

多维效标测量的构成*

周智红 王二平

(中国科学院心理研究所, 北京 100101)

摘要 通过为某机床企业销售员制定一套自我绩效评定迫选量表, 收集判断性效标测量和非判断性效标测量, 研究判断性效标测量和非判断性效标测量与自我绩效评定迫选量表中各工作行为的关系, 验证了两个假设: (1)任一效标测量不可能与所有有效工作行为显著相关; (2)在效标测量中, 既应采用判断性效标测量, 也应采用非判断性效标测量, 二者互相补充, 不可替代。

关键词 效标测量, 判断性效标测量, 非判断性效标测量。

分类号 B849: C93

1 问题的提出

效标 (criterion) 概念的外延很广。本研究专门探讨人事管理领域所用的测量学意义上的效标, 即效标测量, 并采用 Austin 和 Villanova 对效标测量的定义, 即效标测量是一种可直接或间接测量的绩效样本 (行为及其结果), 并体现组织有关预测源 (predictor) 或人事决策的价值^[1]。此定义说明了应以绩效为效标的主要内容, 这种绩效可以是一个个的工作行为, 也可以是行为所带来的结果; 另外, 也说明了绩效的选择体现了组织的价值观, 因为一个组织奉行何种价值观, 就必然会选择体现这种价值观的绩效。

效标测量是多维的。从职务成效的角度来看, 效标测量是衡量任职者成功或失败的标准^[2]。在效标测量出现的早期, 人们就已认识到成功具有多维的特征, 因此影响成功的因素是复杂的, 这就需要多维效标测量来检验预测成功或评价成功的测验的效度。

从绩效的角度来看, 在人事管理中, 通常应用职务说明书 (job description) 对每一种职务的职责、作业范围以及绩效标准作出界定, 作为规范和评定任职者职务绩效的依据^[3]。长期以来, 人们对职务绩效的理解一直限定在职务说明书规定的范围之内。然而, Borman 和 Motowidlo 于 1993 年提出, 应将职务

绩效划分为作业绩效 (task performance) 和关系绩效 (contextual performance)。作业绩效是指任务的完成情况, 关系绩效是指有利于完成组织工作的心理和社会环境^[4]。

Borman 和 Motowidlo 的这一提法引起了学者们的关注, 以后许多研究者的研究也都证明了绩效可以分为作业绩效和关系绩效^[5-7], 并且它们是互相独立的, 这进一步证明了以绩效为内容的效标测量是多维的。

从效标测量划分的角度来看, 任职者的绩效可以从个人的行为、行为的结果及特质三个方面进行评定^[8]。在这三种标准中, 个人特质是最弱的一种标准^[9]。对个人行为及结果评定的效标测量一般包括判断性效标测量 (judgmental measurements) 和非判断性效标测量 (nonjudgmental measurements) 两类^[10]。

非判断性效标测量包括产品数据 (如销售额、产量等) 和人事数据 (如事故、离职、缺勤等) 的测量, 测量的是组织的直接目标, 即行为的结果。判断性效标测量是由评定者对被评定者的个人行为作出判断, 对职务绩效来说, 这些行为非常重要而且必不可少。

非判断性效标测量测的是行为的结果, 由于个体的行为导致的结果是多方面的, 又由于导致某种结果的因素也是多方面的, 因此, 只采用一种行为结

收稿日期: 2000-10-12。

* 国家自然科学基金资助项目 (批准号: 79770090)。

果来评定个体绩效是不全面的,也是不合理的,必须采用多种行为结果,即多维效标测量。判断性效标测量是建立在对个体行为判断的基础之上,由于影响职务绩效的重要和必不可少的个体行为是多方面的,因此,对于判断性效标测量来说,也必须是多维的。

绩效评定是为了某种决策目的。从决策目的来看,在人事管理和研究决策中,是独立地对待多个效标测量,还是将多个效标测量整合起来,组成一个合成效标测量分数?对这个问题的看法还存在争论。

合成效标测量模型是将某具体职务所要求的多个绩效测量线性组合成一个合成分数,来反映任职者总的绩效。多维效标测量模型认为效标测量的各个维度彼此相互独立,不能合成,合成的指标意义含糊,因此必须收集多个职务绩效测量,来确定效标测量的独立维度^[2]。

总之,无论是从职务成效的角度来研究效标测量的多维性,还是从绩效划分的角度、效标测量分类的角度、人事管理和研究决策目的的角度来研究效标测量的多维性,无不表明效标测量是多维的,在实际应用和研究中,应采用多维效标测量。

通过以上对效标测量多维性研究的回顾及分析,本研究提出了两条假设:(1)任一效标测量不可能与所有有效工作行为样本显著相关(多维效标测量的必要性假设)。(2)衡量职务绩效应由非判断性效标测量和判断性效标测量两方面构成(多维效标测量的充分性假设)。

2 方法

本研究拟通过在销售工作行为方面为机床企业销售员制定一套自我绩效评定方案,收集判断性效标测量和非判断性效标测量,来研究判断性效标测量和非判断性效标测量与销售员自我绩效评定量表中工作行为样本的关系。

2.1 编制并实施绩效评定量表

建构一套绩效评定方案,很重要的问题之一是找到可能的绩效评定指标。而要找到合理的绩效评定指标,须对职务进行分析。因此,对机床企业的销售员进行职务分析,目的是要找到评定销售员绩效的可能指标。

职务分析的方法有多种。关键事件技术分析方法是通过对职务绩效起关键作用、最反映职务成功或失败的与职务有关的职务活动进行职务分

析^[3]。与职务绩效有关的职务活动有多种,其他传统的职务分析方法多是对职务活动进行频次分析,但是对于某种职务来说,频次高的职务活动并不一定对职务绩效起重要作用。因此本研究采用关键事件技术对机床企业销售员进行职务分析,找出最能代表职务成效的关键事件,并确定与之相关的销售工作行为。

职务分析的具体方式采用半结构化个别访谈。要求每名被访者主要回答以下问题:(1)任职者理解的职务职责、作业范围;(2)举例说明,任职者认为对产品销售起促进作用的活动;(3)举例说明,任职者认为对产品销售起妨碍作用的活动。

根据职务分析结果,归纳出销售工作行为样本。

根据工作行为样本,制定评定量表。由于销售员一般单独活动,他人不易了解其行为,所以拟采用销售员自我评定的方法。又由于自我评定具有自我美化的倾向,所以拟采用迫选法。因此本研究对销售工作行为样本的评定采用的是销售员自我评定的迫选量表。迫选量表的具体编制程序如下:

(1) 过渡量表的编制与实施

通过查阅报刊工作伦理论述、先进人物事迹报道,编制与工作有关的行为描述,作为迫选量表中匹配的工作行为样本。将匹配的工作行为样本与从职务分析中所得到的工作行为样本混在一起,按随机顺序排列,采用 Likert 6 点评定方法编制成过渡量表,并对机床企业的销售员进行测查。

(2) 迫选量表的编制与实施

根据过渡量表中所得到的各工作行为样本的社会评价价值,将社会评价价值相等或相近的匹配的工作行为样本和从职务分析中所得到的工作行为样本配对,构成自我评定的迫选量表。并对机床企业的销售员进行测查。

2.2 收集效标测量

效标测量的收集是一个独立的研究过程,为了防止被污染,须在工作行为样本收集之后进行。本研究所收集的效标测量包括非判断性效标测量和判断性效标测量。

至于非判断性效标测量,收集实施了迫选量表的每名机床企业销售员的如下资料:97年销售额、97年销售量、98年销售额、98年销售量、98年的计划指标(指组织计划每名销售员应完成的销售额)。又通过计算,转换生成了5种非判断性效标测量,即销售额差、销售量差、销售额增长率、销售量增长率

以及完成计划指标的比例作为非判断性效标。这样,共收集 10 种非判断性效标测量。

至于判断性效标测量,首先,根据访谈所整理出的销售员的职责,编制判断性效标测量,表达为 Likert 5 点或 6 点评定量表。然后,由两至三名销售员的直接上级根据所编制的判断性效标测量量表,对每名销售员作评定。对于每个评定指标,要求评定者一次评完所有的被评定者;其次,在每个评定指标下,被评定者都是按随机顺序排列的;另外,每名评定者评定指标的顺序也是随机排列的。这样就收集到了每名销售员的判断性效标测量。

2.3 工作行为样本与效标的筛选

采用点二列相关统计法,计算各种判断性效标测量和各种非判断性效标测量与自我评定迫选量表中的工作行为样本的相关,并用 t 检验考察其显著性水平。筛选出达到显著相关的工作行为样本和效标测量。

3 结果

3.1 工作行为样本的收集

根据职务分析结果,对每名被访者所提到的对机床产品销售有促进作用和妨碍作用的销售行为进行整理,将凡是两名以上的被访者提到的工作行为描述作为一项工作行为样本,并转换为正面描述。这样,共整理出 59 项销售工作行为样本。

3.2 工作行为样本的评定

(1) 过渡量表

为了与从职务分析中所得到的 59 项工作行为样本配对,编制了 59 项与工作有关的行为样本。将从职务分析中所得到的工作行为样本和编制的匹配的工作行为样本随机排列,构成过渡量表。

对 11 家机床企业(在京的机床企业 7 家,外地的机床企业 1 家,外地驻京办事处 3 家)的 114 名销售员(男性 103 名,女性 11 名)实施过渡量表。求出过渡量表中各工作行为样本的社会评价值的平均值。

(2) 迫选量表

为了充分利用过渡量表中匹配的工作行为样本的结果,让每一个从职务分析中所得到的工作行为样本都有和它配对的匹配的工作行为样本,本研究者采用了以下三条原则:

① 将社会评价相等或相近的从职务分析中所得到的工作行为样本和匹配的工作行为样本配对(最基本的原则)。

② 当一个从职务分析中所得到的工作行为样本和两个匹配的工作行为样本的社会评价都接近时,如果已有一个匹配的工作行为样本和其他从职务分析中所得到的工作行为样本配对,则选另一个匹配的工作行为样本与从职务分析中所得到的该工作行为样本配对。

③ 尽量少用重复的匹配的工作行为样本。

根据以上原则,在配对的迫选量表中,共用了 45 项匹配的工作行为样本与 59 项从职务分析中所得到的工作行为样本。在这 45 项匹配的工作行为样本中,只用过一次的工作行为样本有 33 项,用过两次的工作行为样本有 10 项,用过三次的工作行为样本有 2 项。

由于每家机床企业销售员的人数比较少,为了达到统计要求,选取了各方面情况基本相同的三家机床企业作为研究对象,对他们的 41 名销售人员实施迫选量表。其中男性 38 名,女性 3 名。机床企业 1 有 12 名销售员,机床企业 2 有 19 名销售员,机床企业 3 有 10 名销售员。每名销售员都有 2 年以上的销售经历。

3.3 工作行为样本与效标测量的相关分析

以从职务分析中所得到的工作行为样本为目标行为,采用二分变量,统计出迫选量表中每名被试的自我绩效评定情况。

统计出每名销售员的每种判断性和非判断性效标测量的结果。

采用点二列相关统计法,计算 6 种判断性效标测量和 10 种非判断性效标测量与迫选量表中的 59 项工作行为样本的相关,并用 t 检验检验其显著性水平。结果如表 1。

3.4 效标测量的因素分析

采用主成份分析法抽取因素。对 8 种非判断性效标测量(“97 年销售额”、“98 年销售额”、“销售额增长率”、“97 年销售量”、“98 年销售量”、“销售量增长率”、“98 年计划指标”和“98 年完成计划指标的比例”)和 6 种判断性效标测量(“了解所售产品性能、结构、特点”、“联系客户”、“协调售后服务”、“售前反应是否及时”、“掌握当地市场信息”和“掌握本厂经营销售政策”)进行因素分析。

依据累积贡献率进行因素的选取。由于前 3 个因素的累积贡献率已达 82.420%,其余因素的累积贡献率增加很少,根据结构简单性原则,决定选取前 3 个因素。各因素的解释结果如表 2。

将各效标测量在 3 个因素上的负荷进行斜交旋

考察第 9、34、47 项的工作行为样本与各效标的相关方向,发现“97、98 年销售额差”、“97、98 年销售量差”两种效标测量与这三项工作行为样本相关方向不同于其他效标测量与这三项工作行为样本的相关方向,即造成这三项工作行为样本与各效标测量相关方向存在差异的是“97、98 年销售额差”、“97、98 年销售量差”两种效标测量。

以上情况表明,“97、98 年销售额差”、“97、98 年销售量差”两种效标测量对职务成效的意义与其他效标测量对职务成效的意义不一致。

由于“97、98 年销售额差”、“97、98 年销售量差”与 98 年销售额、销售量呈正相关,与 97 年销售额、销售量呈负相关,而 98 年销售额、销售量,97 年销售额、销售量对职务成效的意义是一致的。因此,对于组织来说,“97、98 年销售额差”、“97、98 年销售量差”是不理想的效标测量,应该删除。

这样,保留下来的效标测量共有 8 种非判断性效标测量和 6 种判断性效标测量。

(2) 工作行为样本的筛选

从表 1 的结果可看出,与各效标测量存在显著相关的工作行为样本中,有 11 项工作行为样本与各效标测量的相关方向为负,它们分别是第 1、9、15、27、32、34、39、44、45、47、59 项工作行为样本。

在迫选量表中,由于每项工作行为样本都是以配对方式出现的,而且是以从职务分析中所得到的工作行为样本为目标行为,因此,如果效标测量与该项工作行为样本呈负相关,那么,当以匹配的工作行为样本为目标行为时,则效标测量一定与该项工作行为样本中匹配的工作行为样本呈正相关。这说明,在这 11 对工作行为样本中,匹配的工作行为样本比从职务分析中所得到的工作行为样本对组织的意义更大。因此,在这 11 对工作行为样本中,应以匹配的工作行为样本作为绩效评定的指标。

由于与各效标测量存在显著相关的工作行为样本有 23 项,即通过不同的效标检验,有 23 项工作行为样本的效度值较高,因此,通过对工作行为样本的筛选,最后的迫选量表由 23 对工作行为样本组成。

4.2 假设一的验证

从表 1 的结果可看出:

与“了解所售产品性能、结构、特点”效标测量达到显著相关的有第 5、26、37 项工作行为样本;

与“联系客户”效标测量达到显著相关的有第 37、47 项工作行为样本;

与“协调售后服务”效标测量达到显著相关的有

第 11、13、39、45 项工作行为样本;

与“售前反应是否及时”效标测量达到显著相关的有第 11、37、39、59 项工作行为样本;

与“掌握当地市场信息”效标测量达到显著相关的有第 11、39 项工作行为样本;

与“掌握本企业销售政策”效标测量达到显著相关的有第 13、39、45、59 项工作行为样本;

与“97 年销售额”效标测量达到显著相关的有第 27、34、37 项工作行为样本;

与“98 年销售额”效标测量达到显著相关的有第 2、27、34、37 项工作行为样本;

与“销售额增长率”效标测量达到显著相关的有第 28、32、35、43 项工作行为样本;

与“97 年销售量”效标测量达到显著相关的有第 2、9、44、47 项工作行为样本;

与“98 年销售量”效标测量达到显著相关的有第 2、37、44、47 项工作行为样本;

与“销售量增长率”效标测量达到显著相关的有第 1、7、12、15、24 项工作行为样本;

与“98 年计划指标”效标测量达到显著相关的有第 27、34、37、39 项工作行为样本;

与“完成计划指标的比率”效标测量达到显著相关的有第 11 项的工作行为样本;

所有的效标测量都至少有一项工作行为样本与之达到显著相关,而且与各效标测量达到显著相关的工作行为样本都不完全相同。这说明以上所有这些效标测量都能用来评定个体的绩效,而且虽然这些效标测量有的与相同的工作行为样本存在显著相关,但是与每个效标测量存在显著相关的工作行为样本又不完全相同。因此,每个效标测量只能反映职务绩效的某一侧面,即支持了假设一。

4.3 假设二的验证

从表 1 的结果可看出,与判断性效标测量显著相关的有第 5、11、13、26、37、39、45、47、59 项工作行为样本,共 9 项;与非判断性效标测量显著相关的有第 1、2、7、9、11、12、15、24、27、28、32、34、35、37、39、43、44、47 项工作行为样本,共 18 项。

既与判断性效标测量显著相关,又与非判断性效标测量显著相关的只有第 11、37、39、47 项工作行为样本,即只有四項工作行为样本既与判断性效标测量又与非判断性效标测量显著相关,其他的 19 项工作行为样本要么只与判断性效标测量显著相关,要么只与非判断性效标测量显著相关。

以上情况表明判断性效标测量和非判断性效标

测量都只是反映了个体绩效的某些方面,这就支持了假设二、

另外,通过对6种判断性效标测量和8种非判断性效标测量进行因素分析,也验证了假设二。从表3各效标测量的主成分分析结果可看出,因素1包含了所有的判断性效标测量,因素2包含了6种非判断性效标测量,因素3包含了2种非判断性效标测量。虽然各因素不完全独立,但从表4的结果可看出,因素3与因素1和因素2的相关很低,分别只有-0.090和0.016,这表明因素3反映的是效标测量的不同方面,不能由因素1和因素2代替。另外,因素1和因素2的相关系数为0.272,而因素1和因素3的相关系数为-0.090,通过Fisher Z转换,两相关系数的标准分数分别为0.277和-0.090,这表明销售员的直接上级对销售员的绩效作评定时,可能受销售员的销售额和销售量效标,即非判断性数量效标的影响,但几乎不受销售员销售额增长率和销售量增长率效标,即非判断性增长率效标的影响。因此,在进行效标测量时,既应采用判断性效标测量,也应采用非判断性效标测量,二者互相补充,不可替代。这说明只有多维效标测量才能全面地反映职务绩效,即支持了假设二。

5 结论

(1) 任一效标测量不可能与所有有效的工作行为样本显著相关。

(2) 在效标测量中,既应采用判断性效标测量,也应采用非判断性效标测量,二者互相补充,不可替代。

参 考 文 献

- 1 Austin J T, Villanova P. The criterion problem:1917—1992. *Journal of applied psychology*, 1992, 77:838
- 2 Cascio W F. *Applied psychology in personnel management*. 2nd ed. Virginia: Reston, 1982. 102, 115—116
- 3 McCormick E J. *Job Analysis: Methods and Applications*. NY: Amacom, 1979
- 4 Borman W C, Motowidlo S J. Expanding the criterion domain to include elements of contextual performance. In: Schmitt N, Borman W C ed. *Personnel selection in organizations*. San Francisco: Jossey-Bass, 1993. 77—98
- 5 Motowidlo S J, Van Scotter J R. Evidence that task performance should be distinguished from contextual performance. *Journal of applied psychology*, 1994, 79:475—480
- 6 Conway J M. Analysis and design of multitrait-multirater performance appraisal studies. *Journal manage*, 1996, 22: 139—162
- 7 Van Scotter J P, Motowidlo S J. Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of applied psychology*, 1996, 81:525—531
- 8 Mckenna E F. *Business psychology and organizational behaviour*. Hove: Lawrence Erlbaum Associated Ltd, 1994. 527
- 9 Stephen P R. *Organizational behavior*. 北京:清华大学出版社,1997. 651
- 10 Landy F J, Farr J L. *The measurement of work performance: methods, theory, and application*. New York: Academic Press, 1983. 27

THE COMPONENT OF MULTI-DIMENSIONAL CRITERIA

Zhou Zhihong Wang Erping

(*Institute of Psychology, Chinese Academy of Sciences, Beijing 100101*)

Abstract

A criterion is a sample of performance (including behavior and outcomes), measured directly or indirectly, perceived to be of value to organizational constituencies for facilitating decisions about predictors or programs. Overlooking precursors' researches about criteria, one conclusion can be drawn that when criteria are used, we should use multiple criteria. According to this, we put forward two hypotheses: (1) Every criterion can't correlate with all work behaviors. (2) Not only judgemental measurement criteria, but also nonjudgemental measurement criteria should be used in performance appraisal. Using critical incident technique, we analysed salesperson's job in the machinery factory, collected a series of anecdotes of job behavior (collected from supervisors and salespersons) that described especially good or especially poor job performance, and summed up relative work behavior samples. Organizing these sale behavior samples and other behavior samples, we developed a self-rating forced-choice scale for salesperson's performance appraisal. After that, we collected six kinds of judgemental measurement criteria and ten kinds of nonjudgemental measurement criteria. Judging the correlation between judgemental measurement criteria, nonjudgemental measurement criteria and behaviors in forced-choice scale, and selecting criteria and work behavior samples, we found that all criteria were correlated with at least one work behavior sample and those work behavior samples which significantly correlated with every criterion were not the same as one another. So it verified the hypothesis (1) We also found that only four work behavior samples were correlated with judgemental measurement criteria and nonjudgemental measurement criteria. Other work behavior samples either correlated with judgemental measurement criteria or nonjudgemental measurement criteria. So hypothesis (2) also was verified. Moreover, the factor analysis of judgemental measurement criteria and nonjudgemental measurement criteria also verified the hypothesis 2.

Key words criteria, judgemental measurement, nonjudgemental measurement.