Introduction

Healthcare is a rapidly developing field with evolving tools, skills, and protocols aimed at improving patient care. Due to these inherent dynamics, medical professionals require frequent training and evaluation of technical skills. The nursing workforce experiences high volumes of patient interactions involving technical expertise ranging from measuring blood pressure to placing a Foley catheter. In addition to the current nursing shortage and need for new hires, the sheer amount of technical skills in the field of clinical nursing demands efficient formal orientation trainings [1]. Providing accurate, efficient, and affordable competency assessment for healthcare providers is essential for continual quality improvement [1-3]. Medical professionals perform tasks every day that carry risk of infection, technical error such as misusing a device, and decreased patient satisfaction due to inefficiency or perceived lack of confidence, all of which can negatively impact patient outcome [4].

Competency assessment in healthcare aims to prevent these negative effects by correctly identifying professionals who are struggling and providing training and education for improvement [5]. However, the assessment of clinical competence remains a “laudable pursuit” that isn’t universally being reached due to a myriad of challenges [2]. Assessing competency is time-intensive and expensive, and requires at least one certified nurse educator to teach/assess relatively simple nursing skills. In our study, over 1,400 nurses being trained at one hospital for a glucometer device received 15 minutes of didactic instruction and 5 minutes of hands on demonstration as part of seven 4-hour training sessions over the course of two months. That training period involved paying 1,400 nurses for their time and paying experts and evaluators to lead the training, all while mandating all nurses spend time in an education session rather than at the bedside. In addition to the costs and time, successful competency assessment in and of itself is challenging due to biased evaluators, inconsistent assessment tools, and no objective way to guarantee that those individuals who need more help receive it [3,4,6]. Some innovations have attempted to address the challenges of competency assessment, such as using video-based self-assessments to give participants and their evaluators accurate views of their performance [6]. The field of nursing and healthcare in general is in need of a competency assessment methodology that is reliable, fast, and affordable.

Crowd sourcing is the practice of obtaining services from a large group of people, typically the general public on an online community [7]. Researchers outside healthcare or within the field frequently...
Kirsch et al. (2016) utilize the ‘power of the crowd’ to perform discrete tasks [8], create content frameworks like Wikipedia [9], produce repositories of public data [10]. Khatib, et al., used crowdsourcing as a method to create an online game in which participants created an accurate model of a protein structure that had previously eluded scientists [11]. Crowdsourcing provides data that is inexpensive, rapid, and objective due to the huge size of participant pool [12-14]. Holst, et al., show that crowdsourcing can be used to obtain valid performance grading of videos of urologic residents and faculty at different levels on basic robotic skills tasks [15,16]. Moreover, Holst, et al., obtained these results in hours, making the use of crowdsourcing for competency assessment of technical skills more valuable. We hypothesize that crowdsourcing can be used as an accurate, cheap, and near-immediate assessment tool adjunct for medical technical skills.

Methods

This study, including all study materials (surveys), was submitted and approved by Seattle Children’s Hospital IRB as project #15251.

Glucometer task video recordings

All faculty nurses at Seattle Children’s Hospital (approximately 1,400 nurses) were being trained to use a new glucometer device to measure patient blood glucose. After Institutional Review Board approval, the crowdsourcing pilot project was introduced for 5 minutes at each of the seven, 4-hour nursing Mandatory Education training sessions. Nurses were informed that they would be filmed and that the videos would be uploaded to a crowdsourcing site without any identifiable information and then were given the option to participate or not participate in the study. The participating nurses were required to run through a series of steps on the glucometer at a station and upon completion, fill out a self-evaluation form stating that they had completed each of the steps in correct order (Figure 1). A camera was set up at one of these glucometer stations and nurses had the option to volunteer for the project. No identifying information (face, badge, written names) was intentionally obtained in the video recording. Over the course of seven Mandatory Education sessions, 34 nurses were video recorded (61.4 total minutes) as they completed the glucometer task (Figure 2).

Video editing and uploading

Videos were uploaded to secure, password-protected computers and edited to the length where each nurse was on camera. Any unintentionally identifiable information (badge) was digitally scrubbed from the video. The videos were completely anonymous and had random codes assigned. Editing and uploading took 5-10 minutes per video and was completed within one week of the final Mandatory Education session.

Crowdsourcing

Evaluators were recruited through the Amazon.com Mechanical Turk™ platform and served as our “crowd-workers.” In order to qualify for the study, the crowd-workers had to have completed 100 or more Human Intelligence Tasks (HITs), the task unit used by Mechanical Turk™, and must have had a greater than 95% approval rating as qualified by the Amazon.com site as described in Chen, et al. [17]. These worker evaluations were labeled only by a unique, anonymous user identification code provided by Amazon.com and no other information was known about them (gender, age, sex, ethnicity, etc). Each Mechanical Turk™ subject was compensated $0.60 USD for participating.

A glucometer task assessment survey was created from the self-assessment used in the Mandatory Education sessions and hosted online on a secure server. The survey consisted of 3 parts. First, the evaluators were shown labeled items to be used in the task (glucometer, badge, vial of fake blood, etc). Second, the evaluators were shown an annotated video of a nurse performing the completed task (Figure 3, annotated screen shot). At the end of this second part, the crowd was presented with a screening question to exclude but still remunerate evaluators who were not paying attention to the annotated video. Third, the evaluators completed the technical skills assessment.
survey. The survey tool was identical to the self-assessment used by the nurses (Figure 4). The survey listed each of the glucometer steps in correct order. The survey also contained an attention question to ensure that the assessor was actively paying attention. Assessors who incorrectly answered the attention question were also excluded from the analysis. The crowd was asked if the nurse completed the step, completed the step in correct order, or if the video was unclear. At the end of the survey, the crowd was allowed to write-in comments tailored to the video they just watched. One of the nurses in the 34 videos was an expert who completed the task in correct order. Crowd worker videos were not shown in a consistent sequence, nor was a single crowd worker guaranteed a chance to review all videos.

Feedback survey

A second survey was created at the end of the study and was emailed to all nurses who attended the Mandatory Education sessions by Seattle Children’s Hospital (Figure 5). The nurses who participated in the study could volunteer to fill out the survey and give feedback about their participation.

Analysis

Using responses from crowd workers who successfully completed the survey and passed the attention question, we calculated percentages of correct steps for each video. Responses from crowd workers were limited to “yes”, “no” or “yes but not in correct order” for each step in the video. Based on these results, we were able to calculate the percentage of correct steps per video and also the percentage of success per step based on all of the videos. The analysis was descriptive proof of principle to assess whether crowds were able to identify correct and incorrect steps.

Results

After excluding subjects based on either the pre-survey screening question or the mid-survey attention question, the 34 nursing glucometer task videos received evaluations from 1300 Mechanical Turk™ crowd-workers (average of 38.2 evaluators per video). We received the 1300 evaluations over the course of 3 hours and 54 minutes. The seven Mandatory Education sessions took 28 hours and used self-assessment, by comparison. Without utilizing the evaluations, we observed a trend of shorter task duration with number of steps completed in correct order. The crowd accurately assessed that the expert (Figure 6, Video ID 1) completed the steps and completed them in order. In comparison, the crowd identified nurses who missed steps and/or completed steps out of order (Figure 6).

The crowd determined that three of the videos (Figure 7, Video ID 14, 25, 6) demonstrated frequent errors (at or above 10% missed and/or incorrect ordered steps). Additionally, the crowd identified certain steps that were frequently missed and/or incorrectly ordered by >20% of the nurses (Figure 8, Step # 2, 3, 4, 5, 7, 12). In particular, steps that Mandatory Education deemed “follow-up” steps were some of the most frequently missed (Figure 9). Finally, the feedback survey sent to the nurses showed that the feedback they would find most valuable in competency assessment would be specific comments on how to improve and a video of the gold standard (Figure 10). Also, 62% of the nurses who completed the feedback survey stated that they would “definitely” and/or “probably” use video recordings for assessment if available (Figure 11).

Discussion

The current methods of assessing competency in healthcare workers such as nurses are not ideal because they are costly, slow, and subjective [18]. Currently, due to the variability between instructors, nurses, new technical skills, and hospital protocols, there is no “gold standard” for assessing nursing technical skills to ensure competency [3]. It isn’t feasible to have 1,400 nurses blindly evaluated by a handful of faculty not only because it would be extremely time consuming for the faculty, but also because it would leave room for subjectivity and unreliable evaluations. CSATS (Crowd Sourced Assessment of Technical Skills) offers an alternative approach. Each nurse received more than 30 anonymous, blinded evaluations in only a few hours, and many of those evaluations include specific criticism. The
crowdworkers on the Mechanical Turk™ site are healthcare-naïve and not affiliated with the hospital training programs, and thus are not biased or influenced by expectations of performance.

CSATS could be a valuable triage tool for nurse educators. Future applications could include building a training kiosk where nurses or other healthcare professionals could demonstrate a dry-run of a bedside technical skill. The video could be recorded, immediately uploaded to a crowdsourcing site along with an evaluation tool, and the individual could have results near-immediately. For example, nurses could take 15 minutes of their shift to read through instructions and record themselves using the glucometer. The education team could be a valuable triage tool for nurse educators. Future

---

**Figure 6**: Crowdsourced evaluations of videos.

**Figure 7**: Videos organized by steps missed.

**Figure 8**: Steps organized by percent success.

**Figure 9**: Compilation of crowdsourced evaluations of follow-up steps.

**Figure 10**: Feedback survey question 1.

---

**Nurses – Glucometer Study**

Follow-up Questions (n=60 per video)

- FU Q1: Did the nurse quickly close the lid on the bottle after obtaining a test strip?
  - Yes
  - Not sure
  - No

- FU Q2: Was the test strip facing up when the nurse inserted it into the glucometer?
  - Yes
  - Not sure
  - No

- FU Q3: Did the nurse apply the drop(s) of blood to the tip of the test strip and NOT on top or on the side of the test strip?
  - Yes
  - Not sure
  - No

- FU Q4: Did the nurse keep the glucometer flat while applying the drop(s) of blood to the test strip?
  - Yes
  - Not sure
  - No

- FU Q5: Did the nurse add a comment after receiving the results of the blood test?
  - Yes
  - Not sure
  - No
could upload an evaluation tool to a crowdsourcing site and establish a passing or failing grade. The crowdsourced feedback could be used to triage the nurses according to their skill level, thus allowing those who demonstrate competency to not waste valuable time with unneeded training, and offering specific, tailored advice to those who require more education. Moreover, as demonstrated in our results, the education team may discover from the crowd-sourced data that many nurses frequently miss a certain step. This could lead to re-education regarding that step for all nurses.

A limitation of this study is that it was a relatively simple task that could all be done in one area, without other healthcare providers, patients, or devices. Many other technical skills performed by nurses are done at the bedside or in clinic. Continuations of this study could broaden the scope of CSATS with assessing more complex technical skills.

Our study showed that healthcare-naïve crowd workers can assess technical skill proficiency rapidly and accurately at nominal cost. CSATS may be a valuable tool to assist educators/supervisors in creating targeted training curricula for nurses in need of follow up while rapidly identifying nurses whose technical skills meet expectations, thus, dramatically reducing the resource burden for nursing training.

References


Figure 11: Feedback survey question 2.