

情绪研究的新工具—图片型量表研究现状

刘 苹¹, 王亚男^{1,2}, 胡江柠¹

(1. 四川大学商学院, 成都 610000; 2. 西藏大学经济与管理学院, 拉萨 850000)

摘 要:文章主要对图片型量表的研究成果进行了文献综述,从情绪的内涵及其识别方法进行了概述。在回顾现有研究成果的基础上,从情绪类型、图片内容、表达形式和题项数量等方面对图片型量表的研究分类进行整理。总结归纳了图片型量表的开发条件及适用人群,梳理了图片型量表的开发及验证流程。最后,提出了目前图片型量表研究中存在的问题,指出了未来研究方向与前景。

关键词:情绪测量;图片型量表;文献综述

中图分类号:B841.2

文献标识码:A

文章编号:1003-5184(2024)01-0081-10

1 引言

情绪是人类的附属品,伴随人的一生,并通过各种有意识或无意识的感知影响人们的思想和行为。情绪测量对于理解人类行为至关重要(Pollak et al., 2011)。过去的数十年里,学者和实践者对情绪研究的热情不断提升,更加关注情绪状态如何以更加有效、更为可靠的方式进行测量,情感科学的最新发展产生了越来越多的可持续、自动评估人类情感状态的方法(Sonderegger et al., 2016),但其识别准确率在非实验情境下通常较低(Lietz et al., 2019)。结合应用实际,目前最常用的情绪测量方法仍是自我报告法(Reisenzein, 1994)。自我报告工具主要以文字量表为主,为满足特殊群体的研究需求,国外学者陆续开发了非语言工具——图片型量表(Minna et al., 2007)。

图片型量表是一种使用基于图像的元素承载条目意义的方法(Sauer et al., 2021)。作为词语量表的重要补充,图片型量表比文字量表更具吸引力(Baumgartner et al., 2019)。Desmet 等人(2016)认为相比于传统耗时较多的自报告方法,简单的图片型量表具有快捷、简单、可靠等特点,适用于更多研究场景。然而,有关图片型量表的研究相对较少,且主要表现为量表的开发和验证存在质量良莠不齐、开发验证流程不规范等问题。

基于此,本研究重点介绍情绪研究的新工具——图片型量表,在比较不同情绪识别方法的基础上,回顾图片型量表的现有研究,论述图片型量表的概念、特点、开发条件及适用人群。从情绪类型、图片内容、表达形式和题项数量等角度进行研究梳

理,对图片型量表的开发验证流程进行了文献综述。最终提出了目前图片型量表研究中存在的问题,并指出其研究方向。

2 情绪及其测量方法

2.1 情绪相关概念区分

情绪和情感是一个极为复杂的话题(Broekens & Brinkman, 2013),包含多个相似概念,如情感、情绪、心情、情绪特质等。现有研究虽然开发了大量的情绪测量工具,学者们也习惯性的将其统称为情绪测量工具,但其内涵仍有较大差异。基于此,首先对情绪的相关概念进行界定,具体差异见表1。

情感(Affect)是指意识到的、可以触及的感觉(Charles & Spence, 2019),情绪(Emotion)则是一种高强度、特殊的情感状态(Desmet et al., 2016)。与情感相比,情绪通常具有一个主体,停留时间相对短,并以离散情绪(如高兴、愤怒)形式展示。而情感则借助情绪维度(如愉悦度、唤醒度)进行表示。

心情(Mood)是一种低强度的、弥散的感觉状态(Desmet et al., 2016),是在没有明确的目标时所体验到的一种情感状态(Guevara & Eerola, 2018),可以持续数小时或数天,是对一系列累积事件的反映(Likamwa et al., 2012)。与情绪的差异主要体现在诱发主体、情绪强度、持续时间等方面。

情感特质(Emotion Traits)是一个人广义的情绪倾向,在多年内甚至整个生命周期内稳定不变(Mehrabian, 1996)。相比而言,情绪是生物体的暂时状态,其状态可能持续变化,也可能在一天中的不同时刻快速变化。

表 1 情绪相关概念区分

相似概念	目标	持续时间	强度	常见分类	参考文献
情感 Affect	可能无主体	一般	一般	愉悦度、唤醒度等情绪维度	Charles & Spence, 2019
情绪 Emotion	对某人/某事的情感状态	非常短	高强度	恐惧、厌恶、开心等离散情绪	Desmet et al. ,2016
心情 Mood	无直接情绪主体	比较长	低强度	积极情绪、消极情绪	Guevara & Eerola, 2018
情感特质 Emotion Traits	/	非常长	/	神经质、外倾性等	Mehrabian, 1996

2.2 情绪测量方法

情绪是个体为了应对环境变化而产生的适应性反应,与特定的面部表情、感觉、认知过程、心理变化、行为准备有关(Broekens & Brinkman, 2013),通过分析语音、面部表情、自我报告、生理或手机数据(Lietz et al. , 2019),可以识别和测量情绪。根据测量方法的不同,情绪测量工具可以分为自动识别法、自我报告法两类。其中,自动识别法通过测量被试的生理信号或外部行为,自主获取被试情绪信息(Broekens & Brinkman, 2013; Desmet et al. , 2016; Sonderegger et al. , 2016)。自我报告法要求被试依托一定工具或途径提供一系列情感信息,如语言、图片、动画、问卷等(Broekens & Brinkman, 2013),代表性工具为文字量表、图片型量表。

2.2.1 自动识别法

自动识别法受具身情绪理论(Embodied Emotion Theory)启发,认为情绪可由特定的生理或行为反应模式来表征(Desmet et al. , 2016)。根据其表征模式不同可以划分为生理指标测量法、外部行为测量法两类。

生理指标测量法是测量一种或多种与情绪相关的身体反应的方法(Minna et al. , 2007),由于人们的情绪感知经常反映在其生理信号上,如心率(ECG)、皮电(EDA)、肌电(EMG)、脑电(EEG)、呼吸(RESP)、脉搏(PPG)、皮温(SKT)等,而生理信号是个人真实情感的客观信息来源(Szwoch, 2020),不受人类主观意愿影响,研究者多借助传感器持续采集被试生理数据,记录被试情绪变化。Bradley 等人(2001)发现,不愉悦事件与皱眉肌的大量活动有关,而愉悦事件会促进颧肌的大量活动。Mcmanis 等(2001)基于实证研究发现皮电与唤醒度呈正相关关系。Li 等(2020)利用情绪视频的脑电反应测量个人特质。

外部行为测量法认为情绪可以由单独的表情和行为趋向所表示(Desmet et al. , 2016),根据行为表达方式的的不同,可以分为面部表情识别、语音表情识别、姿态表情识别三类(Bradley & Lang, 1994)。其

中面部表情识别是目前应用最广的外部行为测量方式,主要通过提取面部几何特征、外观特征,并引入时间维度进行静态和动态的图像识别。语音表情识别主要依靠情感语言或非文字的语音表达(Desmet et al. , 2016),通过分析用词、语气、音量、说话风格及其他细微差别识别情绪(Bradley et al. , 2001)。姿态表情识别则是通过身体姿态和移动来识别情绪(Desmet et al. , 2016)。

2.2.2 自我报告法

自我报告法是由受访者自己评定特定时间内某种情绪体验程度的方法(Stemmler, 1989),根据量表组成形式不同,可划分为文字量表、图片型量表。

(1)文字量表

文字量表是指被试以等级量表或形容词来表示自己情绪体验的自报告方式(周文莉 等, 2020)。目前,心理学领域的大多数量表都是多项量表,其原因在于多项量表可以减少测量误差,对潜在概念进行更为精确的测量(姜媛, 林崇德, 2010),如 PANAS 量表,通过 20 个形容词测量个体在过去一段时间的情感状态,其中 10 项代表正向情绪,如专注的、感兴趣的、警觉的、热情的,10 项代表负向情绪(Juergen et al. , 2018),如内疚的、敌对的、易怒的等。

尽管许多测量情绪体验的问卷都是基于语言的,但文字量表常被诟病(Sonderegger et al. , 2016)。一是翻译问题,成熟的文字量表多为国外研究,使用前会面临一定的翻译问题,翻译过程中可能导致含义丢失(Waston et al. , 1988),或无法在其他语言中找到与英语词语精准等同的词语(Obaid et al. , 2015)。二是认知影响问题,有学者指出多项量表要求被试投入更多的动机和努力(Juergen et al. , 2018),其认知过程可能干预最初的感受(Desmet et al. , 2016)。三是回忆偏差问题,多项量表多应用于研究前或研究后,然而情绪随时间显著变化(Pollak et al. , 2011),回顾性的评估方法容易产生回忆偏差和失真(Stone & Shiffman, 2002),单次测量结果不准。考虑到多项文字量表无法持续或快速重复观察被试情绪(Mendelsohn, 1989),学者们开始使用头

像、笑脸或类人模型等非文字工具测量情绪 (Sonderegger et al., 2016)。

(2) 图片型量表

图片型量表是一种利用图像元素传达项目意义的自报告工具 (Sauer et al., 2021), 主要应用于健康、医疗、人机交互、市场营销等领域。与文字量表相比, 图片型量表具有四个特点: 一是更少依赖语言, 图片型量表以图像为载体表达情绪意义, 避免了翻译的复杂性, 在不同文化间具有一定的可靠性 (Wong et al., 2021), 可应用于儿童和低语言技能参与者。二是测量方式更为直观, 更易理解 (Juergen et al., 2018), 测量过程中被试能够直观、准确地报告他们的情绪 (Sauer et al., 2021)。三是以简短量表为主, 对被试的认知要求低, 带来的脑力负荷相对

较少 (Wissmath et al., 2010), 测量过程不易影响测量结果, 故其准确率相对较高。四是测量简单, 易于填答, 可以提升参与动机, 激发兴趣 (Sauer et al., 2021), 具有接受度高, 退出率低等特点 (Marc & Gregor, 2020)。

回顾情绪测量方法, 本文发现每种类型的方法都有其优势和劣势, 通过表 2 可以发现, 自我报告法广泛应用于真实场景, 简单方便易于管理, 实施成本较低, 受测量方式影响, 被试需中断工作完成测量, 无法获得持续数据, 测量结果较为主观。自动识别法多局限于实验室研究, 由于测量过程中需要使用硬件设备, 应用成本相对较高, 但可以实现持续的实时观测, 测量结果相对客观。

表 2 不同情绪识别方法的比较

方法	适用场景	测量时间	成本	客观性	测量方式	干扰工作
生理指标测量	实验室	整个研究过程	非常高	非常客观	实时持续测量	否
外部行为测量	实验室/公共空间	整个研究过程	比较高	比较客观	实时持续测量	否
文字量表	真实场景	研究前、研究后	非常低	比较主观	非持续测量	是
图片型量表	真实场景	一日多次	非常低	一般	非持续测量	是

3 现有图片型量表简介

情绪是一种复杂的神经和心理现象 (Balsamo et al., 2020), 很难用文字表达, 特别是低强度情绪 (Minna et al., 2007; Desmet et al., 2016)。心理学和语言学的大量研究证实了某些情感表现与特定的视听信号之间的相关性 (Zeng et al., 2009)。Kensinger 和 Schacter 研究发现语言刺激会导致左侧杏仁核激活, 而图像刺激会导致右侧或双侧杏仁核激活, 故与文字刺激相比, 图片刺激可能导致自动加工, 而非精炼的、联想的或概念性的加工 (Kensinger & Schacter, 2006)。由于图像可能诱发一种本能反应 (Lang, 1995), 被试在进行图像情绪测试时形成了先于认知的决定 (Pollak et al., 2011), 运用图片、图画以及隐喻方式展示个人情感状态比语言的界定和描述效果更好 (Diem - Wille, 2001)。

最早用于情绪测量的图片型量表可以追溯到 Lang 开发的 SAM 量表。SAM 量表是一种非语言的图片测量技术, 借助人形漫画图片直接测量与各种情感反应相关的情绪维度, 在愉悦度维度人形漫画图片的面部表情从微笑、开心向皱眉、不开心变化。唤醒度维度人形漫画图片面部表情从睁大眼睛向闭上眼睛变化, 肢体动作表现为身体爆炸图标由大向小变化。支配度维度人形漫画图片从小向大变化 (Bradley & Lang, 1994)。

SAM 量表克服了文字量表的时间、精力、翻译书面语言能力等问题 (Obaid et al., 2015), 其信效

度可媲美语义差异量表 (Szwach, 2020), 研究者多用其快速评估由某一事件所带来的情绪变化。值得注意的是, 虽然 SAM 广泛应用于情绪研究, 但自开发以来该量表就受到了过于粗糙、过于简单、不美观等批评 (Sonderegger et al., 2016), 部分研究者也表示尽管 SAM 量表得到了大量验证, 仍有一个问题尚未得到解决, 即量表使用前需要向被试详细解释三个情绪维度 (Obaid et al., 2015), 这一过程可能会耗费更多的时间。基于此, 部分学者开始开发新的用于情绪测量的图片型量表。为进一步梳理图片型量表的现有研究情况, 本研究分别从情绪类型、图片内容、表现形式、题项数量等角度进行文献梳理 (具体可见表 3)。

3.1 情绪类型分类

图片型量表建立在离散情绪理论和情绪维度理论基础。离散情绪理论认为情绪在情感空间上是不连续的 (Juslin et al., 2013), 可以被分类或归类为一组情绪, 其中每一种情绪都可以与特定的可识别面部表情相关 (Riegel et al., 2016)。情绪维度理论则认为情绪以一种连续方式在情感空间变化 (Juslin et al., 2013), 使用愉悦度、唤醒度可以解释一系列情绪变化。

回顾用于情绪测量的图片型量表, 基于离散情绪理论开发的有 LEM、Sorémo、Ottawa Mood Scales 和 MAAC。其中 LEM 包含 8 个与人机交互有关的情绪, 即快乐、期望、着迷、满意、悲伤、厌恶、无聊、不

满(Huisman & Hout, 2020)。Sorémo 主要用于测量 4~13 岁儿童对在线学习平台的情绪反馈,涉及的情绪有满意、失望、高兴、考虑周到、生气、迷惑、启发、无聊、着迷(Girard, 2009)。Ottawa Mood Scales 侧重评估不同年龄段人群的多种心情,除高兴、悲伤、生气、担心、压力等离散情绪外,还测量了被试的自律水平(Wong et al., 2021)。MAAC 主要用于判断儿童焦虑障碍,量表由放松、无聊、疲惫、惊讶、悲伤、自责、害羞、生气、易怒、嫉妒、害怕、厌恶、高兴、得意、满意等 16 种情绪构成(Manassis et al., 2013)。

离散情绪分类直观,符合人们的日常经验,但由于其分类数量有限,无法涵盖所有的情感表现(Likamwa et al., 2012),大多数研究者使用情绪环模型展示更为真实的情绪状态(Lietz et al., 2019)。回顾现有图片型量表发现,基于情绪维度理论开发的图片型量表共有 13 个,其中仅包含愉悦度维度的有 Characterized face series(Kunin, 1998)、Smileyometer(Read, 2008),同时包含愉悦度、唤醒度的量表有 Russkman IM(Sánchez et al., 2006)、Affect Grid(Russell, 1989)、EmoCards(Carretié et al., 2019)、Feeltrace(Cowie et al., 2000)、PAM(Pollak et al., 2011)、Affective Slider(Betella & Verschure, 2016)、LE-PANAVA scale(Schreiber & Jenny, 2020)、AniSAM(Sonderegger et al., 2016)、AniAvatar(Sonderegger et al., 2016)。基于 PAD(Pleasure-Arousal-Dominance)模型开发的图片型量表有 SAM(Bradley & Lang, 1994)、AffectButton(Broekens & Brinkman, 2013)、Pick-A-Mood(Desmet et al., 2016)。

3.2 图片内容分类

依据图片内容不同,图片型量表可以划分为泛人型图片量表和其他图片量表两类。

3.2.1 泛人型图片量表

泛人型图片量表是指借助类人图片/图像的面部表情、肢体动作表达情绪含义的量表,根据展示主体不同又可以划分为仅有面部表情的图片型量表和同时包含面部表情、肢体动作的图片型量表。

面部表情图片量表受面部单元编码系统启发,代表性量表有 Russkman IM(Sánchez et al., 2006)、AffectButton(Broekens & Brinkman, 2013)、EmoCards(Carretié et al., 2019)、Smileyometer(Read, 2008)、Affective Slider(Betella & Verschure, 2016)、LE-PANAVA(Schreiber & Jenny, 2020)、Characterized face series(Kunin, 1998)。根据其展示方式不同,面部表情图片量表又可以分为漫画脸(如 Russkman IM)和现实绘画脸(如 EmoCards)两类,图片多无性

别差异,以中性形式呈现,除 Characterized face series 外,其他图片型量表均缺乏解剖学细节(如耳朵和头发)(Reynolds-Keefe & Johnson, 2011)。

同时使用面部表情和肢体动作表达情绪的图片型量表有 SAM(Bradley & Lang, 1994)、Pick-A-Mood(Desmet et al., 2016)、LEM(Huisman & Hout, 2020)、Sorémo(Girard & Johnson, 2009)、AniSAM(Sonderegger et al., 2016)、AniAvatar(Sonderegger et al., 2016)、MAAC(Manassis et al., 2013)。比较不同量表的图片内容发现,同时包含面部及肢体动作的量表拥有更明显的性别特征(如 LEM、AniAvatar 图片角色为男性, Sorémo、MAAC 图片角色为女性),并开始借助色彩强化情绪含义,如 Sorémo 使用红色代表消极情绪、橙色代表中性情绪、绿色代表积极情绪(Girard & Johnson, 2009)。

3.2.2 其他类型图片量表

由于大脑会对非拟人化的情感信号做出反应(Stone & Shiffman, 2002),除采用拟人或泛人形象表达情绪含义外,还有部分量表借助非拟人线索测量情绪,如 Russell 等基于情绪轮模型开发了 Affect Grid,要求被试使用 9*9 二维情绪空间(愉悦度、唤醒度)中的某一点表示其情绪(Russell, 1989)。Cowie 等在 Affect Grid 的基础上开发了记录语音情绪的 Feeltrace,并使用深浅不同的橙色、红色、蓝色、绿色依次代表积极-高唤醒、消极-高唤醒、消极-低唤醒、积极-低唤醒象限(Cowie et al., 2000)。Pollak 等(2011)开发了基于真实场景的图片型量表——PAM,测量过程中被试从 16 张不同类型的真实图片中选择 1 张最符合其当前情绪的图片完成情绪测试,从而实现了情绪的快速重复测量。Liu 等(2023)基于 PAM 量表,设计开发了 HDRPS,使用 4 张同类型的真实图片测量个体情绪效价。Wong 等(2021)开发了 Ottawa Mood Scales,综合使用面部表情和场景漫画两类素材测量被试情绪。

3.3 表达形式分类

依据量表表达形式,图片型量表可以划分为纯图片量表和混合图片量表两类,其中纯图片量表是指仅包含图片或李克特数字量表在内的图片量表,如 Russkman IM(Sánchez et al., 2006)、SAM 量表(Bradley & Lang, 1994)、AffectButton(Broekens & Brinkman, 2013)、Affective Slider(Betella & Verschure, 2016)、LFS(或 SLFS)(Obaid et al., 2015),混合图片量表是指同时包含图片和词语线索的图片型量表,代表性量表有 HDRPS(Liu et al., 2023)、EmoCards(Carretié et al., 2019)、Feeltrace(Cowie et al., 2000)、Smileyometer(Read, 2008)、LE-PANAVA

(Schreiber & Jenny, 2020)、Affect Grid (Russell, 1989)、Ottawa mood scales(Wong et al., 2021)。

3.4 题项数量分类

依据测试题目数量,图片型量表可以划分为单条目量表、多条目量表,观察表3可知,除 Affective

Slider (Betella & Verschure, 2016)、Ottawa Mood Scales(Wong et al., 2021)、LE – PANAVA(Schreiber & Jenny, 2020)、MAAC(Manassis et al., 2013)外,其余量表均为单条目量表。

表3 现有图片型量表名称

量表	图片内容	情绪类型	表达形式	题项数量
Russkman IM	泛人型 – 面部	情绪维度 – 2 维	纯图片	单条目
AffectButton	泛人型 – 面部	情绪维度 – 3 维	纯图片	单条目
EmoCards	泛人型 – 面部	情绪维度 – 2 维	混合	单条目
Smileyometer	泛人型 – 面部	情绪维度 – 1 维	混合	单条目
Affective Slider	泛人型 – 面部	情绪维度 – 2 维	纯图片	多条目
LE – PANAVA	泛人型 – 面部	情绪维度 – 2 维	混合	多条目
Charaterized face series	泛人型 – 面部	情绪维度 – 1 维	纯图片	单条目
SAM	泛人型 – 面部肢体	情绪维度 – 3 维	纯图片	单条目
Pick – A – Mood	泛人型 – 面部肢体	情绪维度 – 2 维	纯图片	单条目
LEM	泛人型 – 面部肢体	离散情绪	纯图片	单条目
Sorémo	泛人型 – 面部肢体	离散情绪	纯图片	单条目
AniSAM	泛人型 – 面部肢体	情绪维度 – 2 维	纯图片	单条目
AniAvatar	泛人型 – 面部肢体	情绪维度 – 2 维	纯图片	单条目
MAAC	泛人型 – 面部肢体	离散情绪	纯图片	多条目
Affect Grid	其他类型	情绪维度 – 2 维	混合	单条目
Feeltrace	其他类型	情绪维度 – 2 维	混合	单条目
PAM	其他类型	情绪维度 – 2 维	纯图片	单条目
Ottawa Mood Scales	其他类型	离散情绪	混合	多条目
HDRPS	其他类型	情绪维度 – 1 维	纯图片	单条目
LFS(或 SLFS)	泛人型 – 面部	离散情绪	纯图片	单条目

4 图片型量表的开发条件、适用人群

图片型量表为依赖语言的工具提供了一个便利选择(Cook et al., 2018)。有学者指出,相比于其他领域,情绪研究特别适合图片型量表,对比文字描述,使用绘画或类人模型表达情绪将更为容易(Sauer et al., 2021)。而图片本身具有普遍的情感可读性,可以提供更丰富、更吸引人的情感表现(Pollak et al., 2011)。

图片型量表开发假设人们在选择图片时投入的认知努力越少,选择的图片就越符合用户真实的内心情感状态(Desmet, 2002),结合 Sauer 等(2021)提出的图片型量表适用条件及相关文献,梳理用于情绪测量的图片型量表的开发条件及适用人群。

(1)被试理解能力有限

借助自我报告法进行情绪测试的首要条件是被试能够清晰、准确的理解问卷含义。文字量表使用过程中被试可能具有较差的阅读技能(如受教育水平较低),甚至不具有问卷语言水平(如非本土人)(Juergen et al., 2018)。高度象征性的图片与实际物体有更多感知特征线索,可以更好地唤起情绪反

应(Sauer et al., 2021)。故图片量表可用于目标语言能力不足或阅读能力较差的儿童或成人。

(2)需要多次测量

频繁测量对于行为科学、医学、计算机科学的研究具有重要价值(Pollak et al., 2011)。有学者指出,相较于文字量表,同一构念测试的次数越多,图片型量表越可以减弱其不足或提高其优势。原因在于重复测量同一构念时,图片作为符号可以更好、更快的进行信息编码(Bennett & Walters, 2001)。

(3)需要开发简短量表

简短量表不一定就是较差量表(Burisch, 1984),由于复杂量表填写过程中需要更多的认知需求,可能干扰被试的最初感受,Desmet 等(2016)认为使用简单词语或图片量表可以克服该问题。

5 图片型量表的开发流程

图片型量表要求设计简短,借助非文字形式建立与文字量表同等的效度(Juergen et al., 2018)。由于图片型量表开发时间相对较短,目前尚未形成标准的量表开发流程,用于情绪识别的图片型量表开发仅停留在确定漫画形象、情绪类型两个阶段。

相比而言,文字量表开发已经有了较为科学、完善的开发流程,考虑到并不是所有适用于文字量表的规则都适用于图片型量表(Juergen et al., 2018),故有必要梳理、明确图片型量表的开发流程。

回顾图片型量表开发的相关研究,近几年,部分国外学者开始关注图片型量表开发流程的规范性,如 Juergen (2018) 等将图片型量表开发流程划分为测试条目生成、专家评价、条目理解测试、信度检验以及用户反馈 5 个阶段。Baumgartner 等 (2019) 将图片型量表的开发概括为概念化、实现、验证。Sauer 等 (2021) 基于 Baumgartner 等人的研究提出了包含生成条目、解释检验、量表验证在内的图片型量表开发流程。结合上述研究及已有图片型量表开发实践,本文将图片型量表的开发流程划分为生成量表条目、确定图片型量表、图片型量表验证 3 个步骤。

5.1 生成量表条目

生成测试条目由 3 个子步骤构成。

(1) 确定图片型量表的主体内容及表现形式

依据量表所展示的图片内容,图片型量表可以划分为泛人型(仅包含面部表情,或同时包含面部及肢体动作)和其他类型(表格、纯图片)两类。回顾已有图片型量表可知,现有量表多通过绘制泛人漫画承载情绪意义,如 Schreiber 等使用表情符号开发 LE - PANAVA (Schreiber & Jenny, 2020)。Huisman 和 Hout (2020) 借助面部表情和身体语言描述情绪,并使用漫画形式提高情绪的可识别性。值得注意的是,在图片型量表开发过程中,除了需要确定具体展示内容及表现形式,还应考虑性别、年龄、种族等因素对量表适用性的影响。

(2) 确定量表测量的情绪类别

依据分类标准不同,情绪可以划分为离散情绪和情绪维度。为达到不同的研究目的,学者们采用不同的情绪分类方式开发量表。以 Sorémo 为例,为了解 4 ~ 13 岁儿童对在线学习平台的情感反馈, Girard 和 Johnson (2009) 要求儿童和成人分别列出教育过程中常出现的情绪词语,并将选择率较高者作为情绪类型。为帮助用户表达其即时情绪类型, Sánchez 等 (2006) 参考 Russell 提出的情绪轮理论,确定了 28 种离散情绪。Pollak 等 (2011) 和 Liu 等 (2023) 致力于借助真实图片表达被试情绪状态,选择情绪维度作为情绪标签。

除确定量表的情绪类型外,量表条目还需注意最佳情绪颗粒度的选择和确定。Desmet 等 (2016) 指出图片型量表不仅应该表现出最多的情绪,还需要足够少以保证快速、直接地进行自我报告,实现测量精度和测试速度之间的基本平衡 (Minna et al.,

2007)。

(3) 生成量表条目

依据确定的图片内容、表现形式及情绪类别绘制或采集量表条目时,需考虑色彩对条目含义的影响。色彩心理学研究表明色彩具有意义,关键颜色线索可以影响行为、认知和情感 (Laken et al., 2012)。考虑到色彩会唤起各种情绪感觉,如兴奋、平静 (Ou et al., 2004),学者们在绘制图片型量表时会控制或利用颜色与情绪之间的关系。如 Obaid 等 (2015) 为了避免乐高人偶肤色、脸部阴影对情绪表达的影响,在 LEGO Pictorial Scales (简称 LFS) 的基础上绘制了标准化的 LFS (即 SLFS)。Girard 和 Johnson (2009) 则分别使用红色和绿色表示消极和积极情绪。

5.2 确定图片型量表

图片型量表确定需要经历专家评价、迭代更新、被试测试 3 个环节。一是专家评价,为确保生成的测量条目具有充足的情绪含义,需邀请专家评价量表条目,并采集提升建议,通过计算条目的专家一致性得分,判断测量条目的内容效度 (Sauer et al., 2021);二是迭代更新,针对内容效度偏低的条目展开讨论,结合提升或改进建议进行迭代更新。设计完成后需重复进行专家评价,再次判断条目的内容效度,直至所有条目均达到较好的内容效度;三是被试测试,要求被试使用图片型量表进行测试,写出其对量表的理解,采集被试判断的依据及量表提升建议 (Baumgartner et al., 2019)。测试完成后,依据量表存在的问题及提升建议再次进行迭代更新和被试测试,直至被试测试准确率达到较高水平。

5.3 图片型量表验证

为确定图片型量表具有与文字量表一样的心理特征 (Sauer et al., 2021),图片型量表的开发和设计者们均在其研究中进行了量表验证。根据有效性验证内容不同,图片型量表验证可归纳为信度测试、效度测试两类。

(1) 信度测试

信度测试又称可靠性测试,是关于一致性的测量 (Broekens & Brinkman, 2013),通常通过计算量表的克隆巴哈系数进行测量 (Baumgartner et al., 2019),其最小接受值为 0.5,理想数值在 0.65 至 0.8 之间 (Cook et al., 2018)。由于用于情绪测量的图片型量表多为单条目量表,在量表验证过程中无法计算克隆巴哈系数,有学者在验证量表信度时,首先借助刺激材料(如 IAPS 图片)诱发被试情绪,继而进行图片型量表测量,将每位被试评分与总体评分之间的平均相关系数作为可靠性测试结果 (Broek-

ens & Brinkman, 2013), 其内部一致性标准参照 Loewenthal 和 Lewis (2001) 的相关研究。

(2) 效度测试

效度测试是指一个工具测量它想测量的内容的程度 (Cook et al., 2018), 是图片型量表最主要的验证方式, 包含同时效度、收敛效度、判别效度三类, 一般使用皮尔森相关进行检验。其中, 同时效度又称同期效度, 是指在同一时点, 新开发量表与既有同结构量表测量结果的相关程度 (Cook et al., 2018), 通常同时效度数值越大, 新开发量表的效果越好, 一般大于 0.45 即可证明同时效度显著 (Murphy & Davidshofer, 1998); 收敛效度又称聚合效度, 是指理论上相关的两种测量方法之间的关联度 (Broekens & Brinkman, 2013), 一般来说相关系数越大越好, 达到 0.5 则说明新开发量表具有良好的收敛效度 (Cohen, 1988); 判别效度是指理论上不相关的两个构念之间的关联程度, 与同时效度、收敛效度相反, 其相关系数越小越好, 参考 Cook 等 (2018) 提出的判别标准, 相关系数小于 0.45 则为合格效度, 小于 0.30 则表示良好效度。

6 前人研究局限与展望

6.1 现有研究的不足

(1) 国内的有关研究相对滞后

情绪测量对于情绪科学研究至关重要, 然而国内学者在图片型情绪测量工具的研发和应用方面相对滞后。梳理现有情绪研究工具可知, 国外学者已经开发了数十个用于情绪测量的图片型量表, 其中如 SAM (Bradley & Lang, 1994)、EmoCards (Carretié et al., 2019)、Affective Slider (Betella & Verschure, 2016) 等工具都已得到广泛的应用和检验, 可用于重复测量真实情境下被试的情绪变化。相比之下, 国内的许多研究除沿用国外较为成熟的图片型量表外 (如 SAM 量表), 仍很大程度上依赖于传统的文字量表, 或是现实情境下生态效度较为有限的自主识别方式, 如 Yan 等 (2019) 开发了一种基于脑电图的情绪识别模型, 文虹茜等 (2021) 融合表情和姿态识别情绪。

(2) 图片材料的选择和设计不够严谨

现有量表在图片的选择和设计上往往存在缺乏严谨性的问题。一方面, 尽管学者们在这些量表的开发过程中参考了面部单元编码系统 (Ekman & Friesen, 1978) 及具身情绪理论 (Niedenthal, 2007) 的相关知识, 但在量表开发的具体过程中, 时常疏于阐述这些图片或泛人漫画的选择和绘制依据, 导致其开发和制作显得相对随意, 可能难以精确地承载和传达微妙的情绪差异, 因而无法全面、准确反映真实

的个体情绪。另一方面, 有研究指出, 不同文化背景对情绪表情的解读存在差异 (Matsumoto, 1992), 可见文化差异也可能影响量表的普遍适用性。因此, 通过更加严谨科学的图片选择和设计步骤, 开发更为精细和文化中立的情绪测量工具显得尤为重要。

(3) 可测量情绪范围存在局限

现有图片型量表或依据离散情绪理论进行开发, 使用多种表情图片测量个体情绪 (如 LFS 或 SLFS), 或依据情绪维度理论进行设计, 通过效价、唤醒度、支配度进行情绪画像 (如 AffectButton), 情绪分类方法的不同使得现有图片型量表无法充分覆盖所有的情绪体验, 亦无法充分捕捉不同情绪的复杂性和独特性。

(4) 量表的开发验证尚不规范

当前的图片型量表在信效度方面与成熟的文字量表存在差距。由于发展时间较短, 图片型量表开发尚未形成规范的开发流程, 尤其是与 PANAS 等成熟的情绪量表相比, 这些图片型量表在开发过程中通常未采用严格的统计方法进行信度和效度的综合评估。许多图片型量表开发过程中仅就判别效度或聚合效度进行了有限的验证, 如 PAM (Pollak et al., 2011)。在这方面, 图片型量表的开发需要更加严格地遵循心理计量学的标准, 进行综合性的信效度测试, 并采用更大范围、更多样本的测试, 确保其在不同人群和文化背景下的适用性和准确性。

6.2 未来研究方向和前景

对于行为科学、医学、计算机科学研究来说, 频繁的情绪测量具有很大价值, 未来研究可将图片型量表作为其情绪研究的补充工具, 借助既有图片型量表进行情绪生态性测量, 也可以依据其特定的研究目的、严谨的开发流程开发图片型量表。除此之外, 针对既有研究较多采用泛人形象承载情绪意义, 存在展示方式不严谨、量表使用率不高等问题, 未来研究可以参考 PAM 量表的情绪展示方式, 借助真实图片等高度抽象载体传达情感含义, 探究非人形物品对情绪或情感的表达, 进一步丰富图片与情绪之间的关系研究。

具体而言, 图片型情绪量表未来的研究方向和前景可能包括以下几个方面:

(1) 跨文化研究

情绪表达具有显著的文化差异性。Ekman 和 Friesen 早在 1971 年就在其研究中指出, 不同文化背景下的人们在表达和识别情绪方面存在差异。因此, 未来研究需开发和验证适用于不同文化背景的图片型情绪量表。这要求进行大规模的跨文化比较研究, 并考虑到文化差异对情感认知和表达的影响。

(2) 多模态情绪评估

结合图片型情绪量表与其他类型的情绪评估工具,如语音、文字分析和生理信号监测,可以提供一个更全面的情绪评估系统。Galvão 等人(2021)认为从生理信号,尤其是脑电图(EEG)中识别情绪,是情绪测量中一个日益重要的领域。多模态方法能够从不同维度捕捉情感状态,这对于理解情绪的复杂性至关重要。这种综合方法不仅能提高评估的准确性,还能促进相关工具在临床和教育等更多领域的广泛应用。

(3) 实际应用的拓展

将图片型情绪量表应用于更广泛的实际场景,如临床心理治疗、教育环境、职场环境等,是一个具有重要实际意义的研究方向。例如,在临床环境中,图片型情绪量表可以用于评估和监测心理障碍患者的情绪状态,为心理治疗提供定量数据支持;在教育领域,这种工具可以帮助教师理解学生的情绪需求,促进学生的身心发展;再如在职场环境中,图片型情绪量表可以帮助企业和员工提高对情绪的实时了解和认知,从而以适当方法增强团队协作和工作满意度。

(4) 长期追踪与数据分析

采用图片型情绪量表进行长期的情绪监测和数据收集,可以为情绪障碍的预防和干预提供重要的数据支持。通过对大量个体的情绪数据进行长期追踪和分析,研究人员可以揭示情绪变化的模式和规律,为情绪健康管理提供科学依据。此外,长期数据的收集和分析还有助于改进情绪量表的准确性和适应性。

图片型量表是文字量表的重要补充,现有研究低估了图片型量表的潜力(Sauer et al., 2021),我们认为在不远的将来,除目前应用较多的健康、医学、人机交互、市场营销领域外,图片型量表还将继续从诸多方面发展完善并推广应用,帮助研究者深化其对情绪科学的理解和认识,在多个领域中进一步发挥重要作用。

参考文献

- 姜媛,林崇德.(2010). 情绪测量的自我报告法述评. 首都师范大学学报(社会科学版), (6), 135-139.
- 文虹茜,卿粼波,晋儒龙,等.(2021). 基于表情及姿态融合的情绪识别. 四川大学学报(自然科学版), 58(4), 87-92.
- 周文莉,顾远东,唐天真.(2020). 积极情绪对研发人员创新行为的影响:创造力效能感与工作卷入的中介作用. 科研管理, 41(8), 268-276.
- Balsamo, M., Carlucci, L., Padulo, C., Perfetti, B., & Fairfield, B. (2020). A bottom-up validation of the IAPS, GAPED, and NAPS affective picture databases: Differential effects on behavioral performance. *Frontiers in Psychology*, 11, 2187.
- Baumgartner, J., Frei, N., Kleinke, M., Sauer, J., & Sonderegger, A. (2019). Pictorial system usability scale (P-SUS). *Conference on Human Factors in Computing Systems - Proceedings*, 69, 1-11.
- Bennett, K. B., & Walters, B. (2001). Configural display design techniques considered at multiple levels of evaluation. *Human Factors*, 43(3), 415-434.
- Betella, A., & Verschure, P. F. M. J. (2016). The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS ONE*, 11(2), e0148037.
- Bradley, M., & Lang, P. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
- Bradley, M., Codispoti, M., Cuthbert, B., & Lang, P. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276-298.
- Broekens, J., & Brinkman, W. P. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies*, 71(6), 641-667.
- Burisch, M. (1984). You don't always get what you pay for: Measuring depression with short and simple versus long and sophisticated scales. *Journal of Research in Personality*, 18(1), 81-98.
- Carretié, L., Tapia, M., LópezMartín, S., & Albert, J. (2019). EmoMadrid: An emotional pictures database for affect research. *Motivation & Emotion*, 43(6), 929-939.
- Charles, S. (2019). Assessing the Role of Emotional Mediation in Explaining Crossmodal Correspondences Involving Musical Stimuli. *Multisensory Research*, 1-29.
- Cohen, J. (1988). The effect size index: d. *Statistical Power Analysis for the Behavioral Sciences*, 77-83.
- Cook, A. J., Roberts, D. A., Nelson, K. C., Clark, B. R., & Parker, B. E. (2018). Development of a pictorial scale for assessing functional interference with chronic pain: The Pictorial Pain Interference Questionnaire. *Journal of Pain Research*, 11, 1343-1354.
- Cowie, R., Douglas - Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schroder. (2000). Feeltrace: An instrument for recording perceived emotion in real time. *Proc. isca Wsh. on Speech & Emotion*, 9, 5-7.
- Desmet, P., Vastenburg, M., & Romero, N. (2016). Mood measurement with Pick - A - Mood: Review of current methods and design of a pictorial self-report scale. *Journal of Design Research*, 14(3), 241-279.
- Desmet, P. M. A. (2002). *Designing emotions* (Unpublished Doctoral dissertation). Delft University of Technology.
- Diem - Wille, G. (2001). *A therapeutic perspective: The use of*

- drawings in child psychoanalysis and social science* (pp. 119 – 133). London: SAGE Publications.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124 – 129.
- Ekman, P., & Friesen, E. (1978). Facial action coding system: A technique for the measurement of facial movement. *Palo Alto*, 3(2), 5.
- Galvão, F., Alarcão, S. M., & Fonseca, M. J. (2021). Predicting exact valence and arousal values from EEG. *Sensors*, 21(10), 3414.
- Girard, S., & Johnson, H. (2009). Developing affective educational software products: Sorémo, a new method for capturing emotional states. *Journal of Engineering Design*, 20(5), 493 – 510.
- Guevara, J. C., & Eerola, T. (2018). Music communicates affects, not basic emotions – a constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, 9, 1 – 19.
- Huisman, G., & Hout, M. V. (2020). The development of a graphical emotion measurement instrument using caricatured expressions: The LEM tool. *Emotion in HCI – Designing for People*, 259937797.
- Juergen, B., Andreas, S., & Juergen, S. (2018). No need to read: Developing a pictorial single – item scale for measuring perceived usability. *International Journal of Human – Computer Studies*, 122, 78 – 89.
- Juslin, P. N., Paquette, S., & Pfordresher, P. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4(2), 596.
- Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive Affective & Behavioral Neuroscience*, 6(2), 110 – 126.
- Kunin, T. (1998). The Construction of a New Type of Attitude Measure. *Personnel Psychology*, 51(4), 823 – 824.
- Lakens, D., Fockenberg, D., Lemmens, K., Ham, J., & Midden, C. (2012). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225 – 1246.
- Lang, P. (1995). The emotion probe: Studies of motivation and attention. *The American Psychologist*, 50(5), 372 – 385.
- Li, W., Hu, X., Long, X., Tang, L., Chen, J., Wang, F., & Zhang, D. (2020). EEG responses to emotional videos can quantitatively predict big – five personality traits. *Neurocomputing*, 415, 368 – 381.
- Lietz, R., Harraghy, M., Calderon, D., Brady, J., Becker, E., & Makedon, F. (2019). Survey of mood detection through various input modes. *The 12th ACM International Conference Proceeding Series*, 28 – 31.
- Likamwa, R., Liu, Y., & Lane, N. D. (2012). *Mood scope: Building a mood sensor from smartphone usage patterns*. Proceeding of the International Conference on Mobile Systems.
- 10.1145/2462456.2464449.
- Liu, P., Wang, Y. N., Hu, J. N., Qing, L. B., & Zhao, K. (2023). Development and validation of a highly dynamic and reusable picture – based scale: A new affective measurement tool. *Frontiers in Psychology*, 13, 1078691.
- Loewenthal, K., & Lewis, C. (2001). *An introduction to psychological tests and scales*. Psychology Press.
- Manassis, K., Mendlowitz, S., Dupuis, A., Kreindler, D., Lumsden, C., Monga, S., & Guberman, C. (2013). Mood assessment via animated characters: An instrument to access and evaluate emotions in young children. *Open Journal of Psychiatry*, 3(1), 149 – 157.
- Marc, S., & Gregor, J. J. (2020). Development and validation of the 'Lebender emoticon PANAVA' scale (LE – PANAVA) for digitally measuring positive and negative activation, and valence via emoticons – Science Direct. *Personality and Individual Differences*, 160, 109923.
- Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion*, 16, 363 – 368.
- Mcmanis, M., Bradley, M., Berg, W., Cuthbert, B., & Lang, P. (2001). Emotional reactions in children: Verbal, physiological, and behavioral responses to affective pictures. *Psychophysiology*, 38(2), 222 – 231.
- Mehrabian, A. (1996). Pleasure – Arousal – Dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261 – 292.
- Mendelsohn, G. A. (1989). Affect grid: A single – item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3), 493 – 502.
- Minna, I., Marika, T., et al. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile applications – sciencedirect. *International Journal of Human – Computer Studies*, 65(4), 404 – 418.
- Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological Testing: Principles and Applications*. Prentice Hall.
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316, 1002 – 1005.
- Obaid, M., Dünser, A., Moltchanova, E., Cummings, D., Wagner, J., & Bartneck, C. (2015). *LEGO Pictorial Scales for Assessing Affective Response* (pp. 263 – 280). Springer International Publishing.
- Ou, L., Luo, M., Woodcock, A., Wright, A., Woodcock, A., et al. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3), 232 – 240.
- Pollak, J. P., Adams, P., & Gay, G. (2011). PAM: A photographic affect meter for frequent, in situ measurement of affect. *ACM*, 7 – 12.
- Read, J. C. (2008). Validating the fun toolkit: An instrument for measuring children's opinions of technology. *Cognition, Tech-*

- nology & Work, 2(10), 119 – 128.
- Reisenzein, R. (1994). Pleasure – arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67(3), 525 – 539.
- Reynolds – Keefer, L., & Johnson, R. (2011). Is a picture is worth a thousand words? Creating effective questionnaires with pictures. *Practical Assessment, Research & Evaluation*, 16(8), 1 – 7.
- Riegel, M., Żurawski, W. M., Moslehi, A., Klocek, H. M., & Marchewka, A. (2016). Erratum to: Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). *Behavior Research Methods*, 48(2), 613.
- Russell, J. A. (1989). Affect grid: A single – item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3), 493 – 502.
- Sánchez, A., Norma, P., Hernández, P. J. C., et al. (2006). Conveying mood and emotion in instant messaging by using a two – dimensional model for affective states. *Human Factors in Computing Systems*, 19 – 22.
- Sauer, J., Baumgartner, J., Frei, N., & Sonderegger, A. (2021). Pictorial scales in research and practice: A review. *European Psychologist*, 26(2), 112 – 130.
- Schreiber, M., & Jenny, G. J. (2020). Development and validation of the ‘Lebender emoticon PANAVA’ scale (LE – PANAVA) for digitally measuring positive and negative activation, and valence via emoticons. *Personality and Individual Differences*, 160, 109923.
- Sonderegger, A., Heyden, K., Chavallaz, A., & Sauer, J. (2016). AniSAM & AniAvatar – Animated visualizations of affective states. *ACM*, 4828 – 4837.
- Stemmler, G. (1989). The autonomic differentiation of emotions revisited: Convergent and discriminant validation. *Psychophysiology*, 26(6), 617 – 632.
- Stone, A., & Shiffman, S. (2002). Capturing momentary, self – report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24(3), 236 – 243.
- Szwoch, M. (2020). A Review of Emotion Recognition Methods Based on Data Acquired via Smartphone Sensors. *Sensors*, 20, 6367.
- Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063 – 1070.
- Wissmath, B., Weibel, D., & Mast, F. W. (2010). Measuring presence with verbal versus pictorial scales: A comparison between online – and ex post – ratings. *Virtual Reality*, 14, 43 – 53.
- Wong, M. Y., Paul, E. C., Chen, K. L., & Poh, F. L. (2021). Validation of Pictorial Mood Assessment with Ottawa Mood Scales and the Positive and Negative Affect Scale for Young Adults. *Community Mental Health Journal*, 57(2), 529 – 539.
- Yan, J., Chen, S., & Deng, S. (2019). A EEG – based emotion recognition model with rhythm and time characteristics. *Brain Informatics*, 6(1), 1 – 8.
- Zeng, Z., Pantic, M., Roisman, G. I., et al. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 31(1), 39 – 58.

A New Tool for Emotion Research—Research Status of Pictorial Scale

Liu Ping¹, Wang Ya’nan^{1,2}, Hu Jiangning¹

(1. Business School, Sichuan University, Chengdu 610000;

2. School of Economics and Management, Tibet University, Lhasa 850000)

Abstract: This paper mainly summarizes the research results of the pictorial scale, and summarizes the connotation of emotion and its recognition methods. On the basis of reviewing the existing research results, this paper sorts out the research classification of the pictorial scale from the aspects of emotion type, picture content, expression form and the number of items. It summarizes the development conditions and applicable groups of the pictorial scale, and combs the development and verification process of the pictorial scale. Finally, the problems in the current study of the picture scale are put forward, and the future research direction and prospects are pointed out.

Key words: emotion measurement; pictorial scale; literature review