李艳,刘淑君,李小丽,杨普光(2022).人机协同作文评价能促进写作教学吗?——来自Z校拓展课的证据[J].现代远程教育研究,34(1):63-74.

**摘要：**智能作文评价系统是人工智能、互联网、语言认知等多个领域交叉融合的产物，能为作文提供即时评分和内容反馈，其研发和应用为人机协同写作教学奠定了基础。随着中文智能作文评价系统进入教学应用阶段，人机协同作文评价在中文写作教学中是否有效这一议题越来越受到学界关注。以Z校写作拓展课的28名七年级学生为样本开展的准实验研究将学生分为实验组和对照组，分别接受人机协同作文评价和教师作文评价，用以探究人机协同作文评价对初中写作教学的影响。实验结果显示：（1）接受过评价反馈的作文终稿成绩都较初稿有显著提升，但接受人机协同作文评价和教师作文评价的两组学生的作文成绩之间没有显著差异；（2）相较于教师评价，人机协同作文评价能显著提升学生写作激情，但在自信和毅力维度没有显著差异；（3）人机协同作文评价条件下，教师反馈数量和提问型反馈比率都较教师评价条件下有所下降，但二者在低阶反馈和高阶反馈比率方面不存在显著差异；（4）虽然学生整体上认可人机协同评价的价值，但对继续使用该评价方式持保留态度。总之，虽然人机协同作文评价在促进写作教学效果方面具有潜在优势，但其作用的充分发挥还有赖于教师专业素养的提升、学习环境的支持和系统功能改进等多种因素的共同推动。

**关键词：**智能作文评价；作文成绩；写作动机；教师反馈；人机协同作文评价

**一、引言**

人工智能（Artificial Intelligence，AI）的教育应用正在重塑教育生态，教和学的关系正经历全新的变革（中华人民共和国中央人民政府，2019），教师也面临着角色转换和专业知识结构更新的挑战。人机结合的思维方式可以辅助人类突破个体认知的极限，教师与智能设备的协同教学应是未来教育的理想范式（余胜泉等，2019）。在传统中小学写作教学中，教师批改作文耗时耗力且反馈滞后，学生也普遍存在畏难情绪，写作动机不高。近年来，智能作文评价为写作教学革新带来新契机。智能作文评价系统是人工智能、互联网、语言认知等多个领域交叉融合的产物，能为作文提供即时评分和内容反馈，其研发和应用为人机协同写作教学奠定了基础。

智能作文评价可以追溯到20世纪60年代美国学者Page（2003）研发的第一款智能作文评价系统PEG（Project Essay Grade）。早期的智能作文评价系统只能为作文评分，不提供任何形式的内容反馈。上世纪90年代以后，智能作文评价技术发展迅速，美国研发的Criterion、MY Access！E-rater、Intelligent Essay Assessor（简称IEA）等代表性工具不仅能对作文进行自动评分，还提供多维度反馈，适用于形成性和总结性的作文评价，有的已经被应用在TOEFL（Test of English as a Foreign Language）和GMAT（Graduate Management Admission Test）等大规模英语测试中（唐锦兰等，2011）。近年来，智能作文评价系统的评价对象也逐渐由英语扩展到阿拉伯语、中文、日语等其他语种（Ishioka et al.，2006；Hao et al.，2016；Azmi et al.，2019），IN课堂语文作文智能批改系统（简称IN课堂）、爱语文APP等都是具有代表性的中文智能作文评价系统。先前研究表明，IN课堂的评分效度良好，与教师作文评价在反馈类型与层次上各具特点，二者具有协同教学的基础（刘淑君等，2021）。

人机协同就是“人”与“机器”之间的交互和协作，人与机器做各自擅长的工作，比如机器负责数据计算和部分推理工作，人类负责高阶选择和任务决策，以此实现二者的密切协作与相互协同（钱学森等，1990；刘步青，2016）。在本研究中，“人”和“机”分别指“初中语文教师”和在国内应用较广的“IN课堂语文作文智能批改系统”。IN课堂属于交互式的智能作文批改和反馈平台，于2018年正式投入应用。它借助自然语言处理技术，从语料库中挖掘打分细则、评级参数和偏误规则，不仅可以对学生作文即时评分，还能提供反馈建议。利用IN课堂，学生可随时提交作文并获得即时反馈，教师可通过系统建立虚拟班级、发布作业、补充反馈信息、查看学生写作的行为和结果数据等。IN课堂支持“学生—系统”“教师—系统”和“教师—学生”等维度的多元交互（IN课堂，2018）。“人机协同”指教师和IN课堂密切协作、优势互补，为学生写作提供尽可能多的支持。本研究将借助Z校写作拓展课探究人机协同作文评价在初中写作教学中的应用效果，以期为智能技术与写作教学的深度融合提供借鉴。

**二、相关文献综述**

智能作文评价通过在技术、学生、教师和同伴之间建立交互来支持写作教学。在“学生—系统”的交互中，学生能借助系统完成构思和写作，并自动获得评分和反馈，继而修改和完善文本，这增加了写作练习与反馈的机会，有利于学生写作能力提升（Kellogg，2008）。“教师—系统”的交互包括教师查看学生写作行为与结果数据、监控学生写作进度等，有助于教师及时调整教学计划（Grimes et al.，2010）。借助系统，学生还可以向教师发送信息，教师也能对学生作文做调整性评分和补充反馈。有的智能系统还开通了同伴互评功能，增加学生之间的深度互动（Balfour，2013）。

经过多年的研究和发展，作文的智能评分算法已具备一定的准确性和可靠性。多项研究表明，人机评分一致性已达到较高的水平（Shermis，2014；刘淑君等，2021）。然而，智能作文评价的教学应用效果依然存在争议。一些研究者认为智能作文评价会削弱写作固有的社会性，担心学生仅与机器交互会让其作文变得呆板（National Council of Teachers of English，2013）。因此，研究者持续探究智能作文评价对学生写作质量的影响。近年来，其对学生写作动机以及教师教学行为的影响也被广泛关注。

1.智能作文评价对学生写作质量的影响

研究者发现，不同评价工具、学生样本、教学环境下，智能作文评价系统对学生写作质量的影响效果并不一致（Stevenson et al.，2014）。一方面，Wang等（2013）对57名以英语为非母语的大学生实施为期16周的对照实验（实验组31人、对照组26人），发现智能反馈组较对照组（教师反馈）在文本准确性上有显著优势；Chodorow等（2010）针对463名大学生的作文文本分析表明，智能反馈系统Criterion可以降低非母语学生的英语文本错误率；Choi（ 2010）的研究也发现，当智能反馈融入以英语为非母语的大学生写作教学时，学生作文质量往往会得到提高。另一方面，Shermis等（2008）对美国某地区2017名不同年级的学生借助智能作文反馈系统完成的11685篇文章进行分析，结果发现，经过一段时间的练习，各年级学生作文质量都进步显著，且八年级学生作文成绩提升幅度以及错误率降低幅度都最大，但这可能主要是学生写作修改能力随着年级递增而带来的结果。Ware（2014）的研究则发现，接受同伴反馈、教师反馈和智能反馈（Criterion）的三组初中生在作文准确性、流畅性与整体分数上没有显著差异，教师反馈与同伴反馈较智能反馈对学生文体要素使用有更积极的影响。还有学者将美国某地区114名六年级学生分为实验组（PEG反馈）和对照组（使用谷歌文档编辑文本）开展对照实验，结果也发现，两组学生后测作文质量无显著差异，且较前测都没有显著提升（Wilson et al.， 2020）。可见，当前智能作文评价的教学应用效果尚未获得学界共识。

2.智能作文评价对学生写作动机与写作自我效能感的影响

写作动机是指学生对写作的兴趣、价值认知、努力程度、目标取向等，也有学者将其划分为对写作的兴趣、毅力和激情等维度（Troia et al.， 2012）。写作自我效能感是指学生对自己写作能力的自信程度（Bruning et al.， 2016）。已有研究表明，尽管写作态度并不能直接影响学生写作质量，但具有积极动机和信念的学生倾向于多写多练；学生自我效能感越强，在写作上付出的努力就越多，也会更乐于迎接挑战（Troia et al.， 2013）。写作反馈是影响学生写作动机和自我效能感的关键因素（Dujinhower et al.， 2010）。智能作文反馈对普通学生和学习障碍生的写作动机都有一定的促进作用（Grimes et al.， 2010；Morphy et al.， 2012）。针对美国四所中学师生的一项调查发现，使用智能作文反馈可以提升学生的写作动机和修改行为，提升幅度显著高于使用文字处理软件的对照组（Grimes et al.， 2010）。使用Write-to-Learn反馈系统的学生在创作和修改的时间投入和参与持久性上都有所增加，学生写作动机的毅力维度也有明显提升（Foltz et al.， 2013）。同时，也有研究发现，智能作文评价支持的学生写作自我效能较传统教师作文评价支持的学生更积极，写作自我效能感在一定程度上调节了作文反馈条件对写作成绩的影响（Wilson et al.， 2020）。可见，智能作文评价的应用对学生写作动机和自我效能感有积极影响，后续可以推进更多场景中的研究。

3.人机协同作文评价研究

智能作文评价的教学应用离不开教师的协同支持，这种人机协同也在一定程度上影响了教师的教学实践。Wilson等（2016）以八年级145名学生为样本，探究了“PEG＋教师”协同评价与教师单独评价条件下教师作文反馈数量、类型与层次之间的差异。结果表明，两组教师反馈在数量和类型上没有显著差异，但人机协同评价组教师的高阶反馈信息数量显著高于对照组。Link等（2020）以高校英语课中28名学习者为样本（实验组12人，对照组16人），研究了智能作文评价对教师反馈、学生作文修改和写作质量的影响，结果显示，实验组（“教师＋Criterion”评价）教师的高阶反馈数量较对照组（教师评价）没有显著变化，然而低阶反馈数量较对照组显著减少。可见，智能作文评价会影响教师作文反馈层次，让教师压缩低阶写作技能反馈的时间，从而聚焦高阶写作技能反馈。

那么，人机协同作文评价的教学效果如何呢？在Wilson等（2016）的研究中，“PEG＋教师”评价组的学生作文质量较教师评价组没有显著变化，但学生写作动机在毅力维度得到显著提升。而Link等（2020）的研究发现，“教师＋Criterion”协同评价条件下学生文本准确性没有显著提升，但写作技能留存率要高于对照组。黄静等（2018）将句酷批改网的评价与同伴互评、教师评价等方式相结合，以25名参与大学英语课程的学生为样本开展单组实验，对比学生的作文初稿和终稿质量发现，人机协同评价对大学生英语写作质量有明显促进作用。

综上，虽然当前有关智能作文评价、人机协同作文评价教学应用的研究已取得一定进展，但还存在争议，且多局限于英语语境。中文智能作文评价系统兴起较晚，相关研究还比较薄弱，此类系统与写作教学的融入路径及效果还有待实证探究。目前中小学语文的写作教学中，评价方式单一、批改耗时费力、反馈滞后等问题依然是顽疾。人机协同作文评价能否破解作文教学和评价中的难题，促进中小学写作教学实践，已成为人们关注的重要话题。

**三、研究设计**

1.研究目的与问题

为了探究人机协同作文评价在初中写作教学中的应用效果以及智能作文评价融入初中写作教学的路径，本研究依托Z中学七年级写作拓展课开展了一项教育准实验研究。研究者将选课学生分为实验组和对照组，分别接受人机协同作文评价和教师作文评价的教学实践，以此探究人机协同作文评价对学生作文成绩、写作动机和教师作文反馈的影响，并调研学生对人机协同作文评价的态度。研究问题包括：（1）人机协同作文评价能提升学生的作文成绩吗？（2）人机协同作文评价对学生写作动机的影响如何？（3）人机协同作文评价对教师反馈信息的数量、类型和层次是否有影响？（4）学生对人机协同作文评价的体验和感受如何？

2.研究对象

本研究选取参加Z校写作拓展课的28名七年级学生为研究对象，学生来自同一个教学班，日常学习环境和课程任务安排保持一致。实验课程持续5周，每周2个课时。研究者根据语文成绩和性别均衡原则将学生分为两组，实验组14人（男生7人，女生7人），对照组14人（男生8人，女生6人）。两组学生在完成作文初稿后分别在人机协同评价和教师作文评价的辅助下修改文本。

3.准实验设计

写作拓展课之前，授课教师为师生申请IN课堂账号，并建立虚拟班级，学生在机房练习账号登录、上传作文、查看反馈内容、修改提交等操作。准实验前，所有学生完成写作动机前测问卷调查。在准实验过程中，教师以“倒叙的魅力”和“描写：描出世间景象”为主题开展两次写作活动，每次活动持续两周时间，每周2个课时。第一周，教师对学生进行写作前指导，然后学生完成作文初稿并上交。第二周，学生在机房根据作文反馈信息修改作文；之后教师适时组织交流活动。两组学生写作前指导和撰写初稿的环境保持一致，准实验干预主要表现在作文评价、反馈与修改环节。实验组学生提交初稿后，教师将作文上传IN课堂系统得到智能评分和内容反馈；之后，在审阅智能评价的基础上，教师对学生作文进行评分，并以文内批注和文末总评的方式完成反馈。IN课堂会针对标点、错别字、语法等提供低阶反馈建议，也会针对内容意义、结构、构思等提出高阶提升建议。鉴于机器反馈存在低阶反馈欠全面、高阶反馈欠具体等问题，教师需要在审阅机器反馈的基础上进行补充性反馈。对照组学生提交初稿后，教师直接以文内批注和文末总评的方式书写反馈信息并评分。在修改作文环节，实验组学生可以登录IN课堂平台，上传初稿得到智能评价，之后可多次修改并得到反馈；同时学生还可参照教师反馈信息改进作文。对照组学生仅根据教师反馈内容修改作文。两组学生修改满意后需向教师提交终稿。

准实验过程中，研究者收集学生初稿和终稿成绩数据，整理教师作文反馈信息。实验结束后，学生完成写作动机后测问卷和批改方式体验问卷，研究者对实验组学生开展小组焦点访谈，同时计算学生作文成绩，并对两组教师作文反馈信息进行编码分析。

4.数据收集与分析

本研究的数据包括学生作文成绩、学生写作动机、教师作文反馈、学生体验与感受等内容。

（1）学生作文成绩数据的收集与分析

首先，研究以教师评分和IN课堂评分的均值作为衡量学生成绩的指标。教师评分和IN课堂评分都遵循中考作文评分标准，以40分为满分成绩。研究者将学生作文初稿和终稿都提交IN课堂系统，由此得到智能评分数据；同时教师对两组学生的初稿和终稿成绩进行评分，由此得到教师评分数据。研究者取两组评分的均值作为学生作文最终成绩。其次，研究以学生作文初稿成绩为协变量、以作文评价方式为自变量、以作文终稿成绩为因变量做单因素协方差分析，用以调查在排除前测成绩影响之后，两组学生作文终稿成绩是否存在显著差异；同时研究对两组学生作文练习的初稿和终稿成绩分别执行配对样本t检验，以此调查两组学生作文终稿成绩较初稿的变化情况。

（2）学生写作动机数据的收集与分析

研究者在准实验前后分别调查两组学生的写作动机。学生写作动机量表改编自Piazza等学者提出的量表（Piazza et al.， 2008），共11题，包括学生对写作的自信（3个题项）、毅力（4个题项）和激情（4个题项）三个维度，采用李克特七点量表设计。该量表在研发阶段的克隆巴赫系数为0.89，在前后测中，总量表和三个维度的克隆巴赫系数都处于0.75~0.92，表明量表信度良好。

研究以两组学生写作动机前测为协变量、评价方式为自变量、写作动机后测为因变量执行单因素协方差分析，用以验证经过实验干预两组学生的写作动机是否存在差异；同时对两组学生写作动机前后测得分进行配对样本t检验，以此探究两组学生写作动机在准实验前后的变化情况。

（3）教师作文反馈数据的收集与分析

本研究借鉴Wilson等（2016）的作文反馈信息分类方法，根据国内作文批阅习惯进行改编，比如，将“拼写和大小写”合并为“错别字”，将“语法和句子结构”合并为“句式语法”等，最终构建了适用于国内中学作文评价的反馈信息编码表（见表1）。该表将用于对教师作文反馈信息的编码分析。

**表1　作文反馈信息编码表（改编自Wilson et al.，2016）**

研究首先将两组教师作文反馈信息逐条录入Excel表格，随后将反馈信息拆分为有独立意义的反馈单元。遵循文献中反馈单元的拆分程序（Cho et al.， 2006；Hayes et al.，2010），先抽取5篇作文的反馈信息，研究者和合作教师分别对其做单元拆分，随后比对拆分结果，对不一致处通过协商达成共识。之后，研究者完成全部作文反馈信息的拆分工作，合作教师再对拆分后的反馈单元进行抽样检查。通过以上拆分程序，本研究共得到677条作文反馈单元，其中实验组289条，对照组338条。为了尽可能避免作文长度对教师反馈数量的影响，研究者将每篇作文的反馈单元数转化为每千字的反馈单元数，之后用SPSS 23对两组数据执行独立样本非参数检验，以比较两组间教师反馈数量的差异。

参照已有研究中的编码程序（刘淑君等，2021），研究者与合作教师从反馈类型与层次两个维度对所有反馈单元进行编码，反馈类型的编码一致性为0.91，反馈层次的编码一致性为0.83。编码样例如表2所示。

**表2    教师作文反馈信息的编码样例**

为避免作文长度和作文篇数对教师反馈的影响，研究者计算了每篇作文中直接反馈、提问型反馈、信息型反馈和表扬等4种反馈类型的比率，以及高阶和低阶反馈所占比率，以便对两组教师作文反馈类型与层次进行科学对比。之后，研究者借助SPSS 23对两组教师作文反馈的类型与层次比率执行一系列非参数检验，用以验证两组教师在反馈类型与层次上是否存在差异。

（4）学生体验与感受类数据的收集与分析

准实验结束后，实验组学生完成对人机协同作文评价的体验问卷。该问卷改编自已有研究文献（Grimes et al.，2010；Palermo et al.， 2018），主要调查学生对人机协同作文评价的反馈有效性和接受度等方面的感知。问卷采用李克特五点量表设计，共7题。此外，研究者将实验组学生分为两组，每组7人，围绕学生对人机协同作文评价的体验与感受开展小组焦点访谈，访谈问题包括学生对人机协同作文评分与反馈的态度、针对IN课堂的改进建议等，每场访谈时间为25~30分钟，访谈全程录音。

研究采用SPSS 23对学生问卷各题项得分做描述性统计，之后将其均值与中立分（3分）做均值差异比较，用以检验学生对人机协同作文评价的态度是否显著高于中立分。所有访谈录音都被转录为电子文档，研究者围绕焦点问题对访谈内容分类整理，深入调查学生对人机协同作文评价的体验与感受。

**四、研究发现**

1.人机协同作文评价对学生作文成绩的影响

研究以作文初稿成绩为协变量、反馈方式为自变量、终稿成绩为因变量对学生两次作文终稿成绩进行协方差分析。表3显示，两次作文中，两组学生终稿成绩之间都不存在显著差异（p＞0.05）。

**表3    实验组与对照组的作文成绩的协方差分析（n＝28）**



研究者对两组学生作文初稿和终稿成绩分别进行配对样本t检验（见表4），结果显示在两次作文练习中，实验组和对照组的作文终稿成绩都较初稿有显著提升（p＜0.05）。由此可见，尽管两种评价都有助于提高学生的作文成绩，但人机协同作文评价相对于教师单独评价对学生作文成绩的影响并不显著。

**表4    实验组与对照组两次作文成绩的配对样本t检验（n＝28）**



2.人机协同作文评价对学生写作动机的影响

研究以学生写作动机前测为协变量、评价方式为自变量、后测为因变量对写作动机三个维度的得分进行单因素协方差分析。结果显示，实验组在写作激情维度上得分显著高于对照组（F＝4.30，p＜0.05），在自信和毅力维度上得分与对照组没有显著差异（见表5）。可见，人机协同作文评价可以促进学生写作动机在激情维度上的显著提升。

**表5　实验组和对照组学生写作动机前后测的单因素协方差分析（n＝28）**



3.人机协同作文评价对教师反馈数量、类型与层次的影响

通过分析教师对学生两次作文（共56篇）的反馈内容发现，实验组教师每千字反馈单元的中位数为15.12，对照组为26.09，独立样本非参数检验显示，二者之间的差异具有显著性（Z＝4.13，p＜0.01）。人机协同作文评价条件下，教师的作文反馈数量显著减少，大约是教师单独反馈数量的2/3。

通过对教师作文反馈信息进行拆分、编码和统计，研究者计算了两组教师各类型反馈所占比率，并对两组数据执行独立样本非参数检验。表6显示，实验组和对照组的教师反馈在直接反馈、信息型反馈和表扬等类型的比率上没有显著差异；然而，人机协同评价组的提问型反馈比率（0.13）低于对照组的提问型反馈比率（0.23），且该差异具有显著性（Z＝2.42，p＜0.05）。由此可见，人机协同作文评价会减少教师反馈数量，并显著降低教师提问型反馈的比率。

**表6　实验组与对照组教师反馈类型的非参数检验（n＝56）**

对两组教师反馈层次的比率进行独立样本非参数检验，如表7所示，在高阶反馈的具体类别中，实验组教师反馈在主题立意（Z＝2.09，p＜0.05）和结构构思（Z＝2.11，p＜0.05）层面的比率都显著高于对照组；然而在写作特色层面的反馈比率却显著低于对照组（Z＝2.16，p＜0.05）。从整体上看，人机协同评价组的教师高阶反馈比率（0.75）高于教师评价组（0.70），然而这种差异并不具有显著性（Z＝1.36，p＞0.05）；同时，两组教师低阶反馈比率之间也不存在显著差异（Z＝1.12，p＞0.05）。可见，人机协同作文评价能提升教师对主题立意、结构构思等层面的高阶反馈比率，降低教师对写作特色维度的反馈比率；然而，从整体上看，人机协同评价对教师高、低阶反馈比率的影响并不显著。

**表7　实验组与对照组教师反馈层次的非参数检验（n＝56）**

4.学生对人机协同作文评价的体验与感受

问卷调查发现（见表8），实验组学生7个题项得分都显著高于中立分。可见，学生对人机协同作文评价的态度整体上比较积极。从各题项的描述统计看，前6题学生选择“强烈同意”和“同意”的超过60%。然而，针对第7题（我希望之后继续采用“教师＋IN课堂”的作文评价方式），50%的学生选择“强烈同意”或“同意”，还有50%的学生选择“中立”，可见，一半学生对“继续采用‘教师＋IN课堂’的作文评价方式”持保留态度。

**表8　学生对人机协同作文评价方式的体验与感受（n＝14）**





（注：均分指学生在李克特五点量表各题项的平均得分，其中“强烈反对”＝1分，“中立”＝3分，“强烈同意”＝5分；p指的是各题项均分和中立分3分之间不存在显著差异的概率。）

学生访谈分析显示，有3位同学（21.43%）对人机协同作文评分持肯定态度，10位同学（71.43%）持否定态度，1名同学（7.14%）没有明确表态。持肯定态度的同学认为，智能作文评价系统不仅能给作文进行整体评分，还可以从内容、表达和发展等维度分项评分，能弥补教师只打整体分的不足；系统做到“秒评分”，能让学生产生持续修改提交的动力；另外，用智能评分和教师评分的均值作为作文最后得分，也让评分更公正。持否定态度的同学将矛头指向修改过程中的智能评分，比如有同学认为自己“修改了错别字，疏通了语句，还丰富了内容，作文分数不升反降”；也有同学提出，“有时什么都没改，作文分数会莫名提升”。学生认为系统对文本内容变化的反应不灵敏。

针对人机协同作文反馈，有8位同学（57.14%）持肯定态度，6位（42.86%）持保留态度。持肯定态度的同学认为教师和IN课堂合作给出的反馈信息更全面，能更好地支持学生修改作文；系统善于寻找句段亮点，“不吝表扬”也能提升学生的写作信心和兴趣。持保留态度的同学认为，IN课堂低阶反馈不全面，高阶建议存在笼统、重复和套路化等问题，修改难度较高。有同学以作文《吃遍天下荷兰猪》举例说，系统“建议选取有情感的材料、事件，以增强文章的可读性和感染力”，她并没有找到具体修改方向；而教师建议她“考虑一下，和同伴相比，这只猪的生活发生了什么变化？他满意吗？周围的人满意吗？当他空闲下来时，是否会思考曾经的梦想和人生意义？等等”，看到这些问题，她会不由自主地想去回答，而回答就能让作文内容得到扩展。此外，也有人提出，修改作文过程中既要查看教师反馈信息，也要登录IN课堂查看评价结果，思维需要在二者之间切换，会有点手忙脚乱。学生接收到的信息过多时，需要对信息进行筛选和甄别，这会分散注意力，增加学生的认知负荷。

学生也对人机协同作文评价提出了改进建议，提及频率最高的前三条包括：提高智能评分的准确性和灵敏性；提升智能反馈内容的全面性和具体性；教师需调整写作修改程序，让操作更便捷。这些为人机协同作文评价的未来发展指明了努力方向。

**五、讨论**

通过为期5周的教学准实验，本研究探讨了人机协同作文评价和教师作文评价条件下学生写作表现与教师反馈的差异，获得了一些有意义的研究发现。

1.相较于教师作文评价，人机协同作文评价对学生作文成绩并没有产生显著影响

尽管两组学生的作文终稿成绩都较初稿有显著提升，然而，人机协同作文评价与教师作文评价相比，在提升作文终稿成绩方面并没有表现出明显优势，这与先前类似研究结果具有一致性（Wilson et al.， 2016；Link et al.， 2020）。该结果可以从以下三方面来理解：首先，对照组采用教师作文评价，教师对作文精批细改，这是顺应师生多年习惯的反馈方式，具有较好的效果。其次，作文反馈质量和学生落实反馈的效果有时并不一致。反馈实施的效果不仅受反馈内容本身的影响，也取决于学生对反馈信息的理解和认同程度（Nelson et al.，2009）。高写作水平的学生能理解各类反馈信息并将其落实在写作修改中，由此带来终稿成绩的大幅提升；而低写作水平的学生能落实部分低阶修改建议，对高阶建议的落实程度往往并不理想。因此，即使人机协同评价可以提供更全面和丰富的反馈，如果学生不能有效落实，也不能带来写作质量的显著提升。最后，写作是一项复杂技能，其提升需要长时间的持续干预，本次实验时间有限，学生写作能力的细微进步未必能体现在初稿到终稿的成绩变化中，也难以在与对照组的对比中达到显著水平。

2.相较于教师评价，人机协同作文评价能显著提升学生写作激情

写作能力的理论模型和实证研究都强调了写作动机对提高学生写作质量的重要性（Graham et al.， 2007；Hayes， 2012）。经过实验干预，实验组学生写作动机的激情维度得分显著高于对照组，两组学生在自信和毅力两个维度没有显著差异。这与已有研究结论基本一致。多项研究发现，智能作文评价系统的教学应用，不管是单独使用还是与教师反馈协同作用，都可以让学生写作动机在某些维度得到显著提升（Warschauer et al.，2008；Grimes et al.，2010；Wilson et al.，2016）。学生访谈显示，智能作文评价会带给学生一些新鲜的体验，能激发学生对写作的兴趣；同时，系统提供多次修改和反馈机会，也能提升学生迎接挑战的兴致。

值得注意的是，两组学生写作动机的自信和毅力两个维度都没有显著差异，或许我们还需要慎重推论人机协同作文评价对学生写作动机的积极影响。本阶段教学实验时间较短，写作动机从个别维度提升扩展到全方位的进步需要时间积淀。后续研究可以考虑让学生在更多轮次的“写作—反馈—修改”实践中提升写作的自信心、成就感和解决问题的毅力；教师也可以在长期教学实践中观察学生写作动机某个维度的发展是否会促进其他维度的提升，并研究这种提升与写作质量之间的相关性（Ramineni et al.， 2015）。值得注意的是，如果写作动机的提升是应用智能技术后的新奇效应产物，那么随着时间的推移，写作动机或许会停滞增长或者消退（Cheung et al.， 2013）。

3.人机协同作文评价使得教师反馈数量减少，提问型反馈比率降低

人机协同作文评价条件下教师反馈数量较教师单独评价条件下有所减少，这与Link等（2020）针对英语作文评价的相关研究结果是一致的。值得注意的是，本研究中实验组学生接受的是教师和机器的双重反馈，二者反馈数量之和要高于教师评价组，即学生得到的整体反馈数量是增加的。可以推测，教师审阅智能反馈建议后，会在此基础上作补充性反馈，避免提供重复性信息。教师作文反馈数量减少意味着批改作文花费的时间缩减，批改负担在一定程度上得到缓解。当然，鉴于教师需要审阅和甄别智能反馈信息，这对教师的专业素养和教学智慧提出了更高要求。

从反馈类型看，实验组教师提问型反馈比率显著低于对照组，这与先前研究结果并不一致。Wilson等（2016）的研究发现，人机协同评价和教师评价条件下的教师反馈在类型上不存在显著差异。提问型反馈在激发学生思考和促进师生交流方面有重要作用，需要教师沉浸于文本内容与学生进行深度对话。在人机协同作文评价条件下，教师要审阅并甄别智能反馈的大量信息，并思考如何与之互补，这或许会压缩教师与学生对话的时间和空间，继而导致提问型反馈减少。

先前研究发现，人机协同作文评价会让教师更专注于高阶反馈（Wilson et al.， 2016；Link et al.， 2020），但是本研究中，实验组教师低阶反馈和高阶反馈较对照组没有显著变化，该结果还不能为已有研究提供新证据。IN课堂作文低阶反馈包括对文本中标点、错别字、词语运用、句法等问题的反馈，然而其反馈效果还不理想，会遗漏很多问题；同时，IN课堂提供大量针对内容选材和结构构思的高阶反馈，但反馈的具体性和针对性还有待改进（刘淑君等， 2021）。人机协同评价条件下，教师在低阶反馈和高阶反馈维度都需要对智能反馈进行补充，或许这也是教师高阶比率没有显著提升的重要原因。比较理想的人机协同状态应该是：机器和教师各司其职，前者负责语言质量，后者专注高阶点拨。然而，基于IN课堂的中文智能作文评价对低阶错误的识别率还有待提升，教师对作文低阶反馈的工作还不能被完全代替。未来写作教学实践中，我们既要加强教师人工智能素养、数据素养和测评素养的培养，也要督促产品研发方不断迭代更新评价技术，以此提升人机协同教学质量（汪琼等，2021）。

4.学生基本认可人机协同作文评价的价值，但对继续使用该方式持保留态度

针对人机协同作文评价的体验与感受的问卷调查显示，学生对该评价方式整体上持积极态度，但一半学生对继续使用该评价方式持中立态度，即不确定是否继续使用。该结果看似矛盾，如果结合学生访谈内容来看，会让人有更深入的理解。学生访谈显示，人机协同评价方式会让评价更全面，增加修改和反馈机会，能提升学生修改作文的意愿；然而，双重反馈信息让写作修改过程变得繁琐而复杂，在一定程度上增加了学生的认知负荷。此外，学生也对IN课堂系统提出了改进建议，比如提升智能评分的精确性，增加智能反馈的具体性和全面性，调整教学设计步骤等。智能作文评价系统是计算机科学、语言学、心理学等多个学科交叉融合的成果，它应该不断前进且持续更新（孟翀等， 2021）。学生对人机协同作文评价的“矛盾”态度恰好揭示了该评价方式的潜力所在和改进方向。

在本研究中，教师反馈建立在审阅智能反馈的基础上，受限于客观条件，学生并没有提前与IN课堂互动，只在人机协同反馈条件下完成一轮集中修改，写作修改过程略显仓促。写作修改是一个复杂问题解决的过程，后续研究应该为学生创设更丰富的写作场景，支持学生完成更多轮次的写作修改，优化人机协同写作教学的实施路径。

**六、结论与建议**

随着智能技术在教学中的广泛应用，智能作文评价与写作教学的深度融合是未来写作教育的重要趋势。本研究探究了智能作文评价融入写作教学的重要方式“人机协同作文评价”的教学效果，得出以下基本结论：首先，与教师作文评价相比，人机协同作文评价对作文成绩的影响并不显著，然而它能在一定程度上提升学生写作动机，对学生的“学”有积极的促进作用；其次，人机协同作文评价会对教师反馈产生影响，比如，教师反馈数量减少，提问型反馈比率降低，教师“教”的优势还有待进一步发挥；最后，教师只有重构知识能力结构，并不断进行教学革新，辅之智能作文评价系统的持续改进，人机协同作文评价才能发挥出最大优势，更好地变革传统写作教学。

基于以上结论，本研究围绕人机协同写作教学领域的实践和研究提出以下三点建议：

首先，语文教师需要探索人机协同写作教学的多元路径，找准定位，进一步发挥人机协同中的“人”的优势作用。教师首先需要转变角色，提升人工智能素养，在与智能技术的协同教学中找到最佳着力点，给予学生最切实的反馈和帮助。“人机协同”中的“人”除了语文教师外，还可以扩展为学生本人、同伴甚至家长；“人”和“机”的评价与反馈可以同时呈现，让学生根据两种反馈内容修改作文，也可以是时空分离的，学生在不同轮次的写作修改活动中体验不同的反馈方式。此外，教师对学生的写作支持也是一个由扶到放的过程。如果始终以“精批细改”来“哺育”学生，教师的时间和精力都将难以维系，也有违现代写作教育的初衷。因此，“人机协同”不仅要体现在评价反馈环节，也可以表现为“教师负责策略点拨与信念培育、机器负责即时反馈与过程监控”的协同范式，以此协助学生在与机器互动中提升能力，逐步成长为独立作者。

其次，在写作教学中，教师要好好关注和培育学生的写作修改能力和信念。写作修改是一个复杂认知过程，具有较强的实践性。传统写作教学对写作修改环节不够重视，学生较少得到写作修改的过程指导和质量评价，大部分初中生写作修改能力较弱，写作修改信念低下。如果师生正视这一问题并努力解决，学生就能更好地理解并落实作文反馈信息，人机协同作文评价的实施效果也会得以提升。

最后，产学研合作有利于智能作文评价系统的优化和推广。智能作文评价的研发机构需要和中小学校一线语文教师以及教育研究者通力合作，基于写作教学和评价规律一起探索系统的优化路径。只有立足师生写作教学的需求与评价标准，找准系统在人机协同教学中的定位，在技术上不断优化，智能技术才能在人机协同作文教学中更好地发挥优势。目前，基于IN课堂的智能作文评价在提高语言错误识别率、提升反馈内容具体性、加强写作过程支持等方面还存在提升空间，产学研合作有利于汇集多方力量攻坚克难，最终让人机协同在写作教学中发挥轻负优质的最佳效果。

**参考文献：**

[1]黄静,何华清(2018).人机反馈对学生写作行为的影响研究[J].外语电化教学,(1):19-24.

[2]IN课堂(2018).IN课堂——语文作文智能批改教育迈向智能化阶段[EB/OL].[2021-04-22].http://inketang.com/v8/news\_detail\_00.html.

[3]刘步青(2016).人机协同系统的推理机制及其哲学意蕴[D].上海:华东师范大学.

[4]刘淑君,李艳,杨普光等(2021).智能作文评价的效果研究[J].开放教育研究,27(3):73-84.

[5]孟翀,王以宁(2021).教育领域中的人工智能:概念辨析、应用隐忧与解决途径[J].现代远距离教育,(2):62-69.

[6]钱学森,于景元,戴汝为(1990).一个科学新领域——开放的复杂巨系统及其方法论[J].自然杂志,(1):3-10,64.

[7]唐锦兰,吴一安(2011).在线英语写作自动评价系统应用研究述评[J].外语教学与研究,43(2):273-282,321.

[8]汪琼,李文超(2021).人工智能助力因材施教:实践误区与对策[J].现代远程教育研究,33(3):12-17,43.

[9]余胜泉,王琦(2019).“AI＋教师”的协作路径发展分析[J].电化教育研究,40(4):14-22,29.

[10]中华人民共和国中央人民政府(2019).中共中央、国务院印发《中国教育现代化2035》[EB/OL].[2021-05-15].http://www.gov.cn/xinwen/2019-02/23/content\_5367987.htm.

[11]Azmi, A. M., Al-Jouie, M. F., & Hussain, M. (2019). AAEE-Automated Evaluation of Students’Essays in Arabic Language[J]. Information Processing & Management, 56(5):1736-1752.

[12]Balfour, S. P. (2013). Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review[J]. Research & Practice in Assessment, 8(1):40-48.

[13]Bruning, R. H., & Kauffman, D. F. (2016). Self-Efficacy Beliefs and Motivation in Writing Development[M]// Mcarthur, C. A., Graham, S., & Fitzgerald, J. (Eds.). Handbook of Writing Research. New York, NY: Guilford Press:160-173.

[14]Cheung, A. C. K., & Slavin, R. E. (2013). The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis[J]. Educational Research Review, 9:88-113.

[15]Cho, K., Schunn, C. D, & Charney, D. (2006). Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts[J]. Written Communication, 23(3):260-294.

[16]Chodorow, M., Gamon, M., & Tetreault, J. (2010). The Utility of Article and Preposition Error Correction Systems for English Language Learners: Feedback and Assessment[J]. Language Testing, 27(3):419-436.

[17]Choi, J. (2010). The Impact of Automated Essay Scoring (AES) for Improving English Language Learner’s Essay Writing[D]. Charlottesville, VA: University of Virginia．

[18]Dujinhower, H., Prins, F. J., & Stokking, K. M. (2010). Progress Feedback Effects on Students’ Writing Mastery Goal, Self-Efficacy Beliefs, and Performance[J]. Educational Research and Evaluation, 16(1):53-74.

[19]Foltz, P. W., Streeter, L. A., & Lochbaum, K. E. et al. (2013). Implementation and Applications of the Intelligent Essay Assessor[M]// Shermis, M. D., & Urstein, J. J. (Eds.). Handbook of Automated Essay Evaluation. New York, NY: Routledge:66-88.

[20]Graham, S., Berninger, V., & Fan, W. (2007). The Structural Relationship Between Writing Attitude and Writing Achievement in First and Third Grade Students[J]. Journal of Educational Computing Research, 33:53-80.

[21]Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation[J]. Journal of Technology Learning & Assessment, 8(6):1-44.

[22]Hao, S. D., Xu, Y. Y., & Ke, D. F. et al. (2016). SCESS: A WFSA-Based Automated Simplified Chinese Essay Scoring System with Incremental Latent Semantic Analysis[J]. Natural Language Engineering, 22(2):291-319.

[23]Hayes, J. R. (2012). Modeling and Remodeling Writing[J]. Written Communication, 29(3):369-388.

[24]Hayes, J. R., & Berninger, V. W. (2010). Relationships Between Idea Generation and Transcription: How the Act of Writing Shapes What Children Write[M]// Braverman, C. et al.  (Eds.). Traditions of Writing Research. NewYork, NY: Routledge:166-180.

[25]Ishioka, T., & Kameda, M. (2006). Automated Japanese Essay Scoring System Based on Articles Written by Experts[C]// Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. Sydney, Australia: Association for Computational Linguistics:233-240.

[26]Kellogg, R. T. (2008). Training Writing Skills: A Cognitive Developmental Perspective[J]. Journal of Writing Research, 1(1):1-26.

[27]Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of Automated Writing Evaluation on Teacher Feedback, Student Revision, and Writing Improvement[J]. Computer Assisted Language Learning, (3):1-30.

[28]Morphy, P., & Graham, S. (2012). Word Processing Programs and Weaker Writers/Readers: A Meta-Analysis of Research Findings[J]. Reading and Writing, 25(3):641-678.

[29]National Council of Teachers of English (2013). NCTE Position Statement on Machine Scoring [EB/OL]. [2021-07-15]. http://www.ncte.org/positions/statements/machine\_scoring.

[30]Nelson, M. M., & Schunn, C. D. (2009). The Nature of Feedback: How Different Types of Peer Feedback Affect Writing Performance[J]. Instructional Science, 37:375-401.

[31]Page, E. B. (2003). Project Essay Grade: PEG[M]// Shermis, M. D., & Burstein, J. C. (Eds.). Automated Essay Scoring: A Cross-Disciplinary Perspective. Mahwah, NJ: Lawrence Erlbaum Associates, Inc:43-54.

[32]Palermo, C., & Thomson, M. M. (2018). Teacher Implementation of Self-Regulated Strategy Development with an Automated Writing Evaluation System: Effects on the Argumentative Writing Chock for Performance of Middle School Students[J]. Contemporary Educational Psychology, 54:255-270.

[33]Piazza, C. L., & Siebert, C. F. (2008). Development and Validation of a Writing Dispositions Scale for Elementary and Middle School Students[J]. Journal of Educational Research, 101(5):275-285.

[34]Ramineni, C., Calico, T., & Li, C.(2015). Integrating Product and Process Data in an Online Automated Writing Evaluation System[DB/OL]. [2021-06-15]. https://www.educationaldatamining.org/EDM2015/uploads/papers/paper\_309.pdf.

[35]Shermis, M. D. (2014). State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States Demonstration[J]. Assessing Writing, 20:53-76.

[36]Shermis, M., Garvan, C., & Diao, Y. (2008). The Impact of Automated Essay Scoring on Writing Outcomes[DB/OL]. [2021-06-15]. https://files.eric.ed.gov/fulltext/ED501148.pdf.

[37]Stevenson, M., & Phakiti, A. (2014). The Effects of Computer-Generated Feedback on the Quality of Writing[J]. Assessing Writing, 19:51-65.

[38]Troia, G. A., Harbaugh, A. G., & Shankland, R. K. et al. (2013). Relationships Between Writing Motivation, Writing Activity, and Writing Performance: Effects of Grade, Sex, and Ability[J]. Reading and Writing, 26:17-44.

[39]Troia, G. A., Shankland, R. K., & Wolbers, K. A. (2012). Motivation Research in Writing: Theoretical and Empirical Considerations[J]. Reading and Writing Quarterly: Overcoming Learning Difficulties, 28:5-28.

[40]Wang, Y. J., Shang, H. F., & Briody, P. (2013). Exploring the Impact of Using Automated Writing Evaluation in English as A Foreign Language University Students’Writing[J]. Computer Assisted Language Learning, 26 (3):234-257.

[41]Ware, P. (2014). Feedback for Adolescent Writers in the English Classroom: Exploring Pen-and-Paper, Electronic, and Automated Options[J]. Writing & Pedagogy, 6(2):223-249.

[42]Warschauer, M., & Grimes, D. (2008). Automated Writing Assessment in the Classroom[J]. Pedagogies: An International Journal, 3(1):22-36.

[43]Wilson, J., & Czik, A. (2016). Automated Essay Evaluation Software in English Language Arts Classrooms: Effects on Teacher Feedback, Student Motivation, and Writing Quality[J]. Computers & Education, 100:94-109.

[44]Wilson, J., & Roscoe，R. D. (2020). Automated Writing Evaluation and Feedback: Multiple Metrics of Efficacy[J]. Journal of Educational Computing Research, 58(1):87-125.

收稿日期　2021-09-18　责任编辑　汪燕



