

content; while this varied by region, ranging from 41.4% in the Midwest to 60.3% in the Northeast, the differences did not reach statistical significance [$\chi^2(3, N=283)=6.86, p=0.076$]. No significant associations were found between wellness content and program size ($p=0.304$), age ($p=0.387$), or length ($p=0.807$). Descriptive content analysis revealed substantial heterogeneity in how programs portray wellness to applicants. Common domains among programs providing descriptive information included mental health resources and access to counseling (30.1%), social and community activities (28.1%), retreats (26.7%), and presence of a formal wellness infrastructure or committee(s) (24.0%) (Table 1).

Conclusions: Wellness content on EM residency websites was inconsistently represented, with substantial variation in how support was described. The lack of program characteristic-based differences suggests that wellness communication is unsystematic and may not reflect broader institutional priorities. Clearer and more consistent wellness information could help programs better convey their culture and enable applicants to identify environments aligned with their needs.

48 Optimizing EM Residency Application Reviews a Comparative Study: Faculty versus AI

Robert Steele, Taylor Route, Michael Macias, Lauren Donnelly

Background: Residency application volumes have increased, straining faculty time for holistic medical student application review. Artificial intelligence (AI)-assisted screening may be a solution, but can it really review student applications?

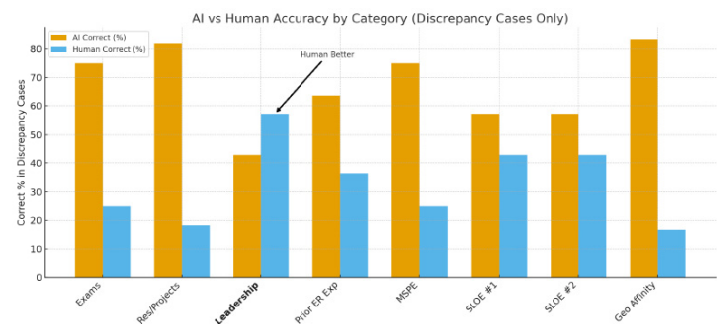
Objectives: To compare agreement and accuracy between an AI-based application scoring tool versus trained faculty application reviewers for EM residency interview screening.

Methods: We developed an AI algorithm using the cloud-based platform Airtable to generate scores in 10 predefined domains (exams, research/special projects/community service, leadership/distinction, prior EM experience, MSPE, two SLOE/letter domains, geographic affinity, exceptional bonus, and red flag). Thirty-four applicants from one recruitment cycle were independently scored by both AI and a faculty reviewer. For each domain we calculated percent exact agreement between AI and faculty. All discrepant scores were adjudicated by a trained independent blinded reviewer, who determined which score (Faculty/AI) was most accurate. We calculated the proportion of discrepant cases using McNemar's test.

Results: Across 340 domain-level ratings, AI and faculty agreed exactly on 252 (74.1%). Agreement by domain ranged from 38.2% for leadership/distinction to 94.1% for red flag scores. Among 88 discrepant ratings, the AI score

was adjudicated correct in 58 (65.9%) and the faculty score in 30 (34.1%) ($p=0.004$). Patterns were similar across most individual domains.

Conclusions: An AI-based application scoring tool demonstrated substantial agreement with faculty ratings and was more likely than the individual faculty reviewer to match adjudicated scores when disagreement occurred. AI-assisted scoring may be a feasible adjunct for initial residency application screening, potentially reducing faculty workload while preserving decision quality.



49 Content Validity Index (CVI) as a Tool for Instrument Development: A Methodological Case Study in Neonatal Lumbar Puncture and Umbilical Vein Catheterization for Emergency Medicine Simulation

Brendan Freeman, Kathryn Zabinski, Darya Ryndych

Background: The Content Validity Index (CVI) is a structured framework for evaluating instrument content. Although well-described in nursing literature, its use is limited elsewhere. Simulation-based procedural tools, particularly for neonatal lumbar puncture (LP) and umbilical vein catheterization (UVC), often lack documented content validity, creating uncertainty about instrument quality.

Objectives: To apply CVI methodology to develop and refine task-specific checklists and global rating scales for neonatal LP and UVC for emergency physicians. We hypothesized that iterative CVI analysis with expert raters would improve item- and scale-level validity metrics.

Methods: Design: Cross-sectional, survey-based instrument validation study using CVI methodology.

Setting: Electronic ratings from experts in neonatal and emergency medicine simulation.

Participants: Six expert raters recruited by purposive sampling completed two rounds of item relevance ratings. Measures: Item-Level CVI (I-CVI) and Scale-Level CVI/Average (S-CVI/Ave) were calculated. Items below accepted thresholds (I-CVI < 0.78) were revised or removed. Descriptive statistics were used for CVI calculations.

Results: Several items showed low I-CVI values initially

and were modified or eliminated. After refinement, S-CVI/Ave values for most checklists and global rating scales approached or exceeded 0.80, indicating stronger content validity

Table 1: Expert Ratings, Item-Level Content Validity Index (I-CVI), and Scale-Level Content Validity Index (S-CVI) from Round 1 Review of Four Assessment Tools: (1) UVC Task-Specific Checklist, (2) OSATS-UVC, (3) Lumbar Puncture Task-Specific Checklist, and (4) OSATS-Neonatal LP. Items rated "relevant" or "quite relevant" were scored 1; items rated "somewhat relevant" or "not relevant" were scored 0. NA = No answer.

TASK-SPECIFIC CHECKLIST-UVC								
Item	Expert #						I-CVI	
	1	2	3	4	5	6		
1	Creates and maintains a sterile field	1	1	0	1	1	1	0.83
2	Ties umbilical tape around the base of the umbilicus	1	1	1	1	1	1	1
3	Selects appropriate sized catheter based on patient's weight	1	0	1	1	1	1	0.83
4	Grasps umbilical cord with forceps and cuts cord perpendicularly with #10 or #11 blade scalpel to expose umbilical vasculature	1	1	1	0	1	1	0.83
5	Correctly identifies the umbilical vein	0	1	1	1	1	1	0.83
6	Inserts catheter into umbilical vein, to appropriate depth, which is when blood freely flows through the catheter, but no more than 5 cm	0	1	1	0	1	1	0.67
7	Applies an appropriate method to secure the umbilical line (transparent dressing, sutures, or H-tape method)	1	1	0	1	1	1	0.83
8	Would you allow this individual to perform an UVC unsupervised and independently on your next patient? Yes or no?	1	0	1	1	1	1	0.83
S-CVI/Ave								0.83
OSATS-UVC								
1	Instrument Handling	1	0	1	1	1	1	0.83
2	Respect for tissue	0	0	0	0	1	1	0.33
3	Time & Motion	0	0	1	1	1	0	0.5
4	Knowledge of specific procedure	0	1	1	1	1	1	0.83
5	Competency Evaluation	0	0	1	1	1	1	0.67
S-CVI/Ave								0.63

TASK-SPECIFIC CHECKLIST LUMBAR PUNCTURE								
0	New Scoring System (1,2,3)	1	1	1	1	1	1	1
1	Plans Insertion Site	1	1	1	1	1	1	1
2	Preparation	1	1	1	1	1	1	1
3	Cleanses	1	1	1	1	1	1	1
4	Maintains sterility	1	1	0	1	1	1	0.83
5	Instructs holder	1	1	1	1	1	1	1
6	Inserts Needle	1	1	1	1	1	1	1
7	Advances needle	NA	1	1	1	1	1	1
8	Makes corrections	1	1	1	1	1	1	1
9	Acquires fluid	1	1	1	1	1	0	0.83
10	Removes needle	0	1	1	1	1	1	0.83
11	Discards Sharps	1	1	1	1	1	1	1
12	Competency evaluation	1	0	1	1	1	1	0.83
S-CVI/Ave								0.95
OSATS-NEO LP								
1	Instrument Handling	1	1	1	1	1	1	1
2	Tissue Handling	1	1	1	1	1	0	0.83
3	Time & Motion	1	1	1	1	1	1	1
4	Knowledge of Procedure	1	1	1	1	1	1	1
5	Decision-Making and Adaptability	1	1	1	1	1	1	1
6	Competency Evaluation	1	1	1	1	1	1	1
S-CVI/Ave								0.97

evidence. Some individual items remained below threshold, but overall scale-level metrics improved.

Conclusions: CVI methodology offered a systematic approach for refining simulation-based assessment tools for neonatal LP and UVC. The resulting instruments demonstrate improved content validity and may strengthen evaluation practices in procedural training. Limitations include the small expert sample and absence of further validation evidence.

50 Compassion Satisfaction, Burnout, and Secondary Traumatic Stress Across Career Stages in Emergency Medicine Physicians

Tuan Vo, Rebecca Jeanmonod, Donald Jeanmonod

Background: Emergency physicians (EPs) work in high-pressure environments where compassion satisfaction (CS) can coexist with burnout (BO) and secondary traumatic stress (STS), the key components of compassion fatigue. This study evaluates CS, BO, and STS across career stages and examines differences between trainees and attendings.

Methods: A cross-sectional survey using the validated Professional Quality of Life (ProQOL) tool was distributed anonymously to practicing and training EPs in the United States. Demographics were collected. Chi-Square and Kruskal-Wallis tests were used, with significance set at $p < 0.05$.

Results: A total of 1,170 EPs participated: 503 male (42.99%), 658 female (56.24%), and 9 identifying as other (0.77%). Respondents included 166 residents (14%), 192 EPs within 5 years post-graduation (16%), 240 with 6–10 years (21%), 240 with 11–15 years (21%), 144 with 16–20 years (12%), and 188 with >20 years (16%). Significant differences across career stages were found for CS ($p = 1.99E-15$), BO ($p = 6.73E-11$), and STS ($p = 9.0E-4$). CS demonstrated a U-shaped trend—highest among residents and >20-year EPs, with lower levels in mid-career. BO and STS followed inverted U-shaped trends, peaking in mid-career and lowest among trainees and late-career EPs.

