心理科学进展 2005，13（6）：739~744
Advances in Psychological Science

创造力测量手段——同感评估技术（CAT）简介*

宋晓辉[1] 施建农[1,2]

（[1]中国科学院心理研究所 心理健康重点实验室，北京，100101） （[2]首都师范大学学习与认知实验室，北京 100037）

摘 要 同感评估技术（consensus assessment technique），又称主观评价法，是 Amabile 于 1982 年提出的评价创造力的一种方法。文章简要回顾了同感评估技术出现之前，基于发散思维测验的创造力评价法及其局限性。人们在寻求新的创造力评价方法的过程中，提出了同感评估技术。文章重点介绍了同感评估技术的理论基础，即它的同感定义和概念定义，然后说明了同感评估技术的具体使用方法、发展，以及在各领域的应用情况。最后，通过与创造性思维测验的比较，对同感评估技术作了评价，指出了 CAT 的优势和存在的问题。

关键词 同感评估技术，发散思维测验，领域特殊性，内隐观。

分类号 B849；G305

自从 20 世纪 50 年代 Guilford[1] 在美国心理学年会发表演说之后，创造力研究重新得到了心理学界的关注。为了科学地评价创造力，研究者们提出了各种各样的测量工具，其中，创造性思维测验是使用地最广泛的工具之一。创造性思维测验的核心为发散性思维测验，测验要求个体对特定刺激做出多种反应，强调的是观念流畅性。例如非常规用途测验（Unusual Usage Test，UUT），就是让被试列举出某物品的多种用途来考察他的发散思维能力。20 世纪 60、70 年代发散思维测验是创造力测量的主导工具，大部分创造力研究使用的都是发散思维测验，其中应用最为广泛的就是 Torrance [2]创造性思维测验（TTCT），它根据被试的回答在流畅性、变通性、独特性和精细性四个维度上进行评分[3]。

但是这个时期的发散思维测验由于其本身的局限性和发展的不完善，遭到了大量的批评，其中，最主要的也是最有争议就是人们对“发散性思维是否等于创造性思维？”的质疑。而这个问题则明显地反映在发散性思维测验的预测效度不高，至少是很不稳定上。尽管 Torrance[4]，Plucker[5]和 Yamada[6]都认为儿童期的 TTCT 成绩对以后的学业成就(school achievement)有较好的预测性。但 Kogan 和 Pankove 在 1972 年所做的一项为期五年的追踪研究，结果却发现只有在较小的学校环境里，发散思维测验才具有一定的预测性[7]。1974 年，Kogan 扩展了 1972 年的实验，结果发现来，十年级学生的发散思维测验成绩对于其大学的创造性活动只有 10%的预测性，而五年级学生的发散思维成绩则对大学期间的创造性活动已不具有预测性[8]。Migram[9]，Runco[10]等

收稿日期：2005-06-08

* 本研究得到国家自然科学基金项目（30370489）、中国科学院知识创新工程重要方向性项目（KSCXZ-SW-221）和北京市重点实验室——首都师范大学《学习与认知实验室》经费资助，以及国家教育科学十五规划项目(GBB010921)的支持。

通讯作者：施建农，E-mail：shijn@psych.ac.cn

人的研究也都分别对发散性思维测验的预测性的稳定性提出了挑战。发散性思维的这种显著弱点，使人们重新思考创造力的评价方法。而这个问题又依赖于对创造力的理解。美国哈佛大学的 Amabile 教授认为，尽管人们对创造力的定义可能不完全一样，但人们，特别是同一领域的专家们对同一作品会有基本一致的看法，这就是同感(consensus)。这种同感可以成为评价创造力的基础。因此，她于 1982 年提出了创造力研究的主观评价法“同感评估技术（consensus assessment technique，简称 CAT）”[11]。本文将对 CAT 做介绍，并对其合理性提出评论。

1 CAT 的理论基础

任何测量技术都是建立在特定理论基础之上的，CAT 也不例外。在 Amabile 看来，一个人的创造力必定反映在其作品所具有的创造性中，而作品的创造性则主要反映在作品的新颖性(newness or originality)和适宜性(usefulness or appropriateness)两个方面。显然，现实中，一个作品的新颖性和适宜性既取决于作品本身的特征，也取决于人们，特别是熟悉该领域的专家们对该作品的评价。在通常情况下，专家们对该作品的评价能够比较好地反映该作品的特征。同一领域的专家们对一个作品的评价通常会有较好的一致性。因此，Amabile 对创造力的“同感(consensual)”定义是创造力评价的一个操作性定义，评价的主要依据是评价者对创造力的内隐标准。不过，Amabile 认为，能够采用 CAT 技术评价的作品必须满足下面两个条件（1）具有新颖性和适宜性；（2）作品生产是开放式的[12]。显然，从广义上看，第一个条件是所有作品都能满足的，因为，所有作品都可以被看成是有新颖性和适宜性的，只是新颖和适宜的程度不同而已。但第二个条件并不是所有作品都能满足的。那些显而易见有统一标准答案的作品是不能用 CAT 来评价的。

同感定义隐含了这样一个假设，即：人们知道什么是创造力。这个假设在一定程度上得到了创造力内隐理论[13,14]的支持[15]。内隐理论是一种内在的信念，人们使用它来解释环境中的事件，做出判断，并计划自己的行动[16]。在创造力内隐理论的研究方面，最有影响的学者是美国 Yale 大学的 R. J. Sternberg。他在 20 世纪 80 年代所做的有关创造力内隐理论的研究具有广泛的影响。他考察了不同领域的专家和外行对创造力的看法，结果发现，无论是专家还是外行，人们对创造力是有共识的，即人们具有共同的内在评价标准。Sternberg 称之为创造力的内隐理论(implicit theory)[15]。在我们看来，这种内隐理论与当前发展心理学界特别关注的心理理论(theory of mind)[17]有类似之处。在发展心理学家看来，即使没有人正式地告诉孩子，年幼的孩子对外部的世界仍然会有自己的看法，而且，这种内在的看法具有普遍性。因此，对创造力的内隐理论是真实的、重要的，是我们的一部分，它会影响到我们对创造力以及创造性行为的认识[18]。正是由于内隐理论的存在，专家在评定产品的创造性时，才会达到较高的一致性。

Amabile 近年来也对观察者的主观评价标准作过考察，发现在对各个评价标准进行重要性排序时，评价者之间有相当高的一致性，10 个评价者中就有 7 个认为想像力（新颖性）是评价产品创造力的最重要的标准[8]，这与她对创造力所作的概念定义有一致性[12]。Christiaans 在 2002 年的研究中也发现，在众多维度中，评价者更重视作品的新颖性（包括独创性和不寻常性）[18]。

2 同感评估技术介绍

2.1 评价者

评价者是同感评估技术实施的重要因素。不过，基于内隐理论的 CAT 技术要求评价者对相关领域比较熟悉。在评价时，要求评价者根据自己对创造力的内隐理论，对产品或反应的创造性做出独立的评价。

2.2 适用于 CAT 评价的作品

能适用于 CAT 技术评价的作品或反应需要满足一定的条件。换句话说，不是所有的作品或反应都可以 CAT 技术来评估。就像前面提到的，具有明显标准答案的作品或反应不适宜用 CAT 评价。如数学计算，或物质的常规分类，如常规动植物的分类，或基本测量，如长度或体积的测量等。一般来说同感评估技术对引发创造性产品或反应的任务有如下要求：（1）能够引发产品或可清楚观察的反应，以供观察者对其评价；（2）所有被试的测试环境、实验材料和指导语是相同的；（3）对应于创造力的概念定义，任务是启发式的，即具有足够的开放度，允许产品具有相当的灵活性和新颖性；（4）在社会心理研究中，任务不能过于依赖某些特殊技能；（5）其引发的产品或反应可以由适宜的评价者对其进行信度较高的评价[12]。

目前，同感评估技术主要用于与言语或语言、艺术或表演或具有多种解决方案的问题解决等有关的作品或反应的创造性。与言语或语言有关作品或反应有：诗歌、故事或为图片定标题等；与艺术或表演有关的作品或反应有：拼贴画、绘画（包括线条画、油画和素描等）、手工艺品和动作表演等；而与问题解决有关的作品或反应有：计算机编程、沙漠求生或建筑设计等。在心理学研究中，CAT 应用得最为广泛的是用于评价儿童言语创造力和艺术创造力。

2.3 评价过程

在对某个作品或某类作品进行评价时，首先要挑选熟悉相关领域的成员组织评价小组，其次要告诉评价者，应该对作品的哪些方面进行评价，最后由评价者对所有被试的产品或反应做出独立评价。所有评价者根据自己对创造力的内隐理论对作品或反应做出评价。通常有三种做法：第一种是按照产品所具有的创造性由高到低对所有产品进行排序；第二种是将所有产品分为五类：创造性很低、创造性稍低、无法判断、创造性稍高和创造性很高；第三种是采用 Likert 五点量表，并在其中三个上分别写上：高、中和低[12]。在这三种方法中，第三种方法的评分者一致性较高，使用较多的也是这种方法。

采用 CAT 评价时，以下几点是非常重要的：（1）评价者必须熟悉该领域，即有该领域的工作经验；（2）所有评价者必须对作品进行独立评价；（3）评价者必须先了解要评价的所有作品，然后根据作品的相对水平对创造性高低做出评价；（4）评价时，应该以随机顺序评价作品[12]。

2.4 CAT 的信度检验

由于 CAT 技术是基于评价者对创造力的内隐理论的，其主观性比较大，因此，评分者一致性的高低就反映了 CAT 技术的信度水平。通常，对 CAT 的一致性检验有 3 种方法：(1)Winer 在 1971 年提出的通过分析内外部变异得到的信度系数，即[19]

$$\alpha = 1 - \frac{MS_{\text{评分者内}}}{MS_{\text{评分者间}}}$$

(2)Spearman-Brown 预测公式[20]，一致性系数：

$$\alpha = \frac{nr}{1 + (n-1)r}$$

其中 n 为评价者数目，r 为评价者之间的相关系数；(3)使用 SPSS 软件计算 Cronbach 系数[12]。从已有的众多研究结果看，用这 3 种方法得到的结果非常相似。由于 SPSS 软件的强大统计功能和使用的广泛性，使得用 SPSS 软件计算 Cronbach 系数成了最简单的方法，因此，在实际研究中也是最常用的一种方法。

评分者一致性系数的高低非常重要，较高的一致性系数表明同感评估技术的有效性。Amabile 对各种任务的评分者信度做过考察，发现都具有较高的一致性系数。我们从 Spearman-Brown 公式可以看出，一致性系数高低还与评价者的数目有关，评价者越多，则评分者一致性越高。通常，CAT 技术能接受的一致性系数为 0.70 以上，因此，对与言语有关的产品的评价需要至少 3 个评分者，与问题解决有关的产品需要至少 4 个评分者，而对艺术领域产品的评价则至少需要 7 个评分者[12]。

3 CAT 技术的应用

同感评估技术由于其接近于真实生活的评价方式，得到越来越广泛的使用 [21,22]，逐渐成为创造力评价的一种重要方法，被广泛应用于社会心理学、个体差异及领域特殊性和领域一般性研究。

Amabile 认为社会和环境因素对创造力表现有重要影响，而同感评估技术的产生正是为了社会心理研究。这些研究包括评价和奖赏对创造力的作用、榜样和动机导向的影响以及其他环境因素（如，学校、工作以及家庭）的影响等[23]。例如她和 Hennessey 在 1982 年做过的奖赏和任务标签对儿童创造力的影响的实验，使用了讲故事的任务，由三位小学教师作为评价者，评分者信度为 0.91，最后得到了有奖赏组的言语创造力高于无奖赏组的结果[12]。国内曲小军使用同感评估技术作过评价和奖赏对于场依存和场独立儿童的语言创造力的影响[24]。

另外，随着越来越多的研究使用同感评估技术，Amabile 发现它也可以用于个体差异的研究，例如她和她的合作者在一项关于工作的内部动机和外部动机的个体差异的研究中发现，使用同感评估技术测量的创造性与个体内部动机相关显著[25]。

在关于创造力是领域特殊性的还是领域一般性的争论中，CAT 也有经常被采用。例如，Baer 在关于发散性思维是否具有任务特殊性的研究中，使用同感评估技术评价被试在讲故事任务和诗歌任务上的创造性表现[26,27]；Han 在关于领域特殊性的研究中，使用同感评估技术研究了语言、艺术和数学领域的创造性表现[21]。

CAT 技术最初只用于对在相同的实验条件，尤其是相同的指导语要求下得到的平行作品进行评定，最近，Baer 经过试验，发现同感评估技术还可以扩展到对非平行产品的评定 [28]。他将 8 年级被试在不同实验条件下的 103 篇故事、102 篇诗歌和 103 篇个人叙述放在一起由 13 位专家进行评定，然后进行评分的一致性分析，结果故事的评分的一致性达到 0.940，诗歌达到 0.968，个人叙述达到 0.957。这表明 CAT 技术受任务和实验条件的限制可以进一步缩小，可将许多作品放在一起进行评定。这就大大扩大了 CAT 技术的适用范围。

4 对 CAT 技术的评价

为了对 CAT 技术有一个客观的评论，我们认为，将它与基于发散性思维的传统创造力测验作比较是比较合适的。

自 Guilford 提出发散性思维是创造性思维的核心以来，很多在国际上有影响的创造力测验，如 TTCT（Torrance Creative Thinking Test） [29]， WCAP （ Williams Creativity

Assessment Packet）[30], TCT-DP（Test for Creative Thinking - Drawing Production）[31]等都是基于发散性思维理论的。在 Guilford 看来，在发散性思维中表现出来的反应的低频性和不寻常性就可以代表独特性，而独特性是创造性的一个核心。然而，事实上，有创造性的东西是独特的，但独特的东西不一定是有创造性的，因为，有创造性的东西必须同时满足例外一个条件，那就是适宜性和合适性（appropriateness）。换句话说，一个新异独特的东西，如果是没有用的，那它只能算是古怪的而不是创造性的。显然，适宜性具有相对性，并且在很大程度上依赖于外界的评价。传统的创造性思维测验对此的考虑比较少，因此，测试结果的生态效度受到严重影响。而 CAT 技术则在这方面考虑的比较多，因此具有很好的生态效度。实际上，CAT 技术对创造性产品的考虑更侧重于质量而不是数量，这正好与传统的发散性思维测验不同。传统的发散性思维测验更重视数量[32]。

在理论基础上，CAT 技术的基础是人们对创造力的内隐理论，也就是人们对客观事物的主观评价是有共性的。这种观点得到了当前内隐理论研究的支持，而且，在实际的研究中也证明 CAT 技术有很好的适用范围。然而，值得提醒的是，正因为这样，CAT 技术也必然存在着致命的弱点，那就是，当一个产品的创造性程度远远超出该时代的认识程度时，该产品就会因为不能被识别而被误认为不具有创造性。历史上有不少伟大的科学发现就是因为不能在当时被评价而被延迟了很多年。不过，不管过了多久，最终能被识别也说明，创造性是要通过别人的评价才能被识别的。这一点也正好是 CAT 技术的主要思想。或许这就是为什么 CAT 技术越来越得到认可和推广的原因。

尽管，国内研究者对主观评定技术并不陌生，也在各种研究场合使用，但从目前的文献来看，用基于主观评价的 CAT 技术来开展创造力研究却并不多见，希望本文能起到抛砖引玉的作用。

参考文献

[1] Guilford J P. Creativity. American Psychologist, 1950, 5: 444~454

[2] Torrance P. The nature of creativity as manifest in its testing. In: R J Sternberg (Ed.). The nature of Creativity – Contemporary psychological perspectives. Cambridge: Cambridge University Press, 1988. 43~75

[3] 施建农 徐凡. 超常儿童发展心理学. 安徽教育出版社, 2004. 276~301

[4] Torrance E P. Guiding creative talent. Huntington, New York: Robert E, Drieger Publishing Company, 1976

[5] Plucker J A. Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. Creativity research journal, 1999, 12(2): 103~114

[6] Yamada H, Tam A Y. Prediction study of adult creative achievement: Torrance's longitudinal study of creativity revisited. Journal of creative behavior, 1996, 30(2): 144~149

[7] Kogan N, Pankove E. Creative ability over a five-year span. Child development, 1972, 43: 427~442

[8] Kogan N, Pankove E. Long-term predictive validity of divergent-thinking tests: Some negative evidence. Journal of educational psychology, 1974, 66(6): 802~810

[9] Milgram R M, Milgram N A. Creative thinking and creative performance in Israeli students. Journal of educational psychology, 1976, 68(3): 255~259

[10] Runco M A. Predicting creative performance from divergent thinking. In: Runco M A (Ed). Divergent thinking. Norwood, NJ: Ablex, 1991. 137~145

[11] Amabile T M. Social psychology of creativity: A consensual assessment technique. Journal of personality and social psychology, 1982, 43: 997~1013

[12] Amabile T M. Creativity in context: Update to the social psychology of creativity. Boulder, GO: Westview, 1996

[13] Runco M A, Bahleda M D. Implicit theories of artistic, scientific, and everyday creativity. Journal of creative behavior, 1986, 20: 93~98

[14] Runco M A, Johnson D J. Parents' and teachers' implicit theories of children's creativity: A cross-cultural perspective. Creativity research journal, 2002, 14(3&4): 427~438

[15] Sternberg R J. Implicit theories of creativity, intelligence and wisdom. Journal of personality and social psychology, 1985, 49(3): 607~627

[16] Romo M, Alfonso V. Implicit theories of Spanish painters. Creativity research journal, 2003, 15(4): 409~415

[17] Wellman H M. The Child's Theory of Mind. Cambridge: MIT Press, 1990

[18] Christiaans H H C M. Creativity as a design criterion. Creativity research journal, 2002, 14(1): 41~54

[19] Winer B. Statistical principles in experimental design. New York: McGraw-hill, 1991

[20] Nunnally J. Psychometric theory. New York: McGraw-hill, 1994

[21] Han K. Domain-specificity of creativity in young children: How quantitative and qualitative data support it. The journal of creative behavior, 2003, 37(2): 117~142

[22] Plucker J A, Renzulli J S. Psychometric approaches to the study of human creativity. In: Sternberg R J (Ed.). Handbook of creativity. Cambridge: University Press, 1999. 3~15

[23] Hennessey B A., Amabile T M. The conditions of creativity. In: Sternberg R J (Ed.). The nature of creativity. New York: Cambridge university press, 1988 [12] Baer J. Performance assessments of creativity: Do they have long-term stability? Roeper review, 1994, 17: 7~11

[24] 曲小军. 评价和奖赏对场依存、场独立儿童语言创造力的影响. 中国心理卫生杂志，2005，19（1）：408~411

[25] Amabile T M, Hill K J, Hennessey B A, et al. The work preference inventory: Assessing in trinsic and extrinsic motivational orientations. Journal of personality and social psychology, 1994, 66: 950~967

[26] Baer J. The effects of task specific divergent-thinking training. Journal of creative behavior, 1996, 30: 183~187

[27] Baer J. The case for domain specificity of creativity. Creativity research journal, 1998, 11: 173~177

[28] Baer J, Kaufman J C, Gentile C A. Extention of the consensual assessment technique to nonparallel creative products. Creativity research journal, 2004, 16(1): 113~117

[29] http://www.indiana.edu/~bobweb/Handout/d3.ttct.htm (2005-10-27)

[30] http://www.indiana.edu/~bobweb/Handout/d16.cap.html (2005-10-27)

[31] Urban K K. Assessing Creativity: The Test for Creative Thinking - Drawing Production (TCT-DP) The Concept, Application, Evaluation, and International Studies. Psychology Science, 2004, 46 (3): 387 - 397

[32] Simonton D K. Genius, creativity, and leadership. Cambridge, MA: Harward University Press, 1984

A Consensual Technique for Creativity Assessment

Song Xiaohui, Shi Jiannong

(KLMH, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China)

(Lab of Learning & cognition, Capital Normal University, Beijing 100037, China)

Abstract: Consensus assessment technique (CAT) is a method proposed by Amabile in 1982 for creativity measurement. After reviewing the features of divergent thinking tests and their limitations, the authors pointed out that it is necessary to develop an alternate technique of measuring creativity and introduced CAT method in the present paper. Meanwhile, the authors discussed the theoretical basis of CAT and exemplified its application in measuring creative articles. At the end of this article, the authors summarized features of CAT and evaluated its value, and also pointed out its advantages.

Key words: consensual assessment technique, divergent thinking test, domain-specific, implicit theory.