

Supplement 3: Medical Education Research Study Quality Instrument (MERSQI) Components

Domain: Item	Response Options: Scores	Operational Definition
Study Design	<ul style="list-style-type: none"> • Single-group cross-sectional or single-group post-test only: 1 • Single-group pretest and post-test: 1.5 • Nonrandomized, 2 group: 2 • Randomized controlled trial: 3 	<ul style="list-style-type: none"> • Survey studies are cross-sectional. • Case-control and cohort studies (2 or more defined cohorts) are considered 2-group nonrandomized.
Sampling: institutions	<ul style="list-style-type: none"> • 1 institution: 0.5 • 2 institutions: 1 • 3 or more institutions: 1.5 	<ul style="list-style-type: none"> • Number of institutions refers to origin of study participants (not study authors).
Sampling: response rate	<ul style="list-style-type: none"> • Not applicable • < 50% or not reported: 0.5 • 50%–74%: 1 • ≥ 75%: 1.5 	<ul style="list-style-type: none"> • Response rate is the proportion of those eligible who completed the posttest or survey. For intervention studies, this is the proportion of those enrolled who completed the intervention evaluation. • Use "not applicable" only if a response rate truly does not apply (e.g., data obtained from a medical record or professional organization database).
Type of data	<ul style="list-style-type: none"> • Assessment by study participant: 1 • Objective: 3 	<ul style="list-style-type: none"> • Observer ratings are considered objective.
Validity evidence for evaluation instrument scores	<ul style="list-style-type: none"> • Not applicable • Content: 1 • Internal structure: 1 • Relationships to other variables: 1 	<ul style="list-style-type: none"> • Relevant content evidence would include using theory, guidelines, experts, and existing instruments to identify or refine the instrument. • Relevant internal structure evidence would include all reliability (internal consistency, interrater, interstation, and test-retest) and factor analysis. • Relevant evidence of relationships to other variables would include expert-novice comparisons and concurrent or predictive correlation with other variables. • Use “not applicable” only if the study does not measure a psychological construct <i>and</i> there is no instrument to rate (e.g., gender as the sole outcome); should be used very rarely.
Data analysis: sophistication	<ul style="list-style-type: none"> • Descriptive analysis only: 1 • Beyond descriptive analysis: 2 	<ul style="list-style-type: none"> • Descriptive analyses include frequency, mean, and median. • Any test of statistical inference is considered “beyond descriptive.”

Data analysis: appropriate	<ul style="list-style-type: none"> • Data analysis appropriate for study design and type of data: 1 	<ul style="list-style-type: none"> • Considered “no” if there is a statistical error or if authors failed to analyze data at all.
Outcome	<ul style="list-style-type: none"> • Satisfaction, attitudes, perceptions, opinions, general facts: 1 • Knowledge, skills: 1.5 • Behaviors: 2 • Patient/health care outcome: 3 	<ul style="list-style-type: none"> • General facts include participant demographics. • Knowledge/skills are in a test setting (paper, computer, simulation, or patients in a nonauthentic setting). • Behaviors are physician actions with real patients in a clinical context, or other activities in a real context. • Patient/health care outcomes are actual effects on real patients, programs, or society.