Comparison of SP and Faculty Member Responses on Long Checklists
MS Clifton, JF Mahoney, PJ Campbell, DB Middleton, PL Rogers
Advanced Clinical Education Center
University of Pittsburgh School of Medicine

Introduction
SP scoring has been shown to be accurate, but reliability decreases quickly with increasing checklist length. A separate observer may be able to reliably score a greater number of items, but adds to the cost of the program. The goal of this study was to assess the reliability of observers vs. SP scoring from memory of long checklists on an extended SP history and physical examination (H&P).

Methods
For three years, all second year medical students were evaluated on their performance of a complete, 60 minute, head-to-toe H&P, after their Clinical Skills Course (CSC).

• Scored by
  • Faculty with 130 item checklist
  • SP with 50 item checklist
  • (subset of 130 items)
• Compared faculty and SP agreement
• Analyzed disagreements by:
  • Individual Faculty Member
  • # H&Ps scored by each Faculty Member
  • Individual SP
  • Item number

For additional information: mclifton@medschool.pitt.edu
phone (412) 648-8714

Results
Number exams 574
Number faculty 119 (range 1-27 H&Ps)
Number SPs 22 (range 2-64 H&Ps)
Data points available 27,849 (97%)
Overall agreement 25,361 (91.1%)
# faculty ≤ 1 disagreement 7 (6%)
# faculty ≥ 8 disagreements 11 (9%)
# items ≥ 95% agreement 18 (36%)
# items ≤ 85% agreement 11 (22%)

Conclusions
• Overall very good agreement between SPs and faculty, even on very long checklists
• Analyzing agreement useful for
  • Identifying tapes to review
  • Identifying checklist items to review