

DOI:10.3724/SP.J.1008.2015.00172

2014年国际坦克大赛乘员选拔心理项目效度初步研究

魏存¹, 邓光辉^{1*}, 沈兴华¹, 张勇²

1. 第二军医大学心理与精神卫生学系军事心理学教研室, 上海 200433

2. 解放军 73049 部队医院, 苏州 215008

[摘要] **目的** 探讨 2014 年国际坦克大赛乘员选拔心理测评项目的有效性。**方法** 选择内田-克雷佩林测验、艾森克人格问卷、EP707A 手指灵活性测试仪、EP706 注意分配仪和情境测试, 对 36 名大赛候选人(21~33 岁)进行逐项心理测量。

结果 心理选拔结果的综合排名与部队排名具有良好的拟合度, $\kappa \geq 0.75$ 。**结论** 5 项测验项目对选拔对象作业绩效有良好的预测作用, 证明测评项目具有良好效度。

[关键词] 军事心理学; 心理选拔; 评价中心; 会聚操作; 效标

[中图分类号] R 395 **[文献标志码]** A **[文章编号]** 0258-879X(2015)02-0172-05

Validity of psychological test items for selecting crew members for the 2014 World Tank Championship

WEI Cun¹, DENG Guang-hui^{1*}, SHEN Xing-hua¹, ZHANG Yong²

1. Department of Military Psychology, Faculty of Mental Health and Psychology, Second Military Medical University, Shanghai 200433, China

2. Hospital of No. 73049 Troop of PLA, Suzhou 215008, Jiangsu, China

[Abstract] **Objective** To study the validity of psychological test items for selecting the crew members for the 2014 World Tank Championship. **Methods** Totally 36 candidates (21-33 years old) were assessed with items in Uchida-Kraepelin test, EPQ, EP707A Fingers Flexibility Instrument, EP706 Attention Distribution Instrument and Situational Judgement test. **Results** The comprehensive rankings of psychological test results well fit with their rankings in the military competence, with kappa value higher than 0.75. **Conclusion** The five tests in this study can well predict the performance of the subjects, indicating that the involved items have satisfactory validity.

[Key words] military psychology; psychological screening; assessment centre; converging operation; criterion

[Acad J Sec Mil Med Univ, 2015, 36(2): 172-176]

心理选拔在军事领域的重要性早已深入人心。在众多心理选拔方法中,“评价中心”(assessment center)由于采用多项技术、多个评估者和将全部信息进行综合分析^[1],有效降低测量和评价误差而得到广泛应用,成为军队心理选拔的重要工具。随着研究和应用的不断深入,虽然评价中心具有良好的内容效度^[2]、效标关联效度^[3]等,但其构想效度问题却日益受到质疑^[4-6]。大量研究发现,评价中心始终没有按照预先设想的那样以维度为基础实现有效测量,其会聚效度低于区分效度,评分变异中活动变异

总是大于维度变异^[7-9],由此引发了评价中心是维度中心取向、活动中心取向还是交互作用取向的争论。虽然至今无定论,但未来评价中心应在传统的维度中心取向的基础上,给予活动中心取向足够重视,并加强两者之间的融合,已成为很多学者的共识。然而,基于任务的评价中心提出后十多年,在实践中虽有一些应用,但国内实证研究非常缺乏。

2014 年我军参加了俄罗斯国际军事坦克大赛,其中乘员的心理选拔采用了以活动或任务为导向的评价中心。从坦克乘员工作分析中得到岗位关键任

[收稿日期] 2014-10-31 **[接受日期]** 2014-12-09

[基金项目] 全军医学科技“十二五”课题(CWS12J015)。Supported by PLA Research Project of “the 12th Five-Year-Plan” for Medical Science and Technology (CWS12J015).

[作者简介] 魏存, 博士, 讲师。E-mail: oyangxueer@yeah.net

* 通信作者(Corresponding author). Tel: 021-81871677, E-mail: bfbedu@126.com

务模型,结合国际比赛高竞争、高压特点,尽可能创设与岗位相关的测评情境。经过专家评估关键工作任务的模拟难度与成本,选择坦克乘员设备操作、竞赛压力下的行为表现、团体协作的行为表现为测评情境,通过模拟操作测验、模拟情境测验、纸笔测验等技术的会聚操作,给予被试者跨情境综合评估。本研究为此次心理选拔项目的效度做初步研究。

1 对象和方法

1.1 对象 通过部队专业技能筛选的36名候选装甲兵,其中通信岗位、射击岗位和驾驶岗位各12人,均为男性,年龄21~33岁,平均(25.33±2.67)岁。

1.2 效标的确定 依据候选人的训练情况、考核成绩、健康状况等,由训练单位根据候选人综合打分情况给予总体评价并进行排名。

1.3 测试指标与工具的确定 前期,通过访谈法进行坦克乘员任务分析,获得岗位关键工作任务包括驾驶操作、射击操作、岗位配合。同时,考虑到作业情境即国际性竞赛高竞争、高压力的特点,通过专家咨询法,最终确定本次选拔主要测试的关键岗位任务为坦克乘员的设备操作、压力下的抗压行为表现、团体协作的行为表现。根据关键岗位任务,选择如下测量方法。

内田-克雷佩林测验(Uchida-Kraepelin test)^[10]:该测验是一种作业法心理测验,也称UK测验,要求被试在时间压力下做一位数的连续加法计算,可反映被测者耐受性、稳定性和作业能力等心

理指标。在本次评估中可较准确评估被测者在压力下的综合作业表现。根据被试完成的作业量、作业曲线和错误率等情况,由5名心理学专业的老师以1~10分来进行等级打分,最终综合5位老师的打分给予优秀、良好、较差三个等级评价。评分结果按7.01~10.00分、4.01~7.00分、1.00~4.00分的标准,分别划分为优秀、良好、较差3个等级。

注意分配测验:选用EP706型注意分配仪,主要反映被测者在压力下短时间内快速辨别声音的能力和声、光注意分配能力。根据单任务和双任务两种情形下被试作业的正确性和复杂反应时两个指标的测量结果划分不同标准:0.50~0.99s为优秀,1.00~1.20s为良好,1.21~2.00s为较差。

手指灵活与协调测验:选用EP707A型手指灵活性测试仪,主要反映被测者手指灵活与协调性,以及在压力下的双手协调反应能力和动作记忆能力。根据手指灵活性测试仪的两次操作时间计算平均时间,划分标准如下:平均时间1.20~1.80s为优秀,1.81~2.50s为良好,2.51~4.00s为较差。

情境测验:该测验要求被测者以12人为一个小组参与团体项目“不倒森林”,以小组为单位完成一项高压任务。根据被测者的行为表现评估其压力情境下的紧张度。每组有5名心理学专业人员同时进行观察并记录。5名评估者分别从攻击性、持久性、出错率、紧张度、团队协作、关照他人6个方面进行评估。评分标准见表1。

表1 情境测试关键工作行为列表与评分标准

Tab 1 Key work behavior lists and gradings in the scenarios test

Key work behaviors	List	Gradings
Collaboration during operation	Take care of others, good teamwork	1: Take no care of companion, no sense of team; 2: Take good care of companion, encourage everyone occasionally; 3: Take excellent care of companion, encourage everyone frequently;
Tension during operation	Stiff action, error rate	1: Very stiff action; high error rate; 2: Mildly stiff action; low error rate; 3: Flexible motion; nearly no error;
Stability during operation	Aggressive words and deeds, be discouraged	1: Very aggressive in words and deeds, very discouraged; 2: Mildly aggressive in words and deeds, a little discouraged; 3: Not aggressive in words and deeds, not discouraged;

艾森克人格问卷^[11]:该测验为自陈测验,由龚耀先等修订,包括E量表(内外向性)、N量表(神经质)、P量表(精神质)和L量表(效度量表)4个分量表,共88项。本次评估中主要用于剔除情绪不稳定人员,将P量表得分或N量表得分 ≥ 16 分者予以淘汰。

1.4 统计学处理 采用SPSS18.0软件,运用描述性统计、相关性分析、kappa检验等进行数据处理。

2 结果

2.1 各种测验方法得分情况 在内田-克雷佩林测试中,通信岗位候选人中优秀、良好、较差人员分别占13.9%、11.1%和8.3%;射击岗位候选人中优秀、良好、较差人员各占11.1%;驾驶岗位候选人中优秀、良好、较差人员分别占8.3%、13.9%和11.1%。见表2。

在注意分配能力测试中,通信岗位候选人单任务得分优秀、良好、较差人员分别占16.7%、13.9%和2.8%;双任务得分优秀、良好、较差人员分别占

16.7%、11.1%和5.6%。射击岗位候选人单任务得分优秀、良好、较差人员分别占16.7%、11.1%和5.6%;双任务得分优秀、良好、较差人员分别占11.1%、13.9%和8.3%。驾驶岗位候选人单任务得分优秀、良好、较差人员分别占13.9%、13.9%和5.6%;双任务得分优秀、良好、较差人员分别占8.3%、16.7%和8.3%。见表3。

表2 不同岗位人员内田-克雷佩林测验结果

Tab 2 Uchida-Kraepelin test results of crew members of different missions

N=36, n(%)

Group	Excellent	Good	Bad
Communication	5(13.9)	4(11.1)	3(8.3)
Shooting	4(11.1)	4(11.1)	4(11.1)
Driving	3(8.3)	5(13.9)	4(11.1)
Total	12(33.3)	13(36.2)	11(30.5)

表3 不同岗位人员注意分配仪测量结果

Tab 3 Attention distribution measurements of crew members of different missions

N=36, n(%)

Group	Single task (sound stimuli)			Dual-task (sound-light stimuli)		
	Excellent	Good	Bad	Excellent	Good	Bad
Communication	6(16.7)	5(13.9)	1(2.8)	6(16.7)	4(11.1)	2(5.6)
Shooting	6(16.7)	4(11.1)	2(5.6)	4(11.1)	5(13.9)	3(8.3)
Driving	5(13.9)	5(13.9)	2(5.6)	3(8.3)	6(16.7)	3(8.3)
Total	17(47.3)	14(38.9)	5(14.0)	13(36.1)	15(41.7)	8(22.2)

手指灵活性测试中,通信岗位候选人中优秀、良好、较差人员各占11.1%;射击岗位候选人中优秀、良好、较差人员分别占13.9%、11.1%和8.3%;驾驶岗位候选人优秀、良好、较差人员分别占11.1%、8.3%和13.9%。见表4。

表4 不同岗位人员手指灵活性测试仪测量结果

Tab 4 Results of finger flexibility test of crew members of different missions

N=36, n(%)

Group	Excellent	Good	Bad
Communication	4(11.1)	4(11.1)	4(11.1)
Shooting	5(13.9)	4(11.1)	3(8.3)
Driving	4(11.1)	3(8.3)	5(13.9)
Total	13(36.1)	11(30.5)	12(33.4)

一致性系数 $kappa \geq 0.75$,综合5位老师的打分确定被试该项的得分。通信岗位候选人中优秀、良好、较差人员分别占13.9%、13.9%和5.6%;射击岗位候选人中优秀、良好、较差人员分别占13.9%、11.1%和8.3%;驾驶岗位候选人优秀、良好、较差人员分别占11.1%、8.3%和13.9%。见表5。

表5 不同岗位人员情境模拟测试结果

Tab 5 Results of scenarios test of crew members of different missions

N=36, n(%)

Group	Excellent	Good	Bad
Communication	5(13.9)	5(13.9)	2(5.6)
Shooting	5(13.9)	4(11.1)	3(8.3)
Driving	4(11.1)	3(8.3)	5(13.9)
Total	13(36.1)	11(30.5)	12(33.4)

评价结果经内部一致性检验,5位老师的评价

艾森克人格问卷在整个选拔中主要用于淘汰不合格者。凡神经质得分,即N量表得分 ≥ 16 分者予以剔除。最终射击和驾驶组各有1名选手被淘汰。

2.2 各测量方法等级评分结果的相关性 对内田-

克雷佩林测验、模拟操作测验以及模拟情境测验的评分结果进行相关分析,除情境模拟与手指灵活性评价结果无显著相关外,其余各项测试均两两相关($P < 0.05$ 或 $P < 0.01$)。见表6。

表6 各项测试评分结果的相关分析

Tab 6 Correlation analysis of results from different tests

	Uchida-Kraepelin test	Attention distribution	Fingers flexibility	Situational judgement
Uchida-Kraepelin test	1	0.50**	0.51**	0.67**
Attention distribution	0.50**	1	0.49**	0.38*
Fingers flexibility	0.51**	0.49**	1	0.32
Situational judgement	0.67**	0.38*	0.32	1

* $P < 0.05$, ** $P < 0.01$

2.3 心理选拔综合排名与部队排名符合度 对艾森克人格测试外的所有他评项目进行一致性检验,剔除掉差异较大评价者的评分,剩下4名评价者评分内部一致性系数 $\kappa \geq 0.75$ 。评分者集中讨论,根据不同岗位胜任能力的特征需求,给予各分测验结果不同的权重,最终综合评分达成一致,并给出心理综合排名。为检验心理选拔综合排名和部队排名的符合度,进一步做两者间一致性检验,结果 $\kappa \geq 0.75$,表明心理选拔综合排名与部队排名一致性较好。见表7。

只要根据关键任务行为的清单^[14-15]及评分标准,观察并记录候选人员在测评中的行为表现,不需要再根据其行为表现在测评维度上评分。总而言之,是以岗位关键任务为蓝本,构建一系列工作样本模拟,并对受测者在各个活动内完成任务的行为表现进行评价^[8,16]。在实施过程中也没有完全摒弃维度中心取向。为了兼顾被试行为背后个人特质因素和情境因素两方面的影响,避免出现跨情境时行为不一致性^[1],我们还是在维度的基础上,列出与测评维度相关的关键任务行为清单。如情境测试中,分别从攻击性、稳定性和协作性3个基本的元维度出发,列出对应的行为列表,将测评维度操作化,从而尽可能提高预测的时效性。

表7 心理选拔综合排名与部队排名一致性分析

Tab 7 Consistency of psychological selection rankings with rankings in military competence

Rankings in psychological evaluation	Rankings in military competence			Total
	Excellent	Good	Bad	
Excellent	10	1	1	12
Good	1	12	1	14
Bad	1	1	8	10
Total	12	14	10	36

3 讨论

一直以来,“评价中心”是西方国家广为流行的评价和选拔管理人员的综合性测评方法^[12]。相对于以维度为导向的评价中心,本次选拔所尝试的以任务为导向的评价中心操作起来更为简单,不需要过多地为了构建测评维度而搜集大量岗位信息,并对信息进行仔细而繁琐的提炼^[13]。在评价过程中,

从结果来看,各种测评方法评价结果彼此验证,存在较高相关。对内田-克雷佩林测验、注意分配操作测验、手指协调性操作测验以及情境模拟测验的评价结果进行相关分析,结果发现,除情境模拟测验与手指灵活性测验评价结果无显著性相关外,其余各项测试均两两呈显著相关。在对各项测验结果进行综合时,评分者们要么通过集中讨论使评分结果达成一致,要么用统计方法对评分结果进行整合^[17]。本次测评的总体评分并非各分项目评分的算术平均值,而是采用评分者主观评分,根据不同岗位胜任能力的需求,对分测验得分在总评价中予以不同的权重,并采取集中讨论使评分结果达成一致的方法。在进一步对最终心理测评评价结果与部队评价结果进行内部一致性检验发现,两者具有很好的一致性($\kappa \geq 0.75$),表明5项测验项目对选拔对象作业绩效有良好的预测作用,证明测评项目具

有良好效度。探究原因,36名候选人是经过部队层层严格的考核而筛选出来的岗位胜任者,而选拔中的测评情境是紧扣实际岗位任务而设计的,这说明创设无比接近真实岗位任务的模拟情境对预测至关重要^[18]。

此外,本次选拔强化“筛选与控制”^[19],兼顾选优与劣汰。由于个体心理特点既有相对的稳定性,又有一定程度的可变性,所以仅凭一次心理选拔,难以全面地评价人,而后期针对比赛进行为期1年的高强度训练也会在很多方面给人带来磨炼甚至改变。因此,我们在本次候选人员的心理选拔中坚持了“筛选-控制”的理念,在一些必备的心理素质上严格筛除掉不合格人员。比如艾森克人格测量中,精神质得分和神经质得分很高的被试者直接给予淘汰。但对待那些可发展的能力和品质,鉴于后期还有1年的针对性教育训练,采取保留的原则和态度。内田-克雷佩林测试、注意分配测试以及手指灵活性测试结果都有着较好的区分度,对于成绩中等或偏下的被试者不盲目淘汰,结合其他几项测验结果,综合评价,强化“筛选-控制”理念,融合选拔与训练,有利于形成动态、连续选拔体系,实现心理选拔的静态与动态平衡。

4 利益冲突

所有作者声明本文不涉及任何利益冲突。

[参考文献]

- [1] 卞冉,高钦,车宏生. 评价中心的构想效度谜题:测量维度还是活动? [J]. 心理科学进展,2013,21:358-371.
- [2] Woehr D J, Arthur W. The construct-related validity of assessment center ratings: a review and meta-analysis of the role of methodological factors [J]. J Manag, 2003,29:231-258.
- [3] Arthur W Jr, Day E A, McNelly T L, Edens P S. A meta-analysis of the criterion-related validity of assessment center dimensions [J]. Personnel Psychol, 2003, 56:125-154.
- [4] Sackett P R, Dreher G F. Constructs and assessment center dimension: some troubling empirical findings [J]. J Appl Psychol, 1982,67:401-410.
- [5] Shore T H, Thornton G C, Shore L M. Construct validity of two categories of assessment center dimension

ratings [J]. Personnel Psychol, 1990,43:101-116.

- [6] Sackett P R, Harris M M. A further examination of the constructs underlying assessment center ratings [J]. J Business Psychol, 1988,3:214-229.
- [7] Lance C E. Why assessment centers don't work the way they're supposed to [J]. Industrial and Organizational Psychology: Perspectives on Science and Practice, 2008a,1:84-97.
- [8] Lance C E. Where have we been, how did we get there, and where shall we go? [J]. Industrial and Organizational Psychology: Perspectives on Science and Practice, 2008b,1:140-146.
- [9] Lievens F, Dilchert S, Ones D S. The importance of exercise and dimension factors in assessment centers: simultaneous examinations of construct-related and criterion-related validity [J]. Human Perform, 2009,22:375-390.
- [10] 邓光辉,孔克勤,内田-克雷佩林心理测验在我国的试用研究 [J]. 心理科学, 1995,4:230-233.
- [11] 龚耀先. 艾森克个性问卷在我国的修订 [J]. 心理科学, 1984,4:11-18.
- [12] Hough L M, Oswald F L. Personnel selection: looking toward the future, remembering the past [J]. Ann Rev Psychol, 2000,51:631-664.
- [13] 周子玉,刘耀中. 以任务为导向的评价中心技术 [J]. 中国人力资源开发, 2009,(8):38-40.
- [14] Sagie A, Magnezy R. Assessor type, number of distinguishable dimension categories, and assessment center construct validity [J]. J Occupat Organizat Psychol, 1997,70:103-108.
- [15] 陈民科,王重鸣. 评价中心的开发程序与构思效度 [J]. 人类工效学, 2002,8:27-30.
- [16] Lance C E, Lambert T A, Gewin A G, Lievens F, Conway J M. Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings [J]. J Appl Psychol, 2004, 89:377-385.
- [17] 王小华,车宏生. 评价中心的评分维度和评分效果 [J]. 心理科学进展, 2004,12:601-607.
- [18] 吴志明,张厚粲. 评价中心的构想效度和结构模型 [J]. 心理学报, 2001,33:372-378.
- [19] 王辉,孟宪惠,武国城,李革新,郭云. 空军飞行学员心理学“筛选-控制”选拔体系 [J]. 航天医学与医学工程, 1992,5:271-273.