

认知反思测验在军人群体中的应用

买跃霞^{1,2} 巨化平¹ 肖 玮² 苗丹民²

1. 武警工程大学乌鲁木齐校区, 乌鲁木齐;

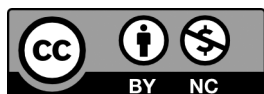
2. 空军军医大学医学心理系, 西安

摘 要 | 理性思维是军事决策中最基本的思维方式, 认知反思是获得理性思维的重要条件。本研究通过文献分析, 编制了认知反思初始测验, 并基于项目分析理论对初始测验进行题目删减与维度确定, 最终形成了单维度的认知反思测验, 并在中国军人群体中进行信效度检验。研究结果表明该测验具有良好的信效度, 符合心理测量学要求, 可以作为测量军人理性思维中反思性倾向的有效工具。

关键词 | 认知反思; 军事决策; 锚定效应; 项目反应理论; 理性思维

Copyright © 2024 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

所谓“运筹帷幄, 决胜千里”, 决策是军事活动中作战指挥的核心内容。在当代高技术战争条件下, 战争节奏加快, 战线流动性增强, 情况更加复杂多变, 对军事决策提出了更高的要求。作战决策的基本心理活动方式是思维。大量的行为学研究揭示, 人类的思维过程存在系统的局限性。卡尼曼认为人们往往无法做到完全意义上的理性, 相反各种认知缺陷、启发式以及惯性思维决定了我们的大多数判断和决策^[1]。锚定效应 (Anchoring Effect) 就是人们在不确定情境下一种决策偏差现象。锚定效应在军事上也屡见不鲜。早在春秋时期, 《孙子兵法》就提出了“能而示之不能, 用而示之不用, 近而示之远, 远而示之近”“增兵减灶”等军事对抗谋略。从现代心理学意义上来看, 这些谋略其实就是为误导敌方的思维判断所设定的“沉锚”^[2]。

面对现代战争中明暗相间、真假难辨、层次交叉的信息, 作战指挥员必须综合运用多种思维方法, 克服惯性思维, 避免决策偏差, 得出正确决策, 指导作战行动。其中理性思维决策因其运用科学的程序和方法, 按照严格的逻辑规则进行分析判断, 能够比较科学地对多变量、多因素的复杂的作战问题进行决策的特点, 具有广泛的适用性, 尤其适用于对较高层次的大规模作战行动进行思考决策, 是指挥员最

通讯作者: 买跃霞, 空军军医大学应用心理学博士, 武警工程大学乌鲁木齐校区教研部军事心理学教研室讲师, 研究方向: 军事决策与军人健康维护。

文章引用: 买跃霞, 巨化平, 肖玮, 等. 认知反思测验在军人群体中的应用 [J]. 中国心理学前沿, 2024, 6 (4): 456-464.

<https://doi.org/10.35534/pc.0604051>

基本和最常用的思维方式。

加拿大多伦多大学应用心理学和人类发展学系荣誉教授基思·斯坦诺维奇 (Keith E. Stanovich) 提出理性思维和智力是两种不同的认知功能,但也可以像智力一样,是可测量的认知能力。根据斯坦诺维奇的观点,认知反思是获得理性思维的重要条件^[3]。到目前为止,最广为接受的衡量这种倾向差异的测验当属弗雷德里克 (Frederick) 编制的认知反思测验 (Cognitive Reflection Test)^[4]。该测验自 2005 年发表以来,已经被引用了 700 多次。CRT 因为测验时间短,同时又是理性思考和决策能力的最佳预测指标之一,激发了许多学者的研究兴趣。研究者们又相继编制了新的认知反思测验:托普拉克、韦斯特和斯坦诺维奇 (Toplak, West, and Stanovich, 2013) 编制的认知反思测验 (CRT)^[5],普里米等人 (Primi et al., 2016) 编制的认知反思测验 (CRT—L)^[6],汤姆森和奥本海默 (Thomson and Oppenheimer, 2016) 编制的认知反思测验 (CRT—2)^[7]。国内学者对 CRT 测验的研究还很少,并且都采用的最早期弗雷德里克 (Frederick, 2005) 年编制的。本研究将基于近几年国外不同版本的 CRT 测验进行翻译,并在项目反应理论指导下对各项目进行功能分析,考察各项的评估效力,进一步精简和修订中文版军人认知反思测验,并检验其信效度。

2 研究方法

2.1 研究对象

样本一:采用整群方便抽样,抽取某部队 440 名军人施测中文版认知反思测验,其中男 418 人,女 22 人,平均年龄 23 ± 4 岁,平均军龄 3 ± 4 年,其中受教育程度为:初中学历占 4.2%、高中学历为 28.1%,中专学历为 18.9%,大专学历 31.9%,本科学历 16%,研究生学历为 0.9%。样本一的数据用于认知反思测验项目分析后的精简和修订。

样本二:采用整群方便抽样,抽取某部队 387 名军人,均为男性,平均年龄为 22 ± 4 岁,军龄 3 ± 4 年,其中受教育程度为:初中学历 3.1%、高中、中专学历为 52.3%、大专学历为 35.3%、本科及研究生学历 9.3%。样本二的数据用于修订后的认知反思测验的信效度分析。

2.2 研究工具与程序

样本一只接受认知反思测验,样本二先完成锚定决策任务,然后接受修订后的认知反思测验。测验采用纸笔团体施测方法,测验过程没有时间要求,测验完成后当场检查完成情况并回收。

中文版认知反思测验:修订托普拉克 (Toplak)、韦斯特 (West) 和斯坦诺维奇 (Stanovich) 编制的 7 个条目的认知反思测验,普里米等人 (Primi et al., 2016) 编制的 6 个条目的认知反思测验,汤姆森和奥本海默 (Thomson and Oppenheimer, 2016) 编制的 4 个条目的认知反思测验 (共 13 个条目,4 个条目是重复的)。本研究具体翻译与修订过程如下:本研究首先征得源量表作者的授权同意,然后严格按照 Brislin 翻译模式将英文版翻译。首先,将这 13 个条目由五位国内心理学决策相关研究者译成中文;其次,由一名留美的博士研究生把量表译回英文,与英文版量表的原文进行对比,修改后确定中文版的认知反思测验。请 3 名研究生完成修改后的测验,并对他们进行深入访谈,确认无理解存在歧义的项目。

锚定决策任务:选取 6 个自编问题和 2 个经典锚定决策问题,共 8 个问题作为锚定决策研究的实验材

料。例如：“非洲国家所占联合国国家比例高于 10%（65%）还是低于 10%（65%）？”。其中自编锚定问题中高低锚值的确定：分别选取预实验中被试在每个问题上估计值最高的 27% 的中位数和估计值最低的 27% 的中位数作为分别作为相应问题的高、低锚。6 个自编的锚定决策问题的适用性和有效性在预实验中均已验证。锚定问题决策研究中，研究者让被试进行标准范式下的锚定判断：先判断“甘地的寿命是长于还是短于 x 岁？”，再估计“甘地的寿命是多少岁？”。8 个锚定决策问题固定顺序呈现，但高低锚值的设置顺序分成两种，一种为高锚、低锚……；另一种为低锚、高锚……，两种条件在被试间平衡。

2.3 数据处理

认知反思测验的计分方法，将被试的答案分为三类：正确答案、直觉答案、非典型答案。如果答案属于正确答案，则该项目记 1 分，被试答对题目的总和为认知反思测验的最后得分。

锚定决策任务中的原始数据处理方法为：首先，用 Tukey's Test 方法即箱线图法检测被试在锚定决策任务中的绝对估计值的极端值。然后，用最大估计值替换大于最大估计值的极端值，用最小估计值替换小于最小估计值的极端值（共有 8 个极端值被替换）。最后，将 8 道锚决策问题的绝对估计值进一步转化为以 0 为平均数，1 为标准差的 z 分数形式，并算出每名被试在高锚值条件下估计值的平均值和低锚值条件下估计值的平均值。

采用 Epidata3.1 对原始数据进行录入，并采用 SPSS 21.0 与 JASO 统计软件对实验数据进行描述统计与信度检验与方差分析；应用 IRTPRO（Cai, Thissen, and du Toit, 2011）软件进行项目分析；采用 Mplus 7.0 软件进行验证性因素分析。

3 结果与分析

3.1 测验条目的初步分析

分析结果表明，大部分被试的反应要么是直觉反应，要么是正确答案。根据弗雷德里克（Frederick）最初 CRT 测验的条目要求：至少 80% 被试要在条目上做出要么是正确答案，要么是直觉反应。被试在条目 1、2、3、5、7、9、10、13 的反应符合要求，条目 4、6、8、11、12，不符合要求，故删除这 5 个条目，如表 1 所示。

表 1 被试在各条目上的作答情况与参数估计

Table 1 The responses and parameter estimates of participants on items

项目	% C (H)	a (SE)	b (SE)
1	45.5 (53.0)	0.72 (0.14)	0.28 (0.15)
2	65.5 (20.7)	0.19 (0.12)	-3.47 (2.26)
3	55.2 (36.8)	1.47 (0.21)	-0.24 (0.09)
5	33.4 (54.0)	1.53 (0.22)	0.65 (0.11)
7	42.5 (36.6)	3.08 (0.52)	0.24 (0.07)
9	41.4 (42.3)	2.33 (0.34)	0.28 (0.08)
10	38.0 (34.5)	0.60 (0.13)	0.88 (0.24)
13	45.9 (36.1)	2.09 (0.30)	0.14 (0.07)

3.2 项目分析

按照项目反应理论要求,编制和修订的量表都必须是单维的,即测验只测量被试的某一种潜在特质,测验结果受其他潜在特质的影响可以忽略不计。被试对测验中任何一个项目的反应是该单一特质 θ 的函数。因此,本研究因素分析法进行单维性假设检验。在单维性检验之前,需要先对数据进行 KMO 和 Barlett 的检验。本研究的 KMO 值为 0.860,且 Barlett 的球形检验的 p 值为 <0.001 ,说明样本数据非常适合进行因素分析。接下来使用因素分析法进行单维性假设检验。分析结果表明,第一因子特征根值与第二因子特征根值分别为 3.31 与 1.11,它们分别能解释的方差为 40.17% 与 11.68%,符合单维性假设的 Hambleton 标准,可以进行 IRT 项目质量分析。

本研究采用双参数模型(2PL)对测验项目进行分析。使用项目反应理论分析各项目时,采取适用于边际极大似然估计法进行参数估计,并根据各个项目的区分度参数(a)、难度参数(b)、和项目特征曲线(ICC)最终挑选出适合的项目。各项目参数具体情况见表 1、图 1 和图 2。

以往研究提示, a 的取值范围一般设置在 0 到 2.5 或 3 的区间。区分度 b 取值与被试的潜在能力 θ 的取值范围相同,均为正负无穷之间,但在 θ 取标准分数的量表中,大部分的 b 取值与 θ 取值范围一样在正负 4 之间。项目特征曲线(ICC)越陡峭,覆盖面积越多,说明该项目对大部分被试能提供比较精确的特质水平估计。本测验项目特征曲线矩阵显示,项目 2 曲线过于平缓,区分能力较差。因此根据表 1 中项目的区分度值和图 1 中的项目特征曲线,删掉项目 2,7。

删除条目后,1、3、5、9、10、13 共 6 个题目组成的认知反思测验中项目区分度参数 a 在 0.59 ~ 2.53 之间,难度参数 b 在 -0.25 ~ 0.90 之间。本测验信息量在 -1 到 2 这个区间在 7 以上,基本能够达到评估效果。

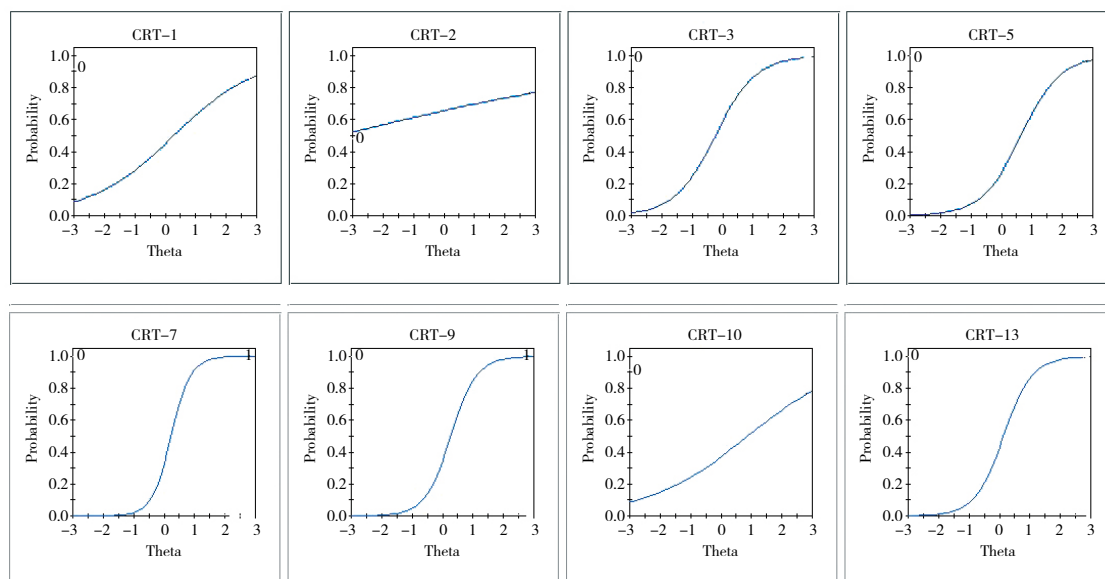


图 1 各条目的项目特征曲线

Figure 1 Item characteristic curve

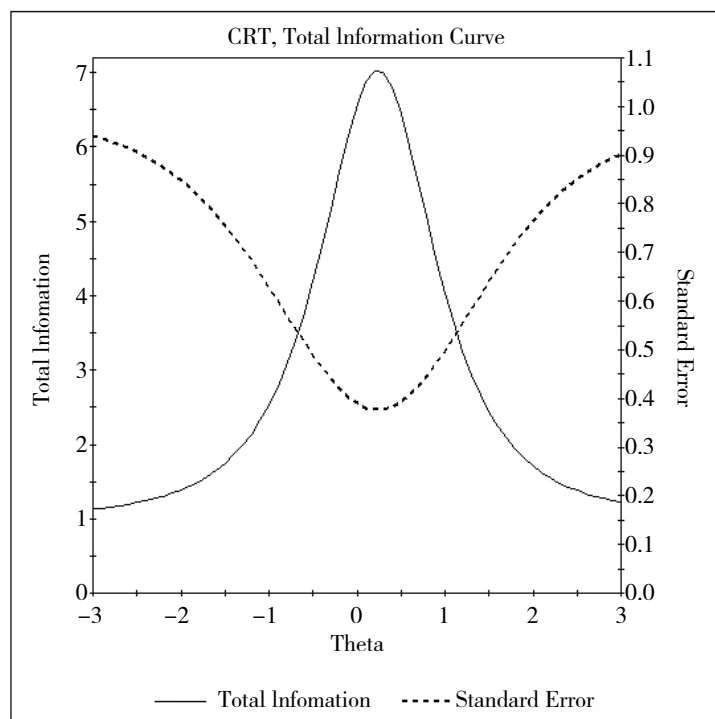


图2 认知反思测验测验信息函数曲线

Figure 2 The information function of cognitive reflection test

3.3 结构效度分析

根据以往研究,验证性因素分析可以作为结构效度的评价。本研究基于样本2的测验数据采用验证性因素分析(CFA),验证测验项目是否具有合理性和结构稳定性。各拟合程度指数见表2。

表2 CFA 模型拟合指数

Table 2 Model fitting index of CFA

模型	χ^2	df	χ^2/df	p	CFI	NNFI	RMSEA (90%)
结构	29.24	9	3.25	0.000	0.940	0.904	0.071 (0.044, 0.101)

在众多的拟合性指标中,拟合指数 χ^2 作为最常报告的一个,是最主要的检验指标。它与自由度一起使用说明模型正确性的概率, χ^2/df 越接近 1,表示协方差矩阵和估计的协方差矩阵之间的相似程度越大,模型的拟合度越好。实际研究中, χ^2/df 接近 2,可以认为模型的拟合性是比较好的。在样本量大的情况下, $\chi^2/df < 5$ 就可以接受。本研究中 χ^2/df 为 3.25,认为模型拟合较好。

比较拟合指数 (comparative fit index, CFI) 作为目前使用最广泛的指标之一,也是最稳健的指标之一 (Fan, Thompson, and Wang, 1999) [8]。本研究中,上述指标都在 0.9 以上,说明模型拟合性很好。

近似误差均方根 (root mean square error of approximation, RMSEA) 是一种简约测量, 值越大, 拟合越差, 可接受的取值范围在 0 ~ 0.1 之间, 当值等于 0 则为完美拟合, 当值小于 0.05 时表示模型精确拟合, 当值在 0.05 ~ 0.08 之间为合理拟合, 当值 0.08 ~ 0.1 之间为拟合不良 (Hu and Bentler, 1990) [9]。本研究结果显示模型拟合指标都符合统计要求, 模型有较好的拟合度和稳定性, 修订后的中文版认知反思测验符合单维度结构。

3.4 区分效度分析

根据认知反思测验的总分进行分组, 总分排序后前 27% 为高分组, 后 27% 为低分组。分组后, CRT 低分组 = 0.42 ± 0.50 ; CRT 高分组 = 4.92 ± 0.68 。在本次研究中, 以被试在高低锚定条件下估计值的 z 分数平均值为因变量, 锚类型以及认知反思水平为自变量, 进行方差分析。理论假设: 如果该测验能够有效区分出被试的认知反思水平, 那么高分组和低分组收到锚定效应的影响程度不一样即锚定效应会随着个体理性思维倾向的提高而降低, 结果如表 3、表 4 所示。

锚值类型主效应显著 ($F(1, 193) = 284.49, p < 0.001$), 认知反思主效应显著 ($F(1, 193) = 18.75, p < 0.001$), 认知反思水平和锚值类型交互效应显著 ($F(1, 193) = 36.33, p < 0.001$)。由于交互效应显著, 进一步对不同认知负荷水平下, 被试的锚定效应进行简单效应分析, 结果表明在低锚值下, 低认知反思水平被试的估计值 ($M = -1.58$) 低于高水平被试的估计值 ($M = -1.21$) ($F(1, 193) = 3.381, p = 0.052$), 达到边缘显著; 而高锚值条件下, 低认知反思水平被试的估计值 ($M = 2.35$) 显著高于高水平被试的估计值 ($M = 0.649$) ($F(1, 193) = 40.80, p < 0.000$), 图 3 为交互效应图。

表 3 不同条件下被试估计值的平均值与标准差

Table 3 The average and standard deviation of estimated values of participants

锚类型	认知反思水平	
	低水平组 (n=112)	高水平组 (n=90)
低锚	-1.58 ± 0.13	-1.21 ± 0.13
高锚	2.35 ± 0.18	0.649 ± 0.19

表 4 不同条件下被试估计值比较

Table 4 Comparison of estimated values of participants

变异来源	df	MS	F	p
锚值类型	1	819.10	284.49	0.000
锚值类型 × 认知反思	1	104.61	36.33	0.000
组内误差	193	2.88		
认知反思	1	43.48	18.75	0.000
组间误差	193	2.32		

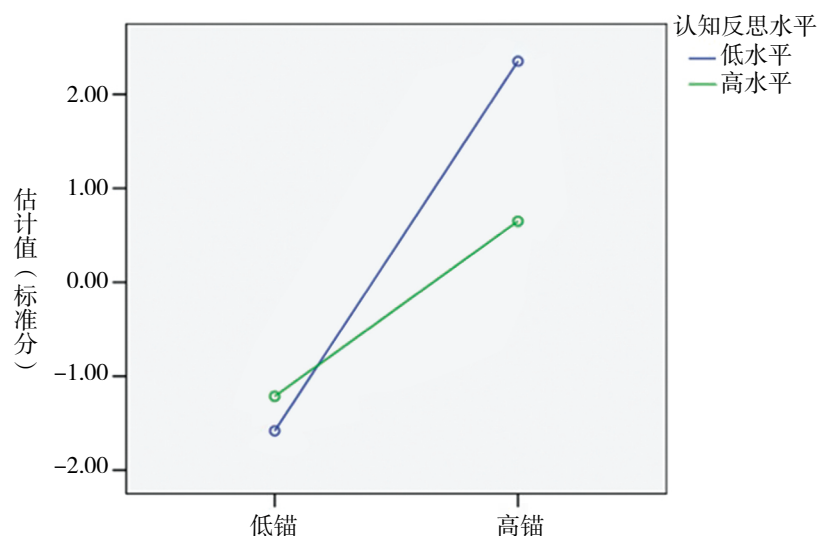


图 3 锚值类型与思维倾向的交互作用图

Figure 3 Interaction diagram between anchor types and thinking tendencies

3.5 信度分析

在本研究中,认知反思测验的Cronbach's α 系数为0.66。以往研究表明,一个含10个左右项目的量表, α 系数应能达到0.80以上;一个4个项目的量表,信度系数 α 可能会低于0.6或0.5。虽然中文版认知反思测验在该研究中为0.66,但已经达到了心理测量学对信度的要求,进一步表明中文版的认知反思测验具有良好的信度。这也与国外研究中CRT测验的信度系数是0.57~0.68相一致。

4 结语

在过去50年的心理与认知科学研究中,认知“懒惰”一直是大多数双加工理论的重要主题。由于人类的认知“懒惰”,人们往往倾向于具有普遍性的默认加工方式,导致在决策中出现非理性偏差。但人类的认知“懒惰”性还具有极大的个体差异性。斯坦诺维奇(Stanovich)和韦斯特(West)致力于研究个体在认知“懒惰”性的差异^[10],他们认为,人们在抑制不正确的优势反应并进行进一步反思以获得正确反应的可能性方面并不相同。也就是说,尽管人们时常不可避免地掉入很多“直觉陷阱”,导致决策偏差,但仍不乏有一些优秀的决策者,在某些情况下能有效克服认知“懒惰”带来的消极影响,做出了比普通人更为理性的决策与判断。

CRT测验反映了个体在收集和处理信息的过程中表现出来的思维反思性和控制性倾向,在衡量理性思维水平高低过程中起到了重要的作用。本研究翻译并修订了国外认知反思测验,在修订过程中,对题目进行多次修改和筛选,并对单维进行了验证,对量表的信效度进行检验,最终确定该量表具有良好的心理测量学指标。其中,为了检验测验是否有效地区分了高低认知反思能力,本研究进一步考察该测验区分出的高低认知反思水平群体在锚定效应中表现,发现高低认知反思水平组都会受锚定效应的影响,但高认知反思水平的被试受到锚定效应的程度要小即锚定效应会随着个体理性思维倾向的提高而降低。结合弗雷德里克(Frederick, 2005),托普拉克、韦斯特和斯坦诺维奇(Toplak, West, and Stanovich,

2011) 关于认知反思能力在其他直觉与偏差任务中的表现, 本研究进一步验证了该 CRT 测验能预测理性思维, 在决策中的表现出更少的非理性偏差。

思维认知是战争对抗的最根本依托、最深层力量^[11]。随着技术的发展及海量信息的出现, 个人的认知能力将不足以确保做出明智和及时的决策, 人类思维已成为战争的终极领域。如何应对人类思维固有的脆弱性、提高决策优势, 是打未来信息战的重要保证。理性思维是研究规避思维局限性的重要方向。但本研究还存在一定的局限性, 从测量角度来看, 理性思维重点考察的是人们的思维活动是否会受到原有的知识、经验和信念的干扰或误导而导致决策偏差, 在这类测验任务中通常会有意创造直觉与理性判断冲突情景, 以考察受试者能否避免或推翻直觉所带来的误导性影响。CRT 测验在衡量理性思维水平高低过程中起到了重要的作用, 但离一个理性心理测量工具还有一定的距离。

参考文献

- [1] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases [J]. Science, 1974, 95 (4157): 1124-1131.
- [2] 宋永健. 指挥决策要防止“锚定效应” [N]. 解放军报, 2015-03-31 (6).
- [3] 基思·斯坦诺维奇. 超越智商: 为什么聪明人也会做蠢事 [M]. 机械工业出版社, 2015: 36-50.
- [4] Frederick S. Cognitive Reflection and Decision Making [J]. Journal of Economic Perspectives, 2005, 19 (4): 25-42.
- [5] Toplak M E, West R F, Stanovich K E. Assessing miserly information processing: An expansion of the Cognitive Reflection Test [J]. Thinking & Reasoning, 2013, 20 (2): 147-168.
- [6] Primi C, Morsanyi K, Chiesi F, et al. The Development and Testing of a New Version of the Cognitive Reflection Test Applying Item Response Theory (IRT) [J]. Journal of Behavioral Decision Making, 2016, 29 (5): 453-469.
- [7] Thomson K S, Oppenheimer D M. Investigating an alternate form of the cognitive reflection test [J]. Judgment and Decision Making, 2016, 11 (1): 99-113.
- [8] Fan X, Thompson B, Wang L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes [J]. Structural Equation Modeling: A Multidisciplinary Journal, 1999, 6 (1): 56-83.
- [9] Hu L T, Bentler P M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives [J]. Structural Equation Modeling: A Multidisciplinary Journal, 1999, 6 (1): 1-55.
- [10] Stanovich K E, West R F. Individual differences in reasoning: implications for the rationality debate? [J]. Behavioral & Brain Sciences, 2000, 23 (5): 665-726.
- [11] 陈东恒. 制认知权: 战争制胜重要支撑 [N]. 解放军报, 2022-04-19 (7).

Application of Cognitive Reflection Test in China Military

Mai Yuexia¹ Ju Huaping¹ Xiao Wei² Miao Danmin²

1. Urumqi Campus, Engineering University of PAP, Urumqi;

2. Department of military medical psychology, Air Force Medical University, Xi'an

Abstract: Rational thinking is the most basic way of thinking in military decision-making, and cognitive reflection is an important condition for obtaining rational thinking. This study developed a cognitive reflection initial test through literature analysis, and based on item response theory, the initial test was subjected to question reduction and dimensionality determination, ultimately forming a one-dimensional cognitive reflection test. The reliability and validity of the test were tested in the Chinese military population. The results show that the test has good reliability and validity, meets the requirements of psychometrics, and can be used as an effective tool to measure the reflective tendency in the rational thinking of soldiers.

Key words: Cognitive reflection; Military decision-making; Anchoring effect; Project response theory; Rational thinking