结果(见表 15 的 M3)  $\Delta RMSEA \leq 0.015$ ,  $\Delta SRMR \leq 0.030$ , 或  $\Delta CFI \leq 0.01$ , 这些结果说明了各题项跨年龄组的截距等值成立, 即强等值成立。然而对于好奇心、乐观、乐群和包容度四项能力, 有些题项的截距在 10 岁和 15 岁两个年龄组是不等的。自由估计好奇心的第 6 题, 乐观的第 5 题, 乐群的第 4 和第 5 题, 以及包容度的第 3 题的截距之后, 模型拟合指标均达到可接受水平, 所以部分强等值成立。

表 15 15 项社会与情感能力跨年龄组测量等值模型拟合信息

| 能力   | 测量等值模型    | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|------|-----------|---------------|----------------------|-------|------------|-------|
| 果敢   | M1. 形态等值  | 626.350 (28)  | 0.077 [0.072, 0.082] | 0.958 | 138489.598 | 0.028 |
|      | M2. 弱等值   | 678.726 (34)  | 0.072 [0.068, 0.077] | 0.955 | 138488.637 | 0.038 |
|      | M3. 强等值   | 2425.287 (40) | 0.128 [0.124, 0.133] | 0.832 | 140181.861 | 0.088 |
|      | M4. 部分强等值 | 1236.660 (38) | 0.093 [0.089, 0.098] | 0.916 | 139011.008 | 0.061 |
| 合作   | M1. 形态等值  | 1024.756 (28) | 0.099 [0.094, 0.104] | 0.931 | 108946.490 | 0.040 |
|      | M2. 弱等值   | 1134.07 (34)  | 0.094 [0.090, 0.099] | 0.924 | 109002.466 | 0.065 |
|      | M3. 强等值   | 1295.308 (40) | 0.093 [0.089, 0.097] | 0.913 | 109110.367 | 0.079 |
| 创造性  | M1. 形态等值  | 326.978 (18)  | 0.069 [0.062, 0.075] | 0.973 | 104543.542 | 0.026 |
|      | M2. 弱等值   | 341.179 (23)  | 0.062 [0.056, 0.068] | 0.972 | 104513.295 | 0.032 |
|      | M3. 强等值   | 891.011 (28)  | 0.092 [0.087, 0.097] | 0.924 | 105018.679 | 0.072 |
| 好奇心  | M1. 形态等值  | 613.486 (16)  | 0.101 [0.095, 0.108] | 0.949 | 102768.291 | 0.036 |
|      | M2. 弱等值   | 647.821 (21)  | 0.091 [0.085, 0.097] | 0.947 | 102758.179 | 0.044 |
|      | M3. 强等值   | 982.870 (26)  | 0.101 [0.095, 0.106] | 0.919 | 103048.781 | 0.069 |
|      | M4. 部分强等值 | 846.100 (25)  | 0.095 [0.090, 0.101] | 0.930 | 102920.899 | 0.052 |
| 情绪控制 | M1. 形态等值  | 778.078 (26)  | 0.089 [0.084, 0.095] | 0.945 | 136429.563 | 0.039 |
|      | M2. 弱等值   | 824.582 (32)  | 0.083 [0.078, 0.088] | 0.942 | 136422.729 | 0.048 |
|      | M3. 强等值   | 986.543 (38)  | 0.083 [0.079, 0.087] | 0.931 | 136531.352 | 0.052 |
| 共情   | M1. 形态等值  | 297.879 (18)  | 0.065 [0.059, 0.072] | 0.970 | 99628.163  | 0.025 |
|      | M2. 弱等值   | 371.366 (23)  | 0.065 [0.059, 0.070] | 0.962 | 99657.203  | 0.049 |
|      | M3. 强等值   | 897.533 (28)  | 0.093 [0.087, 0.098] | 0.906 | 100138.922 | 0.092 |
| 活力   | M1. 形态等值  | 789.217 (28)  | 0.087 [0.081, 0.092] | 0.906 | 138423.544 | 0.043 |
|      | M2. 弱等值   | 839.501 (34)  | 0.081 [0.076, 0.086] | 0.901 | 138419.006 | 0.050 |
|      | M3. 强等值   | 1033.628 (40) | 0.083 [0.078, 0.087] | 0.878 | 138581.155 | 0.055 |
|      | M4. 部分强等值 | 909.613 (39)  | 0.078 [0.074, 0.083] | 0.893 | 138435.975 | 0.052 |
| 乐观   | M1. 形态等值  | 593.979 (28)  | 0.075 [0.069, 0.080] | 0.945 | 132358.939 | 0.035 |
|      | M2. 弱等值   | 707.365 (34)  | 0.074 [0.069, 0.079] | 0.934 | 132444.387 | 0.057 |
|      | M3. 强等值   | 1195.857 (40) | 0.089 [0.085, 0.094] | 0.887 | 133004.224 | 0.075 |
|      | M4. 部分强等值 | 985.669 (39)  | 0.082 [0.077, 0.086] | 0.908 | 132744.821 | 0.067 |
| 毅力   | M1. 形态等值  | 662.562 (28)  | 0.079 [0.074, 0.084] | 0.933 | 116994.560 | 0.036 |
|      | M2. 弱等值   | 756.240 (34)  | 0.077 [0.072, 0.081] | 0.924 | 117043.028 | 0.056 |
|      | M3. 强等值   | 946.494 (40)  | 0.079 [0.075, 0.083] | 0.905 | 117216.351 | 0.077 |
| 责任感  | M1. 形态等值  | 408.879 (18)  | 0.077 [0.071, 0.084] | 0.937 | 105397.155 | 0.035 |
|      | M2. 弱等值   | 537.877 (23)  | 0.079 [0.073, 0.084] | 0.917 | 105521.382 | 0.071 |
|      | M3. 强等值   | 788.599 (28)  | 0.087 [0.081, 0.092] | 0.877 | 105780.713 | 0.087 |
| 自控力  | M1. 形态等值  | 294.854 (18)  | 0.065 [0.059, 0.072] | 0.956 | 108449.542 | 0.034 |
|      | M2. 弱等值   | 305.717 (23)  | 0.058 [0.052, 0.064] | 0.956 | 108411.598 | 0.036 |
|      | M3. 强等值   | 633.680 (28)  | 0.077 [0.072, 0.082] | 0.905 | 108777.077 | 0.061 |
| 乐群   | M1. 形态等值  | 289.837 (18)  | 0.065 [0.058, 0.071] | 0.958 | 113938.074 | 0.030 |
|      | M2. 弱等值   | 350.536 (23)  | 0.063 [0.057, 0.069] | 0.949 | 113971.246 | 0.048 |
|      | M3. 强等值   | 1061.928 (28) | 0.101 [0.096, 0.106] | 0.840 | 114804.302 | 0.087 |
|      | M4. 部分强等值 | 437.091 (26)  | 0.066 [0.061, 0.072] | 0.936 | 114047.943 | 0.055 |

续表 15

| 能力  | 测量等值模型    | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|-----|-----------|---------------|----------------------|-------|------------|-------|
| 抗压力 | M1. 形态等值  | 582.233 (18)  | 0.093 [0.087, 0.100] | 0.943 | 119845.749 | 0.034 |
|     | M2. 弱等值   | 656.531 (23)  | 0.087 [0.081, 0.093] | 0.936 | 119873.020 | 0.045 |
|     | M3. 强等值   | 731.292 (28)  | 0.083 [0.078, 0.088] | 0.929 | 119899.502 | 0.053 |
| 包容度 | M1. 形态等值  | 356.062 (28)  | 0.057 [0.052, 0.062] | 0.943 | 129009.206 | 0.031 |
|     | M2. 弱等值   | 411.726 (34)  | 0.055 [0.051, 0.060] | 0.934 | 129025.385 | 0.044 |
|     | M3. 强等值   | 1179.786 (40) | 0.089 [0.084, 0.093] | 0.802 | 129936.901 | 0.076 |
|     | M4. 部分强等值 | 524.646 (39)  | 0.059 [0.054, 0.063] | 0.916 | 129108.554 | 0.048 |
| 信任  | M1. 形态等值  | 139.153 (18)  | 0.043 [0.037, 0.050] | 0.986 | 107049.837 | 0.017 |
|     | M2. 弱等值   | 215.880 (23)  | 0.048 [0.042, 0.054] | 0.978 | 107111.718 | 0.052 |
|     | M3. 强等值   | 351.613 (28)  | 0.056 [0.051, 0.062] | 0.964 | 107242.855 | 0.075 |

注: 数据来源于OECD社会与情感能力测评项目苏州市10岁和15岁年龄组总体数据。

## (2) 跨性别测量等值检验

多组验证性因子分析结果表明 15 项社会与情感能力量表分别在总体学生样本、10 岁和 15 岁组学生样本中具备跨性别测量等值性, 结果如表 16 到表 18 所示。在形态等值检验中, 允许各种参数自由估计, 得到的模型拟合很好 ( $RMSEA \leq 0.08$ ,  $CFI \geq 0.90$  以及  $SRMR \leq 0.06$ ), 形态等值成立 (见表 16-18 的 M1)。当设定同一题项在男女样本中的因子载荷相等 (弱等值, 见表 16-18 的 M2) 后, 嵌套模型比较分析发现两模型差异不显著 (指标变化参见表 14), 从简约原则考虑应接受简洁模型 M2, 即弱等值成立。进一步设定每个题项在男女样本中的截距相等 (强等值见表 16-18 的 M3), 模型拟合指数变化不显著, 强等值成立。然而对于 15 岁年龄组, 模型修正指数显示活力的第 5 题在男女中的截距不相等, 允许自由估计之后, 模型拟合指标均达到可接受水平, 活力的部分强等值成立。

表 16 15 项社会与情感能力跨性别测量等值模型拟合信息 (总样本)

| 能力   | 测量等值模型   | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|------|----------|---------------|----------------------|-------|------------|-------|
| 果敢   | M1. 形态等值 | 774.330 (28)  | 0.086 [0.081, 0.091] | 0.942 | 141808.167 | 0.034 |
|      | M2. 弱等值  | 790.934 (34)  | 0.078 [0.074, 0.083] | 0.941 | 141771.444 | 0.037 |
|      | M3. 强等值  | 839.029 (40)  | 0.074 [0.070, 0.079] | 0.938 | 141766.213 | 0.039 |
| 合作   | M1. 形态等值 | 983.174 (28)  | 0.097 [0.092, 0.102] | 0.939 | 109720.794 | 0.038 |
|      | M2. 弱等值  | 1003.714 (34) | 0.089 [0.084, 0.094] | 0.938 | 109688.007 | 0.044 |
|      | M3. 强等值  | 1084.555 (40) | 0.085 [0.081, 0.089] | 0.933 | 109715.522 | 0.050 |
| 创造性  | M1. 形态等值 | 356.281 (18)  | 0.072 [0.066, 0.079] | 0.972 | 105969.774 | 0.026 |
|      | M2. 弱等值  | 369.286 (23)  | 0.064 [0.059, 0.070] | 0.971 | 105938.340 | 0.031 |
|      | M3. 强等值  | 380.598 (28)  | 0.059 [0.054, 0.064] | 0.971 | 105905.214 | 0.031 |
| 好奇心  | M1. 形态等值 | 558.699 (16)  | 0.097 [0.090, 0.104] | 0.958 | 103657.400 | 0.034 |
|      | M2. 弱等值  | 568.753 (21)  | 0.085 [0.079, 0.091] | 0.958 | 103623.016 | 0.037 |
|      | M3. 强等值  | 888.375 (26)  | 0.096 [0.090, 0.101] | 0.933 | 103898.200 | 0.057 |
| 情绪控制 | M1. 形态等值 | 1189.31 (28)  | 0.107 [0.102, 0.112] | 0.920 | 137795.865 | 0.042 |
|      | M2. 弱等值  | 1209.009 (34) | 0.098 [0.093, 0.102] | 0.919 | 137762.238 | 0.045 |
|      | M3. 强等值  | 1286.639 (40) | 0.093 [0.088, 0.097] | 0.914 | 137786.540 | 0.046 |
| 共情   | M1. 形态等值 | 280.730 (18)  | 0.063 [0.057, 0.070] | 0.974 | 101005.583 | 0.022 |
|      | M2. 弱等值  | 284.774 (23)  | 0.056 [0.050, 0.062] | 0.974 | 100965.189 | 0.024 |
|      | M3. 强等值  | 315.132 (28)  | 0.053 [0.048, 0.059] | 0.972 | 100951.108 | 0.028 |

续表 16

| 能力  | 测量等值模型    | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC        | SRMR  |
|-----|-----------|---------------|----------------------|-------|------------|-------|
| 活力  | M1. 形态等值  | 756.516 (28)  | 0.085 [0.080, 0.090] | 0.916 | 139744.070 | 0.040 |
|     | M2. 弱等值   | 793.118 (34)  | 0.079 [0.074, 0.083] | 0.913 | 139710.489 | 0.043 |
|     | M3. 强等值   | 1001.059 (40) | 0.081 [0.077, 0.086] | 0.889 | 139884.801 | 0.054 |
|     | M4. 部分强等值 | 896.925 (39)  | 0.078 [0.074, 0.082] | 0.901 | 139767.487 | 0.048 |
| 乐观  | M1. 形态等值  | 563.904 (28)  | 0.073 [0.068, 0.078] | 0.950 | 134041.356 | 0.034 |
|     | M2. 弱等值   | 603.664 (34)  | 0.068 [0.063, 0.073] | 0.947 | 134018.272 | 0.040 |
|     | M3. 强等值   | 699.268 (40)  | 0.067 [0.063, 0.072] | 0.938 | 134057.005 | 0.046 |
| 毅力  | M1. 形态等值  | 638.842 (28)  | 0.078 [0.072, 0.083] | 0.942 | 118584.773 | 0.034 |
|     | M2. 弱等值   | 679.848 (34)  | 0.072 [0.068, 0.077] | 0.939 | 118550.787 | 0.038 |
|     | M3. 强等值   | 764.307 (40)  | 0.071 [0.066, 0.075] | 0.932 | 118569.685 | 0.044 |
| 责任感 | M1. 形态等值  | 427.399 (18)  | 0.079 [0.073, 0.086] | 0.937 | 106536.304 | 0.034 |
|     | M2. 弱等值   | 444.536 (23)  | 0.071 [0.065, 0.077] | 0.935 | 106499.332 | 0.036 |
|     | M3. 强等值   | 484.233 (28)  | 0.067 [0.062, 0.072] | 0.930 | 106477.989 | 0.038 |
| 自控力 | M1. 形态等值  | 288.395 (18)  | 0.064 [0.058, 0.071] | 0.959 | 109457.506 | 0.033 |
|     | M2. 弱等值   | 304.730 (23)  | 0.058 [0.052, 0.064] | 0.957 | 109427.103 | 0.036 |
|     | M3. 强等值   | 328.004 (28)  | 0.054 [0.049, 0.060] | 0.954 | 109394.077 | 0.038 |
| 乐群  | M1. 形态等值  | 321.179 (18)  | 0.068 [0.062, 0.075] | 0.955 | 115909.460 | 0.032 |
|     | M2. 弱等值   | 326.773 (23)  | 0.060 [0.055, 0.066] | 0.954 | 115866.887 | 0.033 |
|     | M3. 强等值   | 390.796 (28)  | 0.060 [0.055, 0.065] | 0.946 | 115884.667 | 0.037 |
| 抗压力 | M1. 形态等值  | 551.516 (18)  | 0.090 [0.084, 0.097] | 0.949 | 121164.649 | 0.032 |
|     | M2. 弱等值   | 604.045 (23)  | 0.084 [0.078, 0.089] | 0.945 | 121157.878 | 0.040 |
|     | M3. 强等值   | 690.646 (28)  | 0.081 [0.076, 0.086] | 0.937 | 121189.502 | 0.047 |
| 包容度 | M1. 形态等值  | 327.257 (28)  | 0.054 [0.049, 0.060] | 0.944 | 130618.457 | 0.030 |
|     | M2. 弱等值   | 355.762 (34)  | 0.051 [0.046, 0.056] | 0.940 | 130598.545 | 0.038 |
|     | M3. 强等值   | 523.306 (40)  | 0.058 [0.053, 0.062] | 0.909 | 130742.962 | 0.049 |
| 信任  | M1. 形态等值  | 150.978 (18)  | 0.045 [0.039, 0.052] | 0.986 | 108272.011 | 0.017 |
|     | M2. 弱等值   | 162.978 (23)  | 0.041 [0.035, 0.047] | 0.985 | 108237.967 | 0.022 |
|     | M3. 强等值   | 195.697 (28)  | 0.041 [0.035, 0.046] | 0.982 | 108225.118 | 0.027 |

注:数据来源于OECD社会与情感能力测评项目苏州市10岁和15岁年龄组总体数据。

表 17 15项社会与情感能力的跨性别测量等值模型(10岁组)

| 能力  | 测量等值模型   | $\chi^2 (df)$ | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|-----|----------|---------------|----------------------|-------|-----------|-------|
| 果敢  | M1. 形态等值 | 263.618 (28)  | 0.068 [0.061, 0.076] | 0.956 | 76086.527 | 0.029 |
|     | M2. 弱等值  | 273.490 (34)  | 0.062 [0.056, 0.069] | 0.955 | 76047.207 | 0.033 |
|     | M3. 强等值  | 294.622 (40)  | 0.059 [0.053, 0.066] | 0.953 | 76019.148 | 0.034 |
| 合作  | M1. 形态等值 | 421.622 (28)  | 0.088 [0.081, 0.095] | 0.946 | 57619.637 | 0.034 |
|     | M2. 弱等值  | 425.981 (34)  | 0.080 [0.073, 0.086] | 0.946 | 57574.804 | 0.037 |
|     | M3. 强等值  | 451.322 (40)  | 0.075 [0.069, 0.082] | 0.943 | 57550.953 | 0.042 |
| 创造性 | M1. 形态等值 | 189.222 (18)  | 0.072 [0.063, 0.082] | 0.964 | 55840.328 | 0.029 |
|     | M2. 弱等值  | 192.076 (23)  | 0.064 [0.055, 0.072] | 0.964 | 55802.189 | 0.031 |
|     | M3. 强等值  | 207.533 (28)  | 0.059 [0.052, 0.067] | 0.962 | 55776.652 | 0.033 |
| 好奇心 | M1. 形态等值 | 287.094 (18)  | 0.091 [0.082, 0.100] | 0.948 | 54451.038 | 0.037 |
|     | M2. 弱等值  | 297.799 (23)  | 0.081 [0.073, 0.089] | 0.947 | 54420.751 | 0.043 |
|     | M3. 强等值  | 388.592 (28)  | 0.084 [0.077, 0.092] | 0.930 | 54470.553 | 0.057 |

续表 17

| 能力   | 测量等值模型   | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|----------|---------------|----------------------|-------|-----------|-------|
| 情绪控制 | M1. 形态等值 | 458.526 (28)  | 0.092 [0.085, 0.099] | 0.926 | 72193.198 | 0.040 |
|      | M2. 弱等值  | 468.805 (34)  | 0.084 [0.077, 0.091] | 0.925 | 72154.286 | 0.043 |
|      | M3. 强等值  | 488.213 (40)  | 0.079 [0.072, 0.085] | 0.923 | 72124.502 | 0.044 |
| 共情   | M1. 形态等值 | 103.950 (18)  | 0.051 [0.042, 0.061] | 0.982 | 52060.537 | 0.020 |
|      | M2. 弱等值  | 108.599 (23)  | 0.045 [0.037, 0.054] | 0.982 | 52024.194 | 0.025 |
|      | M3. 强等值  | 138.630 (28)  | 0.047 [0.039, 0.055] | 0.977 | 52013.233 | 0.033 |
| 活力   | M1. 形态等值 | 372.241 (28)  | 0.082 [0.075, 0.090] | 0.888 | 73295.767 | 0.043 |
|      | M2. 弱等值  | 382.797 (34)  | 0.075 [0.068, 0.082] | 0.887 | 73256.543 | 0.046 |
|      | M3. 强等值  | 436.883 (40)  | 0.074 [0.068, 0.080] | 0.871 | 73258.563 | 0.052 |
| 乐观   | M1. 形态等值 | 210.911 (28)  | 0.060 [0.053, 0.068] | 0.954 | 69738.240 | 0.033 |
|      | M2. 弱等值  | 233.551 (34)  | 0.057 [0.050, 0.064] | 0.950 | 69716.144 | 0.043 |
|      | M3. 强等值  | 262.927 (40)  | 0.055 [0.049, 0.062] | 0.944 | 69692.600 | 0.047 |
| 毅力   | M1. 形态等值 | 289.158 (28)  | 0.072 [0.064, 0.079] | 0.938 | 62214.558 | 0.037 |
|      | M2. 弱等值  | 308.485 (34)  | 0.067 [0.060, 0.074] | 0.935 | 62178.233 | 0.042 |
|      | M3. 强等值  | 344.338 (40)  | 0.065 [0.058, 0.071] | 0.928 | 62159.272 | 0.048 |
| 责任感  | M1. 形态等值 | 196.657 (18)  | 0.074 [0.065, 0.083] | 0.931 | 55763.702 | 0.034 |
|      | M2. 弱等值  | 200.391 (23)  | 0.065 [0.057, 0.074] | 0.931 | 55726.383 | 0.036 |
|      | M3. 强等值  | 218.919 (28)  | 0.061 [0.054, 0.069] | 0.926 | 55696.961 | 0.038 |
| 自控力  | M1. 形态等值 | 81.018 (18)   | 0.044 [0.034, 0.054] | 0.980 | 57601.332 | 0.023 |
|      | M2. 弱等值  | 87.420 (23)   | 0.039 [0.031, 0.048] | 0.979 | 57566.995 | 0.029 |
|      | M3. 强等值  | 101.193 (28)  | 0.038 [0.030, 0.046] | 0.976 | 57538.543 | 0.031 |
| 乐群   | M1. 形态等值 | 78.541 (18)   | 0.043 [0.034, 0.053] | 0.976 | 59961.131 | 0.022 |
|      | M2. 弱等值  | 80.718 (23)   | 0.037 [0.029, 0.046] | 0.977 | 59925.129 | 0.027 |
|      | M3. 强等值  | 106.921 (28)  | 0.039 [0.032, 0.047] | 0.968 | 59912.644 | 0.035 |
| 抗压力  | M1. 形态等值 | 301.952 (18)  | 0.093 [0.084, 0.103] | 0.937 | 64308.899 | 0.036 |
|      | M2. 弱等值  | 319.684 (23)  | 0.084 [0.076, 0.093] | 0.934 | 64281.933 | 0.040 |
|      | M3. 强等值  | 351.705 (28)  | 0.080 [0.072, 0.087] | 0.928 | 64264.592 | 0.043 |
| 包容度  | M1. 形态等值 | 140.204 (28)  | 0.047 [0.039, 0.055] | 0.953 | 69910.986 | 0.029 |
|      | M2. 弱等值  | 147.490 (34)  | 0.043 [0.036, 0.050] | 0.952 | 69870.51  | 0.032 |
|      | M3. 强等值  | 178.092 (40)  | 0.044 [0.037, 0.050] | 0.942 | 69853.022 | 0.036 |
| 信任   | M1. 形态等值 | 87.817 (18)   | 0.046 [0.037, 0.056] | 0.983 | 57432.936 | 0.019 |
|      | M2. 弱等值  | 100.790 (23)  | 0.043 [0.035, 0.052] | 0.981 | 57408.331 | 0.031 |
|      | M3. 强等值  | 120.380 (28)  | 0.043 [0.035, 0.051] | 0.978 | 57386.077 | 0.037 |

注: 数据来源于OECD社会与情感能力测评项目苏州市10岁年龄组数据。

表 18 15 项跨性别测量等值模型分析 (15 岁组)

| 能力 | 测量等值模型   | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|----|----------|---------------|----------------------|-------|-----------|-------|
| 果敢 | M1. 形态等值 | 396.299 (28)  | 0.085 [0.078, 0.093] | 0.958 | 62094.579 | 0.029 |
|    | M2. 弱等值  | 413.012 (34)  | 0.079 [0.072, 0.085] | 0.957 | 62062.149 | 0.036 |
|    | M3. 强等值  | 518.644 (40)  | 0.081 [0.075, 0.088] | 0.946 | 62118.636 | 0.044 |
| 合作 | M1. 形态等值 | 638.732 (28)  | 0.110 [0.103, 0.117] | 0.916 | 51004.350 | 0.047 |
|    | M2. 弱等值  | 663.712 (34)  | 0.101 [0.095, 0.108] | 0.914 | 50980.187 | 0.057 |
|    | M3. 强等值  | 734.262 (40)  | 0.098 [0.092, 0.104] | 0.905 | 51001.592 | 0.062 |

续表 18

| 能力   | 测量等值模型    | $\chi^2$ (df) | RMSEA [90% CI]       | CFI   | BIC       | SRMR  |
|------|-----------|---------------|----------------------|-------|-----------|-------|
| 创造性  | M1. 形态等值  | 167.161 (18)  | 0.068 [0.059, 0.077] | 0.977 | 48509.339 | 0.024 |
|      | M2. 弱等值   | 186.365 (23)  | 0.063 [0.055, 0.071] | 0.975 | 48487.591 | 0.039 |
|      | M3. 强等值   | 204.069 (28)  | 0.059 [0.052, 0.067] | 0.973 | 48464.342 | 0.040 |
| 好奇心  | M1. 形态等值  | 368.130 (16)  | 0.110 [0.101, 0.120] | 0.947 | 47842.149 | 0.039 |
|      | M2. 弱等值   | 373.240 (21)  | 0.096 [0.088, 0.105] | 0.947 | 47806.305 | 0.042 |
|      | M3. 强等值   | 658.401 (26)  | 0.116 [0.109, 0.124] | 0.904 | 48050.513 | 0.068 |
|      | M4. 部分强等值 | 394.770 (24)  | 0.093 [0.085, 0.101] | 0.944 | 47803.263 | 0.041 |
| 情绪控制 | M1. 形态等值  | 536.961 (26)  | 0.104 [0.097, 0.112] | 0.934 | 64070.176 | 0.045 |
|      | M2. 弱等值   | 563.108 (32)  | 0.096 [0.089, 0.103] | 0.931 | 64047.180 | 0.050 |
|      | M3. 强等值   | 647.347 (38)  | 0.094 [0.088, 0.101] | 0.921 | 64082.275 | 0.051 |
| 共情   | M1. 形态等值  | 248.080 (18)  | 0.084 [0.075, 0.094] | 0.947 | 47353.156 | 0.033 |
|      | M2. 弱等值   | 250.253 (23)  | 0.074 [0.066, 0.082] | 0.947 | 47314.376 | 0.034 |
|      | M3. 强等值   | 269.046 (28)  | 0.069 [0.062, 0.077] | 0.944 | 47292.216 | 0.035 |
| 活力   | M1. 形态等值  | 457.915 (28)  | 0.092 [0.085, 0.100] | 0.916 | 64660.174 | 0.044 |
|      | M2. 弱等值   | 478.017 (34)  | 0.085 [0.078, 0.092] | 0.913 | 64620.611 | 0.047 |
|      | M3. 强等值   | 686.034 (40)  | 0.095 [0.088, 0.101] | 0.873 | 64810.817 | 0.062 |
|      | M4. 部分强等值 | 576.251 (39)  | 0.087 [0.081, 0.094] | 0.895 | 64685.309 | 0.052 |
| 乐观   | M1. 形态等值  | 438.105 (28)  | 0.090 [0.083, 0.098] | 0.937 | 62298.892 | 0.039 |
|      | M2. 弱等值   | 456.575 (34)  | 0.083 [0.076, 0.090] | 0.935 | 62259.634 | 0.043 |
|      | M3. 强等值   | 535.678 (40)  | 0.083 [0.077, 0.089] | 0.924 | 62289.755 | 0.053 |
| 毅力   | M1. 形态等值  | 401.041 (28)  | 0.086 [0.079, 0.094] | 0.931 | 54432.920 | 0.038 |
|      | M2. 弱等值   | 414.857 (34)  | 0.079 [0.072, 0.086] | 0.930 | 54389.310 | 0.040 |
|      | M3. 强等值   | 462.126 (40)  | 0.076 [0.070, 0.083] | 0.922 | 54378.391 | 0.045 |
| 责任感  | M1. 形态等值  | 230.499 (18)  | 0.081 [0.072, 0.090] | 0.944 | 49311.781 | 0.035 |
|      | M2. 弱等值   | 248.038 (23)  | 0.074 [0.066, 0.082] | 0.941 | 49286.479 | 0.045 |
|      | M3. 强等值   | 299.235 (28)  | 0.073 [0.066, 0.081] | 0.929 | 49295.603 | 0.053 |
| 自控力  | M1. 形态等值  | 252.925 (18)  | 0.085 [0.076, 0.095] | 0.931 | 50634.302 | 0.046 |
|      | M2. 弱等值   | 265.973 (23)  | 0.077 [0.068, 0.085] | 0.928 | 50603.233 | 0.049 |
|      | M3. 强等值   | 306.397 (28)  | 0.074 [0.067, 0.082] | 0.918 | 50597.566 | 0.051 |
| 乐群   | M1. 形态等值  | 226.359 (18)  | 0.080 [0.071, 0.090] | 0.949 | 53748.626 | 0.038 |
|      | M2. 弱等值   | 237.764 (23)  | 0.072 [0.064, 0.080] | 0.947 | 53716.471 | 0.042 |
|      | M3. 强等值   | 289.408 (28)  | 0.072 [0.065, 0.080] | 0.936 | 53726.907 | 0.044 |
| 抗压力  | M1. 形态等值  | 283.529 (18)  | 0.090 [0.081, 0.100] | 0.951 | 55228.375 | 0.032 |
|      | M2. 弱等值   | 307.250 (23)  | 0.083 [0.075, 0.091] | 0.948 | 55205.668 | 0.040 |
|      | M3. 强等值   | 385.503 (28)  | 0.084 [0.077, 0.092] | 0.934 | 55246.136 | 0.054 |
| 包容度  | M1. 形态等值  | 259.387 (28)  | 0.068 [0.060, 0.075] | 0.934 | 58344.912 | 0.035 |
|      | M2. 弱等值   | 288.044 (34)  | 0.064 [0.058, 0.071] | 0.928 | 58330.327 | 0.052 |
|      | M3. 强等值   | 459.834 (40)  | 0.076 [0.070, 0.083] | 0.880 | 58485.821 | 0.074 |
| 信任   | M1. 形态等值  | 83.818 (18)   | 0.045 [0.036, 0.055] | 0.987 | 49421.869 | 0.018 |
|      | M2. 弱等值   | 92.389 (23)   | 0.041 [0.032, 0.050] | 0.986 | 49390.311 | 0.028 |
|      | M3. 强等值   | 117.523 (28)  | 0.042 [0.034, 0.050] | 0.982 | 49376.614 | 0.031 |

注:数据来源于OECD社会与情感能力测评项目苏州市15岁年龄组数据。



上述分析表明, 15 项社会与情感能力量表不仅在 10 岁和 15 岁组学生样本中跨年龄等值, 而且分别在总体样本、10 岁和 15 岁组学生子样本中男女跨性别等值, 从而确保这些能力的跨年龄和跨性别比较真实有效。

#### (四) 项目反应理论模型

OECD 社会与情感能力研究的目的是为了考察 10 岁和 15 岁组学生社会与情感能力上的表现。由于这些社会与情感能力属于个人特质(personal traits), 是无法直接观察到的潜在结构(latent constructs), 因此必须根据被试对一组题项(item)的回答来估计这些深层的心理结构。与 OECD 开展的其他大规模国际测评所不同的是, OECD 对于社会与情感能力的测量基于学生-家长-教师“三角互证”(triangulation)的模式来进行(Kankaraš et al., 2019), 即在传统的学生自评之外, 加上家长和教师对学生社会与情感能力的间接评价。在正式测评中, 三方评价都通过相应的社会与情感能力量表来进行, 一共包括 15 个单项能力测评(也即 15 个基本量表), 其中学生和家長采用的量表完全相同, 每项能力都采用 8 道题来进行测量。考虑到一名教师要同时评价多名学生, 故而教师问卷中每项能力只在 8 道题中选取 3 道。正是由于学生、家长测评和教师测评之间的题量差别, 所以 OECD 在计算各项能力得分时, 也采用了不同的思路。对于学生自评和家长评价, OECD 基于项目反映理论(Item Response Theory, 简称 IRT), 使用 ConQuest 4 软件来估计(Adams, Wu, & Wilson, 2015), 经过删减题目和多次测量学校校正, 得到跨国可比的标准化分数。对于教师评价, OECD 则简单计算了每个基本量表中 3 道题目的平均分数, 将其作为该单项能力的得分, 而且没有再进行其他测量学的校正。

基于学生和家長测评数据的最终能力值要通过赋分、模型估计、反应风格校正、标准化这四个步骤来实现。首先, OECD 需要对题项(item)的原始选项(response)进行赋分(0—4 分); 其次, 利用 IRT 中的广义分步计分模型(Generalized Partial Credit Model, 简称 GPCM)对合并样本(所有城市的 10 岁组和 15 岁组学生)进行参数估计, 并基于加权似然估计(Weighted Likelihood Estimate, 简称 WLE)算法获得能力初始值; 再次, 校正默认反应风格(Acquiescence Response Style)所带来的估计偏误, 获得能力校正值; 最后, 将能力校正值通过线性变换转化为均值 500 分(所有城市 10 岁组均值)、标准差 100 分(所有城市 10 岁组标准差)的标准分, 也即最终能力值。

##### (1) 原始量表的赋分

社会与情感能力量表是 5 点计分的李克特量表, 在利用 IRT 模型进行处理之前, 需要先行对学生和家長评价的结果进行赋分, 将量表项目的五个选项转化为具体分值(表 19)。从下表中可以看出, 对于正向题, 1—5 选项被赋值成为 0—4 分, 而对于反向题, 则进行相反赋分。

表 19 项目分数设定

| 选项    | 正向题得分 | 反向题得分 |
|-------|-------|-------|
| 非常不同意 | 0     | 4     |
| 不同意   | 1     | 3     |
| 一般    | 2     | 2     |
| 同意    | 3     | 1     |
| 非常同意  | 4     | 0     |

##### (2) GPCM 模型与初始值估计

OECD 社会与情感能力研究基于 IRT 理论来估计学生自评和家長评价的各项社会与情感能力的得分, 具体采用的是 GPCM 模型(Muraki, 1992)。GPCM 模型是一个概率模型, 用于处理分步计分类型的题项, 或者称为有序多分类(ordered polytomous)题项。OECD 社会与情感能力研究中的所有题项都采用 5 点计分的李克特量表, 从“非常不同意”到“非常同意”。每个题项都有多个选项。对于具有  $m_i$  个类别的题项  $i$ , GPCM 模型采用下面式子来定义个体在特定题项中选择某个选项的概率:

$$P(x_i|\theta, \beta_i, \alpha_i, \tau_i) = \frac{\exp \sum_{j=0}^x \alpha_i (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \sum_{j=0}^h \alpha_i (\theta_n - \delta_i + \tau_{ij})} x_i = 0, 1, \dots, m_i$$

此处,  $P(x_i|\theta, \beta_i, \alpha_i, \tau_i)$ 表示第  $n$  个被试在第  $i$  个题项上得到分数  $x_i$  的概率;  $\theta_n$  是第  $n$  个被试的潜在特质的估计值;  $\alpha_i$  是第  $i$  个题项的区分参数 (discrimination parameter);  $\delta_i$  是表示第  $i$  个题项在潜在连续统 (latent continuum) 上的位置参数 (location parameter);  $\tau_{ij}$  是一个表示第  $i$  个题项的第  $j$  个步骤的参数。

通过 GPCM 模型的估计并获得模型中所示参数的基础上, 可以计算能力初始值, OECD 社会与情感能力研究选择生成基于加权似然估计 (WLE) 算法的初始估计值 (Warm, 1989)。在上述 GPCM 模型的估计过程中需要使用综合权重 (senate weight), 学生综合权重依据学生最终权重 (WT2019) 计算得到, 而家长综合权重依据家长最终权重 (WT2019\_PA) 计算得到。通过综合权重的使用, 可以确保参与本次研究的所有城市都具有相同的代表性 (即每个城市加权后的测评人数都为 3000 人)。

### (3) 初始值的 ARS 校正

反应风格 (Response Style) 是指被试回答问卷时不以题目实际内容为依据的一种系统性行为倾向 (郭庆科, 2007)。比如, 对于同一类型的题目, 某一被试都倾向于选择“4-同意”, 而忽略了题目本身所包含的具体信息。反应风格的存在会引起共同方法偏差, 并造成系统的测量误差。对于反应风格造成的测量偏差, 可以采用的方法主要有计数法和模型法两大类 (张纛斌, 王烨晖, 2019)。OECD 社会与情感能力研究采用了正、反向计分题配对法, 从 15 项能力的量表表中挑选了 25 对正、反向计分题, 并用如下公式计算其默认肯定集 (acquiescence response sets, ARS):

$$ARS_n = \frac{1}{2} \left( \frac{\sum_{i=1}^{25} (x_{in}^p) + \sum_{i=1}^{25} (x_{in}^n)}{25} \right)$$

上式中,  $ARS_n$  表示被试  $n$  的默认肯定集;  $x_{in}^p$  表示被试  $n$  对于正向计分题  $i$  的回答;  $x_{in}^n$  表示被试  $n$  对于反向计分题  $i$  的回答。

利用上述公式得到每个被试的 ARS, 可以对 WLE 初始值进行默认反应校正 (也称为 ARS 校正), 以减少由于默认反应风格带来的测量偏误。ARS 校正利用 ARS 对 WLE 初始值的线性回归系数, 估计两者之间的线性关系, 再利用此种线性关系估计含有 ARS 偏误的 WLE 初始值分数, 并将其从 WLE 初始值中减去。具体地, 首先用线性回归模型估计 WLE 分数与 ARS 之间的线性关系, 获得参数  $a$ 、 $b$ , 也即:

$$WLE \text{ score} = a + b * ARS + \varepsilon$$

然后, 利用参数  $a$ 、 $b$  得到含有 ARS 偏误后的 WLE 分数, 也即:

$$WLE \text{ score}_{ARS} = a + b * ARS$$

最后, 将初始估计值减去含有 ARS 偏误的估计值, 得到 ARS 校正后的 WLE 估计值 (即校正值):

$$WLE \text{ score}_{Adjusted} = WLE \text{ score} - WLE \text{ score}_{ARS}$$

### (4) 校正值的标准化

WLE 初始值和 ARS 校正值都是分对数形式 (logit metric), 不便于用来直接呈现各项能力值, 因此需要进行标准化转换。与 PISA 类似, OECD 社会与情感能力研究选择将各项能力值转化为一个均值 500 分, 标准差 100 分的标准化分数。具体公式如下:

$$SS_n = 500 + 100 \frac{(WLE \text{ score}_{Adjusted}^n - \overline{WLE \text{ score}_{Adjusted}})}{\sigma_{WLE \text{ score}_{Adjusted}}}$$

上式中,  $SS_n$  表示个体  $n$  被标准化后的能力值;  $WLE \text{ score}_{Adjusted}^n$  表示个体  $n$  的校正值;  $\overline{WLE \text{ score}_{Adjusted}}$  表示 10 岁组校正值的均值;  $\sigma_{WLE \text{ score}_{Adjusted}}$  表示 10 岁组校正值的标准差。

在标准化的过程中, 10 岁组的均值和标准差被作为参照, 以确保均值 (500 分) 对应着 10 岁组的平均水平, 即超过 500 分表明高于 10 岁组的均值, 而低于 500 分则相反。值得注意的是, 对 10 岁组的均值和标准差的计算, 都需要使用学生综合权重, 以确保计算结果对所有城市都具有相同的代表性。

在经过上一步标准化之后,还需要进行一次中间值(mid-points)校正。中间值校正的目的在于,使得均值 500 分所对应的平均选项为中间选项(3-一般)。在所有被试中,有一类被试倾向于做出中间选择(mid-point),将这部分被试的平均能力值进行计算得到 $SS_m$ 。然后,从 500 中减去 $SS_m$ ,再将其差值加到上一步标准化能力值 $SS_n$ 中,就获得了如下式所示最终能力值:

$$Final\ Score_n = SS_n + (500 - SS_m)$$

### (5) 单项能力值的整体结果

表 20—21 中分别呈现了根据学生、家长、教师数据得到的各项能力值的均值和标准差。总体而言,在同一年龄组内,学生自评的能力值较高,而家长评价的能力值较低,尤其在 10 岁组阶段最为明显。不同年龄组相比较而言,学生自评的结果中,10 岁组在各项能力上的得分要普遍高于 15 岁组,但是基于家长和教师评价的结果,15 岁组在部分能力上的得分要高于 10 岁组。

表 20 单项能力的均值与标准差 (学生自评与家长间接评估)

|      | 学生自评 |     |      |     | 家长评估 |     |      |     |
|------|------|-----|------|-----|------|-----|------|-----|
|      | 10岁组 |     | 15岁组 |     | 10岁组 |     | 15岁组 |     |
|      | 均值   | 标准差 | 均值   | 标准差 | 均值   | 标准差 | 均值   | 标准差 |
| 责任感  | 642  | 116 | 580  | 87  | 566  | 88  | 589  | 94  |
| 毅力   | 651  | 109 | 571  | 77  | 563  | 82  | 582  | 81  |
| 自控力  | 613  | 101 | 572  | 76  | 547  | 82  | 575  | 78  |
| 抗压   | 525  | 113 | 466  | 87  | 551  | 99  | 555  | 94  |
| 乐观   | 625  | 114 | 558  | 84  | 596  | 83  | 588  | 80  |
| 情绪控制 | 596  | 125 | 536  | 82  | 543  | 93  | 562  | 94  |
| 共情   | 675  | 114 | 599  | 82  | 612  | 98  | 614  | 92  |
| 信任   | 653  | 124 | 575  | 86  | 637  | 99  | 628  | 94  |
| 合作   | 675  | 111 | 604  | 83  | 643  | 97  | 639  | 98  |
| 包容度  | 643  | 110 | 613  | 91  | 575  | 88  | 564  | 76  |
| 好奇心  | 645  | 101 | 574  | 75  | 606  | 90  | 588  | 82  |
| 创造性  | 636  | 111 | 567  | 87  | 583  | 89  | 568  | 80  |
| 乐群   | 637  | 107 | 572  | 81  | 607  | 89  | 585  | 79  |
| 果敢   | 536  | 84  | 512  | 74  | 542  | 79  | 529  | 75  |
| 活力   | 592  | 99  | 527  | 76  | 573  | 78  | 548  | 72  |

表 21 单项能力的均值与标准差 (教师间接评估)

|      | 10岁组 |      | 15岁组 |      |
|------|------|------|------|------|
|      | 均值   | 标准差  | 均值   | 标准差  |
| 责任感  | 2.62 | 0.77 | 2.68 | 0.58 |
| 毅力   | 2.60 | 0.90 | 2.72 | 0.66 |
| 自控力  | 2.42 | 0.73 | 2.48 | 0.56 |
| 抗压   | 2.25 | 0.69 | 2.10 | 0.68 |
| 乐观   | 2.76 | 0.67 | 2.66 | 0.57 |
| 情绪控制 | 2.69 | 0.75 | 2.61 | 0.65 |
| 共情   | 2.65 | 0.70 | 2.66 | 0.58 |
| 信任   | 2.72 | 0.61 | 2.64 | 0.53 |
| 合作   | 2.88 | 0.69 | 2.84 | 0.56 |
| 包容度  | 2.25 | 0.80 | 2.30 | 0.68 |
| 好奇心  | 2.73 | 0.79 | 2.69 | 0.63 |
| 创造性  | 2.45 | 0.82 | 2.49 | 0.66 |
| 乐群   | 2.77 | 0.68 | 2.69 | 0.60 |
| 果敢   | 2.11 | 0.85 | 2.17 | 0.71 |
| 活力   | 2.52 | 0.75 | 2.44 | 0.64 |

## 参考文献

- 郭庆科, 王昭, 韩丹, 时勘. (2007). 中国高中学生中的反应风格及其效应. *心理学报*, 1, 176—183.
- 王孟成. *潜变量建模与Mplus应用. 基础篇*. 重庆: 重庆大学出版社, 2014.
- 张纓斌, 王烨晖. (2019). 反应风格的测量与统计控制. *心理科学*, 42, 747—754.
- Adams, R., Wu, M., & Wilson, M. (2015). *ACER ConQuest: Generalised Item Response Modelling Software*, Australian Council for Educational Research.
- Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41—75.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Chen, F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464—504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233—255.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297—334.
- French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 13, 378—402.
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Buehner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cut-off values of fit indices. *Psychological Methods*, 16, 319—336.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1—55.
- Kankaraš, M., Feron, E., & Renbarger, R. (2019). *Assessing students' social and emotional skills through triangulation of assessment methods*. OECD Education Working Papers, No. 208, Paris: OECD Publishing.
- McDonald, R. (2013). *Test Theory*. New York: Psychology Press.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM Algorithm. *ETS Research Report Series*, 1—30.
- Muthén, L. K., & Muthén, B. O. (1998—2017). *Mplus User's Guide: 8th Edition*. Los Angeles, CA: Muthén & Muthén.
- OECD (2019). *TALIS 2018 Technical Report*. Paris: OECD Publishing.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507—514.
- Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427—450.

(责任编辑 王 森)

## The Technical Report on OECD Study on Social and Emotional Skills of Chinese Adolescence

Zhang Jing<sup>1</sup> Tang Yipeng<sup>2</sup> Guo Jiajun<sup>3</sup> Shao Zhifang<sup>3</sup>

(1. Department of Educational Psychology, Faculty of Education, East China Normal University, Shanghai 200062, China;

2. Department of Education, Faculty of Education, East China Normal University, Shanghai 200062, China;

3. The School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China)

**Abstract:** This technical report presents the analysis of Suzhou data in the SSES main study and the assessment of psychometric properties of items and scales that involve a series of iterative modelling and analysis steps. These steps included the calculation of Alpha and Omega to evaluate the subscales' reliabilities; the application of confirmatory factor analysis (CFA) to evaluate constructs; multiple-group confirmatory factor analysis (MGCFA) to review measurement equivalence across groups (age cohorts and gender groups); and using the Item Response Theory (IRT) and Generalised Partial Credit Model (GPCM) to scale items and generate scores for study participants.

**Keywords:** social and emotional skills; reliabilities; confirmatory factor analysis; measurement equivalence; item response theory