

表 1 更新版良好测量学属性的标准

测量属性	等级	标准
结构效度	+	CTT
		CFA: CFI 或 TLI 或类似的指标 > 0.95 或 RMSEA < 0.06 或 SRMR < 0.08 ^a
		IRT/Rasch
		未违背 <u>单维性</u> ^b : CFI 或 TLI 或类似的指标 > 0.95 或 RMSEA < 0.06 或 SRMR < 0.08 ^a
		并且
		未违背 <u>局部独立性</u> : 控制主要因素后各条目之间的残差相关性 < 0.20 或 Q3's < 0.37
		并且
		未违背 <u>单调性</u> : 有足够的图表或条目适宜性 > 0.30
		并且
		足够的 <u>模型拟合</u>
		IRT: $\chi^2 > 0.001$
		Rasch: 加权和未加权均方 ≥ 0.5 和 ≤ 1.5 或 Z 标准化值 > -2 和 < 2
?	CTT: 并非报告了所有有关“+”的信息	
	IRT/Rasch: 未报告模型拟合	
-	不符合“+”的标准	
内部一致性	+	至少有低级证据 ^c 证明足够的结构效度 ^d 和每个单维度量表或子量表的 Cronbach's α 值 ≥ 0.70 ^e
	?	“至少有低级证据证明足够的结构效度”的标准不符合
	-	至少有低级证据 ^c 证明足够的结构效度 ^d 和每个单维度量表或子量表的 Cronbach's α 值 < 0.70 ^e
信度	+	ICC 或加权 Kappa 值 ≥ 0.70
	?	ICC 或加权 Kappa 值未报到
	-	ICC 或加权 Kappa 值 < 0.70
测量误差	+	SDC 或 LoA < MIC ^d
	?	MIC 不明确
	-	SDC 或 LoA > MIC ^d
结构效度的假设检验	+	结果与假设相符 ^f
	?	审查小组未定义假设

	-	结果与假设不相符 ^f
跨文化效度\测量等同性	+	在多组因素分析中，分组因素（如年龄、性别、语言）之间没有发现重要差异，在分组因素方面也没有的重要的 DIF（McFadden's $R^2 < 0.02$ ）。
	?	未进行多组因素分析或 DIF 分析
	-	发现了分组因素或 DIF 的重要差异
效标效度	+	金标准的相关系数 ≥ 0.70 或 $AUC \geq 0.70$
	?	并非报告了所有有关“+”的信息
	-	金标准的相关系数 < 0.70 或 $AUC < 0.70$
反应度	+	结果符合假设 ^f 或 $AUC \geq 0.70$
	?	审查小组未定义假设
	-	结果与假设相符 ^f 或 $AUC < 0.70$

该标准基于 Terwee 等人和 Prinsen 等人的研究

AUC: 曲线下的面积; CFA: 验证性因子分析, CFI: 比较拟合指数, CTT: 经典测试理论, DIF: 项目功能差异; ICC: 组内相关系数; IRT: 项目反应理论; LoA: 一致性界限; MIC: 最小重要变化; RMSEA: 近似误差均方根; SEM: 测量标准误; SDC: 最小可测变化值; SRMR: 标准化残差均方根; TLI: Tucker-Lewis 指数

“+”=足够; “-”=不足; “?”=不确定

a 不同研究工具的因子结构应相同，以便评估其结构效度的相应指标分数

b 单维性是指每个子量表的因子分析，而结构效度是指（多维）患者报告的结果测量的因子分析。

c 根据 GRADE 的方法，对证据进行等级划分

d 这些证据可能来自不同的研究

e 删除了 Cronbach's α 值 < 0.95 的标准，因为这与 PROM 的开发阶段有关，而与评估现有 PROM 量表无关

f 所有研究的结果应综合考虑，然后决定 75% 的结果是否符合假设