Supplementary material

Psychometric evaluation of the survey instrument

Manuscript: Open cholecystectomy training, practice, and confidence in the laparoscopic era: A cross-sectional survey of 391 surgeons

1. Rationale

The 24-item questionnaire was developed for this study and had not previously been validated. Its internal structure was therefore examined in an exploratory psychometric evaluation conducted by an independent statistical consultancy. The evaluation was not used to derive, weight or adjust any of the descriptive results reported in the manuscript; it was performed to determine whether the items could legitimately be combined into a measure of proficiency.

2. Item selection

Of the 24 items, three (current role, residency year and fellowship year) were treated as sample characterization variables. Four further items were excluded before modelling: robotic cholecystectomy volume, which showed almost no variance (94.4% of responses in the lowest category); years of practice, used for characterization; the checkbox item on training formats, whose multiple-response format is unsuitable for factor modelling; and the conditional item on the modality of common bile duct exploration training. Seventeen items were retained.

3. Analytical approach

Because response formats were heterogeneous, a mixed correlation matrix was computed with the mixedCor function of the psych package, applying polychoric coefficients between ordinal variables, tetrachoric coefficients between dichotomous variables and polyserial coefficients between ordinal and dichotomous variables. Smoothing was applied to ensure a positive definite matrix and a continuity correction of 0.5 was applied to 20 cells with zero frequencies. The number of factors was examined by parallel analysis with the minimum residual estimator. A single-factor exploratory analysis was then estimated with a standardized loading of 0.30 as the retention threshold. Retained items

were submitted to confirmatory factor analysis in lavaan using the unweighted least squares mean and variance adjusted (ULSMV) estimator with items declared as ordered categorical. Fit was assessed against CFI at least 0.90, RMSEA below 0.08 and SRMR below 0.08, reported in standard, scaled and robust versions, the last incorporating corrections for non-normality. Analyses were conducted in R version 4.5.2.

4. Results

Parallel analysis nominally suggested six factors, but the first empirical eigenvalue was substantially larger than the second (7.47 vs 0.88) and accounted for 44% of total variance, while factors two through six exceeded their simulated counterparts only marginally. A unidimensional model was therefore tested. In the single-factor exploratory analysis four items fell below the retention threshold and were excluded: formal standardized assessment with OSATS or GOALS, simulation-based training for emergency conversion, intraoperative cholangiography and indocyanine green fluorescence. Thirteen items were carried forward to confirmatory analysis.

Fit indices diverged sharply according to the correction applied. Standard and scaled indices indicated adequate to excellent fit, whereas the robust indices did not support the model (Supplementary Table S1). Because the robust indices incorporate corrections for the non-normality of the observed data, the unidimensional model was not considered to fit acceptably, and no composite or factor score was derived.

Standardized loadings (Supplementary Table S2) were highest for the confidence items (0.840-0.946), followed by the operative volume items (0.778-0.889) and lowest for the dichotomous items (0.342-0.696). The latent variable recovered by the model is therefore better described as a general positive self-appraisal associated with accumulated experience than as technical proficiency in open cholecystectomy.

Supplementary Table 1 Fit of the unidimensional confirmatory model (ULSMV, n = 391)

Fit index	Value	Criterion
Chi-square (df = 65)	104.527	-

Scaled chi-square	132.555	(P < 0.001)	-
CFI, standard	0.996		≥ 0.90
CFI, scaled	0.979		≥ 0.90
CFI, robust	0.813		≥ 0.90
TLI, standard	0.995		≥ 0.90
TLI, robust	0.776		≥ 0.90
RMSEA, standard	0.039	(0.025-0.053)	< 0.08
RMSEA, scaled	0.052	(0.039-0.064)	< 0.08
RMSEA, robust	0.190	(0.171-0.209)	< 0.08
SRMR	0.054		< 0.08

CFI: Comparative fit index; RMSEA: Root mean square error of approximation; SRMR: Standardized root mean square residual; TLI: Tucker-Lewis index; ULSMV: Unweighted least squares mean and variance adjusted. Values in parentheses for RMSEA are 90% confidence intervals.

Supplementary Table 2 Standardized loadings and proportion of item variance explained

Item	**Lambda**	**R-squared**
Confidence: dissection and control of the cystic artery and duct	0.946	0.894
Confidence: obtaining the critical view of safety	0.937	0.878
Open cholecystectomy volume (primary surgeon)	0.889	0.790
Confidence: exploration of the common bile duct	0.869	0.755

Confidence: identification and repair of bile duct injury	0.840	0.705
Open cholecystectomy volume (assistant)	0.837	0.701
Laparoscopic cholecystectomy volume	0.778	0.606
Open cases in residency sufficient for independent performance	0.696	0.484
Perceived usefulness of training formats received	0.638	0.407
Directly observed and evaluated by a supervisor	0.595	0.354
Has ever converted a laparoscopic cholecystectomy to open	0.592	0.351
Formal training in common bile duct exploration	0.581	0.338
Formally evaluated in open surgical skills in the curriculum	0.342	0.117

Four items did not reach the retention threshold in the exploratory step and are not represented in the confirmatory model: formal OSATS or GOALS assessment, simulation-based training for emergency conversion, intraoperative cholangiography and indocyanine green fluorescence.

5. Interpretation

Two explanations for the failure of the four excluded items are compatible and not mutually exclusive. They are dichotomous and ask whether an event occurred rather than what the respondent is able to do, a format that constrains both variance and behavioural specificity. In addition, the experiences they describe were nearly universally absent in this sample: 79.3% had never undergone standardized assessment and 83.6% had never received simulation-based training for conversion, producing a floor effect that leaves little variance available to correlate with anything else. Under either explanation the instrument cannot discriminate levels of formal preparation among these respondents, which is why the manuscript reports the items only as descriptive frequencies. A revised instrument, built from behaviourally anchored items with a uniform response format and

with self-confidence and formal training modelled as separate constructs, is under development and will require content validity review by expert judges, piloting and independent exploratory and confirmatory analyses before any measurement claim is made.