

· 论 著 ·

# 护理专业学生大语言模型使用行为问卷的编制及信效度检验

马应卓<sup>1</sup>, 王俊<sup>1</sup>, 刘彤<sup>1</sup>, 杨阳<sup>2</sup>, 刘尚琴<sup>1</sup>, 赵庆华<sup>1</sup>, 肖明朝<sup>1</sup>

**摘要:**目的 编制护理专业学生大语言模型使用行为问卷,并检验其信效度。方法 以计划行为理论为指导,通过文献回顾、半结构化访谈、课题研究小组讨论、2轮专家函询和预调查,形成初始调查问卷。采用目的抽样法,2024年9—11月选取3所医学院符合条件的护理专业学生进行调查,收集有效问卷634份,进行信度和效度检验。结果 形成的护理专业学生大语言模型使用行为问卷包括态度、主观规范、知觉行为控制、使用意图和使用行为5个维度共28个条目,累计方差贡献率为66.584%,问卷各条目的内容效度为0.800~1.000,问卷平均水平的内容效度为0.914; $\chi^2/df=1.647$ ,RMSEA=0.044,CFI=0.951,IFI=0.951,TLI=0.945,GFI=0.891,NFI=0.885,结果在可接受范围。问卷总Cronbach's  $\alpha$ 系数为0.917,各维度为0.814~0.898;问卷总重测信度为0.823,各维度为0.770~0.864;问卷Spearman-Brown折半信度为0.775,各维度为0.823~0.905。结论 护理专业学生大语言模型使用行为问卷具有良好的信度和效度,可为护理教育者评估护生使用大语言模型意图及行为提供参考。

**关键词:**大语言模型; 生成式人工智能; 计划行为理论; 使用行为; 信度; 效度; 护理教育

**中图分类号:**R47;G455 **DOI:**10.3870/j.issn.1001-4152.2026.10.086

## Development and psychometric testing of the Large Language Model Usage Behavior

### Scale for Nursing Students Ma Yingzhuo, Wang Jun, Liu Tong, Yang Yang, Liu Shangqin, Zhao

Qinghua, Xiao Mingzhao. Department of Nursing, the First Affiliated Hospital of Chongqing Medical University, Yuzhong District, Chongqing 400016, China

**Abstract:** **Objective** To develop a scale for evaluating nursing students' usage behavior of large language models (LLMs) and to test its reliability and validity. **Methods** Guided by the Theory of Planned Behavior (TPB), an initial questionnaire was developed through literature review, semi-structured interviews, research team discussions, two rounds of expert consultation, and a pilot survey. A purposive sampling method was employed to survey eligible nursing students from three medical universities between September and November 2024. A total of 634 valid questionnaires were collected for reliability and validity testing. **Results** The finalized assessment questionnaire for nursing students' LLM usage behavior comprised five dimensions (attitude, subjective norms, perceived behavioral control, usage intention, and usage behavior) with 28 items in total and the cumulative variance contribution rate of 66.584%. The content validity index (CVI) of individual items ranged from 0.800 to 1.000, with an average CVI of 0.914. Confirmatory factor analysis demonstrated good model fit, with  $\chi^2/df=1.647$ , RMSEA=0.044, CFI=0.951, IFI=0.951, TLI=0.945, and GFI and NFI within acceptable ranges. The overall Cronbach's  $\alpha$  coefficient of the questionnaire was 0.917, with dimension-specific coefficients ranging from 0.814 to 0.898. The overall test-retest reliability was 0.823, with dimension-specific values ranging from 0.770 to 0.864. The overall split-half reliability was 0.775, with dimension-specific values ranging from 0.823 to 0.905. **Conclusion** The developed assessment questionnaire demonstrates good reliability and validity, providing a reference for nursing educators and administrators to evaluate nursing students' intentions and behaviors regarding LLM usage.

**Keywords:** large language model; generative artificial intelligence; theory of planned behavior; usage behavior; reliability; validity; nursing education

人工智能的快速发展,促进了基于大语言模型

(Large Language Model)的聊天机器人的出现和更新迭代<sup>[1]</sup>,大语言模型是指包含数十到万亿甚至更多参数的Transform语言模型,该模型主要使用深度学习技术,经过大规模的数据训练,能够执行自然语言处理和人机交互等复杂的任务<sup>[2]</sup>。相关研究表明,大语言模型具有为护理专业学生(下称护生)提供个性化的学习内容、构建临床护理场景、检索护理学相关知识等优势,激发护生的使用意图<sup>[3-4]</sup>。但技术自身的局限(信息的滞后性、误导性、偏倚性等)和护生的使用风险行为(过度依赖等)引发了国内外各医学院

作者单位:1. 重庆医科大学附属第一医院护理部(重庆,400016);2. 简阳市人民医院护理部

通信作者:肖明朝, x mz. 2004@163. com

马应卓:男,博士在读,护士, yingzhuo1999@126. com

科研项目:重庆医科大学2024年校级教育教学改革研究重大项目(JY20240202);重庆医科大学2025年校级研究生人工智能赋能课程建设项目(AI2025009);重医附一院护理科研重点项目(HLPY2025-01)

收稿:2025-12-09;修回:2026-02-26

校和护理教育者的担忧<sup>[5]</sup>。护生使用人工智能工具的意图和行为对其学业成绩和未来执业能力有着显著影响。目前我国相关评估工具较少,尚未有学者对护生的使用行为开展调查。因此,本研究基于计划行为理论,运用德尔非法编制护生大语言模型使用行为问卷并进行信度和效度检验,旨在为护理教育者提供评估工具,评估护生对大语言模型的心理和行为特征。

## 1 对象与方法

### 1.1 对象

**1.1.1 函询专家** 纳入标准:①副高及以上职称或研究生及以上学历;②从事护理教育、医学教育、信息化护理、人工智能和医学交叉领域工作 8 年及以上;③熟悉问卷编制方法,具有严谨科学的态度;④知情同意且自愿参与本研究;⑤了解并使用过大语言模型。本研究共邀请 17 名专家,完成 2 轮咨询 15 名,来自北京市、上海市、广东省、重庆市、四川省、陕西省、河南省、贵州省。男 6 名,女 9 名;年龄 32~61 岁,中位数 42 岁;学历为本科 5 名,硕士 8 名,博士 2 名;职称为中级 4 名,副高 6 名,正高 5 名;研究领域为护理教育 6 名,信息化护理 3 名,医学人工智能 2 名,医学教育 4 名。工作年限 10~40(21.67±9.40)年。

**1.1.2 质性访谈对象** 2024 年 6 月,采用目的抽样法和滚雪球抽样法招募来自中国 13 所医科大学的 24 名护生进行线上半结构式访谈。访谈至 22 人时信息达到饱和,继续访谈 2 人以确认饱和度<sup>[6]</sup>。纳入标准:自愿参与本研究;有大语言模型使用经历;知情同意参与研究。排除标准:毕业或辍学的学生;使用的是未基于大语言模型开发的聊天机器人(如苹果手机的 Siri);有任何心理健康问题或参与其他可能干扰结果的研究。女 16 名,男 8 名;博士 5 名,硕士 11 名,本科 8 名。

**1.1.3 正式调查对象** 2024 年 9—11 月,采用便利抽样法,对重庆市、陕西省、四川省 3 所医学院校的护生开展 2 轮线上和线下问卷调查,纳入排除标准与半结构式访谈相同。依据探索性因子分析样本量至少为条目数的 5~10 倍;验证性因子分析的样本量应>200<sup>[7]</sup>,问卷包括态度、主观规范、知觉行为控制、意向和行为 5 个维度共 34 个条目,考虑 20% 的样本丢失率,总样本量至少为 463。本研究共调查 950 人,其中有效样本 634 人。本研究已获得重庆医科大学附属第一医院伦理审查委员会批准[2023 年科研伦理(2023-334 号)]。

### 1.2 方法

**1.2.1 问卷理论框架** 计划行为理论(Theory of Planned Behavior, TPB)<sup>[8]</sup>认为,意图决定行为,意图由 3 个潜在成分构成:①态度,即个体对执行该行为

的积极或消极情感;②主观规范,即来自其他个人或群体在采用特定行为时的社会压力;③知觉行为控制,即个体对执行某一行为的难易程度的感知,它反映了经验和预期的障碍<sup>[9]</sup>。在本研究中,计划行为理论提供了系统的分析框架。态度指的是护生对大语言模型使用的情绪,包括感知有用等;主观规范为护生感知到的重要他人(朋友、同学等)对自己使用大语言模型的影响;知觉行为控制定义为护生在使用大语言模型中感知到的难易程度。意图在该研究中被定义为护生在各种场景使用大语言模型的意愿程度。行为是护生使用大语言模型具体场景的频率。

### 1.2.2 构建条目池

**1.2.2.1 文献回顾** 以 natural language processing, large language model\*, generative artificial intelligence 等为英文检索词,检索 PubMed、Web of Science 等英文数据库。以大语言模型、生成式人工智能、自然语言处理等为中文检索词,检索中国知网、万方知识服务平台等数据库。首次检索为建库至 2023 年 8 月,2024 年 5 月进行补检。文献纳入标准:①内容涉及大语言模型在护理学领域应用;②可获取全文;③中英文文献。排除标准:①未明确指出人工智能产品是否基于大语言模型;②重复发表、信息及数据不完整的文献;③综述、社评、评论、简讯、观点等类型文章。分析纳入文献内容,依次阅读文题、摘要、全文后纳入 33 篇。研究小组对文献内容进行逐一讨论提炼,并依据计划行为理论框架,最终形成问卷条目池,共 38 个条目。

**1.2.2.2 质性访谈** 访谈提纲:①请分享你对大语言模型的理解和认知;②你认为使用大语言模型有什么优势和缺点;③你认为哪些人对你使用大语言模型造成影响;④你认为哪些因素会促进或阻碍你使用大语言;⑤你目前的主要用途有哪些。采用定向内容分析法对访谈结果进行整理、分析、编码,共提炼出 10 个一级主题 28 个二级主题。根据访谈结果对初始问卷条目进行顺序调整和增加(新增 9 个条目),形成的问卷包含 47 个条目。

**1.2.3 专家函询** 本研究通过线上的方式,于 2024 年 6—8 月通过电子邮件或微信方式发送电子咨询表开展专家函询。采用 Likert 5 级评分法对问卷各维度及条目的重要性进行评分,从“很不重要”到“很重要”依次赋 1~5 分。内容效度评价采用 4 级评分法,评价各条目与相应维度的关联性,从“很不相关”到“非常相关”依次赋 1~4 分。本研究以重要性赋值均数≥3.50 分、变异系数≤0.25 作为指标筛选标准<sup>[10]</sup>,并参考专家意见,对维度或条目进行补充、删除或修改。研究小组根据函询结果及专家意见,经讨论后修改、增减指标,形成第 2 轮函询问卷。经历 2 轮函询后专家意见趋于一致,停止函询。第 1 轮邀请

17名专家参与函询,收回15份,第2轮发放15份,收回15份。2轮咨询的专家的回收率分别为88.24%、100%,判断依据系数为0.91、0.93,熟悉程度系数为0.83、0.88,专家权威系数分别为0.87、0.91,肯德尔和谐系数分别为0.203、0.402。根据2轮专家函询意见,经课题组讨论后,①合并6个条目为3个条目;②删除12个条目;③修改29处,补充2个条目,剩余5个维度、34个条目。

**1.2.4 预调查** 2024年8—9月,采用便利抽样法选取重庆医科大学30名符合条件的护生进行预调查。课题组成员向学生说明本次研究的目的,取得同意后,现场发放问卷,填写完成后,询问学生对各条目是否理解、条目语言表达是否清楚、是否存在歧义,并请提出修改意见。预调查回收30份问卷,有效回收率为100%,完成时间为5~8 min。3名学生表示问卷个别条目表达歧义。课题组成员根据结果对条目语言表述方式进行修改。

**1.2.5 正式调查** 采用便利抽样法,于2024年9—10月开展第1阶段问卷调查,结束后对已推送问卷的班级进行记录,第2阶段(2024年10—11月)调查排除重复班级。问卷调查分为线上和线下同步进行:线上调查主要针对陕西省、四川省2所大学的护生,课题组1名成员使用问卷星软件制作电子问卷,另1名成员负责联系学院的辅导员或班主任进行问卷的审核和推送(问卷主要通过班级微信群、qq群的方式发布),电子问卷以匿名填写的方式以保障学生的隐私;线下调查主要针对重庆市的护生,由4名接受过培训的课题组成员联系重庆医科大学护理学院辅导员或班主任,选取班会,研讨会等学生集中的时段和场合,通过现场发放的方式收集问卷。调查工具:①一般资料调查表,包括性别、年龄、婚姻状况、职业、学历等;②护理专业学生大语言模型使用行为问卷,问卷共34个条目,采用Likert 5级计分,所有条目为正向计分,各维度所属条目的平均分作为该维度的最终得分,分值越高表明被试在该维度上的水平或倾向越高。第1阶段回收500份问卷。剔除未使用过大语言模型的144名护生,以及规律作答、填写时间过短等52份无效问卷,剩余304份问卷,有效率85.39%。304名护生中男43名,女261名;年龄17~37[21(19,22)]岁;硕士及以上98名,本科206名。第2阶段回收450份问卷,剔除未使用过大语言模型的89名护生以及规律作答、填写时间过短等31份无效问卷,剩余330份问卷,有效率91.41%。调查对象中男48名,女282名;年龄17~34[21(19,24)]岁。硕士及以上125名,本科205名。

**1.2.6 质量控制** 问卷均当场填写并回收,回收时检查填写是否完整。在数据收集结束后,对问卷进行筛选,排除不使用大语言模型或者误认为自己使用的聊天机器人是基于大语言模型的问卷以及其他无效

问卷(包括填写一半或规律作答等)。问卷星设置了1个ID号仅填写1次。

**1.3 统计学方法** 使用SPSS27.0及AMOS24.0软件进行统计分析,计数资料采用频率和构成比进行描述,计量资料采用均值和标准差进行描述。①项目分析:采用临界比值法、相关系数法、内部一致性系数法筛选条目<sup>[7]</sup>。②效度分析:第1阶段收集的304份有效样本数据用于探索性因子分析,第2阶段收集的330份有效样本数据用于验证性因子分析。采用主成分分析法(PCA)提取公因子,并采用正交旋转法进行转轴,遵循公因子方差解释率在>50%,特征值>1的原则,保留在其所属公因子上的载荷系数>0.400,在其他公因子上的载荷系数<0.400的条目<sup>[11]</sup>。若条目存在多重载荷且载荷系数相近(载荷系数均>0.400,且差值<0.200),则予以删除。验证性因子分析:各拟合指标满足 $\chi^2/df < 3$ ,RMSEA<0.050、GFI、CFI、IFI、AGFI>0.900或接近0.900<sup>[12-13]</sup>。聚合效度:由平均方差提取值(AVE)和组合信度(CR)决定。当各维度AVE值均>0.500,且所有CR值>0.700,各因子载荷大于0.500,说明聚合效度较好<sup>[14]</sup>。区分度效度:当公因子AVE平方根均超过该公因子与其他任一公因子的相关系数,说明区分度良好<sup>[14]</sup>。内容效度根据函询专家对问卷各条目内容关联性的评价结果,计算条目水平的内容效度指数(ICVI)及问卷平均水平的内容效度指数(S-CVI/Ave)。问卷I-CVI≥0.780,S-CVI/Ave≥0.900,提示内容效度较好<sup>[15]</sup>。③信度分析:采用Cronbach's  $\alpha$ 系数、重测信度(在第2轮问卷收集结束后选取30名护生进行标记,间隔2周再次填写,计算两次测量结果的相关系数以判断问卷重测信度)、Spearman-Brown折半信度分析问卷内部一致性信度。0.700作为可接受的标准,若Cronbach's  $\alpha$ 系数>0.800说明内部一致性信度较好<sup>[16]</sup>。

## 2 结果

**2.1 项目分析结果** ①临界比值法结果显示:各条目决断值4.141~14.919,且高、低分组差异有统计学意义(均 $P < 0.001$ )。②Pearson相关分析结果显示:条目A9(0.316)、B6(0.279)相关系数值均<0.400,考虑删除。其他条目与总分相关系数介于0.431~0.708。③内部一致性系数法结果提示:删除任一条目后问卷总的Cronbach's  $\alpha$ 系数均未较前明显上升,故未删减任何条目。

## 2.2 效度

### 2.2.1 结构效度

**2.2.1.1 探索性因子分析** 因子分析显示,KMO值为0.913,Bartlett's球形检验 $\chi^2 = 6134.622(P < 0.001)$ ,表明适合行因子分析,采用主成分分析法,经最大方差正交旋转,为了最大程度保留数据的原始变异信

息,根据碎石图的拐点和特征值大于 1 的组合来确定问卷公因子的个数,共提取出 6 个公因子,累计方差贡献率为 65.758%。条目 A7(因子 6 载荷=0.734)和 A8(因子 6 载荷=0.661)虽在单一因子上载荷较高,但因子 6 整体仅包含此 2 条,与其他因子间缺乏理论关联性;条目 A10、I6 在所有因子上的载荷均低于 0.500,未达到因子载荷阈值。因此,删除条目 A7、

A8、A10、I6 后,进行第 2 次探索性因子分析,共提取 5 个公因子,累计方差贡献率为 66.584%。28 个条目的因子载荷均>0.500,条目 7 出现交叉载荷现象,但鉴于其主、次载荷差值大于 0.200,且同伴支持是衡量学生主观规范不可或缺的指标,经小组专家讨论本研究予以保留,见表 1。2 轮探索性因子分析最终形成具有 5 个维度、28 个条目的正式问卷。

表 1 探索性因子分析结果(n=304)

| 条目  | 知觉行为控制       | 态度           | 使用行为         | 使用意图         | 主观规范         |
|---|--------------|--------------|--------------|--------------|--------------|
| 1. 我认为大语言模型可以提高护理专业学生的学习效率                    | 0.152        | <b>0.783</b> | 0.070        | 0.317        | 0.171        |
| 2. 我认为大语言模型可以增强护理专业学习的学习体验(学习互动性、临床场景模拟)      | 0.201        | <b>0.820</b> | -0.011       | 0.211        | 0.114        |
| 3. 我认为大语言模型能够为护理决策提供科学辅助支持                    | 0.214        | <b>0.678</b> | 0.170        | 0.043        | 0.241        |
| 4. 我认为大语言模型可以提高护理专业学生解决实际问题的能力                | 0.137        | <b>0.810</b> | 0.068        | 0.102        | 0.251        |
| 5. 我认为掌握大语言模型对护理专业学生的职业发展是重要的                 | 0.114        | <b>0.654</b> | 0.244        | 0.210        | 0.299        |
| 6. 我认为大语言模型能满足护理专业学习的需求                       | 0.220        | <b>0.679</b> | 0.051        | 0.182        | 0.302        |
| 7. 我认为同学(朋友)会推荐(支持)我使用大语言模型                   | 0.211        | 0.401        | 0.092        | 0.253        | <b>0.673</b> |
| 8. 我认为老师会推荐(支持)我使用大语言模型                       | 0.098        | 0.288        | 0.047        | 0.213        | <b>0.794</b> |
| 9. 我认为学校会推荐(支持)我使用大语言模型                       | 0.172        | 0.272        | 0.142        | 0.184        | <b>0.802</b> |
| 10. 我认为网络用户评论会推荐(支持)我使用大语言模型                  | 0.164        | 0.291        | 0.118        | 0.186        | <b>0.716</b> |
| 11. 我能准确描述大语言模型在护理学习中的主要用途                    | <b>0.646</b> | 0.363        | 0.115        | 0.255        | -0.015       |
| 12. 我能评估使用大语言模型时可能出现的用户风险(如泄露隐私)              | <b>0.688</b> | 0.027        | 0.200        | 0.077        | 0.065        |
| 13. 我能独立操作大语言模型完成护理相关任务(如护理病例分析、护理知识检索)       | <b>0.758</b> | 0.289        | 0.102        | 0.148        | 0.006        |
| 14. 我认为大语言模型易于使用                              | <b>0.724</b> | 0.185        | 0.023        | 0.277        | 0.044        |
| 15. 我能评估大语言模型输出的相关信息质量(如信息准确性、逻辑一致性)          | <b>0.685</b> | 0.039        | 0.118        | 0.243        | 0.174        |
| 16. 我有充足的资源和网络条件使用大语言模型                       | <b>0.722</b> | 0.130        | 0.113        | 0.131        | 0.304        |
| 17. 我有丰富的大语言模型使用经验                            | <b>0.665</b> | 0.105        | 0.312        | 0.000        | 0.248        |
| 18. 我有足够的动力学习如何使用大语言模型                        | <b>0.588</b> | 0.289        | 0.264        | 0.225        | 0.098        |
| 19. 我愿意使用大语言模型辅助撰写发言稿、简历设计等普通文本任务             | 0.246        | 0.189        | 0.089        | <b>0.785</b> | 0.220        |
| 20. 我愿意使用大语言模型进行趣味对话和社会事件讨论等日常社交互动场景          | 0.281        | 0.170        | 0.209        | <b>0.651</b> | 0.140        |
| 21. 我愿意使用大语言模型辅助护理专业知识学习、测试、答题等课程相关任务         | 0.193        | 0.289        | 0.205        | <b>0.723</b> | 0.109        |
| 22. 我愿意使用大语言模型辅助论文润色、数据分析、课题申报等护理科研任务         | 0.147        | 0.152        | 0.218        | <b>0.690</b> | 0.269        |
| 23. 我愿意使用大语言模型辅助案例分析、疾病诊断、生成患者术后康复指导手册等临床护理任务 | 0.277        | 0.239        | 0.277        | <b>0.619</b> | 0.196        |
| 24. 我使用过大语言模型辅助撰写发言稿、简历设计等普通文本任务              | 0.167        | 0.070        | <b>0.736</b> | 0.245        | 0.110        |
| 25. 我使用过大语言模型进行趣味对话和社会事件讨论等日常社交互动场景           | 0.145        | 0.071        | <b>0.781</b> | 0.001        | -0.003       |
| 26. 我使用过大语言模型辅助护理专业知识学习、测试、答题等课程相关任务          | 0.190        | 0.119        | <b>0.708</b> | 0.270        | 0.083        |
| 27. 我使用过大语言模型辅助论文润色、数据分析、课题申报等护理科研任务          | 0.063        | 0.068        | <b>0.758</b> | 0.274        | 0.170        |
| 28. 我使用过大语言模型辅助案例分析、生成患者术后康复指导手册等临床护理任务       | 0.232        | 0.089        | <b>0.788</b> | 0.058        | 0.044        |
| 特征根值  | 4.487        | 4.321        | 3.469        | 3.342        | 3.024        |
| 累计方差贡献率(%)                                    | 16.025       | 31.458       | 43.849       | 55.784       | 66.584       |

注:条目已重新排序。

2.2.1.2 验证性因子分析 构建一阶 5 因子模型,  $\chi^2/df=1.647$ , RMSEA=0.044, CFI=0.951, IFI=0.951, TLI=0.945, GFI=0.891, NFI=0.885, 表明问卷模型拟合可接受。各条目标化路径系数>

0.500(均  $P<0.05$ )。模型图见附件 1。

2.2.1.3 内容效度 本问卷 I-CVI 为 0.800~1.000, S-CVI/Ave 为 0.914, 问卷内容效度较好。

2.2.1.4 聚合效度及区分度效度 见表 2。

表 2 聚合效度及区分度效度结果

| 项目     | 相关系数  |       |        |       |       | AVE   | AVE 平方根 | 组合信度  |
|--------|-------|-------|--------|-------|-------|-------|---------|-------|
|        | 态度    | 主观规范  | 知觉行为控制 | 意图    | 行为    |       |         |       |
| 态度     | 1.000 |       |        |       |       | 0.532 | 0.729   | 0.871 |
| 主观规范   | 0.382 | 1.000 |        |       |       | 0.532 | 0.729   | 0.818 |
| 知觉行为控制 | 0.406 | 0.300 | 1.000  |       |       | 0.521 | 0.721   | 0.897 |
| 使用意图   | 0.367 | 0.405 | 0.479  | 1.000 |       | 0.550 | 0.741   | 0.840 |
| 使用行为   | 0.269 | 0.195 | 0.365  | 0.359 | 1.000 | 0.590 | 0.768   | 0.844 |

2.2.1.5 信度 见表3。

表3 信度结果

| 项目     | Cronbach's<br>α 系数 | Spearman-<br>Brown 折半信度 | 重测<br>信度 |
|--------|--------------------|-------------------------|----------|
| 态度维度   | 0.870              | 0.848                   | 0.781    |
| 主观规范   | 0.814              | 0.823                   | 0.770    |
| 知觉行为控制 | 0.898              | 0.905                   | 0.853    |
| 意图     | 0.848              | 0.864                   | 0.864    |
| 行为     | 0.874              | 0.849                   | 0.861    |
| 总问卷    | 0.917              | 0.775                   | 0.823    |

3 讨论

**3.1 问卷具有科学性** 本研究以计划行为理论为指导,紧密结合护生对大语言模型的使用行为特征,在清晰界定问卷各维度概念基础上,通过文献回顾和半结构访谈构建条目池。通过德尔菲专家函询对问卷维度及条目池进行删减、修改、增加、合并,选取15名专家覆盖护理教育、医学教育、信息化护理、医学与人工智能交叉领域,均具有高级职称或研究生学历,工作经验丰富且学术水平较高,具有一定的代表性;2轮函询专家回收率分别为88.24%、100%,权威系数分别为0.87、0.91,这确保问卷修订具有可靠性。问卷经过预测试、项目分析和信效度检验,整体过程较为严谨,符合问卷编制要求,具有较好的科学性。

**3.2 问卷具有较好的信效度** 内容效度分析结果显示,问卷内容效度较好,能恰当反映护生大语言模型使用行为的测评内容。此外,该研究共进行2次探索性因子分析,结果表明,所有条目因子载荷均>0.400,每个条目都在其中1个维度上表现出较高的载荷,总共提取5个因子,累计方差贡献率为66.584%。不仅维度符合预期的理论框架,而且条目归属维度情况符合理论假设。验证性因子分析各模型实际拟合值基本达到统计学要求,各因子的AVE为0.521~0.590,组合信度0.818~0.897,聚合效度较好;此外,各因子AVE平方根值均大于该因子与其他因子间的相关系数,区分效度良好。验证性因子分析结果显示,各条目能较好地度量对应的因子,各个条目标准化路径系数>0.500(均P<0.05),5因子模型拟合良好,说明问卷各因子与相对应的测量条目间的关系符合预期设计。信度分析显示,问卷总的Cronbach's α系数为0.917,各维度为0.814~0.898;问卷总的重测信度为0.823,各维度为0.770~0.864;问卷总的Spearman-Brown折半信度为0.775,各维度为0.823~0.905,表明具有较好的信度<sup>[17]</sup>。

**3.3 问卷具有较好的实用性** 现有研究多聚焦于大语言模型的功能使用态度(信任度、有用性、易用性等)维度评价<sup>[18]</sup>,单一模型的评价(如ChatGPT<sup>[19]</sup>),或普适性问卷开发,并非针对护理专业人员或学生设

计<sup>[20]</sup>。本研究编制的护理专业学生大语言模型使用行为问卷,从学生的角度,围绕其对大语言模型的使用态度、主观规范、知觉行为控制、意图与实际使用行为的层面综合测评其对这类人工智能工具的接受度、熟练度、实际使用场景和频率等。此外,本研究开发的问卷整合了护理专业场景且基于本土化护理教育环境设计(如专业适用性),避免了跨文化适应性偏差。问卷计分简单(Likert 5级计分),条目内容清晰,实施性较强,数量设置合理(5个维度28个条目),平均测评时间5~8 min。该问卷可为评估护生大语言模型使用行为提供标准化工具,既可用于现状调研以识别使用障碍(如技术掌握不足),也可为大语言模型素养培训提供依据。

4 结论

本问卷编制过程严谨科学,具有良好的信效度,条目设置合理,可作为护理教育、管理者评估护生使用大语言模型意图及行为的辅助测评工具,为构建大语言模型培训方案和管理制度提供参考依据。但本研究的调查人群仅限于四川、陕西、重庆,样本代表性受限,未来研究可扩大样本来源,以验证问卷的泛化能力和稳定性。

(特别说明:本文系第一作者学位论文整理而成。在退修过程中采纳专家建议对验证性因子分析模型进行修正与优化,文中数据与学位论文原文略有不同,以本刊发表版本为准)

附件1 验证性因子分析模型图  
请用微信扫码查看



附件1 验证性因子分析模型图

参考文献:

[1] 马应卓,王俊,刘彤,等.大语言模型在护理学领域应用的范围综述[J].护理学杂志,2024,39(19):124-129.  
 [2] Mitchell M, Krakauer D C. The debate over understanding in AI's large language models[J]. Proc Nat Acad Sci,2023,120(13):e2215907120.  
 [3] 李鹏,张源慧,唐龙,等.ChatGPT在护理问题和护理措施自主回答中的应用性研究[J].中华护理教育,2025,22(4):398-402.  
 [4] 汤志杰,孙国珍,李芸霞,等.大语言模型在护理领域应用的机遇与挑战[J].中国护理管理,2024,24(6):929-933.  
 [5] 马应卓,王俊,李雪连,等.护生对大语言模型聊天机器人使用体验的质性研究[J].护理学杂志,2024,39(16):69-72.  
 [6] Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales[J]. Qual Res Sport Exerc Health, 2021,13(2):201-216.  
 [7] 吴明隆.问卷统计分析实务:SPSS操作与应用[M].重庆:重庆大学出版社,2010:166-281.

[8] Ajzen I. The theory of planned behavior[J]. *Organ Behav Hum Decis Process*, 1991, 50(2): 179-211.

[9] Ajzen I. The theory of planned behavior: frequently asked questions[J]. *Hum Behav Emerg Technol*, 2020, 2(4): 314-324.

[10] Schofield R, Chircop A, Baker C, et al. Entry-to-practice public health nursing competencies: a Delphi method and knowledge translation strategy[J]. *Nurse Educ Today*, 2018, 65: 102-107.

[11] Mike L, Guimond S, Kelly S, et al. Social cognition in early course of schizophrenia: exploratory factor analysis[J]. *Psychiatry Res*, 2019, 272: 737-743.

[12] Miketinas D, Cater M, Bailey A, et al. Exploratory and confirmatory factor analysis of the Adolescent Motivation to Cook Questionnaire: a self-determination theory instrument[J]. *Appetite*, 2016, 105: 527-533.

[13] 郑文智, 吴文毅. 结构方程模型拟合评鉴: 整体拟合, 内部拟合与复核效度检验[J]. *心理学探新*, 2014, 34(1): 57-61.

[14] 温忠麟, 侯杰泰, 马什赫·伯特. 结构方程模型检验: 拟合

(本文编辑 赵梅珍)

(上接第 71 页)

[12] 李全磊, 颜美琼, 林岑, 等. 肿瘤患者选择留置 PICC 真实体验的质性研究[J]. *护理学杂志*, 2012, 27(5): 40-42.

[13] 欧美芳, 杨秋燕, 陈小梅, 等. 肿瘤化疗患者 PICC 置管意愿的质性研究[J]. *当代护士*, 2018, 25(3): 85-87.

[14] 秦海肖, 张素兰. 青少年肿瘤患者不同中心静脉通路使用体验的质性研究[J]. *肿瘤基础与临床*, 2024, 37(4): 468-471.

[15] 肖沙璐, 钱海宁, 张晓春, 等. 老年结直肠癌化疗患者 2 种中心静脉置管真实体验的质性研究[J]. *护士进修杂志*, 2022, 37(4): 378-383.

[16] 张庭婕, 朱晟, 赫洋. 淋巴瘤患者对选择中心静脉导管决策体验的质性研究[J]. *保健医学研究与实践*, 2024, 21(S1): 217-221.

[17] Janatolmakan M, Awys L, Khatony A. Cancer patients' experience with implanted venous ports: a qualitative descriptive study[J]. *J Vasc Nurs*, 2024, 42(1): 74-79.

[18] Parás-Bravo P, Paz-Zulueta M, Santibañez M, et al. Living with a peripherally inserted central catheter: the perspective of cancer outpatients: a qualitative study[J]. *Support Care Cancer*, 2018, 26(2): 441-449.

[19] 汪锦芳, 丁金霞. 恶性肿瘤病人选择留置静脉输液港真实体验的质性研究[J]. *全科护理*, 2015, 13(17): 1641-1643.

[20] Wang Y, Wang X, Qiu X. Indwelling experience and coping strategies of upper arm infusion ports in patients with cancer: a qualitative study[J]. *BMJ Open*, 2023, 13(3): 7.

[21] Ryan C, Hesselgreaves H, Wu O, et al. Patient acceptability of three different central venous access devices for the delivery of systemic anticancer therapy: a qualitative study[J]. *BMJ Open*, 2019, 9(7): e026077.

[22] 孔秋焕, 吕林华, 李燕儿, 等. 留置输液港肿瘤患者感受与需求的调查[J]. *中华现代护理杂志*, 2011, 17(6): 662-663.

[23] Depboylu B C, Yazman S. Do Videos sections of Internet search engines provide accurate and adequate information about totally implantable venous access ports? [J]. *J Vasc Access*, 2021, 22(2): 225-231.

[24] Boland L, Graham I D, Légaré F, et al. Barriers and facilitators of pediatric shared decision-making: a systematic review[J]. *Implement Sci*, 2019, 14(1): 7.

[25] Liverpool S, Pereira B, Hayes D, et al. A scoping review and assessment of essential elements of shared decision-making of parent-involved interventions in child and adolescent mental health[J]. *Eur Child Adolesc Psychiatry*, 2021, 30(9): 1319-1338.

[26] 韩金纳, 蔡瑞卿, 刘可. 肿瘤患儿父母共享决策影响因素的质性研究[J]. *护理学杂志*, 2023, 38(9): 52-55.

[27] 曲瑞杰, 程秋泓, 王海播, 等. 中年肺癌化疗患者参与跨专业共享决策方案的制订及实施[J]. *护理学杂志*, 2022, 37(23): 25-28.

[28] 周华璐, 何细飞, 郑云艳, 等. 化疗患者血管通路装置决策辅助工具的应用现状[J]. *护理学杂志*, 2024, 39(10): 126-129.

[29] 程芳, 施欢欢, 王薇. 计划行为理论在国内外护理研究中的应用[J]. *护理学杂志*, 2012, 27(8): 91-94.

[30] Ray-Barruel G, Woods C, Larsen E N, et al. Nurses' decision-making about intravenous administration set replacement: a qualitative study[J]. *J Clin Nurs*, 2019, 28(21-22): 3786-3795.

[31] 李蓉, 张国琴, 周丽娟, 等. 肿瘤患者输液港使用真实体验质性研究的 Meta 整合研究[J]. *实用临床医药杂志*, 2024, 28(16): 128-135.

[32] 柏兴华, 臧爽. 乳腺癌患者放疗期间携带经外周置入中心静脉导管真实体验的质性研究[J]. *中国全科医学*, 2014, 17(21): 2448-2451.

(本文编辑 赵梅珍)