JAANP

Standardizing nurse practitioner student evaluation during high-stakes clinical examinations

Melanie Keiffer, DNP, APRN, ANP-BC, CNE, CCRN Alumnus (Clinical Professor and Vice-chair)[1], Mindi Anderson, PhD, APRN, CPNP-PC, CNE, CHSE-A, ANEF, FAAN (Director of the Healthcare Simulation Program and Professor)[1], Dawn O. Eckhoff, PhD, APRN, CPNP-PC (Assistant Professor)[1], Christopher W. Blackwell, PhD, APRN, ANP-BC, AGACNP-BC, CNE, FAANP, FAAN (Director of Adult-Gerontology Acute Care Nurse Practitioner Program and Associate Professor)[1], Steven Talbert, PhD, RN (Interim Director of Nursing PhD Program and Clinical Assistant Professor)[1], & Ann Marie Parker, BA, ADN (Standardized Patient Educator)[2]

ABSTRACT

Nurse practitioners (NPs) require clinical competence in advanced health assessment skills to differentiate between normal, variations of normal, and abnormal patient findings. High-stakes clinical examinations, using live hands-on simulation scenarios and standardized patients (SPs) or other human role players, are accepted teaching and evaluation tools for NP students. Providing objective, valid, and reliable feedback to students during high-stakes clinical examinations is of considerable value for ongoing skill development. The study examined opportunities to improve the quality of student evaluation in simulation testing modes. A purposive sample of 17 video recordings of health students' comprehensive examination of an SP or physical examination teaching associate (PETA) from a nursing graduate level health assessment course was evaluated. Using a standardized rubric, students were scored live and after a comprehensive examination of a SP/PETA and via a secure web-based video platform by the faculty and an independent reviewer. Evaluator group examination score comparisons revealed that distributions of examination scores within evaluator groups were not similar. Median examination scores were significantly different between groups; faculty median examination scores significantly higher than SPs/PETAs. Efficiency of student evaluation may be increased by improving reviewer training, reducing checklist length, and adopting electronic scoring. Development of an exemplary teaching video providing explanation and detail for expected student skill performance will allow reviewers to practice and improve competence in reliable scoring, reduce time and effort of scorers, and increase accuracy of scoring.

Keywords: Clinical performance evaluation; clinical competence; nurse practitioner; simulated patient encounter; standardized patients.

Journal of the American Association of Nurse Practitioners 00 (2020) 1–7, © 2020 American Association of Nurse Practitioners

DOI# 10.1097/JXX.0000000000000514

Background

High-stakes clinical examinations, using live hands-on simulation scenarios and standardized patients (SPs) or other human role players, are accepted teaching and evaluation tools for nurse practitioner (NP) students (Kowitlawakul et al., 2015; Ryall et al., 2016; Schram, Mudd, 2015). To independently practice as licensed health care providers, NPs require competence in advanced health assessment skills to evaluate normal and abnormal patient findings (National Organization of Nurse Practitioner Faculties [NONPF], 2017). These skills are taught in the Advanced Health Assessment and Diagnostic Reasoning didactic course at the associated university and validated in a summative final practical (performance) examination during the corequisite laboratory course. Examination results frequently determine whether students are "practice ready" and prepared to advance to placement at a clinical site. The simulated laboratory section of this course typically involves human role players, referred to as SPs or sometimes simulated patients or participants (SPs) (INACSL Standards Committee, 2016; Lewis et al., 2017). The NP students are expected to demonstrate clinical competence

[1]University of Central Florida College of Nursing, Orlando, FL. Dr. Keiffer is now at the Department of Nursing, University of Vermont, Burlington, VT, [2]University of Central Florida College of Medicine, Orlando, FL

Correspondence: *Melanie Keiffer, DNP, APRN, ANP-BC, CNE, CCRN Alumnus, Department of Nursing, University of Vermont, 216 Rowell Bldg., 106 Carrigan Drive, Burlington, VT 05405. Tel: 802 656 3399; Fax: 802 656 8306; E-mail: melanie.keiffer@med.uvm.edu*

Received: *15 May 2020; revised: 15 July 2020; accepted 21 July 2020*

in a comprehensive head-to-toe physical examination of an SP or physical examination teaching associate (PETA).

In addition to learning comprehensive examination skills, students practice therapeutic communication, problem-based assessment, documentation of the associated assessment, and adapting the examination per the patient's age (lifespan) in a simulated laboratory setting. These are critical skills for NP students (LaManna et al., 2019) and expected outcomes are similar to the discussion of Advanced Physical Assessment by Anderson et al. (2010).

Students have several opportunities to practice with SPs and/or PETAs within the laboratory course structure. Activities include a practice communication exercise with no physical component where SPs/PETAs provide feedback on communication skills. A focused history and physical examination where a patient presents with acute symptoms, that is, "problem-based examination," is required (Anderson et al., 2010; LaManna et al., 2019).

The NP program is supported by nursing laboratories that include a simulation suite. Five patient rooms with video-recording capabilities; a health assessment laboratory with three private examination rooms; and six private examination stations, two debriefing rooms, and several storage and conference rooms are available to use in the health assessment course. Examination assessment kits are available for checkout for student practice, and wall system examination equipment is available in each room for student use.

At the completion of the course, students participate in a simulated high-stakes head-to-toe examination on a patient presenting for a "well check," followed by a write-up of the objective findings. History-taking skills are not assessed because this skill is evaluated separately using a screen-based, standardized virtual program. Therefore, scripting for the SP/PETA is minimal. As part of the examination write-up, students are assigned a special population, such as an infant/child/adolescent or pregnant woman, and the student must discuss how the physical examination varies for differing populations. The total maximum time is 2 hours overall: 1 hour each for the assessment and write-up portions. On completion of the examination, the student is provided graded feedback regarding clinical performance. Providing objective, valid, and reliable feedback to students during high-stakes clinical examinations is of considerable value for ongoing skill development. To examine opportunities to improve the quality of student evaluation in experiential testing modes, SPs, instruments, and scoring are considered.

Standardized patients (SPs)

According to the International Nursing Association for Clinical Simulation and Learning (INACSL) Standards of Best Practice (SoBP): Simulation℠ Simulation Glossary, an SP is a person who has received specialized training to mimic the signs/symptoms of a patient during a simulated activity (INACSL Standards Committee, 2016), whereas a PETA is someone who is trained on physical examination techniques (Johns Hopkins Medicine, n.d.).

The Association of Standardized Patient Educators (ASPEs) provides SoBP to ensure the effectiveness of simulation and the safety of participants (Lewis et al., 2017). It is recommended that SP training include consistency in the portrayal of roles, the provision of feedback, and documentation of performance on assessment instruments. In high-stakes examinations, ASPE recommends training SPs to behave in a standardized manner to provide students with a quality examination experience (Lewis et al., 2017).

Standardized patients have been used in lieu of actual patients as an objective clinical measure for clinical performance examinations since 1960s (Barrows, 1993). Standardized patients are trained to act as patients with medical problems and are selected based on the demographics requested for a particular event.

Many reasons to use SPs exist, including the provision of experiential learning in a controlled environment and preventing bias (Miller et al., 1998). In the clinical setting, faculty are unable to control the type of patient or diagnoses that make evaluations inconsistent from one student to another (Miller et al., 1998). Clinical performance examinations in a simulated setting allow for direct observation of the student, interactive encounters, and assessment of advanced clinical skills in a controlled environment (Yudkowsky, 2020).

Simulation evaluation instruments

Standardized simulation evaluation scoring tools measure student performance and document achievement of student learning outcomes. Development, refinement, and reuse of evaluation instruments to establish validity and reliability of tools is recommended; however, barriers to instrument development in nursing education exist (Kardong-Edgren et al., 2010). Faculty time and expertise are required to evaluate psychometric properties of instruments.

Studies have evaluated the validity of tools in performance examinations. In a study by Park et al. (2016), the validity of a rubric was evaluated for a performance examination (five cases) in a specific format for licensing examinations. The results showed that reviewers thought they were able to discern between those students considered excellent and those who would not pass, but it was much more difficult to differentiate other levels. The authors suggested keeping a list of examinations difficult to score and final decisions along with the rationales to provide consistency among cases. In addition, notes related to differentiation of levels should be kept and shared among reviewers. Finally, double-scoring examinations were considered beneficial in some cases (Park et al., 2016).

In a study of 499 undergraduate nursing students, the Student Assessment Tool for Standardized Patient Simulations was evaluated in relation to student clinical competence with SPs (Castro-Yuste et al., 2018). The authors found sufficient reliability and validity in measuring communication skills and management of patient safety. Reviewing, providing better descriptors, or eliminating items that have that less capacity for discrimination were recommended (Castro-Yuste et al., 2018). Communicating effectively is a core competency of NPs to enhance safe care for patients at the individual and systems levels (NONPF, 2017). Measuring interpersonal and communication skills during simulated experiences are as important as measuring technical and knowledge skills (Castro-Yuste et al., 2018).

Scoring of clinical performance examinations

Scoring of performance examinations may present with many internal and external errors (Tavakol & Pinner, 2018). One error is assessor bias (Tavakol & Pinner, 2018), and acceptable reliability may also be an issue (Daniels et al., 2014). Reliability may be particularly difficult with extensive, numerous item checklists (Yudkowsky et al., 2004). Pausing lengthy videotaped SP encounters to score more frequently is recommended to improve accuracy of clinical performance reporting (Turner et al., 2016).

Few studies of NP physical examination and faculty/SP scoring have been published. In one study by Miller et al. (1998), rating of faculty members in a clinical performance examination of NPs differed. Types of scenarios and faculty-related factors, such as experience and attitudes, were determined to be the cause. In this study, the harder items to rate were related to psychosocial skills, and faculty factors included leniency when the examination counted as a grade. Of interest, faculty scores differed when assigned as first or second grader, with higher scores occurring with the first grader (Miller et al., 1998). Minimal clinical competence is often judged by another during a performance examination (Tavares & Eva, 2014; Yudkowsky, 2020); however, competence related to NP mastery of skill may be difficult to judge (Clark, 2015).

In another study, scores from an NP Objective Structured Clinical Examination with SPs for history-taking and physical examination were compared between faculty (n = 4) on a 268-item checklist with over 1,000 potential tasks (Clark, 2015). Examinations were scored via videotape of the performance; items were scored on a dichotomous "Done" or "Not Done," and strong agreement was found in grading student competence (Clark, 2015).

Standardized patients have been used as evaluators for NP examinations, particularly to corroborate faculty scoring (Miller et al., 1998). In medicine, examinations may be assessed by a patient instructor, that is, someone who has been trained to complete the checklist, who may also serve as an SP (Yudkowsky et al., 2004). Similar to NP students, clinical performance examinations in medical students assess skills such as history-taking and physical assessment (Park et al., 2016; Yudkowsky et al., 2004). Head-to-toe performance examinations are also evaluated with medical students (Yudkowsky et al., 2004).

Currently, a scant amount of literature regarding the best way to conduct NP student evaluations in high-stakes examinations exists. A standardized process to evaluate NP students is essential to evaluate priority competencies and to provide fair and accurate measurement. An evaluation of the current process regarding high-stakes student evaluations was needed.

Purpose/aims

This study compared summative subjective and objective assessment feedback from an SP and/or PETA reviewer, clinical faculty reviewer, and an independent reviewer (IR). The study measured the extent to which reviewers assigned scores to variables to provide simulation program evaluation through audit and feedback to drive process improvement strategies. The authors aimed to 1) compare mean evaluation scores and qualitative feedback from SPs or PETAs, clinical faculty, and IRs in a high-stakes clinical examination, 2) appraise qualitative comments to discover common themes and attitudes in the evaluation process, and 3) examine opportunities to improve the quality of student evaluation in simulation testing modes.

Methods

At the study site, evaluating student clinical competence involves a practical examination with a written summative assessment at the completion of an advanced health assessment course. The study was reviewed by the university institutional review board, and participants, including students, SPs/PETAs, and faculty, were consented with an explanation of research before data collection. Prospectively, SPs/PETAs were randomly assigned to an examination room. Students entered the assigned examination room and performed the required assessment. Video-recording of the practical examination is standard procedure in the advanced health assessment laboratory course, and students enrolled in the course have a photo/video consent form on file.

On the day of the examination, students were scored live by an SP/PETA via video during their comprehensive examination of another SP/PETA. Clinical faculty scored later via video, and an IR also scored at a later date via video. Clinical faculty were experienced NP faculty who previously taught the advanced health assessment clinical course and used the evaluation tool. IRs were experienced NP faculty who had taught advanced health assessment and/or subsequent clinical courses. These faculty were experienced in evaluating NP students via video or at clinical sites using written clinical evaluation

Table 1. Comparison of clinical examination scores

Group	Mean	95% Confidence Interval of Mean	SD	Median	Interquartile Range
SPs/PETAs	66.84 points	66.96–68.89 points	3.99	66.75 points	64.00–69.88 points
Clinical faculty	70.37 points	68.73–72.01 points	3.18	71.50 points	68.67–73.00 points
Independent reviewer	69.19 points	67.05–71.33 points	4.17	70.00 points	66.13–72.38 points

Note: PETA = physical examination teaching associate; SP = standardized patients.

tools. All scorers were blind to others' scores. Although not all reviewers received specific training, the same SP/PETA training notes were given to all.

Standardized patient/physical examination teaching associate training

For SP/PETA training, an adapted "round-robin" teaching strategy was used (University of Illinois College of Medicine at Urbana-Champaign, n.d.). An SP educator experienced in scoring high-stakes clinical examinations led the training course. Trainees were arranged in a circle, and the process encouraged the contribution from all participants. The trainer used an iterative process of building off consecutive contributions by each trainee. Each SP/PETA offered a verbal thought or reaction to each checklist item until all checklist items were complete. Insights, central points, and item-specific questions were discussed and recorded to assure clarity of scoring. Questions were followed-up via a conference call with the nursing faculty for clarification. After the training session, the checklist notes, instructions, and recommendations on scoring were updated. The notes that were sent to the nursing faculty for approval before the finalized checklist, with grading instructions, were sent to the simulated participants and printed for reference during the grading encounter. The SP educator did not participate in any scoring of students.

Grading

The current grading process includes scoring by a trained SP/PETA who participates in the examination as the patient in collaboration with a SP/PETA who observes from a video monitor outside the examination room. On examination completion, the SP/PETA in the examination room meets with a SP/PETA observer to review the checklist. Clinical faculty and IRs, blinded to student scores, access and score the video-recorded examintions at a later time. The student is graded using a five-page, 88-item, six-column checklist that was adapted, with permission, from a 75-point total, *Head-to-Toe Comprehensive Grading Rubric,* created by Jacqueline Michael, PhD, APRN, WHNP-BC at the University of Texas at Arlington College of Nursing and Health Innovation (LaManna et al., 2019).

The checklist has also been improved with input from the college of medicine. The checklist has items focused on psychomotor examination skills, interpersonal skills, and a comments section for open-ended qualitative feedback regarding performance. Items are scored between 0.5-1.0 point/item; correct items are scored with full credit, incorrect items receive half credit, and failure to complete the item receives no credit. Students are able to complete examination items in any desired order (Yudkowsky et al., 2004); however, points are given for completing in a systematic/orderly manner. Faculty do not delegate grading to the SP/PETA but input from the SP/PETA is considered in the grading process.

Design

This study used a retrospective descriptive design with blinded reviews and random assignment of SPs/PETAs and IRs. Standardized patients/PETAs completed the evaluation rubric after the student assessment. Clinical course faculty reviewed and scored assessment videos for all students assigned to their course section. IRs were randomly assigned the recorded student assessments for scoring. Standardized patients/PETAs, faculty, and IRs were blinded to examination scores of other reviewers.

Sample

A purposive study sample (n = 17) included live performance and video records of students' comprehensive clinical examinations viewed in a secure, web-based, online video-recording platform. Each record was scored by the SP/PETA, clinical faculty, and an IR. All faculty viewing videos were Family Educational Rights and Privacy Act trained (U.S. Department of Education, 2018) to protect the privacy rights of students.

Data analysis

Raw scores of examination results were tabulated, and descriptive statistics for each group was calculated. Scoring was analyzed to determine differences between SPs/PETAs, clinical faculty, and IRs using the Kruskal–Wallis test. Post hoc group pairwise comparisons were made using the Dunn procedure with a Bonferroni correction.

Qualitative data from SPs/PETAs, clinical faculty, and IR feedback were analyzed by the primary author. Using inductive manual coding, the sentence structure of comments were analyzed to extract common themes and patterns (Fereday & Muir-Cochrane, 2006). Text with similar meanings were assigned a code name which captured the essence of the text. The data set was broken into smaller samples and organized creating coded labels that best described the sample. The sample was reread, and new codes were created until all the data were coded.

Results

A comparison between means of overall examination scores showed the following average point score out of a total of 75 points: SP/PETA (66.84 points), clinical faculty (70.37 points), and third-party reviewers (69.19 points). The distributions of assessment evaluation scores were significantly different between groups, $\chi^2(2) = 6.87$, $p = .032$. Post hoc pairwise comparisons found clinical faculty scores to be significantly higher than those of SPs/PETAs (71.50 vs. 66.75, $p = .030$). No other significant differences between groups were noted (**Table 1**).

In qualitative analysis, three themes emerged: feelings and perceptions toward the examiner, missed checklist items, and errors in examination maneuvers. Overall, SPs/PETAs provided a considerably higher number of comments and more detailed written feedback than clinical faculty and IRs. Most comments regarding rapport, quality of therapeutic touch, professional appearance, eye contact, communication skills, order of examination, and examination technique were provided by SPs/PETAs. Clinical faculty provided the least number of qualitative comments and focused primarily on missed checklist items. IRs primarily provided feedback on errors in technique, examination maneuvers, and identified miss examination items.

Discussion

Standardized patients/physical examination teaching associates are most familiar with the examination process and the rubric that may contribute to better scoring accuracy than the faculty and third-party reviewer groups. However, they may be less familiar with clinical knowledge (Tavares & Eva, 2014). Because SPs/PETAs may complete the rubric either during or collaboratively with another SP/PETA after the encounter is finished, the task of accurately documenting what occurs (or does not occur) during the examination is burdensome. The rubric is an 88-item, paper-and-pencil checklist, and SPs/PETAs interact with several students in a sequence that may result in checklist fatigue. Similarly, because clinical faculty are required to view hours of videotape, mental fatigue with cognitive tasks may result. A considerable amount of time was spent in qualitative commentary that may not be necessary. On a practical summative examination, the focus is on skill competence, and errors in examination maneuvers may simply be identified by checking "correct, incorrect, or not done."

Of interest, there was no real difference between clinical faculty and IR scores. Therefore, bias on student performance/grading by the faculty member who knew the students found in previous studies (Miller et al., 1998) may not have occurred.

Although differences in scoring were noted among SPs/PETAs, clinical faculty, and IRs, the average scores on the rubric did not result in differences in the final outcome (pass/fail) on the examination. Clinical experience of the reviewer, familiarity with the rubric checklist, and checklist fatigue may be factors that influenced scoring and should be assessed with future studies.

Limitations

The major limitation to this study was a small sample size. In addition, no interrater reliability training for faculty or IRs was completed beforehand; however, all reviewers had access to the same checklist with specific notes on how to score many of the items. Not all reviewers evaluated every student; hence, no interrater reliability was evaluated. The focus of this study was on grading physical examination skills of the participant and did not include evaluation of the written note nor comparison of individual item scores across reviewers. Interpretation of qualitative data failed to provide multiple perspectives because of interpretation by only one reviewer.

Implications for practice

In conclusion, periodic review of summative subjective and objective assessment feedback is necessary to assure the quality of student evaluation in simulation testing modes. The results of this study suggest the opportunity exists to reduce time and increase efficiency associated with summative practical examination delivery. Improving reviewer training, reducing checklist length, and adopting electronic scoring may increase efficiency. Reviewing feedback provides the opportunity to benchmark outcomes for institutional effectiveness and pinpoint areas where students are struggling.

Significant costs, including time, are associated with training SPs/PETAs and faculty to properly rate students and recognize the appropriateness and correctness of physical examination maneuvers. In addition to current training of simulated participants/clinical faculty reviewers, development of an exemplary teaching video or "gold standard" (Park et al., 2016) to provide explanation and detail for expected student skill performance with associated best practice scoring will allow reviewers to practice and improve competence in reliable scoring. According to Miller et al. (1998), it is imperative to establish reliability in scoring among reviewers. Faculty training must occur (Miller et al., 1998).

Because average scores on the rubric resulted in the same passing outcome, the potential exists for clinical faculty to grade on a pass/fail consensus of competency rather than an assigned letter grade. When provided scoring guidelines and a scoring rubric, faculty are able to differentiate between excellent and nonpassing students (Park et al., 2016) and determine whether the student is competent or not (Clark, 2015).

Reducing checklist length may result in less time and effort by scorers to remember what was asked and what was done and may improve accuracy of scores. Items may need to be reduced to only those that are more clinically relevant and evidence-based (Daniels et al., 2014). It is suggested that, for accuracy when scoring by SPs/PETAs, the number of checklist items are somewhere between 12 and 15 (Vu et al., 1992). Because considerable time was expended by SPs/PETAs in writing qualitative comments, consideration should be made to eliminate the comments section. Qualitative feedback in that column is formative, rather than summative. In this final high-stakes summative examination, skills are rated with the goal of evaluating student learning at the end of the course compared with the standard benchmark. Although providing feedback comments may be appropriate during formative assignments earlier in the course, students may benefit from continuing comments as they progress through the remainder of clinical coursework. Qualitative comments may be very helpful for students to receive feedback from a patient point of view, including how an examination "felt" to the patient is important; a view not necessarily picked up by a faculty or outside reviewer.

The use of electronic rubrics and grading provide advantages of summing results, course reports, course statistics, and analytics to view how effective this testing strategy is. The audit of item analysis after an examination helps faculty decide whether to retain certain items for future use, revise them, or eliminate them from the checklist. Review of checklist analytics after each summative examination deployment is necessary to identify checklist items where course content should be revisited to reinforce learning with students. Student review of video feedback provides an opportunity for self-scoring and comparison to faculty grading as a teaching-learning tool (Miller et al., 1998). In addition, lower scoring items may indicate where students struggle or skills that were not taught or need more attention, or it may pinpoint areas where course faculty need further training on how to teach a specific examination skill.

Sometimes, to ensure reliability, more than one faculty member (Miller et al., 1998), assessor or double scorer may rate an experience (Park et al., 2016). Further studies should evaluate whether this is beneficial. Future research should also occur on evaluation of the written portions of the physical examination and rating of intra-rater and interrater reliability.

Conclusion

High-stakes clinical examinations using live hands-on simulation scenarios are effective learning/evaluation methods for NP students. Evidence-based guidelines and SoBP exist to guide educators in the development of quality student simulated learning experiences. Ongoing audit and feedback by faculty of examination processes improves the quality of student evaluation.

Part of this article was presented as a poster at the National Organization of Nurse Practitioner Faculties 46th Annual Conference in Chicago, IL (moved online per COVID 19 pandemic) in 2020 entitled: Standardizing nurse practitioner student evaluation during high-stakes clinical exams.

Authors' contributions: *M. Keiffer and M. Anderson developed the initial research project and applied for IRB. M. Keiffer, M. Anderson, D. O. Eckhoff and C. W. Blackwell participated in data collection and analysis. S. Talbert assisted with quantitative data analysis and A. M. Parker trained SPs for the exam. All authors contributed to writing the initial draft and revision of the manuscript.*

Competing interests: *M. Anderson has several non-related research grants, is VP of INACSL, is an Associate Editor for Simulation & Gaming, and serves on the editorial board for Simulation in Healthcare, along with being a simulation Program Director. The remaining authors report no conflicts of interest.*

References

Anderson, M., Holmes, T. L., Leflore, J. L., Nelson, K. A., & Jenkins, T. (2010). Standardized patients in educating student nurses: One school's experience. *Clinical Simulation in Nursing*, 6, e61–e66.

Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*, 68, 443–453.

Castro-Yuste, C., García-Cabanillas, M. J., Rodríguez-Cornejo, M. J., Carnicer-Fuentes, C., Paloma-Castro, O., & Moreno-Corral, L. J. (2018). A student assessment tool for standardized patient simulations (SAT-SPS): Psychometric analysis. *Nurse Education Today*, 64, 79–84.

Clark, C. A. (2015). Evaluating nurse practitioner students through objective structured clinical examination. *Nursing Education Perspectives*, 36, 53–54.

Daniels, V. J., Bordage, G., Gierl, M. J., & Yudkowsky, R. (2014). Effect of clinically discriminating, evidence-based checklist items on the reliability of scores from an internal medicine residency OSCE. *Advances in Health Science Education*, 19, 497–506.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5, 80–92.

INACSL Standards Committee. (2016). INACSL standards of best practice: Simulation[SM] simulation glossary. *Clinical Simulation in Nursing*, 12(S), S39–S47.

Johns Hopkins Medicine. (n.d.). *Physical Examination Teaching Associates.* https://www.hopkinsmedicine.org/simulation_center/-training/teaching_programs/physical_exam_teaching_associates.html.

Kardong-Edgren, S., Adamson, K. A., & Fitzgerald, C. (2010). A review of currently published evaluation instruments for human patient simulation. *Clinical Simulation in Nursing*, 6, e25–e35.

Kowitlawakul, Y., Chow, Y. L., Salam, Z. H., & Ignacio, J. (2015). Exploring the use of standardized patients for simulation-based learning in preparing advanced practice nurses. *Nurse Education Today*, 35, 894–899.

LaManna, J. B., Guido-Sanz, F., Anderson, M., Chase, S. K., Weiss, J. A., & Blackwell, C. W. (2019). Teaching diagnostic reasoning to advanced practice nurses: Positives and negatives. *Clinical Simulation in Nursing*, 26, 24–31.

Lewis, K. L., Bohnert, C. A., Gammon, W. L., Hölzer, H., Lyman, L., Smith, C., Thompson, T. M., Wallace, A., & Gliva-McConvery, G. (2017). The Association of Standardized Patient Educators (ASPE) standards of best practice (SOBP). *Advances in Simulation*, 2, 10.

Miller, A. M., Wilbur, J., Montgomery, A. C., & Talashek, M. L. (1998). Standardizing faculty evaluation of nurse practitioner students. *Clinical Excellence for Nurse Practitioners*, 2, 102–109.

National Organization of Nurse Practitioner Faculties. (2017). *NP Core Competencies Content.* http://www.nonpf.org/resource/resmgr/competencies/20170516_NPCoreCompsContentF.pdf.

Park, Y. S., Hyderi, A., Bordage, G., Xing, K., & Yudkowsky, R. (2016). Inter-rater reliability and generalizability of patient note scores using a scoring rubric based on the USMLE Step-2 CS format. *Advances in Health Sciences Education*, 21, 761–773.

Ryall, T., Judd, B. K., & Gordon, C. J. (2016). Simulation-based assessments in health professional education: A systematic review. *Journal of Multidisciplinary Healthcare*, 9, 69–82.

Schram, A. P., & Mudd, S. (2015). Implementing standardized patients within simulation in a nurse practitioner program. *Clinical Simulation in Nursing*, 11, 208–213.

Tavakol, M., & Pinner, G. (2018). Enhancing Objective Structured Clinical Examinations through visualisation of checklist scores and global rating scale. *International Journal of Medical Education*, 9, 132–136.

Tavares, W., & Eva, K. W. (2014). Impact of rating demands on rater-based assessments of clinical competence. *Education for Primary Care*, 25, 308–318.

Turner, R. T., Scerbo, M. W., Gliva-McConvey, G. A., & Wallace, A. M. (2016). Standardized patient encounters: Periodic versus post encounter evaluation of nontechnical clinical performance. *Simulation in Healthcare*, 11, 164–172.

University of Illinois College of Medicine at Urbana-Champaign. (n.d.). *Standardized patient policies and procedures manual UIUC College of Medicine*. https://www.med.illinois.edu/simcenter/sp/SP%20Policy%20and%20Procedure.pdf.

U.S. Department of Education. (2018). *Family educational rights and privacy act.* https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

Vu, N. V., Marcy, M. M., Colliver, J. A., Verhulst, S. J., Travis, T. A., & Barrows, H. S. (1992). Standardized (simulated) patients' accuracy in recording clinical performance check-list items. *Medical Education*, 26, 99–104.

Yudkowsky, R., Downing, S., Klamen, D., Valaski, M., Eulenberg, B., & Popa, M. (2004). Assessing the head-to-toe physical examination skills of medical students. *Medical Teacher*, 26, 415–419.

Yudkowsky, R. (2020). Performance tests. In R. Yudkowsky, Y. S. Park, & S. M. Downing (Eds.), *Assessment in health professions education* (pp. 141–159). (2nd ed.). Routledge.