

## 有关创造力测量的一些思考\*

贡 喆<sup>1</sup> 刘 昌<sup>1</sup> 沈汪兵<sup>2,1</sup>

(<sup>1</sup> 南京师范大学心理学院暨认知神经科学实验室, 南京 210097)

(<sup>2</sup> 河海大学公共管理学院暨应用心理研究所, 南京 210098)

**摘 要** 创造力测量是创造力研究的基础, 然而该领域研究成果却一直饱受质疑, 因此如何准确测量创造力是研究者亟需解决的问题。近几年, 创造力测量领域围绕一些研究热点, 在多方面取得显著进展。例如: 发散思维测验独特性维度的计分问题, 或许可以通过主观计分法解决; 顿悟类测验可能表征个体创造力水平, 但效度仍有待于进一步确认; 创造力成就测验可能带来的共同方法变异问题, 需要通过合理应用测验规避; 同感评估技术或许会引起评定者效应; 研究者开始从语义网络角度测量创造力等等。未来该领域研究应当在: 基本概念问题上达成共识; 从测验内容和施测过程优化测验质量; 采用混合测验的策略以及通过跨领域研究增进测量技术多样化等方面进行努力。

**关键词** 创造力; 创造力测量; 创造力测验; 发散思维测验

**分类号** B842

### 1 引言

创造力概念一般被认为包括两大核心特质: 新颖性(originality)和适宜性(appropriateness)。其中, 新颖性意味着: 创造性思维和产品应相对于创造者自身或创新环境而言具有新颖、独特的品质; 适宜性则可以归纳为: 创造性思维和产品对创造者自身或问题情境具备“有用性”(useful)或是对创新环境“有价值”(valuable)。这两大核心特质是创造力与其它相近概念的根本不同, 亦是概念上判别创造力的基本标准。依托于此类创造力基本概念, 研究者进一步对创造力概念展开量化研究, 其中首要的课题即是测量创造力。创造力测量, 即依据一定的创造力理论, 使用测验对创造力进行定量描述的过程。研究者一般认为, 1883 年高尔顿发表的《对人类能力的探求》(Inquiries into Human Faculty)唤起了人们测量创造力的兴趣, 直到 1950 年吉尔福特在美国心理学

会上的演讲后, 创造力测量领域的研究开始进入高峰期(Plucker & Renzulli, 1999)。这表现在, 一方面, 创造力测量成为创造力研究的方法学基础, 它为进一步研究创造力提供有力支持。一项研究报告表明, 各种创造力量表和问卷是创造力研究中收集信息的第一选择, 远远多于访谈法或文献分析(Hürsen, Kaplan, & Özdal, 2014); 另一方面, 各种创造力测量技术蓬勃发展, 研究者围绕创造力主体(Domino, 1970; Gough, 1979; Kirton, 1976)、创造性思维过程(Guilford, 1967; Mednick, 1962; Torrance, 1972)、创造性产品(Amabile, 1982; Besemer & O'Quin, 1987)以及创造性环境(Amabile, Conti, Coon, Lazenby, & Herron, 1996)四个方面对创造力展开测量研究, 创造力测验从量和质上都获得飞跃式发展。Torrance 和 Goff(1989)在 20 余年前的报告中提到, 在当时学术界已经有不少于 255 种创造力测验。虽然暂时未能发现较新的统计数据, 但可以想见的是, 创造力测验的数量早已远远超过 Torrance 和 Goff(1989)当年的估算。现如今, 创造力测验已经应用于社会生活的各个领域, 例如: 数学(Tan, Mourgues, Bolden, & Grigorenko, 2014); 机械(Colangelo, Kerr, Hallowell, Huesman, & Gaeth, 1992); 漫画(Chou, Chen, & Chou,

收稿日期: 2015-04-27

\* 江苏省第四期“333 高层次人才培养工程”科研项目; 国家自然科学基金(31500870); 中央高校基本业务费项目(2014B15314); 博士学术新人培育(1812000002120)

通讯作者: 刘昌, E-mail: liuchang@njnu.edu.cn

2014); 就业市场(Paprika & Nagy, 2012); 网页设计(Zeng, Proctor & Salvendy, 2012); 科学研究(Ayas & Sak, 2014); 教育实践(Blamires & Peterson, 2014); 工程规划(Charyton & Merrill, 2009)等等。

然而, 创造力测量领域, 特别是各种创造力测验在近几十年的发展中是否交出了一幅满意的答卷呢? 我们或许要先打上问号。研究者普遍认为, 创造力测量领域相较于创造力其他研究领域发展更为滞后(Plucker & Makel, 2010; Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012)。作为心理测量领域和创造力研究领域的交集, 创造力测验一些关键的测量指标饱受诟病。例如: Simonton (2003)曾经激烈抨击现有创造力测验: “那些推荐的创造力测量手段没有一个能跨过能力测验所必需迈过的心理测量障碍。例如: 创造力各种分测验的分数与一般智力相关太高, 这意味着区分效度(divergent validity)低下; 而这些分测验之间相关却又太低, 这又说明聚合效度(convergent validity)低下; 创造力测验分数和客观的创造力行为指标相关很弱, 这还代表预测效度(predictive validity)低下。”在诸多讨伐声中, 创造力测验缺乏预测效度或许是最为严重的问题(Kaufman & Baer, 2012; Zeng, Proctor, & Salvendy, 2011), 这甚至可能意味着许多创造力测验事实上是无效的。

既然如此, 那么如何更准确的测量创造力? 为了回答这个问题, 研究者从未停止努力脚步, 他们围绕创造力测量的热点问题进行研究, 提出了多种改进方案。例如: 创新测验方法(Prabhakaran, Green, & Gray, 2014); 丰富计分手段(Benedek, Mühlmann, Jauk, & Neubauer, 2013); 完善统计措施(Primi, 2014); 评估测验合理性(Lee, Huggins, & Therriault, 2014)等等。这些创造力测量技术的新进展, 大大推进了创造力测量的客观性与实用性。鉴于国外有关创造力测量技术的权威综述发表较早, 且内容类似(见 Plucker & Renzulli, 1999; Plucker & Makel, 2010), 国内更缺乏与之相关的前沿介绍, 因此极有必要对近几年创造力测量的最新进展进行综述。

需要说明的问题是: 第一, 由于近几年创造性思维过程测验和创造性产品测验的应用更为广泛深入, 因此本文主要围绕这两个角度选取运用最多、争议最大的4类主流测量技术进行综述, 其中既包括创造力测验, 即: 发散思维测验、顿悟

类测验、创造力成就测验; 也包括评分技术, 即: 同感评估技术。同时值得注意的是, 从严格意义上看, 这4类主流测量技术并不能被认为只隶属于创造性思维测验或创造性产品测验的范畴, 它们在实际应用中往往较为灵活。例如: 发散思维测验一般可归类为创造性思维测验, 但是发散思维的成果也可视为创造性的产品, 研究者是根据思维的产物评定其是否有创造性的。因此如果从这个角度出发, 把发散思维测验归于测量创造性产品的测验亦无不可。又如: 同感评估技术作为一种评分技术, 既可应用于多种发散思维测验, 亦可运用于某类产品的创造性评估乃至创造性环境的测评。第二, 本文无意于细致罗列各种创造力测验的具体操作方法, 该部分内容可见由 Kaufman, Plucker 和 Baer (2008)编著的《Essential of Creativity Assessment》。

下文依次对4类主流创造力测量技术的最新进展及热点问题进行评述, 最后展望创造力测量未来的发展方向。

## 2 发散思维测验

发散思维测验(divergent thinking test)一般以开放性问题的形式呈现, 要求被试尽可能多的根据题目要求罗列答案。发散思维测验有多种分类方法, 例如, 从测验形式将其分为: 语言测验、图形测验以及动作测验; 或从测验内容将其分为: 灵活使用任务或多用途任务(alternate use task) (给普通物品设计多种新颖用法), 举例任务(instance task) (在普通种类中例举新颖物品)以及结果任务(consequence task) (设想某种虚拟条件下可能产生的新颖结果) (Silvia, 2011); 又或从领域一般性及特殊性(domain general-specific)角度将其区分为: 领域一般性发散思维测验, 或领域特殊性发散思维测验。发散思维测验适用于个体或团体施测, 且普及各个年龄段, 是应用最为广泛的创造力测验形式。使用频率较高的发散思维测验是: Guilford (1967)的智力结构测验(Structure of the Intellect, SOI); Torrance (1972)的托兰斯创造力测验(Torrance Tests of Creative Thinking, TTCT)等等。

发散思维测验的基本理论假设是: 创造力水平高的个体相较于普通人, 最显著的特点在于他们能够从多角度思考并解决问题。因此, 通过开放性问题评定个体发散思维水平, 就能够预测个

体创造力潜能。然而,该假设并未准确切中创造性思维的两大核心特质,即新颖性和适宜性。首先,个体能够对某问题从多角度进行思考和解答,未必是创造性思维新颖性的体现,反映的可能只是个体在长时记忆中存储的与该问题相关的信息数量(Leon, Altmann, Abrams, Gonzalez Rothi, & Heilman, 2014)。同样,个体从多角度思考问题也不意味其思维具有适宜性。理论层面的缺失可能进一步给实证研究带来不利影响。例如:研究者在采用发散思维测验时就经常忽视适宜性维度,缺乏适宜性的限制,一些不合逻辑的“新颖”想法就会和创造性思维混为一谈。当该类测验应用于特殊群体时,问题可能会集中出现。Fink, Slamar-Halbedl, Unterrainer 和 Weiss (2012)的研究发现创造力与精神质(psychoticism)存在显著正相关,原因可能在于他们在应用发散思维测验时并没有考虑适宜性维度,这就会把一些明显不合理的答案判定为高新颖性,最终可能导致高估高精神质个体的创造力水平。虽然该问题能够通过改良新颖性计分的方式弥补(Silvia et al., 2008),但由此可见,发散思维测验基本理论假设存在的不足,可能引发一系列应用问题,这需要引起研究者的足够重视。

发散思维测验的应用最为广泛,在某种程度上或许可称之为创造力测验的“代言人”。然而,发散思维测验在获得更多关注的同时也遭到更多指责。发散思维测验的客观性、信度以及效度问题都饱受争议(Baer, 2011; Benedek, Könen, & Neubauer, 2012; Zeng et al., 2011)。Zeng 等人(2011)曾罗列了发散思维测验的六大“罪状”,主要包括:缺乏结构效度、未能整合创造性思维过程、对领域特殊性 & 专业知识的忽视、预测能力较弱以及缺乏生态效度和区分效度。Baer (2011)还曾在美国心理协会会议上呼吁尽量避免使用诸如“易拉罐不同寻常的用法”之类的发散思维测验,他认为这种测验根本不是创造性思维合理的测量方式,继续使用它们只会令创造力研究步入泥沼。

但同时也有大量研究表明发散思维具备良好的信、效度(Beaty, Smeekens, Silvia, Hodges, & Kane, 2013; Hong, Milgram, & Gorsky, 1995; Kim, 2008)。由于发散思维测验信、效度指标,特别是效度指标积累了大量矛盾结论,因此,目前仍然无法确定发散思维测验是否“完全”可信且有效,

围绕此问题的争论势必一直存在。不过于此同时,研究者开始逐渐意识到发散思维测验只是针对创造力潜能的一种估计(Runco & Acar, 2012),潜能的发挥除了依靠能力本身,还需要知识、技能、心理健康乃至机遇(Cropley, 2000)。不仅如此,越来越多研究者倾向于支持创造力具有领域特殊性(Baer, 2011; Pretz & McCollum, 2014),简言之即创造力在不同领域具有差异化表现,创造力测验亦是如此。因此,一种更为合理的推测是:发散思维测验能够在一定程度上有效评定和预测创造力水平,但有赖于不同领域。例如:Clapham, Cowdery, King 和 Montang (2005)的研究就发现,智力结构测验(SOI)中与机械能力有关的分量表,比智力结构测验(SOI)测验总分对创造力行为和专利数量有更好的预测作用;Runco, Millar, Acar 和 Cramond (2010)一项针对托兰斯创造力测验(TTCT)长达 50 年的纵向研究表明,托兰斯创造力测验(TTCT)与个体一些创造力行为有中等相关,但和个体社会层面的创造性成就无关。据此可见,从整体上判定发散思维测验是否有效实际上并不妥当,从不同角度出发,归纳多测验在多领域中的表现,分别总结其信、效度数据或许是更为合理的方法。

对发散思维测验计分方法的争论与改进是近年来创造力测量领域的热点问题,也是发散思维测验研究的重要突破。发散思维测验通常包括对被试答案流畅性(fluency)、变通性(flexibility)以及独特性(uniqueness)或新颖性(originality)的计分。其中,由于发散思维测验中多将独特性维度和新颖性维度视为等同(Benedek et al., 2013),因此本文统一采用“独特性”表征该维度。学术界对流畅性和变通性的计分较为统一,前者即清点被试答案的个数,后者即计算被试答案囊括的不同种类。而发散思维测验独特性维度的计分方法则一直存有争议。较早的经典计分方法为:研究者将所有答案汇总,如果某个答案是特殊的,即唯一的,那么该答案计 1 分,其他答案记 0 分。该计分方法直观明了,较好的反映了创造力“独特性”的含义,得到了一定程度上的运用。其他独特性计分方法还有:列出一张非独特性答案对照表,如果被试答案不在该表上,即可以在独特性维度上计分;还有方法给样本中少于 5% 的答案记 1 分,给其它答案记 0 分。这一类方法采用一种计分模式,即被试凭借独特性答案得分,分数最后相加

成为独特性指标(Silvia et al., 2008)。

然而此类方法的问题在于：第一，流畅性和独特性分数间相关过高，二者的含义会因此发生混淆(Clark & Mirels, 1970)。Silvia (2008)一项潜变量分析研究就发现，潜在流畅性和独特性变量相关极高( $r=0.88$ )。依照研究者对流畅性和独特性的定义，如果说流畅性是发散思维量的指标，那么独特性即为发散思维质的指标，质需要具有其相对独立性。但既有研究结果却揭示二者间存在不可忽视的高相关，这导致独特性分数难以越过流畅性分数的影响对其他变量进行解释(Silvia et al., 2008)，因此有研究者甚至称流畅性为污染变量(Hocevar, 1979; Kaufman et al., 2008)。第二：独特性分数极易受到样本量误差的干扰。一个可以预见的情况是，在一项发散思维测验里，如果实验只有30个样本，那么被试的很多答案都可能被认定为“特殊”，但当样本扩充到300个乃至更多时，原先“特殊”的答案可能将只是一个普通答案。当样本数量极大时，甚至可能出现没有独特性答案的现象(Nusbaum & Silvia, 2011)。第三，对答案独特性的判定需要同时考量其是否合理，因为创造力还内在具有适宜性的含义。依照传统计分方法，一些古怪、奇特的答案很可能被纳入独特性得分的范畴，例如：砖头的用途，某个答案是“吃”。该答案诚然非常特殊，但这显然并不合理，因此不当属于创造力的范畴。

Silvia, Martin 和 Nusbaum (2009)认为可以借助主观计分法解决传统独特性计分方法的问题，他们要求被试在限时发散思维任务中给出有创意的答案，并要求评定者依照不寻常、远距离以及聪明三个指标，对被试给出的答案在5点量表上打分(Silvia et al., 2008)。采取主观计分法的优点在于，第一：流畅性和独特性指标间不再混淆。多项研究表明采用主观计分法后流畅性和独特性相关程度显著减小(Benedek et al., 2013; Primi, 2014; Silvia et al., 2008; Silvia, Nusbaum, Berg, Martin, & O'Connor, 2009)。第二，独特性水平直接由评定者判定，因此独特性分数不再受到样本量的干扰。第三，一些明显不符合创造力的古怪答案在主观计分中可以进行更为准确的判断并加以剔除。

发散思维测验的主观计分方法主要包括：平均得分法(average scoring)和最优项法(Top scoring)。

平均得分法即对被试某一道题目多个答案的得分进行平均，得到其创造力的整体评价，用商数代替求和可以较好的规避流畅性和独特性计分的混淆问题。最优项法即要求被试在回答完每道问题后自行圈出其认为最有创造力的几个答案(答案的数量由实验者确定，另根据 Benedek 等人(2013)的研究表明，选择3~6个最优答案测验效度最佳)，评定者只对这几个答案进行评价，评价后求平均数以表征其创造力。Silvia 等人(2008)的研究表明，当采用传统计分方法时，数据信度表现最差，只有使用不低于15个测验任务时，测验的可靠性指数才能达到0.80水平。但通常情况下，研究者在实验中使用的测验任务不会超过4个；而平均得分法和最优项法在信度上都表现较好；在效度上最优项法比平均得分法更佳，其解释力为平均得分法的两倍，原因可能在于最优项法省略了许多缺乏创意的答案，因此可以传递更多有价值的信息。Benedek 等人(2013)认为，平均得分法由于对被试的所有答案进行平均，可能导致一些关键信息的丢失。例如：某高创造性个体在5个答案中有两个是特别有创意的答案，其余3个答案只是随意写下的，其创造力平均得分将会是居中水平；而某个普通创造性个体写下的5个答案可能既没有富有创意的答案，但同时也没有随意写下的答案，因此其创造力平均得分同样是居中。最优项法突出了其最具创意的答案，省略了相对次要的内容，因此在实际操作中，最优项法保留了被试创造力信息中更有代表性的部分。虽然也有研究表明(Plucker, Qian, & Wang, 2011)平均得分法在几类主观计分方法中表现最好，但原因可能在于 Plucker 等人(2011)的研究中未要求被试自行圈出最有创意的答案，被试前10个答案或后10个答案不等于其最优答案。总之，究竟哪一种方法更为合适，仍有待于进一步研究佐证。

除此之外，为进一步提高主观计分法的效率，Silvia 等(2009)还创制快照评定法(snapshot scoring)对发散思维测验进行计分。不同于一般主观计分方法，快照评定法只需要评定者对被试某道题目的答案进行整体评判，而不需要对被试的每一个答案进行打分，这无疑能节省大量的时间，且根据现有研究成果，快照评定法具备较好的可操作性与可信度，得到了不少研究者的采纳。但同时需要注意的是，根据 Silvia 等(2009)的研究报

告表明,快照评定法相较于一般主观评定方法可能会令测验结果在一定程度上受损。

以上研究虽然倾向于支持主观计分方法,但同时也有研究者提出质疑, Lee (2008)就对 Silvia 等人(2008)的研究报告提出了批评,他认为主观评价法实际上无法真正避免传统计分方法的弊端。例如:最优项法要求被试自行选取最有创意的答案,但当要求被试写出更多答案的时候,出现更特殊答案的可能性同样提高了,此时流畅性和独特性并未真正分离。不仅如此,主观自评的计分方式还会带来更多问题。例如:最优项法容易导致积极偏见(positive bias),即通过该计分方法得到创造力分数可能远高于其真实水平,在效标参照测验中这将会带入系统误差。Primi (2014)也指出,当采用主观计分法时,评定者的标准可能各有高低,当一组标准较高的评定者对被试进行打分时,被试的得分将系统性降低,这势必会引入误差。因此, Primi (2014)认为,应当通过统计手段评估主观计分法可能带来的误差,如借助项目反应理论(item response theory)中的罗舒模型(Rasch model)对数据实现标准化处理,以便比较评定者之间标准的高低。

总而言之,尽管发散思维测验在信、效度问题上承受了很大责难,但是学术界对发散思维测验的革新远未停止。因此研究者依然对发散思维测验抱有极大信心(Kim, 2011; Runco & Acar, 2012)。可以说,迄今为止,发散思维测验仍然是测量创造力的第一选择。

### 3 顿悟类测验

发散思维是由意识引导,解决界定清晰问题的创造性思维过程,执行功能对发散思维起到关键作用。与之相反,创造性思维还具备通过无意识自动联想,解决界定模糊任务的能力(Mednick, 1962)。某问题界定模糊意味着,解决该问题的方法和途径不明确,此时个体需要借助一系列认知加工实现对问题的解决,并继而产生顿悟的“啊哈”体验。研究者围绕该创造性思维设计了一系列创造力问题和测验,本文将这一类问题和测验概述为顿悟类测验。顿悟类测验的基本理论假设是:高创造性个体能够凭借诸如:远距离联想、思维重组、原型启发等认知加工过程整合并处理与目标有关的信息,从而更好的解决顿悟问题,并产

生顿悟体验。该假设逻辑与创造性思维两大核心特质新颖性与适宜性的结合较好,不难想象,能够解答某顿悟问题的答案和策略,势必是较为新颖且适用于该问题情境的。其中应用最广的是 Mednick (1962)的远距离联想测验(remote associates test, RAT)。

Mednick (1962)认为创造性思维的本质是联想。他提出创造力包含“将关联的要素结合以满足某种特定需求或使之有用”的过程,且个体间创造力的差异往往表现在操作联想思维过程上的不同,因此, Mednick (1962)假设个体有联想等级(associative hierarchy)的差异。具体表现为:创造力高的个体能够连接距离更遥远的要素,据此理论假设, Mednick (1962)创立了影响深远的远距离联想测验。远距离联想测验一般由 30 个项目构成,每个项目包括三个词汇,要求被试根据所呈现的三个词汇进行联想,填入与之相关的新词。以英文版远距离联想测验为例,如:光(light),生日(birthday),蜡(wax),正确答案应为:蜡烛(candle)。远距离联想测验在过去几十年中于创造力、联想、顿悟乃至记忆、精神病理学等研究领域得到了广泛运用(Aiello, Jarosz, Cushen, & Wiley, 2012; Fodor, 1999; Storm, Angello, & Bjork, 2011),我国也有系统介绍其使用方法的综述出现(王烨, 余荣军, 周晓林, 2005),同时中文版远距离联想测验的修订工作也有所进展(李良敏, 罗玲玲, 刘武, 2015)。

相较于发散思维测验,远距离联想测验最鲜明的特点是,测验中的题目具备唯一的正确答案。如上文所述,发散思维测验要求被试针对某一问题进行多角度的回答,被试答案的流畅性、变通性是极其重要的考量指标。远距离联想测验并不接受多种可能的答案,因为题目中已经暗含了唯一且最佳的答案。然而,远距离联想测验这一特性引发的争执随之产生。早已经有研究指出,具有“唯一解”的远距离联想测验实际上与智力测验的结构更为类似,且远距离测验与发散思维测验相关程度很低(Laughlin, Doherty, & Dunn, 1968; Taft & Rossiter, 1967),因此远距离联想测验是否能反映个体的创造力水平?为阐明这一问题,有研究者通过分析远距离联想测验的内、外部结构效度以明确该测验真实反映的心理机能(Chermahini, Hickendorff, & Hommel, 2012; Lee et al., 2014)。例

如: Lee 等人(2014)的研究发现, 远距离联想测验的结果不仅与发散思维测验的相关程度很低( $r = -0.05 \sim -0.13$ ), 还与经验开放性人格呈负相关( $r = -0.02$ )。在过往研究中, 经验开放性人格能够显著预测个体发散思维水平; 形成鲜明对比的发现是, 远距离联想测验同流体智力、言语智力则具备较高正相关( $r = 0.33 \sim 0.42$ )。Lee 和 Theriault (2013) 在结构方程模型中将远距离联想测验当做聚合思维这一潜变量的指标, 他们的研究发现, 联想类、聚合类以及发散类思维表征创造力不同的认知结构, 诸如工作记忆、智力等变量比起发散思维水平能够更好的预测个体的聚合思维水平。这些研究继续验证了已有观点, 即远距离联想测验和发散测验的确测量了不同的心理品质。更为重要的是 Benedek 等人(2012)的研究发现, 联想思维实际在发散思维的运用中起到基础作用, 他们的研究表明 4 项联想类任务能够解释发散思维半成的变异, 而在过去, 由于远距离联想测验和发散思维相关低下, 研究者往往认为联想思维与发散思维毫无关系。该研究在一定程度上说明, 远距离联想测验或许很接近传统智力测验, 但它仍然是创造力测验。现在, 越来越多的研究者倾向于认为, 远距离联想测验或许不涉及传统产生性、发散性的创造力机制, 它更着重于聚合性的创造力品质。

但同时, Lee 等人(2014)也指出, 远距离联想测验与智力测验有极高相关仍可能存在两种迥然不同的结论: 第一, 远距离联想测验确实表征聚合性创造力; 第二, 与之完全相反, 远距离联想测验可能仍旧只是与智力测验类似的分析思维测验。因此, 远距离联想测验是否真能反映个体创造力水平? 有关该测验预测效度的数据, 特别是远距离联想测验与创造力成就之间的关系就十分重要。然而该研究领域目前仍然非常缺乏, 所以研究者在使用和解释远距离联想测验时仍然需要非常谨慎。

除了远距离联想测验, 其他顿悟类测验还有诸如经典的邓克尔蜡烛问题(Duncker, 1945); 九点问题; “脑筋急转弯式”顿悟问题(DeYoung, Flanders, & Peterson, 2008); 我国研究者设计的三字谜顿悟、原型启发顿悟等等(沈汪兵, 刘昌, 张小将, 陈亚林, 2011; 罗俊龙等, 2012)。但通过仔细分析可以发现, 一些广泛使用的联想顿悟问题

实际最早属于问题解决范畴, 主要用于探讨定势对解决问题的作用, 虽然打破定势是创造性思维的重要环节, 但其并非是创造力水平可靠的评定标准。Beaty, Nusbaum 和 Silvia (2014)的研究就发现, 解决顿悟问题的能力与真实创造力行为水平相关程度极低。而诸如三字谜、原型启发等顿悟问题的设计初衷在于, 采用 ERP、fMRI 等认知神经科学手段研究顿悟思维本身的脑定位、思维进程等问题。严格说, 它们并非是针对创造力水平的测验。虽然有研究表明, 联想顿悟水平与发散思维有中等程度相关(Gilhooly & Murphy, 2005), 但结合以上所述, 顿悟类测验能否成为创造力水平的合理评估手段, 仍然有待于更多研究予以验证和说明。

#### 4 创造力成就测验

创造力成就是个体在其一生中所创造出新颖且有价值的产品的集合(Carson, Peterson, & Higgins, 2005)。创造力成就测验的理论假设最为直观明了: 如果有什么能够预测未来的创造力潜能, 那么最有可能的是已经存在的创造力成就。对创造力成就的测量方式主要为: 可证实的成就或荣誉的数量; 业已存在的创造性产品所获得的评价以及自述创造力成就清单(Wang, Ho, Cheng, & Cheng, 2014)。自述创造力成就清单是创造力成就测验的主流, 经常被使用的测验包括: Hogevar (1979)的创造力行为清单(Creative Behavior Inventory, CBI); Ludwig (1992)的创造力成就量表(Creative Achievement Scale, CAS); Carson 等人(2005)的创造力成就问卷(Creative Achievement Questionnaire, CAQ); Batey 和 Furnham (2008)的创造力行为传记清单(Biographical Inventory of Creative Behaviours, BICB)等等。Silvia 等人(2012)对广泛使用的创造力成就问卷(CAQ)、创造力行为清单(CBI)等创造力成就测验的可信度进行了分析, 他们采用经典测验理论和项目反应理论系统考察了其项目水平和测验水平, 结果发现这些成就自评量表不仅在测量指标上表现较好, 几种量表之间还维持了很好的共变关系。另根据 Wang 等人(2014)对现有自述创造力成就清单应用情况的调查报告, Carson 等人(2005)的创造力成就问卷(CAQ)由于囊括了艺术和科学领域 10 个大类的成就自评而得到了最为广泛的应用。

自述创造力成就清单采取主观自评的形式,然而,这种形式本身可能带来一些问题,其中最为重要的是共同方法变异问题(common method variance)。共同方法变异指的是,实验结果来源于测量方法而非测量方法所表征的结构(Podsakoff, MacKenzie, Lee, & Podsakoff, 2003)。Ng 和 Feldman (2012)指出,当采取主观自评的方式测量创造力和与创造力相关的变量时,共同方法变异问题极易产生。具体表现为,当对创造力和其他变量进行相关分析时,共同方法变异会导致两个变量间产生虚高的相关,从而极易诱使研究者出现弃真错误。出现这类问题的原因在于,第一:被试会倾向于在一系列主观评价任务中保持认知层面和态度层面的一致性,这会导致变量间相关被人为增高;第二:从内隐理论角度出发,被试可能会认为接受的一系列测验存在一定程度上的共变;第三:被试可能出于社会赞许效应故意夸大其创造力水平,特别当主观自评创造力应用于高风险情境中,例如:招聘、决策等,被试更有可能“装好人(Faking good)” (Silvia et al., 2012);第四,被试的情绪可能会连续影响其在一系列测验中的态度,导致被试在多项测验中采取相同回答策略。Kaufman, Evans 和 Baer (2010)的研究就发现,学生自评创造力水平和其真实创造力表现毫无关联。

为说明共同方法变异在主观创造力测量中可能带来的影响,Ng 和 Feldman (2012)系统比较了主观自评创造力测验与非主观评定创造力测验以及情绪、人格等多个变量之间的关系。结果发现,主观自评测验的确比非主观自评测验产生更大的共同变异效应;不过,当变量为“客观自评”,即要求被试自评其过往生活史等客观事件时,该变量与创造力间的共同方法变异效应较小;研究还发现,积极情绪和主观自评创造力不会出现虚高的相关。这提醒研究者一方面在运用自述创造力成就清单对创造测量时需要合理评估可能出现的误差,另一方面可以采取其他方法减小这种误差的出现。例如:同时汇报主观自评创造力和非主观自评创造力;分时段对创造力和其他相关变量进行测量等等。可以说,通过合理利用测验和汇报必要测量指标,创造力成就测验普遍存在的共同方法变异问题能够得到妥善解决。

除此之外,研究者在应用创造力成就测验时还应注意:首先,根据实验目的妥善选取恰当的

创造力成就测验,有些创造力成就测验为领域一般性,例如:创造力行为清单(CBI),创造力行为传记清单(BICB);而有些测验为领域特殊性,例如:创造力成就问卷(CAQ)。一些测验着重考察较高水平或显著人群的创造力成就,有研究者称之为 Pro-C 或 Big-C 创造力(Kaufman & Beghetto, 2009),例如:创造力成就问卷(CAQ);而另一些测验着重于考察一般水平或普通人群的创造力,即 little-c 创造力,例如:创造力行为清单(CBI),创造力行为传记清单(BICB)。其次,由于该类测验往往产生分类或计数数据,这极易导致数据整体成正偏态分布,因此研究者在处理该类数据时需要首先借助统计方法实现其正态化(Silvia et al., 2012)。

## 5 同感评估技术

同感评估技术(consensual assessment technique),是当前创造力测量领域中最主要的评分技术。其基本理论假设是:某领域内的专家对该领域创造力评估存在一种共识。因此,当评价某产品的创造力水平时,只需要借助该领域专家业已形成的标准对产品进行评定即可。与创造力成就测验大多采用自评不同,同感评估技术本质上为专家评价法。该技术与之对应的基本要求为:第一,测验评定者需为该领域专家;第二,评定者依靠其对创造力的内隐态度对产品进行评判;第三,评定者应独立对产品进行评估(Amabile, 1982)。研究发现,同感评估技术不仅适用于实验室环境中的即时创造性产品,同时也适用于业已存在的准实验产品(Baer, Kaufman, & Gentile, 2004)。由于适用于多种创造性产品的评价且具备良好的信、效度指标,同感评估技术得到了广泛应用。有研究表明,在西方 5 种创造力主流学术期刊中,约 1/4 的创造力测量方法采取了同感评估技术或类似的专家评价方法(Long, 2014),我国也不乏介绍并应用该技术的研究(宋晓辉, 施建农, 2005)。

但也有一些研究者提出异议,他们认为该技术应用成本过高(必须采用专家进行评定);过程繁复(专家必须分别对每个项目进行评定),评价耗时(Kaufman, Baer, Cole, & Sexton, 2008)。因此,有研究者希望可以用非专家代替专家实施同感评估技术。然而有研究表明,尽管非专家在特定类型的研究或产品的评定中可以胜任,但他们表现

并不稳定,和专家间的评价差异很大(Kaufman & Baer, 2012)。例如:Kaufman 等人(2008)的研究表明,当非专家采用同感评估技术时,其评分者一致性系数很低,且评价结果和专家评定结果差异很大。为改进这一问题,Cropley 和 Kaufman (2012)编制了创造性解决方案诊断量表(creative solution diagnosis scale)。该量表包括 30 个项目,每一个项目对应一个评价维度,要求评定者根据项目对该产品在该维度上的水平进行打分。他们的研究发现,当施用该量表对产品创造性水平进行评价时,即使评定者为非专家,评价结果仍具备较高的内部一致性系数。但值得注意的问题是,在使用该量表时,评定者的评价标准实际上直接受到了该量表明确、外显标准的引导,这与 Amabile (1982)强调评定者个体凭借内隐态度进行评判的理念已经背道而驰,与基本理论要求形成矛盾。

同感评估技术的一个突破性贡献在于其巧妙地解决了创造性产品的标准问题。在同感评估技术提出之前,研究者始终无法在创造性产品的标准上达成共识。虽然研究者普遍认同创造力应具备“新颖性”和“适宜性”,但各领域创造性产品仍然具备不同的标准,特别当标准在跨领域的具体情境中时,我们无法认定某个领域的标准就适用于另一个领域的标准。同感评估技术提出直接采用该领域专家对该产品进行评定,因为只有该领域的专家才清楚怎样的产品是具有创造性的(Plucker & Makel, 2010)。不难发现,同感评估技术实际上回避了对于创造力普适性标准的讨论,即不同领域的标准无需互相比,该领域的专家才有实际的发言权。然而这在一定程度上引入了一种混乱,即忽视了评定者可能对测验结果带来的不良变异。有研究者称之为评定者效应(rater effects) (Wolfe, 2004),评定者效应体现在两个层面,表层效应为评定者给出的原始分数,潜层效应为评定者的思维过程和评定认知,这两个层面的效应共同作用最终可能影响测验的信、效度(Wolfe & McVay, 2012)。

尽管同感评估技术会在研究中采用科隆巴赫 alpha 系数报告其评分者一致性信度,以说明评定者打分的可靠性,然而相关系数高并不意味着评分者实际上采用了相同的标准,只能说明他们给出了类似的分数,或者只是评定者数量较多带来的虚假高相关。由于经典测量理论将被试分数理

解为真实分数和误差的结合,所以评定者标准的高低等因素无法通过经典测量理论进行估计(Stemler, 2004)。为阐明评定者效应可能带来的消极影响及其原因,Long 和 Pang (2015)结合定量研究与定性研究的优势,一方面采用概化理论(generalizability theory)技术明确评定者效应对数据可靠性带来的影响;另一方面采取框架分析(framing analysis)技术进一步挖掘其原因。结果发现:量化研究显示,当评定者数量较少时,概化系数和依存系数的指标均不能达到满意水平;质化研究进一步发现,当评定者更强调知识在创造过程中的功能时,他们往往更关注创造性产品的适宜性且倾向于采用更严格的标准。Long (2014)的另一项定性研究也发现,评定者不仅采取新颖性和适宜性标准对创造性产品进行评定,他们还会考虑被试的答案是否有思想,有趣味以及是否聪明。

评定者效应的根源在于评定者作为不同个体具备不同的心理特质,因此他们在各种情境中难以维持一致的评价标准。有研究发现,评定者效应受到评定者背景、知识经验以及内心准则(Bejar, 2012; Cumming, 1990),甚至创造性产品作者的性别以及名字特殊性的影响(Lebuda & Karwowski, 2013);且当评价任务改变,评价标准就会发生变化(Long, 2014);接受培训后的评定者依然存在评定者效应(Hung, Chen, & Chen, 2012)。Hempel 和 Sue-Chan (2010)还从东西方文化差异的角度考虑评定者可能带来误差,他们认为,即使东西方学者对创造力概念持有相同意见,但由于不同文化情境中主流价值观存在巨大差异,该差异可能会直接决定东西方评定者具备不同的评价标准和评价体系,这将最终影响对产品创造性的判定。例如:相较于西方人,中国人在判定某个产品是否具备创造性时,倾向于看重其对社会的进步是否有推动作用(Rudowicz & Hui, 1997)。

虽然研究者同时也认为,如果评定者接受恰当培训,测验包含明确定义的评价体系,评价者效应能得到有效控制(Long & Pang, 2015)。然而,这与同感评估技术的基本精神存在一定矛盾。如果评定者接受了培训的外在标准,那么评定者也就不再是凭借创造力的内隐态度进行评判。理论的基本要求和应用的客观限制之间该如何权衡,或许是同感评估技术下一个需要解决的问题。



## 6 展望

### 6.1 基本概念问题

通过分析前文介绍的几种主流创造力测验技术,不难发现它们在创造力概念上具有不同的内涵。例如:发散思维测验认为创造力就是多角度解决问题的能力;顿悟类测验侧重于创造力无意识顿悟的一面;创造力成就测验则更看重个体现实中的创造性实践和成果。由于测量方式直接依赖于概念,这极易导致:由于基本概念不同,有大量碎片化的创造力测验被提出,测验之间聚合效度低下,因此缺乏互相比较的基础。例如:Mishra和Henriksen (2013)认为除了创造力的新颖性和有效性维度,还应当增加整体性(whole)维度,即相对于问题解决情境的审美标准,他们在此基础上编制同时测量该三个维度的创造力测验。可以想象,如果某研究者的创造力概念不包含整体性维度,那么其测验方法就会与Mishra和Henriksen (2013)的方法差异很大,这不仅导致测量方法之间无法互换使用,还会进一步导致测量结论难以直接互相印证,最终可能陷入各种测量方法杂乱不堪,研究结果相互矛盾的境地。有研究者就将这种现状形容为:进入到Hennessey和Amabile称之为“繁荣却破碎”的研究阶段(Corbalán, Martín-Brufau, Limiñana, & García-Peñas, 2014)。破碎意味着诸多创造力测验彼此分离,无法合力对创造力研究作出贡献。已有很多研究者指责创造力研究领域的许多矛盾结论源头在于创造力测量手段的混乱(Dietrich & Kanso, 2010; Lillard et al., 2013)。但实际上,混乱归根结底可能还是在于研究者无法从根本上对创造力概念达成共识,这种缺乏共识从领域一般性创造力的核心特质到不同领域创造力概念中都有体现。虽然本文开篇提到,大部分研究者认同创造力定义的核心标准是具备“新颖性”和“适宜性”,但仔细分析,其中仍有较大分歧,特别是适宜性维度。例如:适宜性可能从社会环境角度被理解为“有道德或社会价值”(Forgeard & Mecklenburg, 2013);也能从问题解决的角度被理解为“有效或有用”(Sternberg & Lubart, 1996)。如果从前者出发,创造力应当只包括具有积极社会意义的想法和行为,但如果从后者出发,创造力同样也可以是一些“为我所用”的恶意行径,如近几年涌现的恶意创造力(malevolent

creativity)研究就主要探究具备负性社会价值的创造性行为(Harris, Reiter-Palmon, & Kaufman, 2013)。

因此,为解决这一问题,有研究者指出应追根溯源,从定义角度规范创造力理论,并在此基础上统一创造力测量方法(Batey, 2012)。Treffing, Renzulli和Feldhusen (1971)也曾指出,创造力测量的困难很大程度上来自于我们无法对创造力概念形成“理论统一(theoretical unity)”。在某一学科研究领域,实现完全的理论统一或许是不可能同时也是没有必要的。在一定概念框架的基础上,通过不断运用创造力测验,并积极向创造力概念进行反馈,令学术界对创造力的核心特质及各领域内的基本概念达成真正共识,或许才是解决创造力测量混乱局面的根本途径。如上文所述,研究者现已普遍认同创造力概念既包括发散思维,也包括聚合思维,过去发散思维测验就能完全表征创造力的观念被彻底摒弃。研究者现今往往在一项创造力研究中选取多种创造力测验,以避免方法上的片面,这无疑是研究者在创造力基本概念日益形成共识的基础上,进一步于测验定位和选取上获得的进步。借助此类测验运用的推进,反过来会有助于研究者对创造力概念进一步达成共识。通过这种概念到测验,测验再回到概念且不断循环上升的研究逻辑,最终将有助于理清当前创造力测验领域纷乱现状。

### 6.2 测验优化问题

如前文所述,创造力测验的一些基本的心理学测量指标仍然饱受批评。例如:发散思维测验的信、效度问题;远距离联想测验的效度问题等等。因此不断从测验内容、施测过程等方面推进创造力测验的优化工作是当务之急。从测验内容看,研究者需通过对创造力测验的不断应用和改进提高创造力测验的信、效度水平,体现心理测验的科学性。例如:当采取主观计分方法测量创造力时,要求被试在2~3分钟左右完成一道测验时其新颖性分数效度最好,时间无需过长,因为即使时间继续延长,被试答案的信、效度也没有显著提高(Benedek, et al., 2013),同时,时间也不宜过短,因为被试倾向于在创造力测验的后期表现更好(Beaty & Silvia, 2012);又如:Silvia (2011)的研究发现,作为发散思维测验经常使用的三种测验形式:灵活使用任务、举例任务以及结果任务在信度上有较大差异,彼此并不能互相替代;

Gupta, Jang, Mednick 和 Huber (2012)建议在远距离联想测验中避免使用高频答案的题目等等。

从施测过程看,传统创造力测验时间大多较长,耗时费力;施测采用纸笔测验的形式,测验的信息需要通过转录到电脑中才能进行处理,后期工序十分繁复,增加了诸多主观不确定因素。因此,一方面,传统测验的精简化在有条不紊的开展中,且收到了良好效果(Althuizen, Wierenga, & Rossiter, 2010);另一方面,随着信息技术的发展,越来越多的研究者试图在电脑上完成对创造力的测量,以提高创造力测验的标准化和客观化(Diener, Wright, Smith, & Wright, 2014; Hass, 2015; Kwon, Goetz, & Zellner, 1998)。Lau 和 Cheung (2010)就比较了电子版和纸笔版 Wallach-Kogan 创造力测验的不同,结果发现电子版和纸笔版测验都展现出相似水平的信度。不仅如此,多元方差分析还显示电子版和纸笔版在各项分测验上均无显著得分差异。虽然同时有研究表明电子版和纸笔版测验在某些方面并不相同,特别当该创造力测验中要求被试进行画图等操作时,其在电子版创造力测验中的表现要比纸笔版更差(Diener et al., 2014; Kwon et al., 1998),但原因可能在于在电脑中操作鼠标进行画图比纸笔画图更难操作,与之类似的发现还有, Hass (2015)的研究表明,电子版创造力测验不会导致被试答案的创新性减少,但会在一定程度上降低其答案的流畅性。虽然创造力电子版测验尚不成熟,但其在测验标准化和客观化,计分自动化以及建立创造力测验常模大数据(big-data)等方面具有不可忽视的潜力。因此,创造力测验的电子化取向有很大上升空间,未来应在此方向努力。

总而言之,创造力测验需要通过不断地优化和完善,才能真正成为科学可靠的测量工具,从而为创造力研究提供坚强助力。

### 6.3 测验运用问题

创造力测验数量庞大,种类繁多,测验如何运用同样是难点问题。根据研究现状,首先,研究者在应用测验时宜采取混合测验的策略。混合测验即在运用创造力测验时,至少选取两种或两种以上的测验方式,且测验来自不同角度。主要原因在于:第一,从理论角度看,创造力绝非某种人格特质,某类思维习惯或某处环境单独决定的。众多心理学家都持有类似观点: Amabile (1982)

提出,创造性问题解决是在任务、动机、领域相关知识和能力、与创造力有关的技能共同作用的条件下达成的; Plucker 和 Renzulli (1999)曾提出创造性产品的内隐理论(implicit theory),他们认为个体具备某种认知或人格结构,二者互相结合才能产生创造性产品; Csikszentmihalyi (1999)的系统理论(system theory)也认为,创造力来源于个体、场,以及领域的交互作用; Sternberg 和 Lubart (1992)认为,创造力包括6个要素:智力、知识、思考方式、人格、动机以及环境。因此,创造力基本理论不支持应用单一的测量技术。第二,从现实角度看,据上文所述,一方面,目前尚存在多种创造力概念,因此创造力测量决不能仅仅使用某一种,采取多种测验同时施测是较为稳妥的策略(Cropley, 2000)。另一方面,当今流行的各种测验形式和计分方式各有其利弊,唯有结合运用才能扬长避短。

当前,较好的混合测验策略是选取恰当的多角度测验,并且确保几类测验都有一定程度的重合且能够分别测量创造力的不同方面(Pinheiro & Cruz, 2014);或者采用囊括创造力多角度品质的测验,例如: Sen, Acar 和 Cetinkaya (2014)认为,创造性活动中的个体和环境从来就不是孤立存在的,据此他们编制了个人环境适应量表(Person-Environment Fit Scale, PEFSC),从个体和环境两个角度同时测量创造力。

除此之外,从推进创造力测验发展的角度看,一些过往研究中针对某测验的争议问题还应当通过有意识的运用该测验加以解决和验证。例如:研究者可以尝试在较长的纵向时间内对不同群体施测远距离联想测验,并在恰当的时间节点选取多类效标,从而为该测验类型的有效性提供有力说明。如果缺乏主动的验证,由于绝大多数的创造力测验在应用时只是为其他研究目标服务,因此即使某测验在研究中获得了广泛的应用,研究者也可能无法获取有关该测验的重要信息。特别是测验的有效性往往还存在中西方差异,本土化的创造力测验更需要此类基础性工作。一些国外研究者在这些方面做出了很好的尝试(Runco et al., 2010; Silvia et al., 2012),这值得国内研究者借鉴。

### 6.4 测量技术多样化

虽然正如前文多次提及,创造力测量技术数量巨大,然而当前应用最广的测量技术仍然是几

十年前提出的诸如: 发散思维测验、远距离联想测验以及同感评估技术等等。这些经典测量技术固然有其优势, 但亦有诸如: 独特性维度计分问题、评分者效应等顽疾。因此除了对原有测量技术的改进, 创造力测量领域依然需要新想法和新突破, 这就要求研究者持续推进创造力测量技术的多样化发展, 以求从根本上解决问题。

近几年, 研究者开始从语义网络角度测量创造力(Bossomaier, Harré, Knittel, & Snyder, 2009; Prabhakaran et al., 2014), 或改进创造力计分方法(Wang, Chang, & Li, 2008)。Harbison 和 Haarmann (2014)就从语义表征角度提出了称之为“自动化计分”(automated scoring)技术对发散思维测验中的特殊性水平进行打分。根据过往研究, 个体在行为中采用的语义表征模式能够对一系列人类行为起到预测作用。因此, 他们借助潜在语义分析和共同逐点信息(pointwise-mutual information)技术对被试的答案和题目之间的语义相似度(similarity)进行分析。结果发现, 相似度越高, 被试答案的新颖度越低。据此, 他们提出语义相似度可以作为答案新颖性的有效预测指标。由于潜在语义分析等技术采取的是大数据支持下的客观计算(网页版测评可见 <http://lsa.colorado.edu/>), 因此通过应用该类技术, 上文提到主观评价法可能带来的评定者效能在一定程度上避免。用客观化的新指标表征原有的独特性维度, 也可能有助于从彻底规避前文讨论的独特性维度计分问题。但这种方法是否合适, 仍有待于进一步检验。

从认知神经实验的角度, Prabhakaran 等(2014)还认为, 传统创造力测量手段并不适用于脑成像研究。传统创造力脑成像技术研究过于依赖被试对任务的反应时或反应数, 无法准确捕捉到创造性思维的瞬间。为解决这一问题, 他们基于语义网络理论提出了便于进行脑成像研究的创造力单字发音(single-word utterances)测验。即让被试看到一个名词, 让其口头报告一个与该名词有新颖联系的动词, 并利用潜在语义分析技术(Latent Semantic Analysis)分析动词与名词的语词距离, 语义距离越大, 创造性水平越高。他们的研究结果发现, 语义距离与言语、非言语创造力相关程度很高( $\beta = 0.50$ )。

除此之外, 教育学领域提出了更具生态效度的创造力测量方法(Blamires & Peterson, 2014);

功能设计领域的创造力研究也为创造性产品测量提供了新的视角。如: Sarkar 和 Chakrabarti (2011)设计了客观严格的模型和公式对创造力的新颖性和有效性进行计算, 以得出对某产品创新水平的判断。例如: 新颖性指标=重要程度×受欢迎程度×使用频率×使用持续时间或使用期间受益。Redelinghuys 和 Bahill (2006)从工程设计的角度将创造性产品理解为设计价值(design value), 探究某一产品的价值则需要考察设计过程中的资源消耗和努力投入, 他们通过 REV(resources-effort-value)技术构建数理模型, 并将量化的资源、努力以及价值指标带入演算, 以求得创造力指数。这一系列成果提醒创造力测量领域研究者决不能闭门造车, 从其他领域吸取灵感或许能给未来研究打开一片新天地。

## 参考文献

- 李良敏, 罗玲玲, 刘武. (2015). 客观化创造力测量工具: 《中文远距联想测验》编制. *东北大学学报(社会科学版)*, 17(1), 19-24.
- 罗俊龙, 覃义贵, 李文福, 朱海雪, 田燕, 邱江, 张庆林. (2012). 创造发明中顿悟的原型启发脑机制. *心理科学进展*, 20(4), 504-513.
- 沈汪兵, 刘昌, 张小将, 陈亚林. (2011). 三字字谜顿悟的时间进程和半球效应: 一项 ERP 研究. *心理学报*, 43(3), 229-240.
- 宋晓辉, 施建农. (2005). 创造力测量手段——同感评估技术(CAT)简介. *心理科学进展*, 13(6), 739-744.
- 王烨, 余荣军, 周晓林. (2005). 创造性研究的有效工具——远距离联想测验(RAT). *心理科学进展*, 13(6), 734-738.
- Aiello, D. A., Jarosz, A. F., Cushen, P. J., & Wiley, J. (2012). Firing the executive: When an analytic approach to problem solving helps and hurts. *The Journal of Problem Solving*, 4(2), 116-127.
- Althuizen, N., Wierenga, B., & Rossiter, J. (2010). The validity of two brief measures of creative ability. *Creativity Research Journal*, 22(1), 53-61.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997-1013.
- Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, 39(5), 1154-1184.
- Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity*, 13, 195-205.
- Baer, J. (2011). How divergent thinking tests mislead us: Are the Torrance Tests still relevant in the 21st century? The

- Division 10 debate. *Psychology of Aesthetics, Creativity, and the Arts*, 5(4), 309–313.
- Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal*, 16(1), 113–117.
- Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, 24(1), 55–65.
- Batey, M., & Furnham, A. (2008). The relationship between measures of creativity and schizotypy. *Personality and Individual Differences*, 45(8), 816–821.
- Beaty, R. E., Nusbaum, E. C., & Silvia, P. J. (2014). Does insight problem solving predict real-world creativity?. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 287–292.
- Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 309–319.
- Beaty, R. E., Smeekens, B. A., Silvia, P. J., Hodges, D. A., & Kane, M. J. (2013). A first look at the role of domain-general cognitive and creative abilities in jazz improvisation. *Psychomusicology: Music, Mind, and Brain*, 23(4), 262–268.
- Bejar, I. I. (2012). Rater cognition: Implications for validity. *Educational Measurement: Issues and Practice*, 31(3), 2–9.
- Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), 273–281.
- Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts*, 7(4), 341–349.
- Besemer, S. P., & O'Quin, K. (1987). Creative product analysis: Testing a model by developing a judging instrument. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 367–389). Buffalo, NY: Bearly.
- Blamires, M., & Peterson, A. (2014). Can creativity be assessed? Towards an evidence-informed framework for assessing and planning progress in creativity. *Cambridge Journal of Education*, 44(2), 147–162.
- Bossomaier, T., Harré, M., Knittel, A., & Snyder, A. (2009). A semantic network approach to the creativity quotient (CQ). *Creativity Research Journal*, 21(1), 64–71.
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17(1), 37–50.
- Charyton, C., & Merrill, J. A. (2009). Assessing general creativity and creative engineering design in first year engineering students. *Journal of Engineering Education*, 98(2), 145–156.
- Chermahini, S. A., Hickendorff, M., & Hommel, B. (2012). Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach. *Thinking Skills and Creativity*, 7(3), 177–186.
- Chou, H. W., Chen, Y. L., & Chou, S. B. (2014). A method for evaluating the creativity of comic strips. *Thinking Skills and Creativity*, 14, 11–19.
- Clapham, M. M., Cowdery, E. M., King, K. E., & Montang, M. A. (2005). Predicting work activities with divergent thinking tests: A longitudinal study. *The Journal of Creative Behavior*, 39(3), 149–166.
- Clark, P. M., & Mirels, H. L. (1970). Fluency as a pervasive element in the measurement of creativity. *Journal of Educational Measurement*, 7(2), 83–86.
- Colangelo, N., Kerr, B., Hallowell, K., Huesman, R., & Gaeth, J. (1992). The Iowa Inventiveness Inventory: Toward a measure of mechanical inventiveness. *Creativity Research Journal*, 5(2), 157–163.
- Corbalán, J., Martín-Brufau, R., Limiñana, R., & García-Peñas, V. (2014). CREA. 10 years in response to the question: Is it really possible to measure creativity?. (Abstract). *Personality and Individual Differences*, 60, S29.
- Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using?. *Roeper Review*, 23(2), 72–79.
- Cropley, D. H., & Kaufman, J. C. (2012). Measuring functional creativity: Non-expert raters and the creative solution diagnosis scale. *The Journal of Creative Behavior*, 46(2), 119–137.
- Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 313–335). Cambridge, UK: Cambridge University Press.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), 31–51.
- DeYoung, C. G., Flanders, J. L., & Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: An individual differences model. *Creativity Research Journal*, 20(3), 278–290.
- Diener, M. L., Wright, C. A., Smith, K. N., & Wright, S. D. (2014). Assessing Visual-spatial creativity in youth on the autism spectrum. *Creativity Research Journal*, 26(3), 328–337.
- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136(5), 822–848.
- Domino, G. (1970). Identification of potentially creative persons from the Adjective Check List. *Journal of Consulting and Clinical Psychology*, 35, 48–51.
- Dunker, K. (1945). On problem-solving. *Psychological Monographs*, 58, 5.
- Fink, A., Slamar-Halbedl, M., Unterrainer, H. F., & Weiss, E.

- M. (2012). Creativity: Genius, madness, or a combination of both?. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 11–18.
- Fodor, E. M. (1999). Subclinical inclination toward manic-depression and creative performance on the Remote Associates Test. *Personality and Individual Differences*, 27(6), 1273–1283.
- Forgeard, M. J. C., & Mecklenburg, A. C. (2013). The two dimensions of motivation and a reciprocal model of the creative process. *Review of General Psychology*, 17(3), 255–266.
- Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, 11(3), 279–302.
- Gough, H. G. (1979). A creative personality scale for the Adjective Check List. *Journal of Personality and Social Psychology*, 37(8), 1398–1405.
- Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1), 3–14.
- Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science*, 23(3), 288–294.
- Harbison, J. I., & Haarmann, H. (2014). Automated scoring of originality using semantic representations. *Proceedings of COGSCI*, 2327–2332.
- Harris, D. J., Reiter-Palmon, R., & Kaufman, J. C. (2013). The effect of emotional intelligence and task type on malevolent creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 7(3), 237–244.
- Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Computers in Human Behavior*, 46, 85–93.
- Hempel, P. S., & Sue-Chan, C. (2010). Culture and the assessment of creativity. *Management and Organization Review*, 6(3), 415–435.
- Hocevar, D. (1979, April). *The Development of the creative behavior inventory (CBI)*. Annual Meeting of the Rocky Mountain Psychological Association.
- Hong, E., Milgram, R. M., & Gorsky, H. (1995). Original thinking as a predictor of creative performance in young children. *Roeper Review*, 18(2), 147–149.
- Hung, S. P., Chen, P. H., & Chen, H. C. (2012). Improving creativity performance assessment: A rater effect examination with many facet Rasch model. *Creativity Research Journal*, 24(4), 345–357.
- Hürsen, Ç., Kaplan, A., & Özdal, H. (2014). Assessment of creative thinking studies in terms of content analysis. *Procedia-Social and Behavioral Sciences*, 143, 1177–1185.
- Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative?. *Creativity Research Journal*, 24(1), 83–91.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171–178.
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, 13(1), 1–12.
- Kaufman, J. C., Evans, M. L., & Baer, J. (2010). The American idol effect: Are students good judges of their creativity across domains?. *Empirical Studies of the Arts*, 28(1), 3–17.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (Eds.). (2008). *Essentials of creativity assessment*. Hoboken, NJ: Wiley.
- Kim, K. H. (2008). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *The Journal of Creative Behavior*, 42(2), 106–130.
- Kim, K. H. (2011). Proven reliability and validity of the Torrance Tests of Creative Thinking (TTCT). *Psychology of Aesthetics, Creativity, and the Arts*, 5(4), 314–315.
- Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, 61(5), 622–629.
- Kwon, M., Goetz, E. T., & Zellner, R. D. (1998). Developing a computer-based TTCT: Promises and problems. *The Journal of Creative Behavior*, 32(2), 96–106.
- Lau, S., & Cheung, P. C. (2010). Creativity assessment: Comparability of the electronic and paper-and-pencil versions of the Wallach-Kogan Creativity Tests. *Thinking Skills and Creativity*, 5(3), 101–107.
- Laughlin, P. R., Doherty, M. A., & Dunn, R. F. (1968). Intentional and incidental concept formation as a function of motivation, creativity, intelligence, and sex. *Journal of Personality and Social Psychology*, 8, 401–409.
- Lebuda, I., & Karwowski, M. (2013). Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, 25(1), 137–142.
- Lee, C. S., Huggins, A. C., & Theriault, D. J. (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the Remote Associates Test. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 446–460.
- Lee, C. S., & Theriault, D. J. (2013). The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence*, 41(5), 306–320.
- Lee, S. (2008). Commentary: Reliability and validity of uniqueness scoring in creativity assessment. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2), 103–108.
- Leon, S. A., Altmann, L. J. P., Abrams, L., Gonzalez Rothi, L. J., & Heilman, K. M. (2014). Divergent task performance in older adults: Declarative memory or creative potential?.

- Creativity Research Journal*, 26(1), 21–29.
- Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, 139(1), 1–34.
- Long, H. Y. (2014). More than appropriateness and novelty: Judges' criteria of assessing creative products in science tasks. *Thinking Skills and Creativity*, 13, 183–194.
- Long, H. Y., & Pang, W. G. (2015). Rater effects in creativity assessment: A mixed methods investigation. *Thinking Skills and Creativity*, 15, 13–25.
- Ludwig, A. M. (1992). The creative achievement scale. *Creativity Research Journal*, 5(2), 109–119.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232.
- Mishra, P., & Henriksen, D. (2013). A new approach to defining and measuring creativity: Rethinking technology & creativity in the 21st century. *TechTrends*, 57(5), 10–13.
- Ng, T. W. H., & Feldman, D. C. (2012). A comparison of self-ratings and non-self-report measures of employee creativity. *Human Relations*, 65(8), 1021–1047.
- Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different?: Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, 39(1), 36–45.
- Paprika, Z. Z., & Nagy, V. (2012). Assessment of creativity on the job market. *Procedia Economics and Finance*, 3, 166–181.
- Pinheiro, I. R., & Cruz, R. M. (2014). Mapping creativity: Creativity measurements network analysis. *Creativity Research Journal*, 26(3), 263–275.
- Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48–73). New York: Cambridge University Press.
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 35–61). Cambridge, UK: Cambridge University Press.
- Plucker, J. A., Qian, M. H., & Wang, S. J. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *The Journal of Creative Behavior*, 45(1), 1–22.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
- Prabhakaran, R., Green, A. E., & Gray, J. R. (2014). Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods*, 46(3), 641–659.
- Pretz, J. E., & McCollum, V. A. (2014). Self-perceptions of creativity do not always reflect actual creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2), 227–236.
- Primi, R. (2014). Divergent productions of metaphors: Combining many-facet Rasch measurement and cognitive psychology in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 461–474.
- Redelinghuys, C., & Bahill, A. T. (2006). A framework for the assessment of the creativity of product design teams. *Journal of Engineering Design*, 17(2), 121–141.
- Rudowicz, E., & Hui, A. N. (1997). The creative personality: Hong Kong perspective. *Journal of Social Behavior & Personality*, 12(1), 139–157.
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75.
- Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal*, 22(4), 361–368.
- Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies*, 32(4), 348–383.
- Sen, S., Acar, S., & Cetinkaya, C. (2014). Development of the Person–Environment Fit Scale (PEFSC): A new measure of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4), 433–445.
- Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, 20(1), 34–39.
- Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*, 6(1), 24–30.
- Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, 4(2), 79–85.
- Silvia, P. J., Nusbaum, E. C., Berg, C., Martin, C., & O'Connor, A. (2009). Openness to experience, plasticity, and creativity: Exploring lower-order, high-order, and interactive effects. *Journal of Research in Personality*, 43(6), 1087–1090.
- Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 19–34.
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., ... Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2), 68–85.
- Simonton, D. K. (2003). Expertise, competence, and creative

- ability: The perplexing complexities. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 213–239). New York: Cambridge University Press.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved April 12, 2015 from <http://PAREonline.net/getvn.asp?v=9&n=4>
- Sternberg, R. J., & Lubart, T. I. (1992). Creativity: Its nature and assessment. *School Psychology International*, 13(3), 243–253.
- Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, 51(7), 677–688.
- Storm, B. C., Angello, G., & Bjork, E. L. (2011). Thinking can cause forgetting: Memory dynamics in creative problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1287–1293.
- Taft, R., & Rossiter, J. R. (1967). The Remote Associates Test: Divergent or convergent thinking? *Psychological Reports*, 19(3), 1313–1314.
- Tan, M., Mourgues, C., Bolden, D. S., & Grigorenko, E. L. (2014). Making numbers come to life: Two scoring methods for creativity in aurora's cartoon numbers. *The Journal of Creative Behavior*, 48(1), 25–43.
- Torrance, E. P. (1972). Predictive validity of the Torrance tests of creative thinking. *The Journal of Creative Behavior*, 6(4), 236–262.
- Torrance, E. P., & Goff, K. (1989). A quiet revolution. *The Journal of Creative Behavior*, 23(2), 136–145.
- Treffinger, D. J., Renzulli, J. S., & Feldhusen, J. F. (1971). Problems in the assessment of creative thinking. *The Journal of Creative Behavior*, 5(2), 104–112.
- Wang, C. C., Ho, H. C., Cheng, C. L., & Cheng, Y. Y. (2014). Application of the Rasch Model to the measurement of creativity: The creative achievement questionnaire. *Creativity Research Journal*, 26(1), 62–71.
- Wang, H. C., Chang, C. Y., & Li, T. Y. (2008). Assessing creative problem-solving with automated text grading. *Computers & Education*, 51(4), 1450–1466.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46(1), 35–51.
- Wolfe, E. W., & McVay, A. (2012). Application of latent trait models to identifying substantively interesting raters. *Educational Measurement: Issues and Practice*, 31(3), 31–37.
- Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity?. *Creativity Research Journal*, 23(1), 24–37.
- Zeng, L., Proctor, R. W., & Salvendy, G. (2012). User-based assessment of website creativity: A review and appraisal. *Behaviour & Information Technology*, 31(4), 383–400.

## Several thoughts on measuring creativity

GONG Zhe<sup>1</sup>; LIU Chang<sup>1</sup>; SHEN Wangbing<sup>2,1</sup>

(<sup>1</sup> Lab of Cognitive Neuroscience and School of Psychology, Nanjing Normal University, Nanjing 210097, China)

(<sup>2</sup> School of Public Administration and Institute of Applied Psychology, Hohai University, Nanjing 210098, China)

**Abstract:** Creativity measurements are the foundation of creativity research. However, the accuracy of the results of creativity measurements have been long questioned by scholars in the field, making precise creativity measurements an urgent topic for exploration. For example, the divergent thinking test is central in the study of creativity, but its traditional scoring systems have faced well documented problems. Insight tests may represent individual creativity levels, but their validity still remains to be confirmed. Creative achievement tests are not always used legitimately, which may cause common method variance, and consensual assessment techniques may cause rater effects.

Fortunately, however, in recent years, remarkable progress has been achieved in some creativity measurement hotspots. For example, new subjective scoring methods for assessing the uniqueness dimension of divergent thinking tests may improve assessment reliability and validity. This paper hopes to contribute to progress in this area, and researchers explore the measurement of creativity from semantic network aspect. Future research should focus on the following points: unifying basic conceptions, optimizing tests from content and process aspects, using mixed tests, increasing the diversification of tests through intersection research, etc.

**Key words:** creativity; measurement of creativity; creativity test; divergent thinking test