

A Research Review of Validity of Language Testing of Twenty Years

Xiaoyan Liu

School of Foreign Language, Lanzhou University, Lanzhou Gansu
Email: liuxy16@lzu.edu.cn

Received: May 12th, 2017; accepted: May 22nd, 2017; published: May 25th, 2017

Abstract

Language testing is an important field in linguistic research and a significant step in language teaching. Researches on language testing play a vital role in the contents and forms of testing. This article, based on the researches theoretically and empirically in validity of language testing, summarizes the research results during 20 years in order to help more researchers who are interested in this field to better understand validity of language testing and provide practical uses for future teaching and scientific researches.

Keywords

Language Testing, Validity, Construct Validity, Language Competence, Unitary Validity Theory

语言测试效度二十年研究综述

刘晓燕

兰州大学外国语学院, 甘肃 兰州
Email: liuxy16@lzu.edu.cn

收稿日期: 2017年5月5日; 录用日期: 2017年5月22日; 发布日期: 2017年5月25日

摘要

语言测试是语言学研究的一个重要领域, 是语言教学的一个重要环节, 研究语言测试, 对于测试的内容和形式具有重要的指导意义。本文基于研究者们具体对语言测试效度这一具体方面从理论和实证两个方面出发, 综合概括近二十年对语言测试效度的研究成果, 以期帮助更多研究者或相关领域的学者更好地认识理解语言测试的效度, 对将来的教学和科研工作有一定的帮助。

关键词

语言测试, 效度, 构念效度, 语言能力, 效度整体观

Copyright © 2017 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

语言测试, 就中国传统的“应试教育”而言, 即测试者参与设计试卷, 目的是测量学生们的语言能力, 通俗来讲, 就是学生们对于语言知识的掌握水平。随着科学文化的发展, 科学的外语测试一方面可以对学生的语言能力和外语教育教学进行正确的评估, 另一方面通过对被测试者语言能力的评估对测试的设计、教学过程及测试者起到一定的反拨作用。本文通过对过去 20 年 CNKI 中国期刊网全文数据库以及部分国外期刊文献中相关论文的检索, 对语言测试中有关效度方面的研究进行综述。

2. 研究综述

在过去几十年中, 由于全球经济化的发展以及英语语言的全球普及化, 对英语语言测试的发展研究逐渐细化到各种语言测试的内容、测试工具、结果评估、试题内容有效性、测试的反拨作用等方面。正如卢晓丹[1]指出一套设计比较合理的测试应该包括效度、信度、实用性和良好的反拨作用。在一项测试中, 近年来许多学者们一直热衷于对信度和效度的研究。而本文拟梳理语言测试效度理论的演变进程, 以及对于效度验证的方法, 当前对效度理论的研究成果, 对语言测试效度的主要研究者、未来语言测试效度的发展与展望等方面内容展开。

2.1. 文献检索范围及基本趋势

本文综述选取以“语言测试效度”为关键词和篇名在 CNKI 中国期刊网全文数据库上搜索, 从过去 20 年(1996~2016)所发表的论文中共筛选出几十篇相关论文进行对近 20 年语言测试效度研究进行综述。对于英语语言测试的研究始于 20 世纪 40 年代的科学前语言测试。随着科学技术的发展, 人们逐渐转入对语言测试不同方面的研究, 对于语言效度的研究也逐渐受到语言学家们的重视。许多研究者通过对效度理论发展的演变过程、效度验证理论的发展等方面的探讨外, 一些研究者更具体化到研究效度的社会性、计算机辅助语言测试下的效度、英语专业四、六级考试中的效度分析等方面的具体分析。由此可见, 国内学者们越来越重视对不同语言测试效度方面的研究, 对于效度的研究也慢慢进入系统化、稳定化、科学化的研究阶段。

2.2. 研究内容综述

从普通计量学引入的效度概念自 20 世纪 30 年代提出以来, 其定义内涵不断发展变化, 经过近半个多世纪的演变, 取得了重大进展。效度是一种表明一项考试是否测量了所要考察的内容。正如 Bachman [2] 指出: Validation must therefore consider, in addition to the test's content and method, how test takers perform. Validity, on the other hand, is concerned with identifying the factors that produce the reliable variance in test scores. That is, validation addresses the question, what specific abilities account for the reliable variance, in test scores? The process of validation is a continuous one, involving both logical analysis and empirical investiga-

tion。杨惠中[3]指出效度是对考试质量最重要的要求，指出有效测试指的是测试的效度要高。对于效度的探究近年来越来越受到越来越多学者的重视。

从效度理论的发展来看，李清华[4]指出 Kane 把 20 世纪 50 年代以来的效度理论从教育与心理测量学领域视角按照时间先后分为 3 个发展阶段：(1) 基于标准的效度模式；(2) 构念模式；(3) 整体效度模式；并且李清华[4]指出按照人们对效度的普遍认识的演变，把效度理论发展归纳为三个时期：(1) 20 世纪 50 年代以前的单一效度观时期；其就是通过 Binet, Simon, Anastasi, Kelly, Cureton 等语言学家们对效度的定义演变指出当时的效度概念仅仅是在能够找到适当测量标准的前提下被看作是一种相关系数，因为对于所谓的标准的参照存在问题，导致这一定义理据的推翻。(2) 20 世纪 50 年代至 80 年代中期的分类效度观时期；其通过 Guilford, Cronbach, Angoff, Lather, Cumming, Guion 等语言学家们对效度这一概念的理解认识，李清华[4]指出最后形成了 Guion 的“三位一体”的三足鼎立的效度观，即内容效度，构念效度，效标关联效度。(3) 80 年代中期至今的效度整体观时期；由于之前所提倡的“三分法”效度的不全面性，对于测试的标准不能进行全面的参考，通过 Guion, Cronbach, Messick, Tenopyr, Kane, Meehl, Challman, Loevinger, Anastasi 等语言学家对于对于“三分法”效度以及效度理论发展优缺点的分析探讨，Messick 提出的“效度整体观”明确了效度是对分数理论和经验的解释。李清华[4]指出“整体效度观”也存在理论与经验效验的验证仍需继续进行探究。贾贻东[5]通过 Bachman 对语言能力本质与语言能力本质的探讨，指出 Bachman 阐述“整体效度观”与测试开发、解释和适用、对考试分数的解释和适用密切相关。宋月霜[6]通过 Henning, Messick, Bachman 等语言学家对效度的定义解释指出了效度是证据和理论支持测试分数解释的程度，而测量一门考试是否有效就要对该项考试进行效验。贾贻东[5]指出 Bachman 的效验的“测试有效性模式”，该模式包括信度、构念效度、真实性、互动性、影响和实用性这六个方面。同时指出 Bachman 提出的“AUA 框架”(Assessment Use Argument)把抽象的效验融入到测试设计、开发和使用的过程中，突破了测试有用性的框架，侧重于对考生语言能力方面进行转移，将效度与考试的后效结合起来，在测试中不仅是对分数的解释，也包括考生语言能力的文字描述和学习评价的过程等，增强了效验的验证力度。Bachman 和 Palmer 指出“测试有用性”框架用来测试语言测试的开发和使用，这种模式在过去的语言测试效度验证模式方面起到了一定的作用。韩宝成[7]总结出“测试有用性模式”是一种突出操作性，牺牲理论连贯性，偏离“效度整体观”模式。经过对效度验证模式的探索与扬弃，Bachman 所提出的“测试使用论证 AUA”模式不仅仅对测试的开发很重要，对于效度的验证提供了更加切实可行的实践与理论的验证。

从效度的构成成分来说，赵海永[8]等人指出效度既与某一特定的目的相联系，又是一个相对的概念。赵海永[8]指出 Weir 将传统的信度看作效度证据的一个方面而纳入效度的范畴下，并指出了 Weir 对不同效度类别的分析，即理论效度(传统称之为结构效度)、语境效度(传统称之为内容效度)、效标关联效度、后试效度及评分效度(传统称之为信度，且在该分类下把信度归入效度依据下)这五类。袁家麟[9]指出效度有结构效度，内容效度，预期效度，等级相关效度，表面效度和信度这六部分。宋月霜[7]通过对 Alderson 对测试效度的认识分类从内部效度证据(有表面效度、内容效度、应答效度)和外部效度证据，也称之为效标关联效度(有同期效度、预期效度)这两个视角分析了语言测试效度的证据分析。陈雨珣[10]指出根据传统的效度习惯，效度的类别有表面效度、内容效度、标准效度和结构效度。

随着语言测试理论的演变发展，学者们对语言测试效度和构成成分的阐述理解，在语言测试的过程中从自然心理科学视角逐渐转向社会文化心理视角的演变，构念效度逐渐在语言测试效度方面受到很多学者的青睐。杨惠中[3]指出构念效度的内容方面包括内容的相关性和代表性，涉及的是考什么的问题，构念效度的实体方面即考生的作答过程是否反映了考生的语言能力等。宋月霜[6]指出构念效度中的构念指的是有关人类行为某一方面的心理构建或理论定义，构念是不能够被测量或被直接观察到。聂建中[11]指出构念是一个人潜在的心理特质，是不能直接被观察和测量的一种或一组能力，但是能够通过测试成

绩作出推断。对于构念效度的测量,宋月霜[6]总结了学者们对这一效度的测量有效总共有三个成分因素:构念必须是有效的,即被测量的理论构念作为前提是有效的才能进行到下一阶段;测量必须是有效的,测量有效即当测量了想测量的能力后才能反过来证明效度有效;对于分数的测量或者解释必须是有效的。以上这三种效力的结合,才能是对一个构念效度的完整阐述。聂建中[11]通过阐述构念的定义,通过Hughes, Cronbach, Meehl, Bachman, Davies等人对于构念效度的理解阐述,得出结构效度是在很大程度上代表了潜在的一种学习的理论。高怀勇[12]指出了通过效度的发展,许多学者们都对构念效度是涵盖其它效度的一个统一概念,并且指出构念效度作为统一概念这一阐述得到了1999年版《标准》的认可:不再区分效度种类,效度只有一种,那就是构念效度。在构念效度被认可发展的过程中,影响最大的是Messick的“效度整体观”观点的发展对于效度构念地位的巩固可谓是起到了相得益彰的作用。韩宝成[7]指出了Messick的“分层效度框架”,Messick称之为“Progressive matrix”用来对测试分数进行效度分析。高怀勇[12]指出了在效度整体观框架下,构念效度的核心地位得以确立;对于测试评价逐渐转入动态系统以及测量的社会视角转变;构念效度的效验是一个连续的过程,指出对于效度的验证需要多方面的证据。杨惠中[3]指出效度只能通过外部的、独立的标准进行评价,效度研究的方法包括观察法、内省法、访谈和统计分析方法等。对于效度的研究,是一个长期积累效度证据的过程。

语言测试效度的研究对于语言测试在教育中扮演的角色是不言而喻的。随着人们对效度研究的深入,越来越多的学者将自己的关注视角具体化,或者将语言测试的效度问题跟现代科学技术结合起来,这些对于效度的理解与验证也是极为有利的。金艳[13]指出在测试领域,技术的使用已不仅仅是为了提高测试效率,更重要的是提高测试的效度。金艳[13]通过阐述我国著名语言学家桂诗春关于测试和计算机的评估对于效度的分析,分析了计算机化语言测试(CBLT)的发展阶段并结合分析了大学英语四六级的笔考和网考成绩效度分析。而与此同时,考生的计算机熟练程度对于分数效度的分析也有影响。所以对于考试模式的设计对于测试分数效度的测量也是有影响的。除了对于测试测试中成绩本身效度分数的测量,许多学者们的注意力近几年来也逐渐转移到效度理论中的社会观视角。陈建林[14]指出近些年来随着效度理论的演变发展,对语言测试社会属性的研究也逐渐得到许多学者的重视。陈建林[14]指出最早在效度理论中关注社会属性的是学者Cronbach指出测试分数的评估师是在一定的社会背景下进行的。语言测试的社会性影响不仅仅是针对开发者产生影响,对社会方方面面的影响也逐渐显露出来,比如图书出版商、培训机构、国家教育部门以及相应教育文件的发布等等。陈建林[14]指出在对一项测试的评价过程中,测试者需要向受试者、家长、测试使用者、以及教育和社会机构等其他相关负责说明该测试从设计、开发、实施、到分数的解释和报道等环节的效度,以此来保证测试的效度。

2.3. 研究方法

本文所设计的英语语言测试的效度相关研究中,大多数学者们的研究方法主要是在对于效度的综述介绍、理论分析以及应用与实证研究(实证研究主要是多数学者们具体到对于比如听力测试、英语专业四、八级考试)这三个方面。

2.4. 研究者

研究测试效度的学者主要集中于一些本科院校从事外语教学科研的老师和学生们(主要是硕士生、博士生),也有一些从事外语工作的海归派学者们。总的来看,对于语言测试效度的研究是这些研究者们致力于在教的过程中能够得到分数的高效度从而对教学、对社会产生积极的反馈作用。

3. 反思与展望

通过对效度理论发展与效验的综述,许多学者从早期到如今的效度概念,都会将受试者的语言能力

作为切入点进而探讨测试者的测量分数效度问题。刘壮等人[15]指出从1961年TOEFL创始人Carroll提出语言能力是所谓的二维模型即语言成分和语言技能的构建,到Chomsky(1965)语言能力概念的提出,再到Bachman[2]在《语言测试要略》中提出了交际语言能力(Communicative Language Ability, CLA)的发展,逐渐体现出了语言能力的重要性以及语言能力的交际作用。对于语言测试效度是否测量了想要测量的受试者的能力,测试是否有效,越来越多的学者们也逐渐将研究的重点转移到在交际语言能力下的效度问题。除此之外,机助测试也在未来的测试效度研究中受到重视。更重要的是近年来(尤其是国内学者们)对于语言测试效度的社会性后效也应该加入更大的力度来研究论证。

虽然对于语言测试效度的研究已经得到很多学者的综述论证,但是还是存在以下问题:

从研究方法来说,多数学者们的研究主要是以综述、理论的介绍和应用等这些定性的研究为主,对于实证的研究还是很缺乏。实证研究还是最有说服力的;对于研究问题的验证,除了要有充分的理论依据来支持所综述的问题,也要建立在大量的实证研究中,即大量的定量分析对于问题的阐述也是很重要的。其次也缺乏新理论的产生。就语言测试效度的理论研究来说,近几十年来一直是对先前语言学家们对于所述问题理论的反复阐述,并没有产生能够在当今研究状况下轰动语言学界、能够作为新世纪语言测试方面的一个分水岭。

从研究内容来看,机助测试和语言测试及效度的社会性视角的研究应该得到更多的重视。一方面,随着科学技术的飞速发展,计算机在语言测试中的使用对于测试者和被测试者都是有利的,更重要的是对于语言测试效度的测量精确度也有很大帮助;另一方面,语言测试社会性视角的转变是未来的一大研究方向。辜向东[16]总结了近半个世纪以来语言测试从测试的有效性到测试的社会性,并且指出(McNamara & Roever, 2006)语言测试的社会性的方向主要包括语言能力的社会维度及其测量、测试的社会应用和影响。

总的来说,对于语言测试效度的研究最终都要回归到语言测试这个总方向上来。语言测试的效度结果分析对整个教学过程中的“教”与“学”有指导反思作用。就中国现行的教育体制而言,“教考合一”的测试体质仍然在国内风行,且这种测试体质不利于正确考核学生语言能力,缺乏真实场景的实际应用能力。就未来测试的方向,“教考分离”的测试模式在语言测试方面应该更加得到重视。

参考文献 (References)

- [1] 卢晓丹. 论信度与效度关系下语言测试的取向[J]. 科教文汇旬刊, 2014(13): 104.
- [2] Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. 上海: 上海外语教育出版社.
- [3] 杨惠中. 有效测试·有效教学·有效使用[J]. 外国语, 2015, 38(1): 2-26.
- [4] 李清华. 语言测试之效度理论发展五十年[J]. 现代外语, 2006, 29(1): 87-95.
- [5] 贾贻东. Lyle. F. Bachman 语言测试理论评述[J]. 北京科技大学学报(社会科学版), 2010, 26(1): 108-113.
- [6] 宋月霜. 语言测试效度理论研究[J]. 学理论, 2011(10): 230-232.
- [7] 韩宝成. 语言测试效度及其验证模式的嬗变[J]. 外语教学与研究, 2013(3): 411-425.
- [8] 赵海勇. Weir 的语言测试效度理论[J]. 语文学刊, 2006(20): 131-134.
- [9] 袁家麒. 语言测试效度研究[J]. 无锡南洋学院学报, 2003(4): 54-55.
- [10] 陈雨珣. 对语言测试效度及其证据来源的一些认识和思考[J]. 语言研究, 2009(16): 75.
- [11] 聂建中, 汤晓媚. 试论结构效度的发展演变[J]. 山西大学学报哲学社会科学版, 2006, 29(3): 104-107.
- [12] 高怀勇. 语言测试构念效度理论: 发展、贡献与挑战[J]. 东北师大学报(哲学), 2013(5): 135-138.
- [13] 金艳. 计算机化语言测试的效度研究[J]. 外语电化教学, 2012(143): 11-15.
- [14] 陈建林. 语言测试社会属性研究综述[J]. 外语测试与教学, 2014(2): 1-10.
- [15] 刘壮. 对任务式语言测试的探索[J]. 语言文字应用, 2008(2): 58-66.
- [16] 辜向东. 语言测试的社会属性: 《语言评测季刊》十年名家访谈录解析及启示[J]. 中国外语, 2015, 12(1): 67-74.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：ml@hanspub.org