ORIGINAL REPORTS

Nonspecialist Raters Can Provide Reliable Assessments of Procedural Skills

Oria Mahmood, BMSc,[*,†] *Julia Dagnæs, BMSc,*[*,†,‡] *Sarah Bube, MD,*[†,§] *Malene Rohrsted, MD,*[‡] *and Lars Konge, MD, PhD*[*,†]

[*]Copenhagen Academy for Medical Education and Simulation, Copenhagen, Capital Region, Denmark; [†]University of Copenhagen, Copenhagen, Denmark; [‡]Department of Urology, Rigshospitalet, Copenhagen, Denmark; and [§]Department of Urology, University Hospital Zealand, Roskilde, Denmark

BACKGROUND: Competency-based learning has become a crucial component in medical education. Despite the advantages of competency-based learning, there are still challenges that need to be addressed. Currently, the common perception is that specialist assessment is needed for evaluating procedural skills which is difficult owing to the limited availability of faculty time. The aim of this study was to explore the validity of assessments of video recorded procedures performed by nonspecialist raters.

METHODS: This study was a blinded observational trial. Twenty-three novices (senior medical students) and 9 experienced doctors were video recorded while each performing 2 flexible cystoscopies on patients. The recordings were anonymized and placed in random order and then rated by 2 experienced cystoscopists (specialist raters) and 2 medical students (nonspecialist raters). Flexible cystoscopy was chosen as it is a simple procedural skill that is crucial to master in a resident urology program.

RESULTS: The internal consistency of assessments was high, Cronbach's $\alpha = 0.93$ and 0.95 for nonspecialist and specialist raters, respectively ($p < 0.001$ for both correlations). The interrater reliability was significant ($p < 0.001$) with a Pearson's correlation of 0.77 for the nonspecialists and 0.75 for the specialists. The test-retest reliability showed the biggest difference between the 2 groups, 0.59 and 0.38 for the nonspecialist raters and the specialist raters, respectively ($p < 0.001$).

CONCLUSION: Our study suggests that nonspecialist raters can provide reliable and valid assessments of video recorded cystoscopies. This could make mastery learning and competency-based education more feasible. (J Surg Ed ▮:▮▮▮-▮▮▮. © 2017 Published by Elsevier Inc. on behalf of the Association of Program Directors in Surgery)

KEY WORDS: competency-based learning, procedural skills, reliable assessments, nonspecialist raters, rater training, rater competencies

COMPETENCIES: Practice-Based Learning and Improvement, Medical Knowledge

INTRODUCTION

Medical education is changing rapidly, and the way doctors train procedural skills is shifting from traditional apprenticeship and time-based learning to competency-based attainments of skills.[1,2] Competency-based learning is being favored as it implies that trainees will pass when they are competent and not after a certain prescribed time or when a certain number of procedures have been performed, which does not necessarily reflect competence.[3-6] Competency-based learning require specialist assessment of procedural skills. Despite the advantages of competency-based learning, there are still challenges that need to be addressed, e.g., the limited availability of faculty time.[7-10] Some studies also suggest that knowing the identity of the trainee can influence assessment.[11,12] Technology holds some promise, as video recordings of performances create more flexibility and reduce the risk of bias.[13-15]

Studies show that rater training is beneficial and even suggest that 1-hour frame-of-reference training sessions are able to sufficiently train raters to use a simple evaluation instrument for the assessment of procedural skills.[16,17] Studies have shown that medical students can be used in teaching settings instead of professors,[18-20] and this could be translated to competency-based assessment where the use of

Correspondence: Inquiries to Oria Mahmood, BMsc, Copenhagen Academy for Medical Education and Simulation, C. F. Richs vej 116 street th, 2000 Frederiksberg, Copenhagen, Capital Region, Denmark; e-mail: Oriamahmood@gmail.com

nonspecialist raters could be implemented to further reduce the time spent by specialists on assessment. The common perception is currently that specialists need to assess procedural skills, but previous studies have shown that even nonmedically trained individuals can be used to assess surgical skills.[21,22]

Using nonspecialist raters would not only decrease the workload of specialists and minimize interpersonal bias but also provide a more beneficial economical solution in areas where competency-based assessment is needed. The use of nonspecialist raters needs to be proven reliable and valid before it can be implemented as part of competency-based learning assessments.

The aim of this study was to explore the validity of assessments of video recorded procedures performed by nonspecialist raters.

MATERIAL AND METHODS

Design

This study was a blinded observational trial. Novices (senior medical students) and experienced doctors were video recorded while performing 2 flexible cystoscopies each. The recordings were anonymized and placed in random order and then rated by 2 experienced cystoscopists (specialist raters) and 2 medical students (nonspecialist raters). Flexible cystoscopy was chosen as it is a simple procedural skill that is crucial to master in a resident urology program.[23]

Participants

The novices participating in this study were senior medical students who had completed their fourth year of medical school from the University of Copenhagen. They had no experience with flexible cystoscopy, and therefore practiced the procedure at the Simulation Center before performing the procedure on patients.[24] The experienced doctors were all urologists and had performed more than 100 flexible cystoscopies within the last year. Every participant performed 2 video recorded cystoscopies on different patients.

Recordings

The videos were recorded directly through the lens of a flexible cystoscope, and the participants were not identifiable and no sound was recorded. The recordings began at the introduction of the scope in the bladder and stopped when the scope was removed through the urethra. All recordings were anonymized and placed in random order using random.org, and then put into an integrable web-based solution for easy assessment,[13] and thereafter rated by the 2 pairs of raters.

Raters and Assessment Tool

In this study, we had 2 pairs of raters. One pair of experienced raters had performed more than 100 flexible cystoscopies. The other pair was the nonspecialist raters who had limited experience with flexible cystoscopy and assessments of procedural skills. The nonspecialist raters were recruited through their work at the simulation center[24] and were both senior medical students who had completed their third year of medical school. The 2 pairs of raters (nonspecialists and specialists on separate occasions) met and individually watched and assessed 2 pilot videos, then they compared the results and discussed their ratings. After the assessment of the first 2 pilot videos, no further communication was allowed between the raters.

Both pairs of raters used the same slightly modified assessment tool based on global rating scale (GRS) with previously established evidence of validity.[23] The assessment tool is composed of the following 5 different parameters: respect for tissue, time and motion, handling of endoscope, flow of procedure, forward planning, and knowledge of procedure. For each parameter, a score from 1 to 5 was given, 1 being the poorest and 5 the best (Appendix 1).

Statistics

Internal consistency reliability for the assessments was determined by calculating Cronbach's α for both specialist and nonspecialist raters. Pearson's correlation was used to estimate the interrater reliability for both pairs of raters and used to determine the test-retest reliability.[25,26] We established the pass/fail criteria using the contrasting group's method based on the specialist raters' assessments and explored the consequences of using nonspecialist raters. We used statistical software SPSS version 22.0 (IBM, Armonk, NY).

Ethical Considerations

All patients were already scheduled for a cystoscopy at the outpatient clinic of the Urology Department at Rigshospitalet. A urology specialist supervised all procedures. All participants had provided written consent before participation.

RESULTS

Twenty-three novices and 9 experienced doctors participated in this study giving a total of 64 video recordings all rated by a pair of specialist raters and a pair of nonspecialist raters (256 ratings in total). The internal consistency of assessments was high, Cronbach's $\alpha = 0.93$ and 0.95 for nonspecialist and specialist raters, respectively ($p < 0.001$ for both correlations). The interrater reliability was significant ($p < 0.001$) with a Pearson's correlation of 0.77 for