
远程开放教育形成性考核模式及其效度验证

——以上海开放大学 2013 春英语专业专科自开课程为例

陈荷男

【摘要】：本文运用“基于论证的效度验证模式”对 2013 春上海开放大学英语专业专科自开课程形成性考核模式的效度进行研究。研究首先对各课程形成性考核模式的使用频率进行统计，并按重要性递减的顺序依次对排名前四位的网上记分作业、平时作业、学习表现和小组活动四种形成性考核模式进行了验证。研究发现以上形成性考核模式的效度均较低，主要有以下四大缺憾：（1）缺乏具体的实施细则导致概念模糊执行不力；（2）缺乏具体的评分标准导致评分带有主观倾向；（3）缺乏有效的监督机制，诚信度不高；（4）教师对学生表现反馈不足，导致形成性考核对教学的促进和反拨作用不强。研究希望能通过本文的效度验证，相关人员能提出解决问题的有效措施，切实提高形成性考核的效度以促进教学。

【关键词】：远程开放教育，形成性考核，效度验证

研究背景

自 1999 年中央广播电视大学启动“中央广播电视大学人才培养模式改革和开放教育试点”项目以来，形成性考核一直是全国远程教育工作者研究和探索的一个重点。它不仅是远程开放教育质量保证金体系的重要环节，同时也是学习者自主学习过程管理的主要手段，与终结性考核共同构成对学习课程学习成果的整体评价。实施形成性考核的目的是要加强对学习过程的指导和管理，及时反馈学习信息，指导教学，提高学习者的综合素质和能力。目前，形成性考核的比重一般占到课程总成绩的 20%—50%，但个别学校个别课程的形成性考核比重已经超过 70%，甚至达到 100%，形成性考核的重要性由此可见一斑。问题是，作为一种考核方式，形成性考核的效度究竟如何？其在多大程度上反映了学习者的学习效果和学习态度，并真实地记录了学习者的学习过程？本文运用“基于论证的效度验证模式”，对 2013 春上海开放大学英语专业专科自开课程形成性考核模式的效度进行研究，以验证其有效性。

研究方法

（一）理论基础

所谓效度，就是测验对构念进行测量的有效程度。（谢小庆，2002）而构念，是指一种测验所要测量的概念或特征。（吴根洲，2008）通俗地讲，效度也叫有效性，是指一套试题所考的是否就是设计者想要考查的内容；落实到形成性考核中，也即形成性考核所考的是否就是形成性考核推出者想要考查的内容，并真正体现了形成性考核的价值。结合形成性考核的目的和意义（杨晨，2005；张睿，2007；杨婷婷，2012），形成性考核所要考查的内容可以概括为：学习者对知识的掌握程度，学习者的学习态度和学习过程；同时，形成性考核是否及时向学习者反馈了学习信息，指导了学习者的学习，提高了学习者的综合素质和能力等。如果形成性考核考查了全部以上信息并实现了全部以上功能，可认为形成性考核的效度较高；如果只考查了部分内容、实现了部分功能，或者完全没有考查应考查的内容、实现应实现的功能，可认为形成性考核的效度较低。

为了对效度进行验证，Kane 建立了基于论证的效度验证模式（Kane, 2002, 2004, 2006; Kane, Crooks&Cohen, 1999）。这一论证模式的中心是两个论证：解释论证和效度论证。解释论证需要详细说明对分数的解释和考试用途，然后再一一呈现在得出这一分数解释和考试用途过程中所使用的推论和假设。其中的推论可以用一个论证链来表示。Bachman 于 2005 年对这一论证链作了延伸，增加了“决定”这一环节，如图 1 所示。（转引自胥云，2011）

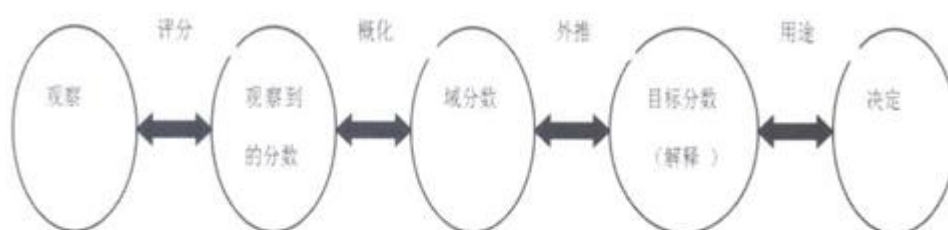


图1 效度验证中延伸后的论证链
(Bachman, 2005:8, 转引自胥云, 2011)

根据胥云（2011）对这一论证链的阐释，将这一论证链运用到形成性考核中去，则其推论过程如下：首先从“观察”到的学习者在形成性考核中的表现经由“评分”推论出“观察到的分数”，也即形成性考核的得分。要使这一推论成立，则其中的假设是有适当的评分标准，并在评分过程中始终坚持该评分标准。其次，“观察到的分数”经由“概化”得到“域分数”。所谓“域分数”就是将某一次观察到的分数看作一次抽样，从这一抽样可以推断出被试在不同时空且评分人员不同时得到的分数。这一过程中的假设是形成性考核观察到的表现相对于观察域来说具有代表性，且观察取样足够大以控制误差。第三步是通过“外推”由“域分数”推出“目标分数”。就形成性考核而言，就是从学习者在形成性考核中的表现推断他在目标情景中或者说非考试情景中的表现，也就是对形成性考核分数的解释。这一推论的假设是衡量目标能力的形成性考核形式与现实中使用该能力的任务相似，也就是形成性考核的真实性。最后一个推论是“考试用途”，一是判断学习者形成性考核合格，二是形成性考核对教学存在促进作用。这一推论的假设是教师对学习者的形成性考核成绩做出及时反馈，并且学习者在收到教师反馈后对学习方法和学习过程做出及时修正，同时教师对教学安排和教学方法做出一定程度的修正以进一步促进教学。该论证链中的推论和假设可以用图 2 来具体呈现。

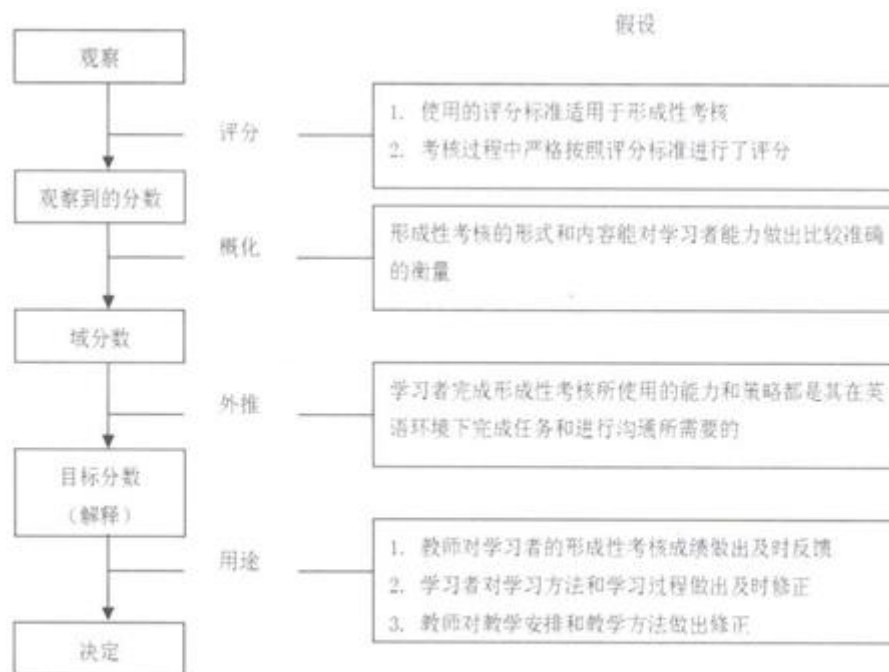


图2 形成性考核的解释论证

推论和假设是解释论证的核心，这些推论和假设需要通过一系列的分析和实证研究进行检验，这就是效度论证。因此，我们需要搜集对解释论证中包含的推论进行验证所需要的所有证据。如果解释论证中所有的推论和假设都经得起严格的检验，则判断该解释论证成立。（胥云，2011）

（二）研究对象

本研究将运用“基于论证的效度验证模式”对上海开放大学英语专业专科自开课程形成性考核模式的效度进行研究。研究选取2013春上海开大英语专业专科全部自开课程，首先对这些自开课程所采用的形成性考核模式进行统计，其次运用“基于论证的效度验证模式”对其中每一种形成性考核模式建立解释论证，并详细地罗列推论得以成立的各种假设，再根据假设搜集相关证据并对所有模式进行——验证。证据的收集渠道包括资料查询、师生调查问卷和个别访谈等。

研究分析与研究结果

（一）形成性考核模式统计

2013春上海开放大学英语专业专科实际开设课程16门，除毕业作业没有形成性考核以外，其余15门课程均设有形成性考核，且形考比重占课程总成绩的30%到100%不等。形成性考核的具体模式包括网上记分作业、书面作业、期中测试三种比较固定的模式，以及课程责任教师可自行设定的其他类别，包括课堂表现、学习表现、平时口语作业、小组活动等等（见表1）。如果对以上所有形考模式按照课程所采用的频度排序，排序过程中合并类似项，如合并实践小组活动、小组活动和小组学习为小组活动，合并平时作业、平时口语作业和书面作业为平时作业，合并课堂表现和学习表现为学习表现，合并期中口试和期中测试为期中测试，则形成如图3所示的柱形图。

表1 2013春上海开大英语专业专科自开课程形成性考核模式与比例

序号	课程名称	形成性考核内容比例					形成性考核占 总成绩比例
		网上记分作业	书面作业	期中测试	其他1	其他2	
1	剑桥商务英语(初级)	40%	40%		课堂表现:20%		30%
2	剑桥商务英语(中级)	40%	40%		期中口试:20%		30%
3	社会实践	20%			虚拟英语实训室:40%	实践小组活动:40%	100%
4	英语国家概况	30%		40%	学习表现(小组活动和学习光盘):30%		30%
5	英语口译(初级)	40%			平时作业:40%	课堂表现:20%	30%
6	英语视听说(1)	40%			平时口语作业:40%	课堂及小组活动:20%	50%
7	英语视听说(2)	40%			平时口语作业:40%	课堂表现:20%	50%
8	英语视听说(3)	40%			平时口语作业:40%	课堂表现:20%	50%
9	英语写作基础	20%	40%		小组学习:40%		30%
10	英语语法	75%		25%			40%
11	英语阅读A	70%			小组活动:30%		30%
12	英语阅读B	50%		10%	课堂表现:30%	小组学习:10%	30%
13	综合英语(1)	40%	20%		学习表现(课堂表现、学习态度、小组活动):40%		30%
14	综合英语(2)	40%	20%		学习表现(课堂表现、学习态度、小组活动):40%		30%
15	综合英语(3)	40%	40%		课堂表现:20%		30%

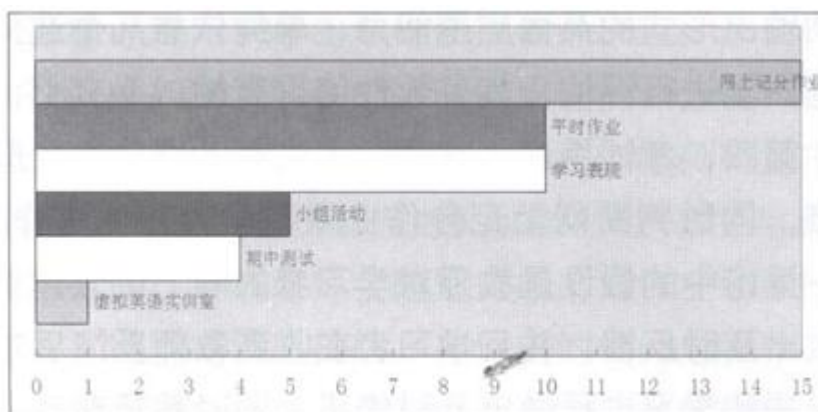


图3 英语专业专科自开课程形成性考核模式频数

由此可知,在英语专业专科开设的这15门课程的形成性考核模式中,使用频数最高的是网上记分作业,频数为15,也就是说15门课程都采用了这一考核模式。其次是平时作业和学习表现,频数为10,然后是小组活动(频数5)和期中测试(频数4)。频数最低的是虚拟英语实训室,频数为1。可见,使用频数最高、重要性显而易见的三种形成性考核模式分别为网上记分作业、平时作业和学习表现,其余3种形成性考核模式个别化倾向明显,在效度研究中选择有代表性的做典型研究。因此,下文对以上形成性考核模式的效度验证将依循重要性递减的顺序进行。

(二) 形成性考核模式效度验证

1. 网上记分作业的效度验证

根据“基于论证的效度验证模式”,笔者首先对网上记分作业这一考核模式建立解释论证框架,详细罗列各种推论和假设,如图4所示。其次,对照解释论证中的推论和假设搜集相应的证据以验证其有效性。

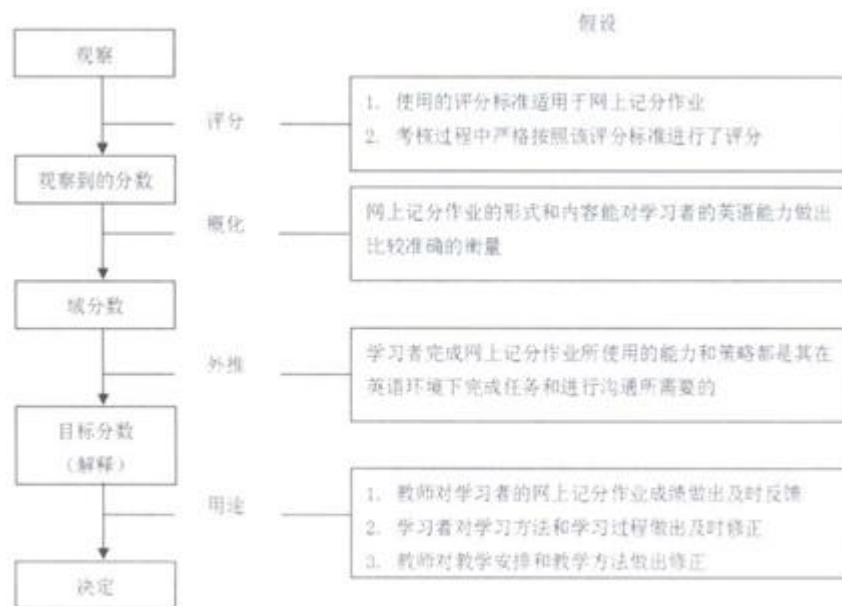


图4 网上记分作业的解释论证

首先第一个推论，由“观察”到的学习者在网上记分作业中的表现经由“评分”推论出“观察到的分数”。事实上，由于学习者是在课外环境下完成的网上记分作业，评分者或者说教师是看不到他们在完成记分作业时的表现的，也就是说教师对学习者的于何时、何地、以何方法完成网上记分作业不甚清楚。那么，我们是否有理由存有这样的疑问：网上记分作业真的是由学习者本人完成的吗？完成记分作业时会有诸如抄袭等作弊现象存在吗？如果存在以上现象，那么网上记分作业的信度就大打折扣。因为“信度高……是效度高的必要条件……”（吴根洲，2008），缺乏信度的测试，其效度也势必大受影响。而关于网上记分作业的评分标准问题，因网上记分作业全为客观题，且由系统自动打分，不存在人为的评分不公或因评分人员不一致而产生评分不公问题，因此该假设成立。

其次，由观察到的网上记分作业成绩经过概化推论出“域分数”。这里的假设是网上记分作业的形式和内容能对学习者的英语能力做出比较准确的衡量。这里要搜集的是基于测验的形式和内容的凭证，也即某些研究者所说的“内容效度”。内容效度是指测试任务对目标使用域（target language use domain）或真实语境中语言任务的代表程度（content coverage）和相关程度（content relevance）（Bachman, 1990）。对内容效度的验证可以通过下述两种方式进行：一是通过专家或有经验的教师判断所出试题的内容是否符合真实语境的真实任务，二是试题是否反映了教学大纲的要求。鉴于前者相对而言比较主观，笔者采取第二种方式，即对照每门课程的教学大纲考察网上记分作业内容的关联性和覆盖面。根据关联性的定义可知，如果网上记分作业的内容与教学大纲的关联度较高，说明网上记分作业具备较高的关联性；反之则关联性较低。同理，如果网上记分作业考察了教学大纲中所要求的所有教学点，则说明网上记分作业的内容覆盖面较广。

这里要说明的一个问题是，上海开放大学英语专业专科自开课程的记分作业次数为2—7次不等，绝大部分课程的记分作业次数在3次左右，且每次作业的题量也不等。据统计，各课程中题量最大的是综合英语（1），7次作业题量总数达到304题，而题量最少的课程只有45题。题量少这一事实在客观上影响了网上记分作业的内容覆盖面。由此，笔者得出如表2所示的数据。可见，上海开放大学英语专业专科自开课程网上记分作业的内容关联性总体而言较高，而内容覆盖面却差强人意。

表2 2013春上海开大英语专业专科自开课程网上记分作业内容效度

序号	课程名称	内容关联性	内容覆盖面
1	剑桥商务英语（初级）	较高	较低
2	剑桥商务英语（中级）	较高	较低
3	社会实践	较高	中
4	英语国家概况	较高	较高
5	英语口语译（初级）	中	中
6	英语视听说(1)	较高	较低
7	英语视听说(2)	较高	较低
8	英语视听说(3)	较低	较低
9	英语写作基础	较高	较低
10	英语语法	较高	中
11	英语阅读A	较高	较高
12	英语阅读B	中	中
13	综合英语(1)	较高	较低
14	综合英语(2)	较高	较低
15	综合英语(3)	较高	较低

第三，根据“域分数”经过“外推”得出“目标分数”，即从学习者在网上记分作业中的表现推论他在目标情景中或者说非考试情景中的表现。这一推论中的假设是学习者完成网上记分作业所使用的能力和策略都是其在英语环境下完成任务和进行沟通所需要的。这里需要解决两个问题：其一，完成网上记分作业的是否就是学习者本人？其二，学习者完成网上记分作业所使用的能力和策略都是其在英语环境下完成任务和进行沟通所需要的吗？对于第一个问题的回答，前文已经说过，网上记分作业是在学习者不受监控的状态下完成的，因此我们不能保证完成作业的就是学习者本人，也就不能保证其有效性。对于第二个问题，英语专业的各门课程，除了“英语国家概况”主要考察识记能力以外，其余课程基本上围绕听说读写译等应用能力进行。而上海开放大学网上记分作业的题型限定为单选题、复选题或是非题，全部属于客观题，如果说听、读等被动接受能力能够通过客观题的形式得到考察的话，说和写，尤其是口译这些属于主动输出形式的英语应用能力，单纯依靠几个选择题和判断题进行评估，其有效性值得商榷。

第四，考试用途，判断学习者网上记分作业成绩合格，同时判断网上记分作业对教学存在促进作用。这一推论中的假设是教师对学习者的网上记分作业成绩做出及时反馈，并且学习者在收到教师反馈后对学习方法和学习过程做出及时修正，同时教师对教学安排和教学方法做出一定程度的修正以进一步促进教学。那么事实究竟如何？上海开放大学的网上记分作业全部在网上进行，由系统自动打分并记录成绩。无论课程责任教师还是面授教师，在整个教学过程中不需要对网上记分作业做任何点评。通过学习者访谈笔者还了解到，面授教师很少顾及网上记分作业，除了偶尔提醒学生在截止日期之前完成记分作业以外，对记分作业内容几乎不做讲解。由此可知，教师对网上记分作业缺乏反馈，学习者和教师都没有因网上记分作业而对自己的学习和教学活动做出过改进。

因此，从论证链的四个环节来看，除了第一个环节的评分标准适用于网上记分作业以外，其余三个环节都存在或多或少的缺憾：第二个环节由于记分作业的覆盖面较低导致内容效度较低；第三个环节，我们既不能保证完成网上记分作业的就是学习者本人，也不认为英语各项能力、尤其是主观输出能力能够通过客观题型得到测量；第四个环节，课程责任教师和面授教师都

没有对网上记分作业成绩做出及时反馈，记分作业对教学的促进作用得不到体现。综上所述，笔者认为上海开放大学英语专业专科自开课程网上记分作业的效度较低。

2. 平时作业的效度验证

笔者对平时作业以及其它形式的形成性考核模式都做了如图 4 所示的解释论证图，详细罗列了推论的各个步骤和其中的假设。考虑到版面问题，这里只直接呈现对于各种形成性考核模式的证据搜集和验证过程。

对平时作业进行效度验证需要搜集的第一个证据是平时作业的评分标准，需要论证的是评分标准适用于平时作业，且在评分过程中严格按照该评分标准进行了评分。据统计，上海开放大学英语专业专科 15 门自开课程中有 10 门课程采用了平时作业这一形成性考核模式，其中 6 门为书面作业，3 门为平时口语作业，1 门为平时作业。课程责任教师在课程形成性考核实施细则中就已经明确规定作业内容的有 4 门，其余 6 门均给了一定的指导性意见，具体的作业内容以及成绩评定方式都由课程面授教师决定。那么，课程面授教师如何对学生的平时作业进行评分呢？通过对部分面授教师的访谈，笔者了解到，教师布置书面作业后通常会给学生留一到两周的时间，然后收缴作业进行批阅；有的教师会随堂完成书面作业和平时口语作业，随堂或课后批阅。在作业的批阅过程中，基本上能做到一个标准，客观公正。但让教师比较为难的是迟交作业者的作业评阅问题。部分学生出勤率不理想，教师布置作业的时候不在场，不知道要做作业也不知道作业内容；有的学生虽然在场，但在完成作业方面存在拖沓现象。这两种学生往往要等到教师或辅导员短信或电话通知或再三催促才会上交作业，而彼时教师已经评阅完该次作业且已公布参考答案。这种情况下，教师往往无法确认学生是独立完成的作业还是参考了公布的答案，所给分数真实性存疑。还有一种情况是，部分学生的作业直接以打印的形式并委托同学或辅导员上交，教师无法判断完成作业的是否是学生本人。因此，笔者认为平时作业的评阅过程存在诸多公正性问题，亟待解决。

平时作业效度验证的第二个假设是平时作业的形式和内容能对学习者的英语能力做出比较准确的衡量，需要验证的是平时作业的内容效度。这一点跟网上记分作业相似。据统计，10 门有平时作业的课程中，6 门课程建议完成 2—5 次作业，4 门课程没有规定作业次数。规定次数的基本上是布置书面作业的课程，而没有次数规定的课程布置的是口语或口译作业（表 3）。从平时作业的内容关联性来看，凡是规定了具体作业内容的课程，或者给出了具体作业建议的，其内容关联性都较高；但内容覆盖面依然显得不够广，如“综合英语（2）”要求教师至少批改 2 次学生的形成性考核作业册，“综合英语（3）”要求学生完成 5 次写作。笔者认为，无论是形成性考核作业册还是写作，都不足以覆盖综合英语课程要求学生掌握的各大知识点和各项英语应用能力。理想的状态应该是由课程面授教师在授课过程中根据学习者的具体情况经常性地布置各种作业，并给予及时批阅和反馈。但因为缺乏制度的制约，没有规定次数和具体内容的平时作业成了学生考核的死角，面授教师布置平时作业随意性极大，期末时只有成绩而所做作业无从考证的现象不在少数。

表3 2013春上海开大英语专业专科自开课程平时作业内容效度

序号	课程名称	规定作业次数	规定作业内容	建议作业内容	具体作业内容
1	剑桥商务英语（初级）	4	/	/	/
2	剑桥商务英语（中级）	4	/	/	/
3	综合英语(1)	3	是	/	形考作业册
4	综合英语(2)	2	是	/	形考作业册
5	综合英语(3)	5	是	/	写作
6	英语视听说(1)	/	/	口语	/
7	英语视听说(2)	/	/	口语	/
8	英语视听说(3)	/	/	口语	/
9	英语写作基础	4	是	/	写作
10	英语口语译（初级）	/	/	听力/口语/口译	/

注：“/”表示没有该内容。

平时作业效度验证的第三个假设是学习者完成平时作业所使用的能力和策略都是其在英语环境下完成任务和进行沟通所需要的。和网上记分作业相同，这里要搜集的证据一是完成平时作业的是否就是学习者本人，二是学习者完成平时作业所使用的能力和策略是否是其在英语环境下完成任务和进行沟通所需要的。对于第一个问题，如前文所述一直存在疑问；对于第二个问题，如果平时作业衡量的不是英语听说读写译等各大能力的合集，也就是说如果平时作业只考核了英语听说读写译等能力的其中一种或几种，则其有效性将受到影响。

平时作业效度验证的第四个假设是教师对学习者的平时作业成绩做出及时反馈，并且学习者在收到教师反馈后对学习方法和学习过程做出及时修正，同时教师对教学安排和教学方法做出一定程度的修正以进一步促进教学。通过对学习者的访谈，笔者得知，某些课程的平时作业上交后再也没有回到过学习者手上，学习者甚至都不知道自己的成绩。问其原因，部分面授教师和辅导员表示“是为了保存文件”的需要，不发给学生是“担心作业回到学生手上后收不回来”。

可见，平时作业由于评分过程公正性较难把握、内容效度偏低、对学生各项能力的考察不够全面、部分面授教师没有对平时作业做出及时反馈以促进教学，从而导致整体效度较低。

3. 学习表现的效度验证

对学习表现进行效度验证，笔者建立的假设是：使用的评分标准适用于学习表现，考核过程中严格按照该评分标准进行了评分；学习表现的具体指标能对所要评估的内容做出比较准确的衡量；学习表现能体现学习者真实的学习能力和英语运用能力；教师对学习者的学习表现做出及时反馈，学习者对学习方法和学习过程做出及时修正，教师对教学安排和教学方法做出修正以进一步促进教学。

首先，笔者对各课程学习表现的具体指标进行了统计，发现课程责任教师对学习表现的界定不一：有的教师单纯考核学生的课堂表现，包括学生的出勤、课堂纪律、课堂上参与教学活动的程度以及表现等；有的教师考核的范围较广，学习表现涵盖学习者的课堂表现、学习态度、参与小组活动以及 BBS 活动的情况等；而有的教师则把学习者参与小组活动的情况单独作为考

核指标。由于这样做的课程不少，因此，为了使分类更为清晰，笔者把学习者的组活动参与情况从学习表现中剔除，作为一种形成性考核模式单独进行验证。

其次，对各课程的形成性考核实施细则进行研究后发现，课程责任教师对于学习表现并没有规定具体的考核细则和评分标准。比如，学生出勤多少次才算合格？在 BBS 上如何表现才算优秀？不上线肯定不算优秀，但在线不发言或所发的帖子跟 BBS 主题不相关的情况又如何评定？据了解，面授教师在教学过程中很少对以上问题进行记录和统计，在期末成绩评定时通常根据的是教师本人对学生的大概印象，“觉得这个学生不错就给个较好的分数，不太认识的学生学习表现分就较低”。因为是从印象出发，所以最终教师考虑的评分点都会落到课堂表现上，从而在实际操作中忽略了 BBS 表现等其它要点。

第三，学习表现能否体现学习者真实的学习能力和英语运用能力呢？研究发现，学习表现更多体现的是学习者的学习态度和参与学习的程度，如出勤、课堂纪律、课堂上参与教学活动的程度、BBS 活动的参与度等；学习能力和英语运用能力的体现较少，仅指学生课堂上参与教学活动的表现和 BBS 的表现，同时，没有客观衡量的指标。而且，各指标的情况也较复杂，比如学生出勤低、BBS 活动参与较少是否一定说明学生学习态度欠佳、参与学习程度较低？如果学生是因为面授与工作时间相冲突导致未能到课、而本人在课外进行了较好的自学，这种情况又如何评判？所以笔者认为此处“参与学习的程度”指的应该是“参与教师看得见的学习的程度”，是无法对教师所不知道的学习进行评判的。

总之，课程责任教师对学习表现内涵的理解不一致，学习表现缺乏具体的考核细则和评分标准，未能有效体现学习者真实的学习能力和英语运用能力，整体效度较低。

4. 组活动、期中测试和虚拟英语实训室的效度验证

在上海开放大学英语专业专科的全部自开课程中，组活动、期中测试和虚拟英语实训室是使用频数最低的三种形成性考核模式，分别只有 5 门、4 门和 1 门，个别化倾向明显。但如果把上文所说的纳入学习表现范畴的组活动提取出来，则组活动的频数增加到 8 门。因此，本文仅以频数最高、具有代表性的组活动为例来验证该形考模式的效度。

同样，对组活动的效度进行验证，笔者所要搜集的证据是：组活动的评分标准适用于该形考模式，且教师在评分过程中严格按照该评分标准进行了评分；组活动的形式和内容能对学习者的英语能力做出比较准确的衡量；学习者完成组活动所使用的能力和策略都是其在英语环境下完成任务和进行沟通所需要的；教师对组活动的成绩做出及时反馈、学习者对学习方法和学习过程做出及时修正、教师对教学安排和教学方法做出修正。

经统计，所有设定组活动的课程除了在教学实施细则中对组活动的人员安排和活动内容做了指导性建议外，都没有规定组活动的具体形式和内容。经过学生访谈，笔者得知每个班级都设有学习小组，每个学习小组都有组活动记录本；小组成员和组长根据学生意愿和辅导员安排相结合的形式确定。对于组活动的组织情况和效果，有学生坦言，因为并不清楚具体的要求，所以只要小组成员基本到齐，一起活动一下就算完成任务了。也有部分组活动“因为上面不检查”而流于形式，甚至出现“闭门造车”的情况，组长或个别活跃分子仅仅依靠一支笔就凭空造出一个组活动来，然后任课教师根据学生上交的活动记录本打分。对于小组成员的考核，有的教师采取打整体分的形式进行，即小组的得分就是小组成员个人的得分，这种打分方式看似公平，实则有失公允，一定程度上打击了多作贡献者的积极性。尤其对于组活动的情况考察，若非教师亲自参与活动，是不可能了解每个学生的真实情况的。

可见，组活动由于缺乏具体的引导和有效监控，并没有达到课程设计者所期望达到的效果，从而导致整体效度较低。

结语

本文运用“基于论证的效度验证模式”，对2013春上海开放大学英语专业专科自开课程形成性考核模式的效度进行了研究。研究对该学期各形考模式使用频数的高低进行排序，按重要性递减的顺序依次对网上记分作业、平时作业、学习表现和小组活动四种形考模式进行了验证；其余两种形考模式，即期中测试和虚拟英语实训室由于采用课程较少而未计入验证之列。研究发现，上述四种形考模式的整体效度都较低，除了网上记分作业因为全部在网上进行，由系统自动打分而比较特殊以外，余下三者概括起来有以下几个共同的缺憾：①缺乏具体的实施细则导致概念模糊，执行情况良莠不齐；②缺乏具体的评分标准导致评分带有主观倾向；③缺乏有效的监督机制，诚信度不高；④教师对学生表现反馈不足，导致形成性考核对教学的促进和反拨作用不强。而网上记分作业除了具备后两条不足之外，还由于内容覆盖面较小而导致内容效度欠佳。

形成性考核是远程开放教育教学质量保证体系的重要组成部分；是强化素质教育，加强教学过程管理，反馈学习信息，改进课程考核方式的重要措施；是测评学生学习效果，促进学生自主学习，提高学生综合素质和能力的重要途径。形成性考核的重要意义不言而喻。但我们也不得不承认，在形成性考核日渐成熟的过程中会存在这样那样的问题，有些制度还来不及完善，有些问题还没来得及解决。这就需要从事远程开放教育的教学和管理人员共同努力，查漏补缺，提出能解决实际问题的有效措施，保证形成性考核的效度以切实促进教学。

本文是对目前采用的形成性考核模式的效度进行验证的一种尝试，希望能揭示形成性考核执行过程中存在的一些问题，引起远程开放教育教学和管理人员的重视，切实采取改进措施以进一步完善与形成性考核有关的各种机制和具体环节。本文对完全基于网络进行远程教学的部分网院，亦有现实参考价值。

*本文系上海远程教育集团课题“远程开放教育形成性考核模式与效度研究——以上海开放大学英语专业专科自开课程为例”（项目批准号：JF1119）的阶段性研究成果。

参考文献:

- [1]Bachman, L. F.1990. Fundamental Considerations in Language Testing. Oxford, UK: Oxford University Press.
- [2]Bachman, L. F.2005. “Building and supporting a case for test use”. Language Assessment Quarterly 2/1.
- [3]Kane, M. T. “Certification testing as an illustration of argument-based validation”. Measurement: Interdisciplinary Research and Perspectives 2/3.2004.
- [4]Kane, M. T., T. Crooks&A. Cohen. 1999. “Validating measures of performance”. Educational Measurement: Issues and Practice 18/2.
- [5]Kane, M. T. “Validation”. Educational Measurement.2006.
- [6]Kane, M. T. “Validating high-stakes testing programs”. Educational Measurement: Issues and Practice 21/1. 2002.
- [7]“中央广播电视大学人才培养模式改革和开放教育试点”项目研究工作实施意见，（高教厅[1999]4号）。
- [8]吴根洲. 高考效度研究[M]. 湖北：华中师范大学出版社，2008.
- [9]谢小庆. 对测验效度的一些新认识，《考试研究》，天津人民出版社，2002：26.

-
- [10]胥云. 语言测试中基于论证的效度验证模式述评[J]. 外语教学理论与实践, 2011, (4) .
- [11]杨晨. 远程开放教育形成性考核模式架构探析[J]. 开放教育研究, 2005, (S1) .
- [12]杨婷婷. 我国开放教育形成性考核研究综述[J]. 中国成人教育, 2012, (17) : 14—18.
- [13]张睿, 于忠党. 远程开放教育中形成性考核的探究[J]. 科技信息, 2007, (17) .