

赵琪凤◎著

构想效度研究



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

一个世纪以来，围绕效度问题展开的研究占据着心理测量研究领域的重
地，构想效度作为效度研究的重心，在理论上

非外借

2014年度北京语言大学青年学者文库出版基金资助

构想效度研究

赵琪凤 / 著



北京语言大学出版社

BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

le 40/07

© 2016 北京语言大学出版社, 社图号 16353

图书在版编目 (CIP) 数据

构想效度研究 / 赵琪凤著. -- 北京 : 北京语言大学出版社, 2016.12

ISBN 978-7-5619-4789-0

I. ①构… II. ①赵… III. ①心理测验—研究 IV. ①B841.7

中国版本图书馆 CIP 数据核字 (2016) 第 307747 号

构想效度研究

GOUXIANG XIAODU YANJIU

排版制作: 北京创艺涵文化发展有限公司

责任印制: 姜正周

出版发行: 北京语言大学出版社

社 址: 北京市海淀区学院路 15 号, 100083

网 址: www.blcup.com

电子信箱: service@blcup.com

电 话: 编辑部 8610-82303647/3592/3395

国内发行 8610-82303650/3591/3648

海外发行 8610-82303365/3080/3668

北语书店 8610-82303653

网购咨询 8610-82303908

印 刷: 北京九州迅驰传媒文化有限公司

版 次: 2016 年 12 月第 1 版

印 次: 2016 年 12 月第 1 次印刷

开 本: 710 毫米 × 1000 毫米 1/16

印 张: 16.75

字 数: 341 千字

定 价: 49.80 元

PRINTED IN CHINA

序（一）

2009年秋季开学，琪凤以优异的成绩考入北京语言大学语言学及应用语言学专业，跟我攻读博士学位。按理说，让她跟我做汉语语法研究，于她于我都是最自然也是最省心、省力的事情。而我，当时正在北京语言大学汉语水平考试中心主任的岗位上，对语言测验学科多少有了一些了解和认识。就当时汉语水平考试中心学科建设和人才培养状况而言，我深深地感到，培养一名语言测验方向的博士，要远比培养一名语法学方向的博士紧要得多。考虑到琪凤硕士阶段的专业背景以及她在汉考中心多年的工作经历，经与她商量，我决定把她往语言学及应用语言学专业语言测验方向去培养。

依据当时学校博士生导师梯队建设的有关政策，我郑重邀请我校语言测验方向的专家张凯教授与我合作，来共同指导琪凤在博士阶段的学习及博士论文的写作。这种由同一学科专业内跨度较大的不同方向的两名导师合作，共同指导一名博士生的做法，在当时可以说是人才培养的一种创新模式。而这，在今天看来，也是符合人才培养模式改革大趋势的。

我当时之所以敢于做出这样的一种选择，除了有张凯教授加盟这一重要因素外，也是出于对当时汉语水平考试中心所具有的语言测验的良好学科基础的信心。这一良好的基础，在我看来，主要体现在三个方面：第一，20多年从事汉语水平考试（HSK）的理论与实践的探索已相当丰富；第二，对国内外语言测验专业文献资料的收集整理已成规

模；第三，一支以中青年学者为主体的语言测验学科的优秀团队业已形成。可以说，这一良好的学科基础，在全国也是独一无二的。我相信，这种依托学科整体实力的人才培养方式，要远胜于一两个专家学者的个别指导。

就这样，琪凤在入学后，一方面，跟我学习汉语语法方面的课程；另一方面，在张凯教授指导下，系统研修语言测验方面的课程，大量阅读语言测验的理论著作。最终，她的博士论文选题确定为《构想效度研究》。

这里，我对琪凤博士论文本身的内容不做过多的评述（这一点张凯教授会有更为精到的论述），但作为琪凤的导师之一，我看到了琪凤在钻研构想效度这一极具难度的理论问题的过程中所表现出来的巨大勇气和刻苦精神。作为导师，我是感佩至深的。

琪凤的博士论文《构想效度研究》，是一篇以文献综述为基础，以理论探索为主旨的博士论文。这一选题，表面看来难度不大，实则不然。其难度主要体现在以下三个方面：

第一，要阅读大量的相关专业方向的英文文献，而这些文献都以探索心理测量这一高度抽象和复杂的理论问题为主旨。我们不难想见，这一工作对于一个普通中文系的毕业生来说，其难度有多大。琪凤在张凯教授的悉心指导下，发扬“蚂蚁啃骨头”的精神，利用三年多的时间，硬是把一本本、一篇篇语言测验的英文文献啃了下来。其中的甘苦，唯有琪凤自己才能完全体味。

第二，要在大量阅读文献的基础上，对构想效度理论近百年来产生、发展、变化的基本脉络进行清晰的梳理，才能使构想效度理论研究的基本成果系统化、体系化。从这篇博士论文所展现出来的基本内容来看，琪凤是出色地完成了这一任务的。她在论文五个主体章节中，对构想效度理论的几个重点问题分别进行了全面的归纳、深入的讨论和公允的评述。这为学界，特别是中国语言测验学界认识、了解、反思这一领域的研究，做出了自己独特的贡献。

第三，这篇博士论文，是要在展示前人研究成果的基础上，提出

作者自己对构想效度研究的理论主张，从而为今后构想效度进一步的理论探索指出发展趋势，指明发展方向。显然，这一目标，对于一名年轻的博士生来说，不可谓不宏大，但是这篇博士论文也初步做到了。我认为，琪凤的博士论文所梳理的构想效度研究框架，对于重构构想效度研究的理论范式，是很有借鉴价值的。

由以上三个方面，我们不难看到琪凤在完成这篇博士论文的过程中所付出的巨大努力。当然，这份努力也得到了学界的认可：琪凤的博士论文得到了答辩委员会的充分肯定，也很快得到了北京语言大学青年学者文库出版基金的资助。

在北京语言大学，语言测验学科曾依托汉语水平考试（HSK）而成为一门极富特色和优势的学科专业方向，也曾经涌现出一批优秀的语言测验专家。斗转星移，沧海桑田，如今，北京语言大学虽已不再从事汉语水平考试（HSK）的研发和考务工作，但是作为富有特色和发展潜力的语言测验学科还在，新一代的中青年优秀语言测验专家也在迅速、健康地成长。这是令人深感欣慰的。

琪凤的博士论文即将正式出版发行。我写上这些话，既是对琪凤的衷心祝贺，也是对北语、对中国语言测验学科未来发展的美好祝愿。

是为序。

张旺熹

二〇一六年十月

序（二）

1950年代早期，美国心理学会（APA）下属的心理测试委员会认为有必要拓宽效度的定义，于是该委员会的两个成员 Paul Meehl 和 Robert Challman 就提出了“构想效度”这个概念。而后，这个委员会的成员经过讨论对这个概念做了修改。经大家同意后，这个概念就被写进了 APA 在 1954 年推出的技术推荐标准。第二年（1955 年），Cronbach 和 Meehl 合写了一篇文章，第一次对构想效度做了专门的论述。

构想效度是心理测量和语言测试中最重要的概念，对于构想效度概念的出现，Rogers 曾给予很高的评价。他说，“构想效度的引进，可以看成是自比奈发明智力测验以来，测验领域中最为显著的变化”，它标志着测验“和历史传统与实用传统彻底决裂”，而且“至今仍以显著的方式改变着测试事业”。然而，并非所有人都像 Rogers 那样正面评价这个新出现的概念，在这个概念引进之初，Anastasi 对此是持怀疑态度的；及至上世纪 80 年代，Davies 还把构想效度叫作“安乐椅上的幻想”。

构想效度概念提出已有 60 多年了，在这 60 多年的时间里，对构想效度问题的研究、诠释虽说已是汗牛充栋，但其中也有很多误解和混乱。对于构想效度这个问题，测量界至今也没有形成一致的意见。

赵琪凤女士的《构想效度研究》一书首先对构想效度研究的历史和现状做了详尽的梳理，对其中的错误、争论及混乱做出了客观、中肯的分析和评价。美国教育测量委员会和美国教育委员会从上世纪中叶开

始推出一本书,书名是《教育测量》,这是教育测量领域的一本工具书。到2006年,这本书已经出到了第四版。在这四个版本中,每一版的第一章都是关于效度问题的,其中1989年第三版中的“效度”一章是由Messick写的,其篇幅近100页(16开本),这简直就是一本书的篇幅了。尽管四个版本的《教育测量》都把效度问题放在最重要的位置,但是,它们均没有对构想效度问题中的错误和混乱做出评论。从这个角度看,琪凤女士所做的全面的梳理工作是开创性的。

在全面梳理和分析的基础上,琪凤女士把几十年来的构想效度研究分为两种视角:测量学视角和社会学视角。她提出,当初Cronbach和Meehl是从测量学视角提出构想效度概念的。他们的本意是通过科学测量和相关的定律实现对假设的心理构想的证明,即把“对心理特质的证明纳入到理论证明的过程中”。这是自然科学的研究视角。然而,上世纪七八十年代以来,构想效度研究出现了一个转向,以Messick等人为代表的研究者把社会文化、道德、测验的有用性等外部因素统统和构想效度扯上了关系,这就形成了构想效度研究的社会学视角。而且,这个视角成了目前构想效度研究的主流。琪凤女士正确地指出,这样一个视角的转换,实际上是一种倒退。

在最后的结论部分,琪凤女士指出:“构想效度是从纯理论的角度提出的”,“构想效度的本质是证明理论有效”。这个结论也是正确的。

这本书是琪凤女士的博士论文。在心理测量、教育测量、语言测验领域里,也有很多人以效度为题做博士论文,但那些论文做的都是实证研究,即找一些实测数据,对某个测验做一个效度检验。当初琪凤女士也想做一个构想效度的实证性检验。她来跟我商量,我认为只做实证检验恐怕分量不够。我还是希望她就构想效度问题在理论上做一些深入的研究。做理论研究比做实证研究的难度要大多了。在琪凤女士论文答辩的时候,北京师范大学脑与认知科学研究院的边玉芳教授说:“看到这篇博士论文后,我眼前一亮,很久没有看到这样的文章了。”边教授的评价表明,琪凤女士已经基本上可以驾驭构想效度这样一个重大的理论

问题了。这部书是琪凤女士一个良好的起点，我希望她将来能够拿出更有震撼力的成果。

是为序。

张凯

二〇一六年十月

提 要

近一个世纪以来,围绕效度问题展开的研究占据着心理测量研究领域的重地。构想效度作为效度研究的重心,在理论上存在着各种分歧和论争,学界至今尚未达成一致。所谓效度(validity),就是测验在多大程度上测到了它想要测的东西。多年来,许多心理测量的研究者围绕着测验效度问题,从多个角度进行了理论层面的探讨。尤其针对构想效度这个核心问题,学界一直存在着两个截然不同的研究视角,即研究者分别从测量的角度和社会学的角度对心理构想的本质展开探索和研究,但研究现状总体看来还比较零散,普遍缺乏全面和系统的理论解析。我们认为这种格局不利于构想效度研究获得突破性进展,更无益于实现对心理构想的探究和证明。有鉴于此,本研究的研究目标为,通过评述构想效度的现有研究,为构想效度的系列研究正本清源,消除混乱,提出可行的解决方案。我们的基本研究思路是,依托大量的国内外相关研究文献,将文献评析与理论对比结合起来,力图公允地评价构想效度研究现状。我们相信,此项研究对构想效度研究的理论梳理具有重要作用,特别是可为实现心理构想的科学研究提供参考和改进意见。

本研究通过对构想效度的产生背景以及其本质的梳理和分析,明确提出,构想效度是在逻辑实证主义思想的影响下,由 Cronbach 和 Meehl 两位学者从纯理论的角度提出的,它的首要任务是界定测验测的是什么。在这个问题上,构想效度的倡导者从理论证明的角度入手,对所测心理特质提出“构想”,将对心理特质的证明纳入理论证明的过程中,试图通过科学测量、定律来实现对构想的证明。因此,构想效度的提出,是采

纳了自然科学的研究视角和研究思路，来解决人类心理特质的测量问题。

基于上述对构想效度本意和本质的探究，我们集中探讨了构想效度研究的测量视角和社会学视角的研究现状。我们发现，Stevens 等心理测量学家忽略了探索与心理构想相对应的客观实体，证明心理构想的客观实在性，而是关注于统计赋值等细节操作。这一现状导致心理测量的相关研究并没有真正解答“测量了什么”这个问题，因此也没有实现对心理构想的探究和科学测量。相比之下，基于社会学研究视角的效度整体化理论则备受欢迎，且在效度研究领域长期占据主导地位。这些主流效度研究者从社会学方面入手，将社会文化、道德、测验的有用性等外界因素纳入效度框架内，形成了效度证据多样化的研究格局。值得注意的是，当前取得丰硕成果的主流研究仅限于实际应用层面，这些成果对于构想效度的科学研究层面并无贡献，可以说，构想效度科学理论层面的研究从二十世纪七八十年代以来是呈现倒退局面的。

在上述分析和评价的基础上，我们提出了本研究的理论主张，即构想效度研究要注重心理构想的实在性和效度检验的理论性方面的探究。我们从构想效度的核心定义出发，确认构想效度的本质是证明关于心理构想的理论有效。对于无法直接观测的心理构想来说，找到与之对应的可见实体是实现科学测量的唯一途径，于是对心理构想的实在性的要求也突显出来，我们只有通过可见实体的已知信息来确定心理构想的客观存在，当然这二者之间需要有定律来进行关联。同时，在构想效度实际分析方法和思路方面，我们也主张从实在论和实体性的角度出发，着力于探寻客观存在的物理或生理机制与抽象的心理构想之间的关联，从而摆脱以往构想效度研究的主观性和不确定性。

我们相信，本研究所梳理的构想效度研究框架，将有助于效度研究学者对构想效度的全面认识和深入探究。本研究提出的构想效度研究主张对相关理论探讨具有一定的借鉴意义。同时，我们对构想效度分析方法的建议，也将引发效度研究学界的思考。当然，本研究作为构想效度理论探讨的一个初步探索，研究的内容、广度和深度仍然有待进一步完善，这还需要我们做一步的思索和研究。

目 录

第 1 章	绪论	1
1.1	选题缘起	1
1.2	研究思路	8
1.3	研究意义	10
第 2 章	效度概说	13
2.1	效度概念的发展演变	13
2.2	效度研究领域中的争议	31
2.3	本章总结	39
第 3 章	构想效度的提出及其本质	42
3.1	Cronbach 和 Meehl 从纯理论的角度提出构想效度理论	42
3.2	Cronbach 和 Meehl 试图提升心理测量的理论高度	59
3.3	Cronbach 和 Meehl 构想效度研究倾向的演变	70
3.4	本章总结	85
第 4 章	构想效度研究的测量视角	88
4.1	Stevens 忽略构想, 关注赋值	89
4.2	Michell 开创定量科学的两类作业	100
4.3	张凯对心理构想的进一步探究——区分两种测量	112
4.4	本章总结	116

第 5 章 构想效度研究的社会学视角	119
5.1 构想效度概念呈现“大而全”的局势	120
5.2 效度证据丰富多样	132
5.3 本章总结	170
第 6 章 心理构想的实在性与构想效度研究的理论性	174
6.1 构想效度是证明构想理论有效	176
6.2 构想效度研究离不开科学测量	181
6.3 构想效度研究发展方向	206
6.4 本章总结	218
第 7 章 结语	223
7.1 基本观点和结论	223
7.2 不足与后续研究	229
参考文献	231
后记	250

第 1 章

绪 论

1.1 选题缘起

1.1.1 效度理论研究现状

自 1921 年效度概念正式提出以来，效度理论一直是教育和心理测量界（包括语言测验领域）长期关注的核心问题。近一个世纪以来，围绕效度问题展开的研究占据着心理测量研究领域的重地。构想效度作为效度研究的重心，在理论研究上存在着各种分歧和论争，学界至今尚未达成一致。可见“效度研究迫切需要理性的进步”（Lissitz, 2009: 3）。

所谓效度（validity），就是测验在多大程度上测到了它想要测的东西。多年来，许多心理测量的研究者围绕着测验效度问题，从多个角度进行了广泛的研究。其中，研究重点集中在效度理论研究和效度检验的实证研究两大方面。效度理论研究通过理论分析，建构理论框架，关注一般性的“测什么”以及“如何测”一类的核心问题；有关效度检验的实证研究，几乎都是针对具体的测验是否有效而做的检验。随着各种统计模型和数学运算的引入，相比之下，关于效度检验的实证性研究更是研究的热点。

与开发引进各种数学模型的热度以及进行实证性研究的规模相比，效度理论研究显得相对冷清和稀疏。然而，效度理论问题，即“测什

么”的问题在心理测量研究中异常重要。我们知道，心理测量所测的心理特质是无法直接观察到的，要想使心理测验有效、准确地测量到心理特质，必须首先明确心理特质的属性和结构。在实现科学、准确地回答“测量了什么”这个难题之后，心理测验的效度问题也就迎刃而解了。因此，在效度实证研究大行其道的环境下，效度理论研究依然不容忽视，因为“测什么”的问题尚未真正解决。

半个多世纪以来，针对心理学中的构想效度问题，一直存在着两个截然不同的研究视角，即研究者分别从社会学^①的角度和测量的角度对心理构想的本质展开探索和研究。虽然同以构想效度为研究对象，但两个研究视角分别以不同的命名加以探索，一方的研究者甚至对另一方的研究知之甚少。我们认为这种格局不利于构想效度研究获得突破性进展，更无益于构想效度研究实现对理论的证明。

针对这两种相互对立的观点，根据所提出观点的受关注度以及研究成果的规模，我们将其称为效度研究的主流观点和非主流观点^②。主流观点所主张的效度理论以及效度验证方法被大多数心理学家所接受，是效度研究领域最为常见的观点和思路，并且该领域的相关概念明确以“效度”或“构想效度”命名；而非主流观点多从心理测量的视角展开研究，虽然尚未以“效度”为明确的命名和定义，但是实质上同样在进行对心理构想的探究，同样属于构想效度的研究范围。这一类的研究为极少数研究者所关注，且研究成果较少，但是这类研究同样得到了教育

① 社会学(sociology)，一种人类行为科学，用以考察个人之间以及个人与群体之间各种社会关系的性质、原因及结果(《不列颠百科全书》(国际中文版)，1999，15：457)。我们将构想效度的一类研究视角划归社会学研究视角，是因为此类效度研究关注于测验与外围的社会环境各因素之间的关联，着力于探讨效度与各种社会关系之间的联系。

② “效度研究的主流观点和非主流观点”这一说法，我们借鉴自 Robert W. Lissitz (2009) 主编的论文集 *The Concept of Validity: Revisions, New Directions, and Applications* 中的提法。主流研究即以“效度”为名义展开的系列研究。非主流研究是指本质上是对心理构想的探索，但表面上并没有以效度研究为明确的研究核心，甚至在论述中几乎不提及“效度”这个关键名词。

和心理测量学界的重视。由此可见,针对效度理论研究存在着多方面立场,并且各自争鸣。

具体来说,效度研究的主流观点是从社会学视角入手展开的,更多的是关注效度和测验使用的实用性和操作性。主流观点深受美国心理学家 Messick 为代表的一批心理学家的影响。他们主张效度“是一种综合性的价值判断”。Messick (1989) 强调“效度验证的并不是测验或观察工具本身,而是根据测验分数或其他指标做出的推断,种种有关分数意义或分数解释的推断,以及有关根据这种解释采取行动可能产生的后果的推断”。也就是说,主流观点认为效度研究是对来自分数解释的各种关系进行的价值判断,研究者需要搜集多方面的证据来证明分数解释和分数使用的准确性。即效度不是测验的固有属性,而是测验分数解释和使用的属性。(Sireci, 2009; Kane, 2009; Zumbo, 2009)

效度研究的非主流观点则关注于测量理论在心理测量领域的应用。研究者们试图遵循物理测量的方法,将心理测量与经典物理测量相比较,找到心理测量目前存在的差异,推进心理测量向科学测量迈进。在这一研究方向上具有突出贡献的研究者为数不多,我们集中讨论了 Stevens 的测量定义及其量表理论。Stevens 旨在促进测量操作在心理测量研究中的实际应用;另一位学者 Michell 则从定量角度探究心理特质的定量结构。非主流这一派学者并没有关注于心理构想的证明问题,他们的理想是提出一套适用于所有测量的理论。如果非主流一派的学者的理想得以实现,那么心理构想的测量问题也就随之解决了。可见,在非主流观点研究的内部,不同的学者之间也存在着研究视角和理论主张的差异,但是他们都在努力探寻心理测量中的效度核心问题,并希望从物理测量中找到答案。他们认为效度是测验的特定属性,与外部的解释和评价毫无关系。(Michell, 2009; Borsboom 等, 2009; Mislevy, 2009)

依据对“效度”这个核心概念的理解,我们将两大流派之间的效度观点总结如下:

表 1 围绕“效度”研究的两大流派

	主流观点	非主流观点
研究倾向	把效度当作学科专业的一部分来研究，注重应用 ^① 层面。	把效度当作科学的一部分来研究，注重理论 ^② 层面。
关注点	从社会学研究视角入手，将涉及效度的所有因素都纳入研究，包括分数解释和社会评价等。	从自然科学研究视角入手，关注于科学测量领域，对效度的研究环境进行净化。
效度定位	效度是分数解释和评价的属性。	效度是测量工具的属性。
研究现状	成果丰硕，为众多学者认可。	隶属小众，值得探讨。
研究局限	满足实用需求，缺乏理论的科学性。	关注理论的抽象层面，缺乏理论的具体实现。

可见，两大研究派别在研究倾向、关注点、效度定位等方面都存在严重分歧，究竟孰是孰非，不能贸然决断。科学研究不能因为主流观点成果丰硕，占据效度研究的主导地位就将之奉为典范，也不能因为非主流观点少人问津而将它束之高阁。

我们认识到，构想效度探索的这一现状，是一个很有价值的研究议题。当一门学科不再表现出思想学派相互对抗的特征时，即大多数的科学家在理论和方法论问题上达成一致的时候，这门学科的发展就进入了一个更加成熟和领先的阶段。此时，整个研究领域由一个共同的范式或模型来界定。

在物理学史中，我们可以看到范式的作用。例如，伽利略与牛顿的机械主义观念在 300 年的时间里一直被物理学家所认可：

在那段时间里，实际上所有的物理学研究都是在那个框架中进行的。然后，当大多数物理学家转而接受爱因斯坦的模型——一种审视研究对象的新方式时，伽利略和牛顿的方法就被取代了。一个范式对另一个范式的取代被库恩解释为科学的革命。（舒尔兹等，2005：19）

① “应用”是一种技术、系统或产品的使用。转引自 <http://baike.baidu.com/view/220910.htm>。

② “理论”指人们由实践概括出来的关于自然界和社会的知识的有系统的结论。《现代汉语词典（第 5 版）》，商务印书馆，2005：836。