

Crowd-Sourcing Assessment of Technical Skills (C-SATS): Faculty Experts vs. Amazon.com Mechanical Turk Project™ vs. Facebook™

Carolyn Chen¹, Lee White, Timothy Kowalewski, Bryan Comstock, Thomas S. Lendvay

1. University of Washington School of Medicine 2.

✉ **Corresponding author:** Carolyn Chen, carochen@u.washington.edu

Categories: Quality Improvement

Keywords:

How to cite this poster

Chen C, White L, Kowalewski T, et al. (2012) Crowd-Sourcing Assessment of Technical Skills (C-SATS): Faculty Experts vs. Amazon.com Mechanical Turk Project™ vs. Facebook™. Cureus 4(9): e140.

Abstract

Purpose Current methods of surgical performance evaluations require significant resources and can contain inherent biases when surgeons grade their trainees. The purpose of this study is to develop an alternative method that mitigates this resource burden and ensures prompt, bias-free, 'expert-like' feedback. **Method** Three groups of subjects were recruited for this study: Amazon.com Mechanical Turk users, Facebook users, and surgeon experts. Recruitment emails to the experts were sent and Mechanical Turk and Facebook announcements were posted on the respective websites. A surgical skill assessment survey called C-SATS was adapted from a validated robotic surgery rating tool and disseminated online. C-SATS is a two-step survey that includes a qualification question determining the assessor's ability to discriminate between high and low performing surgeons and the criterion test containing the video of the surgeon being rated. For the criterion test, the subjects rate the surgeon on a Likert scale of 1-5 for bimanual dexterity, depth perception, and efficiency. The global performance rating of the surgeon was obtained by summing the ratings of the 3 domains with a scale of 3-15. The averaged global performance rating from the Mechanical Turk and Facebook study were compared to the averaged rating by the expert surgeon and equivalence was calculated. **Results** Expert raters graded the skills assessment video with a mean score 12.11, yielding an equivalence window of 11.11 to 13.11. Mechanical Turk and Facebook graders produced mean scores of 12.21 (95% CI: 11.98 to 12.43) and 12.06 (95% CI: 11.57 to 12.55) respectively. It took 28 days to obtain responses from 10 expert surgeons, 5 days to obtain 501 responses from Mechanical Turk subjects, and 25 days to receive 110 responses from Facebook subjects. **Conclusion** We discovered that crowd-sourced surgery-naïve graders provided assessments equivalent to trained expert surgeon evaluators, and at a faster rate.

Open Access

Published 09/23/2012

Copyright

© Copyright 2012

Chen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 3.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Distributed under

Creative Commons CC-BY 3.0

