

基于即兴问答的 EFL 演讲者思辨能力研究

张娜

北方工业大学, 北京

摘要 | 为了探究英语演讲中即兴回答环节所需要的思辨能力或思辨标准, 本研究基于保罗和埃尔德提出的三元结构模型中普遍适用的思辨标准, 采用定量与定性相结合的方法, 对不同水平的英语学习者在英语演讲的即兴回答环节进行分析。研究发现, 选手对思辨标准的掌握程度存在明显差异。所有选手对思辨标准的掌握程度由高到低依次为: 语言的清晰性、论据的充分性、论证的关联性以及观点的深刻性。在这些标准上, 高水平选手的表现明显优于低水平选手。此外, 本研究还试图识别选手在思辨标准上的不足之处, 并提出相应的改进建议。

关键词 | 思辨能力; 思辨标准; 英语演讲; 即兴回答; 三元结构模型

Copyright © 2024 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

在高等教育领域, 思辨能力被广泛认可为其核心要素之一。迄今为止, 对思辨能力的探讨已深入至高等教育的各个学科领域, 包括大学英语教学中听、说、读、写等技能的培养与学习。2018年, 中国教育部发布的《中国英语能力等级量表》明确指出, 达到七级及以上的高级语言学习者, 在口语表达方面应能就社会话题进行有效、深入的交流与讨论, 并能进行充分、准确、恰当的阐述、论证与评析。其中, “阐述、论证

与评析”作为思辨能力的关键维度, 凸显了在英语口语能力的最高水平中, 语言技能与思辨能力的有机结合是不可或缺的。

英语演讲作为口语表达的一种重要形式, 自然要求演讲者具备相应的思辨能力。目前, 英语演讲的研究主要聚焦于演讲焦虑、演讲策略及演讲教学等方面^[1, 2], 而对思辨能力的研究则相对匮乏。Schreiber等人^[3]提出的“演讲能力评分标准”以及Lucas^[4]的评分体系中, 演讲者的思辨能力并未被作为独立的评价要素。然而, 国才英语演讲考试将学生的思辨能力纳入考核指标, 其评价标准不仅关

基金项目: 2024年度北方工业大学研究生教育教学改革研究项目“线上线下混合教学模式下研究生学术英语课程建设的路径探析”(项目编号: YJS2024JG21)。

作者简介: 张娜, 北方工业大学文法学院, 讲师, 研究方向: 英语教学, 应用语言学。

文章引用: 张娜. 基于即兴问答的EFL演讲者思辨能力研究[J]. 教育研讨, 2024, 6(5): 1385-1390.

<https://doi.org/10.35534/es.0605187>

注语言的准确性与流利性，还关注表达的思想是否具有创新性、内容是否丰富且紧扣主题、论述是否严密、条理是否清晰以及阐述是否充分^[5]。尽管这些评价标准提出了对思辨能力的评估，包括准确性、清晰性、相关性与逻辑性等，但由于缺乏相应的实证研究支持，这些评价标准在实际应用中可能会面临一定的挑战与困难。

在探讨思辨能力的理论框架时，保罗和埃尔德提出的三元结构模型，亦称为批判性思维三元模型，是该领域内的一个重要概念^[6]。该模型由思维元素、标准和智力特征三个维度构成。每个思维元素都应通过九条标准进行衡量或检验，这些标准包括清晰性、准确性、精确性、相关性、深度、广度、逻辑性、重要性和公正性。分析思维要素是基础，发展智力特征是目标，而标准则是自我检查或检验他人的工具。这些标准不仅有助于自我评估和提升思辨能力，也有助于评价他人的思辨能力。

文秋芳教授在构建外语类大学生思辨能力层级理论模型时，认为传统的十条标准过于繁琐，因此将其精简为五条，包括清晰性、相关性、逻辑性、深刻性与灵活性^[7]。孙旻则根据大学阶段学习者的特点，将体现在“概念”这一思维元素上的“清晰性”“准确性”“逻辑性”和公正性等衡量标准凝练为“一致性”^[8]。这表明，研究者会根据研究对象的特点对传统的思辨标准进行修订，使其更贴合特定研究领域的需要。

首先，本研究以保罗和埃尔德的三元结构模型中的思辨标准为理论基础，收集不同英语水平选手在即兴回答中的语料，并确定其中体现的思辨标准。其次，通过分析具体的语料来识别选手在思辨标准方面存在的问题，并提出相应的建议。本研究旨在回答以下两个研究问题：

(1) 即兴回答中需要哪些思辨标准？

(2) 不同水平的英语作为外语（EFL）选手在满足思辨评价标准方面存在哪些问题？

2 研究步骤

2.1 研究对象

本研究选取了30名选手的即兴回答作为分析材料，将这些选手分为高水平组（H组）和低水平组

（L组），每组各15人。H组由进入决赛的选手组成，而L组由未能晋级决赛的选手组成。每位选手的回答均由外籍教师提出的问题引导，旨在评估其思辨能力。

2.2 研究方法

2021年比赛采用线上形式进行，本研究对整个比赛过程进行了屏幕录制。随后，笔者将30个问题及其对应的回答进行了转录。在此基础上，笔者与一位具有多年英语演讲教学经验的资深教师合作，基于现有文献和教学实践，共同确定了四个思辨能力的评价维度：清晰性、充分性、关联性和深刻性。随后对每个回答从这四个维度进行了评分。

为确保评分的可靠性，两位评分者对选手的回答进行了独立评分，并计算了一致性系数（Cohen's Kappa）。结果显示，两位评分者在四个评价维度上的一致性较高， κ 值分别为0.82（清晰性）、0.83（充分性）、0.85（关联性）和0.82（深刻性），表明评分结果具有较高的信度。在评分完成后，笔者对选手们在回答中表现出的问题进行了系统的梳理和分析，旨在识别选手们在思辨能力方面的不足之处，为今后的教学和训练提供参考和指导。

3 不同水平 EFL 演讲者的思辨标准

3.1 即兴回答的四个思辨标准

在对选手即兴回答语料进行深入分析的基础上，本研究借鉴文秋芳教授及其他学者的研究成果，以及保罗和埃尔德的三元结构模型，针对即兴回答的特定语境，对原有的九大思辨标准进行了精简和调整。

首先，从语言层面来看，清晰性是即兴回答的基础标准。在英语演讲中，选手需要在语言表达上追求准确性，避免出现严重的词汇、语法和发音错误，并尽量减少中式英语的使用。只有准确表达自己的观点，才能使评委理解其意图，并进一步评价其他标准。

其次，论据的充分性是另一个显著的维度。通过语料分析，我们发现选手在回答中是否能够提出充分的论据，以及是否能够提供具体细致的论据来支撑自己的观点。

再次，论证的关联性是评价即兴回答的另一个重要标准。关联性要求选手的回答内容与问题紧密相关，逻辑严谨。缺乏关联或关联性不强的论证会使得回答显得苍白无力。

最后，观点的深刻性是评价即兴回答的另一个关键维度。借鉴文秋芳教授的分类方法，本研究将“广度”和“深度”纳入“深刻性”这一维度。在广度上，深刻性体现在选手是否能够从不同角度和方面来回答问题。在深度上，深刻性体现在对问题的思考是否足够深入，是否触及问题的核心要素。

3.2 选手的即兴回答在四个思辨标准方面存在的问题

在即兴回答的语境下，选手们所展现的思辨能力可以根据四个标准从高到低进行排序：语言的清晰性、论据的充分性、论证的关联性以及观点的深刻性（包括广度和深度）。在高层次的比赛中，高水平选手应当致力于达到这四个标准。相比之下，高水平选手在这些维度上的表现普遍优于低水平选手，特别是在前三个标准上表现尤为显著。至于最后一个维度，即观点的深刻性，这不仅是对所有选手的挑战，尤其是在深度方面，很少有选手能够给出令人信服的回答。

表 1 选手在即兴回答中体现的思辨标准

Table 1 Critical thinking standards in the improvisational Q&A for EFL speakers

选手	高水平选手 (H)				低水平选手 (L)			
	清晰性	充分性	关联性	深刻性 (广度 / 深度)	清晰性	充分性	关联性	深刻性 (广度 / 深度)
百分比	100%	90%	85%	50%/20%	75	70%	55%	40%/0

值得注意的是，并非所有回答都必须完全满足上述所有标准。在某一标准上的突出表现可能会对其他标准的不足起到补偿作用。例如，针对问题“寒假即将来临，你认为疫情有可能会再次爆发？还是政府已经彻底控制了疫情？”选手的回答可能如下：“We might have another outbreak, but the people who are affected by the virus will be under control. We already have a comprehensive prevention and control system. Under the guidance of Present Xi, we may mobilize more forces to prevent this disease so as to win the battle (H5).”该选手的回答在论据的充分性方面并未提供详尽的细节支持，同时在观点的深刻性维度上也缺乏显著的广度和深度。然而，从语言表达的角度来看，选手在用词方面表现出极高的准确性和恰当性，句与句之间的语义连贯性亦十分突出。换言之，在语言清晰性这一标准上，选手的表现堪称卓越，成功地满足了高标准的要求。因此，在最终的评价中，该选手也获得了较为优异的成绩。

(1) 言之无误——语言的清晰性

在即兴回答中，意义的清晰性和易于理解性是至关重要的。作为英语演讲比赛的一部分，评委能否准确理解选手所传达的意义是评判过程中的关键因素。因此，语言表达的准确性是至关重要的。在

这一维度上，高水平选手（H组）几乎均能达到所设定的标准（100%），而低水平选手（L组）中仅有75%能够完全达到此标准。这一现象表明，在即兴回答中，流畅地表达意义是区分选手水平的重要指标。

在语言使用方面，选手可能会在词汇、语法和发音上出现错误，特别是那些显著的错误，如关键词的误用，会严重影响评委对选手的整体评价。例如，L4选手在回答中将关键词汇“pandemic”（瘟疫）错误地发音为“academic”（学术）。这种关键术语的错误不仅会降低选手的可信度，也可能使评委对选手的专业性和准备程度提出质疑，从而在一定程度上影响选手的最终成绩。

(2) 言之有据——论据的充分性

论据的充分性在即兴回答中是一个关键的评价维度，它主要反映在选手对论据的选取和使用上。选手是否能够提出有说服力的论据，或提供充分的论据来支撑其观点，是实现论点有据可依的基础。这一点是区分高水平选手与低水平选手的一个重要指标。

在某些情况下，选手提出的观点缺乏相应的证据支持。例如，L3选手在回答“世界该如何统一起来来对抗疫情？”这个问题时，他提出了一个观点：“Also we should raise the public attention for the

pandemic.”然而，该选手并未进一步展开论述，导致观点缺乏深入的阐释和具体的论据支持。当评委听到一个新的观点或概念时，他们通常期待选手能够提供进一步的解释和论证，但由于各种原因，一些选手未能满足这一期待。

相比之下，表现更为出色的选手往往能够通过一个或多个具体的例子来阐释并支持其观点。此外，还有一种较为普遍的情况是，选手的回答中缺少足够的细节来充分解释其观点。例如，L12选手在回答“政府在情感和心理上对百姓给予了足够的帮助吗？”这个问题时，其回答未能提供充分的细节来支撑其观点。他说：“I think our government has taken some measures. They've sent some psychologists to the urban and other areas. They've made a strong effort to help the public in our country.”显然，提供充分的细节作为证据是至关重要的。若论据过于笼统或缺乏具体性，不仅可能无法满足至少一分钟的回答时间要求，而且也难以对听众产生持久的印象。

(3) 述之有力——论证的关联性

第三个能够有效区分高水平选手和低水平选手的重要参数是关联性。关联性越强，论证的说服力就越高。在本研究中，大多数H组选手的回答能够直接或间接地与问题相关联。相比之下，只有大约一半的L组选手的回答展现出一定的关联性，而有些选手甚至出现了完全偏离主题的情况。在这些情况下，选手最常出现的问题是其表达的观点与提出的问题之间缺乏必要的关联。

首先，一些选手在面对问题时显得不知所措，采取了回避策略，未能正面回答提问。例如，L1选手面对的问题是：“对于像你表哥这样的人来，最大的心理问题会是什么？”尽管在回答前，该选手与教师确认了“心理”这一术语的含义，但她的回答并未直接针对问题的核心。她的回答如下：

“It's hard for them to do that, because the family may lose the bread-winner. But after discussion and internal battle, they need to support the country's interest. It's the fight, it's the war for the human race. As a policeman, it's his highest responsibility to protect the people.”在整个回答中，选手仅表达了其表哥如何理性地看待与疫情作战的问题，并未具体阐述他内心的挣扎。因此，即便选手在语言表达上没有明显失误，且能提供相关解释，其回答却未能与问题保持直接关

联性。

其次，选手可能未能完全理解问题，而是根据个人感受进行回答。在常规情况下，如果选手未理解问题，他们会请求外籍教师重复问题。若重复后仍不理解，则可能只能根据自己对问题的理解勉强作答，这通常导致回答与问题的关联性不强。这种情况可能会给评委留下不良印象，从而影响选手的得分。

(4) 言之有物——观点的深刻性

深刻性作为评价即兴回答的一个重要维度，由“广度”和“深度”两部分组成，主要反映了选手的知识广度和深度。“广度”是指选手能否从多个角度或方面来考虑和探讨问题。考虑到即兴回答的时间限制，通常为一分钟，因此对于选手在“广度”方面的要求不宜过高。在本研究中，通过分析选手的回答发现，只有接近一半的H组选手能够满足这一要求，而L组选手中能够做到这一点的则明显较少。

以H8选手为例，在讨论疫情期间制作视频的话题时，她能够从三个不同的角度来阐述原因，

“We made so many videos to protect China's image. Because, firstly, with the help of social media, the effect will grow exponentially... Secondly, in this globalized world, we can't defend ourselves or isolate ourselves from these social media... Finally, we know that it's the word of mouth effect...”在限定的时间内，选手能够从多个角度或方面对问题进行回答，这本身就是一项值得肯定的成就。尽管在论据的适切性和严谨性方面可能存在不足，但能够在有限的时间内展示出对问题多维度的思考，这种能力在即兴回答中尤为宝贵。

尽管某些选手未能完全达到“广度”这一标准，但他们若能从一个特定角度出发，对问题进行深入而充分的阐释，这在即兴回答中同样是值得认可的策略。例如，H3选手面对的问题是：

“In 2020, the joint effort of the pandemic has brought out the best or the worst in us?”事实上，对于这个问题应该有正反两方面回答，但选手给出了这样的尝试：“I think it brings the best in us. I saw on TV that a lot of businessmen donated money to the key areas such as Wuhan and Qingdao, ... Also, they bought medical supplies abroad and sent them back to China. Now, we are also doing the same thing to other countries.”尽管选手

仅从一个角度审视问题，但如果他能够提供两个或两个以上的具体事实或例证来支持其观点，这便满足了“论据的充分性”这一标准。

在即兴回答的“深度”方面，鉴于回答时间通常限制在一分钟左右，选手们确实面临着在有限时间内深入探讨问题的挑战，即深入分析问题的复杂性和关键因素。通过对两组选手回答的分析，发现只有少数H组选手能够提供具有一定深度的回答，这些回答超越了普遍认知的常识性信息。相反，L组选手的回答往往停留在常识层面。

部分选手的回答内容较为表面化。例如，在描述疫情期间居家活动的情况时，L4选手的回答可能如下：“When the pandemic broke out, the first thing I was eager to do after I got up is looking through how many new cases were tested, how many patients were recovered from the virus...”这种回答虽然内容属实，但由于内容缺乏深度或高度，无法给人留下深刻印象。相比较而言，H9的回答：“We made so many videos to protect China’s image, and sent them to our alumni. They would put these video on facebook and other social media that can’t be used in China, so as to spread the idea to the whole world.”该选手的回答超越了一般性常识，她通过制作旨在保护中国形象的视频，并将其分发给世界各地的校友，展现了对问题的深入思考。这表明，在面对疫情这一共同挑战时，选手们对于自身责任的认识存在差异，这些差异反映了他们思想的深度或层次。思想的深度或层次越高，其回答越能给人留下深刻印象。与此相比，那些广为人知的常识性观点则难以产生同样的影响力。

4 结论和建议

在本研究中，基于三元结构模型的思辨标准，通过对即兴回答的语料分析，对不同英语水平学生的思辨能力进行了研究，并探讨了他们在这些标准上的表现及其存在的问题。研究表明，根据学生掌握的熟练程度，思辨标准依次为语言清晰性、论据充分性、论证关联性和观点深刻性。高水平英语选手在这些标准上的表现普遍优于低水平选手，这与《中国英语能力等级量表》中对英语口语能力最高等级的要求相吻合，即英语口语能力的最高等级需要将语言和思辨能力进行有机融合。

针对即兴回答中发现的问题，本文提出以下

建议：

(1) 教师应向学生讲授思辨标准的知识，使学生明白一个好的回答应尽量满足这些标准。此外，事先了解这些标准，还能帮助学生知道如何准备即兴回答环节，例如，为了满足论据的充分性和观点的深刻性，他们可能会尽量多地准备细节论据，从各个角度分析观点，以及提出有深度的观点。

(2) 在即兴回答中，语言清晰性是评价学生英语水平的一个重要维度。虽然严格说来它并不属于思辨层面的维度，但却能成为决定选手比赛成绩的最重要指标。也就是说，选手在语言上要尽量做到表达无误。当然，英语毕竟是同学们的外语，在表达过程中，不太影响意思表达的词汇和语法上的失误可以接受。

(3) 在论据的充分性上，建议学生在准备相关话题时，更多了解掌握相关领域的知识，对话题有自己明确的观点，以及支撑观点的细节论据，如具体的典故、数据或者名人名言等。这个维度不仅是学生思辨能力的重要标准，也成为学生在回答问题时有话可说的保障。

(4) 论证的关联性也是一个重要的思辨标准。关联不大或者没有关联往往与英语听力薄弱有关，而听力是英语水平的一个重要指标。选手在回答过程时，可以主动重复问题，或者重复问题中的关键词。另外，在回答的开头或者结尾部分，可以考虑提供概括语句，这些策略都可以帮助听众确认选手回答的关联性。要注意的是，如果实在无法给出直接关联的回答，选手也可以尝试给出间接关联的回答，对其分数评判方面并不会造成较大的影响。

(5) 观点的深刻性，即“广度”和“深度”，它虽然没有前面两个维度那么重要，但却是评委给出高分的重要参考指标。与论据的充分性一样，这个标准也对选手的知识面提出挑战。建议同学们在拿到演讲题目之后，需要把演讲题目作为一个研究课题来对待。同时，也需要透过表层挖掘深层的内涵，只有这样，才能给出有启发性、令人印象深刻的回答。

总而言之，研究发现，不同英语水平的选手在即兴回答中主要体现出上述四个思辨标准。而这些评判标准的确定也必定会有助于英语演讲和口语教

学, 因为本研究以具体的比赛语料为分析对象, 能够为教师的教和学生的学提供更清楚的参考标准。当然, 由于本研究搜集的语料有限, 发现的思辨标准可能并不全面和准确, 需要有更多的实证研究进行验证和补充。

参考文献

- [1] Lucas S. *The Art of Public Speaking* (10th Ed.) [M]. Beijing: Foreign Language Teaching and Research Press, 2010.
- [2] Schreiber L M et al. The development and test of the public speaking competence rubric [J]. *Communication Education*, 2012 (3): 205–233.
- [3] Paul R, Elder L. *Critical Thinking Concepts and Tools* (7th Edition) [M]. Beijing: Foreign Language Teaching and Research Press, 2016.
- [4] 任文. 再论外语专业学生的思辨能力: “缺席”还是“在场”? [J]. *中国外语*, 2013 (1): 10–17.
- [5] 张德禄. 论辩和评价在外语学生思辨能力培养中的作用 [J]. *中国外语*, 2018 (2): 57–64.
- [6] 伍彩芬, 朱东华. 演讲思辨一体化模式建构——以国际人才英语考试商务公关演讲为例 [J]. *西南交通大学学报*, 2019, 20 (5): 113–121.
- [7] 文秋芳, 王建卿, 赵彩然, 等. 构建我国外语类大学生思辨能力量具的理论框架 [J]. *外语界*, 2009 (1): 37–43.
- [8] 孙旻, 俞露, 王晶. 英语演讲实践中的思辨分项技能 [J]. *中国外语*, 2015 (5): 49–56.

A Study on Critical Thinking Skills of EFL Speakers Based on Improvisational Q&A

Zhang Na

North China University of Technology, Beijing

Abstract: In order to identify the unique critical thinking standards required by improvisational Q&A of English speaking, the author analyzed the responses of EFL speakers at different levels based on the universal critical thinking standards in the model proposed by Paul and Elder using a mixed method. It is found that there are significant differences in the mastery of critical thinking standards among the contestants, from high to low for all contestants being as follows: clarity of language, sufficiency of evidences, relevance of argumentation, and depth of viewpoints. High-level contestants perform significantly better than those low-level ones in these standards. In addition, this study also proposes corresponding improvement suggestions.

Key words: Critical thinking skills; Critical thinking standards; English speech; Improvisational response; Paul and Elder Model