

# 医学生信息素养标准化 测评工具研制(二): 测试题项

□ 许卫卫 / 北京大学深圳医院信息科 深圳 518036

杜建 张士靖 / 华中科技大学同济医学院医药卫生管理学院 武汉 430030

**摘要:** 医学生信息素养测评理论框架包括信息意识、信息知识、信息能力、信息伦理与道德4个测量维度。通过专家咨询法评价测试题项的内容效度(题项的准确性、清晰度和适用性),对整个测试体系在数量上进行了删减,并对部分题项的内容作了更科学的处理。通过综合指数法建立了各测试题项的权重系数。应用因子分析方法对测评体系进行结构效度检验,共提取4个公因子,能够与之前建立的测评理论框架中确定的4个维度相一致。方差分析结果表明,医学生与临床医学八年制学生、卫管专业本科生的信息素养水平的差异在统计学上有显著意义( $p < 0.05$ ),区分效度良好,测评体系整体信度较高。效应尺度统计量(ES)分析结果表明,《医学文献检索与利用》课程干预后,医学生的信息素养整体以及各维度的水平均有显著性改善。最终形成的标准化测评工具能够用于医学生信息素养水平的测量。

**关键词:** 医学生, 信息素养, 标准化测评工具

DOI: 10.3772/j.issn.1673—2286.2013.05.008

在医学生信息素养标准化测评工具研制(一)的基础上,针对医学生信息素养的不同维度和指标的特点设计测试题项,包括单项、多项选择题以及量表等形式,对医学生信息素养水平实施横向面上评价和纵向立体评估(前测与后测),并从心理测量学的角度对测评工具进行项目分析、信度、效度(内容效度、结构效度和区分效度)和反应度检验等。所设计并证实的标准化测评工具既可作为医学生和医生测量其信息素养水平的自评工具,也可作为对医学信息素养教育机构的运作状况、效果、效率和效用进行科学评价的依据;同时也期望能为相关教育管理部门制定国家医学信息素养评价标准与干预策略提供决策依据。

## 1 医学生信息素养标准化测试题项的拟定

### 1.1 测试题项初步建立

将涉及信息意识维度的测试题设计为5级Likert量表(Likert Scale, LS),涉及信息知识、信息能力、信息伦理与道德的题项多采用单项选择题(Single Choice, SC)和多项选择题(Multiple Choices, MC)的形式。初拟的测试题项共45个,其中包括3个人口统计学特征题(性别、目前所读年级、专业)、13个Likert量表题、20个单项选择题和9个多项选择题。

### 1.2 测试题项的进一步发展

#### (1) 专家深度访谈

针对初拟的测试题的类型、表述形式、内容特征等方面进行专家深度访谈,咨询了专门研究医学生

信息素养问题的4位专家。经过专家深度访谈结果,进行以下几个方面的处理:

①删除自我评价类的问题。如“您认为您利用互联网获取信息的能力怎样?”不要出现在问卷中。这一点在RRSA-health的测评中也得到体现,即自我主观评价往往与实际客观测评数据不一致。

②因不同医学院校图书馆信息资源和信息服务的差异,去掉了3个涉及使用过的文献数据库和图书馆公共书目系统,以及之前通过哪种方式接受过有关信息素养的培训或学习的多项选择题。

③删除涉及信息需求类型和信息查询行为特征的题项。根据专家意见,初拟的题项从45个减至39个,其中主要对多项选择题进行了处理,从9个减为3个。

#### (2) 焦点小组

选择5名情报学(医学信息管理)专业的硕士研究生新生,他们本科的专业背景有临床医学、医学信息学、社会医学与卫生事业管理等。邀请这5名学生表达自己对改进的测试题项的看法、评价语言和概念表述是否清楚,并参与测试。这5名学生均表示测试题项及其备选答案表达的清晰度比较好,他们能够看懂并根据自己的知识和经验作出反应,测试时间约为8分钟。第一次修改后测量的维度和能力指标涉及上述“医学生信息素养目标体系”的各个方面。

## 2 医学生信息素养标准化测试题项的检验与修订

### 2.1 内容效度检验(Content Validity)

内容效度是指测评工具中的各个测试题项是否测定了其希望测量的内容,即测定对象对问题的理解和回答是否与条目设计者希望询问的内容一致。内容效度一般通过邀请专家评议打分给出意见,说明测评工具在多大程度上达到预期的研究目的。

#### (1) 专家咨询的基本情况

共遴选出长期在医学信息素养领域从事教学和研究工作的专家和学者16名,分别来自中国医学科学院医学信息研究所/图书馆、北京大学医学图书馆、复旦大学医学图书馆等机构。有效回收14位专家咨询问卷,咨询专家从事专业技术年限最低的为4年,最长的为30年,平均17.8年,有11位专家的年限在15年以上。

#### (2) 内容效度结果

采用准确性、清晰度和适用性

3个指标评价各测试题项的内容效度,其中专家对各题项在这3个指标上的判分范围(1~5分)表示其连续程度:①准确性:该测试题项以及备选答案在多大程度上准确描述了测试对象在某一方面具体的信息素养水平。1分为最不准确,5分为最准确。②清晰度:该测试题项以及备选答案表达的清晰度(清楚的程度)。1分为最不清晰,5分为最清晰。③适用性:该测试题项和备选答案对于不同高校(机构)医学生的适用程度。1分为适用性最差,5分为适用性最好。专家评议数据显示,只有SC2和SC4两个题项的准确性、清晰度和适用性分值均比较低,结合专家咨询过程中对之提出的意见和建议,对这两个题项作了修订。

通过上述内容效度分析结果和专家评议意见,医学生信息素养测试题项得到更进一步的发展。此时的测评工具共包含31个测试题项,其中Likert量表题13个,单项选择题17个,多项选择题3个。

### 2.2 项目分析(Item analysis)

主要进行各测试题项的难度(difficulty)、鉴别度(discrimination)以及干扰项(distractor)分析。其中最重要的是鉴别度,即每一个测试题项的“临界比率”(Critical Ratio, CR)。其计算方法是将所有被测试者的得分总和依高低排列,以总分最高的前27%和最低的后27%作为高低分组界限,进行高低分组被测试者在每个题项得分的平均数差异的显著性检验。如果题项的CR值达显著水准( $p < .05$ ),即表示这个题项能鉴别不同被测试者的反应程度,以此作为删除题项的重要依

据。因此,首先应建立各测试题项的标准化分值。

#### 2.2.1 测试样本的选取

采用整群抽样的方法,抽取本科毕业刚刚进入华中科技大学同济医学院研究生学习阶段,且选修了《医学文献检索与利用》课程(以下简称《课程》)的硕士生。分别在《课程》之前和之后进行前测和后测。其中前测共发放问卷220份,回收有效问卷201份,有效回收率为91.36%。选取的样本具有如下特征。

(1) 所采用的医学生信息素养测评理论框架中的《全国医学教育最低基本要求》和《本科医学教育国际标准》均针对医学本科毕业生,而选取的样本均是本科刚刚毕业的硕士研究生。

(2) 3/4以上的测评对象的专业是临床医学,且加上16.9%的基础医学,医学专业的测评对象的比例达92.0%,与测试对象“医学生”相一致。

(3) 测评对象中28.4%的学生生源为华中科技大学同济医学院的本科生,而超过70%的医学生来源于外校。经与华中科技大学研究生招生办公室知情人咨询,根据经验,考入华中科技大学同济医学院的硕士研究生绝大部分代表了同济医学院及其以下层次医学院校(全国范围内)的教育水平。因此,本次抽样选取的医学生具有一定的覆盖面,可在一定程度上覆盖全国范围内所有本科医学院校的教育水平。

#### 2.2.2 题项标准化分值的确定

采用综合指数法建立各测试

题项的权重(标准化分值)。综合指数法是把不同性质、不同类别、不同结构、不同计量单位的工作指标进行指数化并进行综合比较的方法。考虑到医学生的信息素养是一个多维度多指标的概念,反映了个体或群体作用于医学信息的综合态度和能力,在各维度、各指标的基础上发展形成的测试题项间的相对重要性是不相同的,所以通过赋予不同的权重来反映各测试题项间的相对重要性程度,以此来反映每个维度、每个指标对信息素养的贡献率,从而构建出医学生信息素养综合指数模型。医学生信息素养测试题项的权重主要由主观权重和客观权重组成。

#### (1) 主观权重

在对测试题项的内容效度进行专家咨询评议的过程中,同时邀请专家对每一个测试题项的重要性作出评判,即为主观权重,表征该测试题项所表征的能力指标在多大程度上代表了信息素养水平,即对信息素养的贡献度。1分为重要性最低,10分为重要性最高。

利用专家咨询法中得到的各题项重要性得分均数,对其进行归一化处理,计算各题项在整个测评体系中的权重系数。

#### (2) 客观权重

采用信息量权数法来确定指标的客观权重。应用每个题项观测值的变异系数(CV)的大小来衡量该题项的平均信息量。变异系数越大,该题项所蕴含的平均信息量也越大,其客观权重也应越大。标准差与平均数的比值称为变异系数。利用每个题项的测评值的变异系数,并对其进行归一化处理。

#### (3) 组合权重

对于综合评价来说,主观权重

反映了评价者的主观判断或直觉,但可能受到评价者的知识或经验缺乏的影响。而客观权重虽然通常利用比较完善的数学理论和方法,却忽视了评价者(或决策者)的主观信息。所以有必要将两种权重进行结合。

若 $p_i$ 为主观权重, $q_i$ 为客观权重,则指标的组合权重

$$w_i = (p_i + q_i) / \sum_{i=1}^n (p_i + q_i)$$

$w_i$ 为通过综合指数法建立的医学生信息素养测试题项的标准化分值。

### 2.2.3 Likert量表5点选项、多选题备选答案标准化分值的确定

在确定了33个测试题项的标准化分值后,对于17个单项选择题,被测评者回答正确则按照该题项的标准化分值赋分,回答错误则赋为0分。但对于5级Likert量表,由于1~5代表着由低到高、由弱到强的连续程度,被测评者选取“5”则赋分为该题项的标准化分值(满分),选取“1”、“2”、“3”、“4”则应在该题项满分分值的基础上,取其20%、40%、60%和80%的分值赋分。由于MC1、MC2和MC3为多项选择题,且备选项较多(5~6个),在专家咨询的过程中,同时邀请专家对多项选择题各备选答案之间的相对重要性程度进行了评价,评分级别为1~5分,1分表示最不重要,5分表示最重要。篇幅所限,结果从略。

### 2.2.4 项目分析结果

在项目分析中,主要对各题项的鉴别度、难度和干扰项进行了分析。其中鉴别度又称区分度,是测评题目分析的重要内容。难度是专门针对单项选择题的检验指标,是指该题项的正答率。按照201个被测试者的得分总和由高到低排列,以总分最高的前27%(54人,高于56.03分)和最低的后27%(54人,低于44.28分)作为高低分组界限,进行高低分组被测试者在每个题项得分的平均数差异的显著性检验,即两独立样本的t检验,如果题项的CR值达显著( $p < 0.05$ ),则此题具有鉴别度。

分析可见,SC2、SC3、SC17、MC1和MC3五个题项的鉴别度不高。同时,对17个单项选择题的难度和干扰性分析结果如表1。可见,SC2在所有单选题中的难度最低,达0.76,即3/4以上的医学生均能回答正确。相反,SC3的难度最高,200人中仅有4人回答正确,61.7%的医学生不知道PubMed、Biosis Previews、Elsevier ScienceDirect、EMBASE四大生物医学、药学领域常用文献数据库的类型(文摘库和全文库)。同时,SC17的难度也较低,仅次于SC2。删去SC2、SC3、SC17三个单选题。

另外,对于鉴别力较低的下列MC1和MC3两个多选题,同时根据专家意见,这些题项没有统一正确的解答,只是一种行为方式的调查,与其他题目不一致,会影响标准化测评,将其删除。

项目分析后,医学生信息素养测评工具共包含31个题项,其中3个人口统计学特征题,13个Likert量表题,14个单项选择题和1个多项选择题。删除5个题项后剩余题项按之前的顺序依次重新编号,其

表1 各检索平台基本情况

测试题项	正确答案	难度	干扰项分析					
			A	B	C	D	E	F
SC1	B	.62	3.0	62.1	2.6	25.4	—	—
SC2	D	.76***	2.5	0	15.4	75.6	5.0	—
SC3	C	.02***	30.8	3.0	2.0	2.0	61.7	—
SC4	B	.16	3.0	15.9	0.5	0.5	79.6	—
SC5	A	.36	35.8	7.5	7.0	6.0	41.2	—
SC6	B	.08	20.4	8.0	1.5	4.5	9.5	55.7
SC7	A	.58	57.7	12.4	7.5	21.4	—	—
SC8	A	.57	56.7	5.0	22.4	3.0	12.4	—
SC9	D	.17	20.4	3.0	5.0	16.9	53.7	—
SC10	B	.51	1.0	51.2	10.9	9.0	27.4	—
SC11	B	.06	7.5	5.5	5.5	6.5	73.6	—
SC12	A	.13	13.4	1.5	10.0	10.0	63.7	—
SC13	C	.72	1.0	18.9	72.1	5.5	1.0	—
SC14	A	.38	38.3	14.4	4.5	41.3	—	—
SC15	B	.29	6.0	28.9	46.3	15.9	—	—
SC16	A	.36	36.3	6.5	6.5	17.9	31.3	—
SC17	A	.73***	72.6	15.9	8.0	3.0	—	—

注：— 表示该题目没有设置该选项；\*\*\*表示难度过高或过低的题目；题项中备选答案的①、②、③、④、⑤、⑥分别用A、B、C、D、E、F表示。

中, 单项选择题从MC1至MC14, 原MC2改为MC1。

### 2.3 结构效度检验 (Construct Validity)

应用探索性因子分析方法对剩余的28个测试题项进行结构效度检验。首先计算这28个题项的标准化分值, 然后应用SPSS15.0进行因子分析的适当性检验, 结果KMO (Kaiser-Meyer-Olkin) 统计量为0.775, 大于0.5, 适合进行因子分析; 同时Bartlett's球形检验卡方值为1360.545 (自由度为378) 达显著 ( $p=.000<.01$ ), 表明测试题项间有公共因子存在, 适合进行因子分

析。利用主成分分析法进行因子分析, 选择正交最大变异法作为转轴方法, 依据Kaiser提出的选取特征值大于1作为确定因子数目的标准, 共有10个因子, 但Kaiser准则应用时, 因子分析的题项数最好不要超过30项 (该测评为28项), 题项的平均共同性最好在0.70以上 (但该测评题项的平均共同性为0.629)。因此, 还应结合陡坡图选取因子数目。陡坡图显示在第5个因子处形成碎石, 从此以后陡坡线比较平坦, 因而保留3~5个因子比较适宜。因此抽取前4个因子, 这4个因子可解释总变异的37.067%, 接近40%。表2为转轴后各因子所含题项在各因子中的载荷及共同性。

(1) 第一个因子, 所聚集的题项主要涉及从不同的数据库和数据源中检索、收集、组织和分析有关公共卫生和生物医学信息; 例如专题数据库、专业网站、核心期刊或权威学者; 利用图书馆获取本专业相关信息; 从信息获取到利用过程中, 对信息的处理; 怎样评价高质量学术研究成果等方面。包含了信息获取能力、信息评价能力、信息组织与管理能力、信息利用、信息创新能力等方面, 将其归为信息能力因子。

(2) 第二个因子, 题项主要涉及对信息和信息技术在学习、生活或工作中的重要作用的认识; 所学习的专业领域前沿信息很容易吸引

注意力; 有利用网络获取本专业领域相关信息的习惯; 对已有的证据(科学研究结论、知识)在医疗或公共卫生实践中的作用的认知、主动地共享自己拥有而别人没有的信息和知识的愿望; 在网上浏览信息或阅读期刊论文时, 通过题目发现对该信息感兴趣, 强烈的知情愿望会促使认真阅读全文来探究其实质内容等方面, 将其归纳为信息意识因子。

(3) 第三个因子, 题项主要涉及对布尔逻辑表达式(AND,OR,NOT)的掌握; PubMed/CBM加权检索的掌握; 文献数据库高级检索途径的掌握; PubMed主题检索途径的掌握; 对英文截词检索的掌握; 区别图书、图书章节、期刊论文、会议论文等不同文献载体; 对期刊影响因子概念的掌握等方面, 均是信息源或信息检索相关知识的测试, 将其归纳为信息知识因子。

(4) 最后一个因子, 只有两个题项, 包括对在撰写科学论文过程中是否应该详细标注出参考文献的识别、在转载或引用他人文章或研究成果时通常的做法, 将其归纳为信息道德因子。

通过上述分析, 所有的测试变量降维后, 涵盖的考察点与医学生信息素养测量的概念定义和操作定义相比仍然较为完全, 能够与之前建立的测评理论框架中确定的4个维度相一致, 说明该测评工具的结构效度良好。

### 2.4 区分效度检验 (Discrimination Validity)

区分效度是指测验或量表题项区别被试水平能力的度量。由于在医学院校中, 卫生事业管理专业

表2 转轴后各因子所含题项在各因子中的载荷及共同性

测试题项	因子				共同性
	1 (信息能力)	2 (信息意识)	3 (信息知识)	4 (信息道德)	
LS10	0.791				0.687
LS13	0.786				0.611
LS12	0.753				0.679
LS7	0.701				0.640
LS5	0.553				0.616
LS8	0.512				0.528
LS1	0.400				0.534
MC1	0.364				0.350
SC1	0.311				0.562
SC10	0.203				0.679
LS2		0.799			0.716
LS3		0.793			0.716
LS4		0.572			0.647
LS9		0.540			0.578
LS11		0.537			0.576
LS6		0.527			0.612
SC14		0.350			0.605
SC8		0.321			0.758
SC3			0.602		0.693
SC4			0.476		0.629
SC7			0.447		0.683
SC2			0.393		0.683
SC5			0.354		0.692
SC13			0.313		0.588
SC6			0.312		0.711
SC9			0.200		0.661
SC12				0.609	0.554
SC11				0.598	0.611
平均值					0.629

Extraction Method: Principal Component Analysis. (抽取方法: 主成分分析法)

Rotation Method: Varimax with Kaiser Normalization. (转轴法: Kaiser正规化最大变异法)

a. Rotation converged in 8 iterations.

表3 多重比较方差分析结果

	(I) 组	(J) 组	均差(I-J)	正确答案	P值	95%置信区间	
						下限	上限
最小显著差法 (LSD)	八年制	医学生	5.87915(*)	1.61589	0.000	2.6984	9.0599
		卫管生	1.47767	2.24354	0.511	-2.9386	5.8939
	医学生	八年制	-5.87915(*)	1.61589	0.000	-9.0599	-2.6984
		卫管生	-4.40148(*)	1.85742	0.018	-8.0577	-0.7452
	卫管生	八年制	-1.47767	2.24354	0.511	-5.8939	2.9386
		医学生	4.40148(*)	1.85742	0.018	0.7452	8.0577
Dunnett t 双侧检验法(2-sided)(a)	八年制	卫管生	1.47767	2.24354	0.699	-3.4345	6.3899
	医学生	卫管生	-4.40148(*)	1.85742	0.032	-8.4683	-0.3347

\*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

本科生可视为与医学本科生的差异性最大，同时，临床医学八年制学生目标是培养具有医学博士学位的高层次、高素质的临床和科研人才，与传统的5年制医学本科生在培养模式和能力素质上亦有较大差别。因此，按照整群抽样的方法抽取卫生事业管理专业大三年级本科生与临床医学八年制学生分别为35人和48人的样本量，采用单因素方差分析方法，比较其与医学生信息素养得分的差异是否具有统计学意义，从而判断测评工具是否具有区分效度（假设检验的结果有统计学意义就表明该测评工具有区分不同属性人群的能力）。

三组之间的差异显著 ( $p=0.000$ )，医学生与八年制学生、卫管专业本科生的信息素养水平的差异在统计学上有显著意义 ( $p=0.000$ ,  $p=0.018$ )。同时，从均差上还可以看出，医学生的信息素养水平要低于八年制学生，也低于卫管本

科生。根据上述方差分析结果，我们可以认定该测评工具具有良好的区分效度。

### 2.5 信度检验 (Reliability)

信度是对相同条件下重复测量结果的近似程度的一种度量，用于评价测量工具的稳定性和一致性。由于在多选题量表 (Multiple-item Scale) 中内在信度特别重要，

主要考察该测评体系的内在信度。采用计算Cronbach's  $\alpha$  系数和折半信度系数反映指标的内在信度。Cronbach's  $\alpha$  系数是反映测试题项的一致性程度和内部结构的良好性最常使用的一种信度估计方法，通常情况下取值在0~1之间。折半信度是将测试题项按奇、偶分半，求出其折半信度系数，由于各维度下的测试题项数目较少，故只计算整个测评工具的折半信度系数。

从表4、表5可见整个测评工

表4 Cronbach's  $\alpha$  信度系数

维度	包含题项数N	Cronbach's $\alpha$ 系数
信息意识	6	0.759
信息能力	9	0.563
信息知识	11	0.508
信息伦理与道德	2	0.305
Total	28	0.706

表5 折半信度系数

Spearman-Brown Coefficient (斯皮尔曼-布朗公式)	Equal Length (a,b)	0.780
Guttman Split-Half Coefficient (Guttman分半信度系数)	Unequal Length (a,b)	0.776

具的Cronbach's  $\alpha$  信度系数为0.706, 斯皮尔曼-布朗折半信度系数(Spearman-Brown Coefficient)为0.780, Guttman折半信度系数为0.776。其中, 信息意识、信息能力和信息知识维度的Cronbach's  $\alpha$ 系数分别为0.759、0.563和0.508, 但信息伦理与道德维度由于只有2个测试题项, 其信度系数只有0.305。按照Guilford, JP (1965) 提出的具体测量标准, 如果 $\alpha < 0.35$ , 则为低信度, 不宜采用; 如果 $0.35 \leq \alpha \leq 0.7$ , 则为中等信度, 该工具可以接受; 如果 $\alpha > 0.7$ , 则信度较高。因此信息意识维度的信度较高, 信息能力和信息知识维度的信度可以接受, 但信息伦理与道德维度的信度较低, 应通过增加题项、或对题项内容词句修饰等方式作为未来进一步研究的重点。从测评工具整体的信度系数来看, 可认为该测评工具的信度较高。

## 2.6 反应度检验 (Responsiveness)

反应度 (Responsiveness或Responsibility to change) 就是用来考查量表的灵敏性或反应性的指标, 即量表能够反映自测某一特征在时间上变化的能力和程度, 故亦称在时间上变化的灵敏度<sup>[1]</sup>。临床医学用的量表常用于评价不同治疗措施的治疗效果比较, 因此量表必须能反映出研究对象细微的疗效差别, 即具有一定的反应度<sup>[2]</sup>。同样, 医学生的信息素养是可干预的, 同一群体的信息素养水平在《医学文献检索与利用》课程干预前后的效果应该是有显著差异的。反应度指测评工具能测出不同对象在不同时间目标特征变化的能力。一般从以下两个方面考察测评工具的反应

度。第一, 使用测评工具分别在施加干预措施前后对同一群体进行测量, 用配对t检验比较干预前后得分的差异性是否显著。第二, 应用效应尺度统计量 (Effect Size, ES) 来评价反应度。

效应尺度 = (干预后得分 - 干预前得分) / 干预前得分的标准差

一般来说, 效应尺度应大于0.2, 0.2~0.5为较小效应, 0.5~0.8为中等效应, 0.8以上为较大效应。同样应用整群抽样的方法, 在《医学文献检索与利用》课程结束后, 我们应用这一测评工具对已参与前测的硕士研究生新生进行后测。回收有效问卷139份, 将其转化为标准化分值后, 4个维度以及整个测评工具的效应尺度统计量分析结果如表5。

从表6可见, 测评工具整体的效应尺度统计量为2.29, 说明该测评工具具有较强的反应度。从维度层面分析, 信息知识、信息伦理与道德维度包含题项的反应度最高, 信息意识次之, 且这3个维度的ES值均大于1; 信息能力维度的ES值为0.73, 也趋近于较大效应的临界

表6 效应尺度统计量

维度	包含题项数 (N)	前测平均值 (I)	后测平均值 (J)	前后测均差 (J-I)	前测标准差 (Std. Deviation)	效应尺度 (ES)
信息意识	6	13.2816	15.7409	2.4593	2.2112	1.11
信息能力	9	20.4175	24.8905	4.4730	6.1415	0.73
信息知识	11	9.9784	25.1036	15.1252	4.8021	3.15
信息伦理与道德	2	0.2343	4.6006	4.3663	1.0528	4.15
Total	28	43.976	67.1999	23.2239	10.1296	2.29

a. The items are: LS1, LS3, LS5, LS7, LS9, LS11, LS13, SC2, SC4, SC6, SC8, SC10, SC12, SC14.

b. The items are: LS2, LS4, LS6, LS8, LS10, LS12, SC1, SC3, SC5, SC7, SC9, SC11, SC13, MC1.

值(ES=0.8)。同时,将干预前和干预后的观察值作独立样本t检验,前测的平均值为43.9759,后测的平均值为67.1999,  $t=-18.819$ ,  $p=0.000<0.05$ ,差异有统计学意义,两个结果均说明该测评工具能够反映出干预前后医学生信息素养水平的改善。

## 2.7 医学生信息素养标准化测评工具(最终版)

医学生信息素养标准化测评工具(最终版)由标准化测试题项和标准化分值两部分组成。其中,测试题项由3部分组成,即个人基本信息,测评量表和单项、多项选

择题。其中,第一部分包括3个人口统计学特征题,第二部分由13个5级Likert量表题组成,第三部分包括14个单项选择题和1个多项选择题。具体如下。

### 第一部分 个人基本信息(请在表格中打勾“√”)

- 1 性别:           ① 男                           ② 女
- 2 目前所读年级:
- ① 本科一年级      ② 本科二年级      ③ 本科三年级      ④ 本科四年级
- ⑤ 本科五年级      ⑥ 硕士研究生      ⑦ 博士研究生
- 3 专业:           ① 临床医学      ② 基础医学      ③ 药学                ④ 公共卫生
- ⑤ 医学信息管理   ⑥ 卫生事业管理   ⑦ 法医学            ⑧ 护理学

### 第二部分 测评量表(请在表格中打勾“√”)

1. 完全不能够; 2. 基本不能够; 3. 不能确定; 4. 基本能够; 5. 完全能够。

序号	测评指标	1	2	3	4	5
LS1	我了解“信息素养”或“信息素质”这一术语					
LS2	我认为信息或信息技术在我学习、生活或工作中所起的作用非常重要					
LS3	我所学习的专业领域前沿信息很容易吸引我的注意力					
LS4	我有利用网络获取本专业领域相关信息的习惯					
LS5	我有利用图书馆(包括图书馆校园网)获取本专业领域相关信息的习惯					
LS6	我在上网浏览信息或阅读期刊论文时,通过题目发现对该信息感兴趣,强烈的情愿望会促使我认真阅读全文来探究其实质内容					
LS7	我了解与自己专业领域相关的主要信息源,例如专题数据库、专业网站、核心期刊或权威学者					
LS8	我了解循证医学(或循证护理、循证卫生保健、循证实践)这些术语的内涵					
LS9	我认为已有的证据(科学研究结论、知识)在医疗或公共卫生实践中的作用非常重要					
LS10	我能做到从不同的数据库和数据源中检索、收集、组织和分析有关公共卫生和生物医学信息					
LS11	我愿意主动地共享自己拥有而别人没有的信息和知识					
LS12	我能够对已获取的、感觉有价值的、且需要保存的信息科学地归类存储,建立自己的信息库并使之组织有序					
LS13	我能对检出的信息进行综合归纳,并能清楚地陈述自己的观点					



SC8 您的导师建议您认真阅读一下下面这篇文章,并给出了关于这篇文章的相关信息,即Cole BL. Caution: coloured medication and the colour blind. Lancet, 2009, 374(9691):720.

下列哪种方法可以使您在全文检索系统中最确切地查到这篇文章?

- ①  作者检索: Cole BL
- ②  期刊来源检索: 期刊名=Lancet; 年份=2009; 卷=374; 期=9691; 页码=720
- ③  题名检索: Caution:coloured medication and the colour blind
- ④  主题检索: coloured medication与the colour blind
- ⑤  不知道怎样查

SC9 科学引文索引 (Science Citation Index, SCI) 收录期刊的影响因子 (Impact Factor, IF) 是国际上通用的期刊质量评价指标,其计算方法是:

- ①  某刊某年发表的论文当年被引用次数与该刊年发表论文篇数之比
- ②  某刊前两年发表的论文在当年被引用次数与该刊前两年发表论文总篇数之比
- ③  某刊前三年发表的论文在当年被引用次数与该刊前三年发表论文总篇数之比
- ④  某刊某年发表全部论文中引用的参考文献,按发表年排序,取其最近的50%文献的发表年限
- ⑤  不知道影响因子到底是怎么回事

SC10 根据对文献信息内容加工的深度,可将文献分为零次文献、一次文献、二次文献和三次文献,以下与之对应的顺序正确的是:

- ①  实验记录 科技论文 索引 综述
- ②  索引 科技论文 综述 实验记录
- ③  实验记录 索引 科技论文 综述
- ④  实验记录 索引 综述 科技论文
- ⑤  不确定

SC11 您认为在撰写科学论文过程中是否应该详细标注出参考文献?

- ①  没必要
- ②  有必要,但可以灵活处理
- ③  非常有必要,这是对前人劳动成果的尊重
- ④  不清楚,没考虑过这个问题

SC12 您在转载或引用他人文章或研究成果时,通常的做法是:

- ①  总会在致谢中予以说明或作为参考文献列出
- ②  大部分情况下会在致谢中予以说明或作为参考文献列出
- ③  偶尔会在致谢中予以说明或作为参考文献列出
- ④  不做任何标识
- ⑤  没写过文章,不好确定

SC13 下面是来自Elsevier ScienceDirect的一条检索记录,请问其文献类型是:

Mental Health Promotion. International Encyclopedia of Public Health, 2008, Pages 406-418

- ①  一本书
- ②  书中的某个章节
- ③  一篇期刊论文
- ④  一篇会议报告或论文

SC14 如果了解某种疾病的诊疗技术的发展情况,要检索的最适宜的文献类型是\_\_\_\_\_。

- ①  综述(文献综述和系统综述)      ②  随机对照试验  
 ③  meta分析                                      ④  临床实践指南  
 ⑤  不知道

MC1 从信息获取到利用的过程中,您对新信息采取的处理方式是:(可多选)

- ①  会对新旧知识进行对比  
 ②  将之前的检索目标(信息需求)逐步优化  
 ③  能够较果断地决定是否接受或摒弃新信息  
 ④  会将新信息与个人经验结合,并试图在实际运用中尝试新知识  
 ⑤  不知道该如何处理

## 3 结论与建议

### 3.1 结论

医学生信息素养标准化测评工具的设计与实现,怎样体现“标准化”?笔者认为主要有两点:一是测试题项的发展是基于国内外权威的高等教育领域信息素养评价标准以及医学教育国际标准中与信息素养能力相关的指标要求,即测评工具的“基础”是“标准化”的;二是体现在科学的研究设计与系统的检验路径,即测评工具的“过程”是“标准化”的。本研究在国内医学信息素养标准化评价领域迈出了实质性一步,首次通过系统性的文献研究方法提出了医学生信息素养的概念定义和操作定义,为界定测量维度和发展测量指标提供了重要支撑。同时,突破了国内现有的素质测评工具大多为量表形式的局限,针对“医学生信息素养”这一概念的特殊性,同时设计了单项和多项选择题的形式,以期系统、全面、有

效地测量医学生信息素养的各个维度和具体目标。最后,从心理测量学的角度对整个测评体系进行了系统的检验,最终建立了基于标准的、有较高信度、效度和反应度的测评体系。

### 3.2 不足以及继续研究的建议

由于人力、时间以及经费等因素的限制,仍存在不足之处,这些不足之处也正是今后进一步深入研究的方向。

#### 3.2.1 测试题项的全面性和特异性需要进一步加强

信息能力维度方面还应增加信息的传播、交流以及信息有效利用方面的题项,以及增加有关信息管理、分析能力的题项,比如文献管理软件、某些数据库(SCI-E、SinoMed)的分析功能等。信息伦理与道德的题项主要集中在了参考

文献和抄袭方面,还应增加有关数字资源合理使用,比如私设代理服务器、恶意下载等方面的内容。

同时,题项所表征的医学生的信息素养除通用层次大学生的基本信息素养指标外,还应更注重医学的学科特殊性和医学的职业特殊性,医学生在未来的医疗实践中如何有效地获取、分析和利用“非物化信息”的能力(如患者的肢体、表情、语言等信息)以及向患者和公众表达、交流和传播健康信息的能力应该作为医学生信息素养测评体系的重要内容。

#### 3.2.2 专家咨询的学科覆盖面需要进一步拓展

本次进行的专家咨询,其学科背景多是医学生信息素养研究、医学信息检索、信息组织与分析、医学图书馆学等领域。这些专家均是医学信息素养研究的主体,评议结果难免带有领域的主观性。而医学信息素养是一个临床医学、医学信息

学、图书馆与情报学和教育学等多学科交叉的研究领域。专家的学科背景还应包括医学教育、临床医学等,特别是临床医生在实际的医学科研与医疗实践中对医学信息的获取、评价和利用方面的体会更加深刻,将其列入专家评议小组能够有

效地增强医学信息素养研究的客观性和可操作性。

### 3.2.3 实际测评的样本量及其覆盖面需要进一步增加

样本选取自刚刚本科毕业进入

硕士研究生阶段的“医学本科毕业生”。但直接进入工作岗位的医学本科毕业生未在其中,这一部分群体未能检验该测评体系的各项指标。因此,有待在今后的研究中扩大样本的选取范围及其覆盖面。

#### 参考文献

- [1] 许军,郭榕,刘永生,等.自测健康评定量表修订版的反应度研究[J].中国卫生统计,2003,20(5):272-275.  
[2] 安胜利.统计学系列讲座——第7讲:量表的评价[J].护理学报,2006,13(8):94-95

#### 作者简介

张士靖(1958-),女,通讯作者。研究方向:情报语言学、医学信息检索、医学信息素质教育。E-mail: zhangsj9999@163.com

#### The Development of a Standard Measuring Tool for Medical Students' Information Literacy (Two): Test Items

Zhang Shijing, Du Jian, Tang Yuanyuan / Tongji Medical College of Huazhong University of Science & Technology, Wuhan, 430030

Abstract: The measuring theoretical framework of Medical students' information literacy consists of four dimensions; they are information awareness, information knowledge, information ability, and information ethics morals. After evaluating the test items validity (its accuracy, clarity and applicability) through expert consultation method, researchers cut the amount of the test system and scientific processing for particular items is done. They also build different weighting coefficients for each test item with comprehensive index method. Factor analysis method is being used to test the construct validity and four common factors are put forward, which are in conformity with the four dimensions of the measuring theoretical framework. The variance analysis shows that the differences in the level of information literacy amongst medical students, eight-year clinical medical students and health management undergraduates are statistically significant ( $P < 0.05$ ), and have a good discriminant validity. The overall reliability of the evaluation system is relatively high and the effect scales show that after the intervention of taking Medical Literature Retrieval and Utilization, the information literacy of medical students achieve an overall improvements greatly as well as each dimension level. The final standard measuring tool turns out to be possibly used to measure the information literate level of medical students.

Keywords: Medical students, Information literacy, Standard measuring tool

(收稿日期: 2012-09-25)