Face, Content, and Construct Validation of the Bristol TURP Trainer

James Brewin, MSc, FRCS(urol), Kamran Ahmed, MRCS, PhD, Muhammed S. Khan, FRCS(urol), FEBU, Peter Jaye, BSc, MRCP, and Prokar Dasgupta, MSc, MD, FRCS(urol), FEBU

Kings Health Partners, Guy’s Kings and St Thomas’ NHS Foundation Trust, London, UK

INTRODUCTION: Validation studies are an important part of simulator evaluation and are considered necessary to establish the effectiveness of simulation-based training. The widely used Bristol transurethral resection of prostate (TURP) simulator has not been formally validated.

OBJECTIVES: Evaluation of the face, content, and construct validities of the Bristol TURP simulator as an endourology training tool.

DESIGN: Using established validation methodology, face, content, and construct validities were evaluated. Face and content validities were assessed using a structured quantitative survey. Construct validity was assessed by comparing the performance of experts and novices using a validated performance scale and resection efficiency.

PARTICIPANTS AND SETTING: Overall, 8 novice urologists and 8 expert urologists participated in the study. The study was conducted in a dedicated surgical simulation training facility.

RESULTS: All 16 participants felt the model was a good training tool and should be used as an essential part of urology training (face validity). Content validity evaluation showed that most aspects of the simulator were adequately realistic (mean Likert scores 3.38-3.57/5); however, the model does not simulate bleeding. Experts significantly outperformed novices (p < 0.001) across all measures of performance, therefore establishing construct validity.

CONCLUSIONS: The Bristol TURP simulator shows face, content, and construct validities, although some aspects of the simulator were not very realistic (e.g., bleeding). This study provides evidence for the continuing use of this simulator in endourology training.

KEY WORDS: simulation, urology, education, transurethral resection of the prostate, training

COMPETENCIES: Patient Care, Practice-Based Learning and Improvement

INTRODUCTION

Transurethral resection of the prostate (TURP) is considered the gold standard surgical procedure for the management of bladder outlet obstruction secondary to benign prostatic hyperplasia, and it remains one of the index procedures to determine core competency for urology trainees. Learning to perform a TURP is challenging; trainee urologists are therefore faced with having to become competent in an operation with a long learning curve, but for numerous reasons, they are currently performing fewer operations than earlier cohorts of trainees.

In response to changes in training, and a need to optimize patient safety, there has been a trend toward using simulation as an adjunct to training in the operating room (OR). To shorten the learning curve, simulators must be realistic and teach the skills needed in the OR. They must therefore be valid, reliable, and relevant. Several virtual reality TURP simulators have been described, and several studies assessing validity of these simulators have been published (Table 1). However, the most commonly used TURP simulators are physical bench-top prostate models such as the Bristol TURP trainer (Limbs and Things Ltd., Bristol). This has never been formally assessed or validated in accordance with educational principles.

Validity is defined as “the property of being true, correct and in conformity with reality” and is considered an essential part of simulator assessment. For a simulator to be a valid training tool, it must teach the skills required in the OR. Validation is an ongoing process and definitions of validity vary in the literature; however, a number of validation benchmarks have been established and applied...
to the assessment of surgical simulators. Face and content validities are subjective assessments of how realistic and useful a simulator appears and are usually assessed with structured questionnaires. Face validity evaluates the overall impression of whether a simulator tests or teaches what it was designed to test or teach. Content validity is a more detailed examination of the individual components of a simulator.

Construct validity is defined as “a set of procedures for evaluating a testing instrument based on the degree to which the test items identify the quality, ability, or trait it was designed to measure.” It is usually assessed by establishing whether a simulator can differentiate between experts and novices and is considered mandatory before adoption of a simulator for training or assessment. Several published studies have evaluated face and content validities of virtual reality TURP simulators but only a few have also assessed construct validity (Table 1).

The aim of this study was to assess the face, content, and construct validity of the Bristol TURP trainer.

METHODS

Study Design

Mixed methods using established validation methodology were used. A qualitative questionnaire was used to evaluate face and content validities and an experimental design to evaluate the performance of experts and novice groups was used to evaluate construct validity.

Study Participants

Overall, 8 urology trainees (<10 TURPs, median 1.6) and 8 expert urologists (>50 TURPs, median 134) performed a simulated TURP using the Bristol TURP trainer. All participants gave written consent. All participants completed a single simulated TURP on the training model. Following the simulation, participants completed structured questionnaires assessing face validity; the experts completed an additional structured questionnaire to assess content validity.

The Simulator

The Bristol TURP simulator is a synthetic prostate model within a latex bladder mounted on a plastic base (Fig. 1). The simulator is used with a real resectoscope and monopolar or bipolar diathermy with irrigation to resect the physical prostate model. All study participants performed the resection using monopolar diathermy using a digital camera head (Olympus). The prostate model includes the bladder base, ureteric orifices, a verumontanum and the prostate capsule (colored red); it is made from a synthetic conductive material that can be cut with a diathermy loop.

Outcome Parameters

Surgical performance was evaluated by measuring resection efficiency, and by using scores from structured observational assessment tools. Resection efficiency was calculated by dividing the total weight of the resected model (chip weight...
in grams) by the total operating time. The validated Objective Structured Assessment of Technical Skills (OSATS—min 0, max 35) and a global rating scale (GRS—min 1, max 5) were used by expert observers in real time. The global rating scale was adapted from the compulsory work-based assessment tools used by UK urology trainees (www.iscp.ac.uk). Owing to technical error, only 5 video recordings from experts could be analyzed; therefore a sample of 10 video recordings (5 trainees and 5 experts) was analyzed by a blinded expert urologist (SK, >1000 TURPs) to reduce bias.

**Statistical Analysis**

To compare the performance of the expert and novice groups, nonparametric analyses were performed using Statistical Package for the Social Sciences version 16; a 2-tailed significance level of p < 0.05 was considered significant.

**RESULTS**

**Face Validity**

Experts and trainees felt that the simulator should be used as part of the urology training curriculum (median Likert score of 5/5, range 3-5) and that it was a useful training tool for specialty trainees (median Likert score 5/5, range 4-5). A Likert score of 4/5 has been reported to be adequate to demonstrate face validity and consequently our study demonstrates face validity of this model as a training tool. Although most experts and novices felt the model could be used for formal assessment (median 4/5), there was disagreement amongst raters (range 2-5).

**Content Validity**

Content validity was assessed by the 8 experts using qualitative surveys. Median Likert scores varied from 3 to 4 with the distribution of ratings shown in Table 2. Although some experts felt that aspects of the model were acceptable (4 or 5 on the Likert scale), there was considerable variability in expert rating of content validity with some aspects of the simulator not appearing realistic. Furthermore, the model does not bleed.

**Construct Validity**

Experts performed significantly better than trainees in all measures of assessment including the OSATS (median 26.3 vs 14.5, p < 0.001), the GRS (median 5.0 vs 2.0, p < 0.001) and resection efficiency (median 2.19 vs 0.90, p < 0.001).
p < 0.001) (Fig. 2). Blinded video assessment of performance using the OSATS correlated well with real-time analysis by the nonblinded observers (Pearson r = 0.85, p < 0.01). Therefore experts outperformed trainees across all subjective and objective measures of assessment, and this has clearly established construct validity.

**DISCUSSION**

To our knowledge, this is the first study to evaluate a physical model TURP trainer, and it has clearly established face and construct validities. Most aspects of the simulator were considered realistic in the content validation assessment, however there were aspects of the model that were not considered particularly realistic. Additionally the model does not bleed. Despite these limitations, the study participants felt the model was a useful training tool and should be used as part of the urology curriculum.

There is debate in the surgical and educational literature regarding the level of fidelity (realism) that is required to make a simulator a useful training tool. Several studies have found that training on low-fidelity models can improve real OR performance, and one study has shown that a low-fidelity training on a high fidelity model. It is likely that the level of simulator fidelity required is dependent on the task being taught and on the context of the training. The results of this study suggest that although some aspects of the model are not realistic, it is still useful for training in the early part of the learning curve—further research is needed to evaluate how realism affects surgical training and to examine how best to integrate different simulators into the surgical curriculum.

Construct validity is considered mandatory before using a simulator as an assessment tool and is desirable before using it for training. Taking all measures of performance together, there is a clear difference between expert and novice performances, and this study has therefore established construct validity. The results of the study are strengthened by the use of blinded video moderation.

Accurate assessment to provide feedback is an important part of training as it can focus learning and help the learner to undertake “deliberate practice” to continually improve. This study shows that OSATS, the GRS, and resection efficiency are all valid measures that could be used to provide feedback and guide training. However, this should not be at the expense of supervision by experienced trainers—qualitative feedback from trainees highlighted the importance of expert supervision during simulation training. These performance measures could also be used to designing performance-based training—ideally, trainees would have to reach a set standard before operating on real patients rather than only completing a given number of simulator training hours.

This study provides evidence to support the continued use of the Bristol TURP trainer. However, the advantages and disadvantages of using a given simulator for training must be considered by individual trainers and adapted to individual training programs. As discussed before, there are commercially available virtual reality (VR) TURP simulators (Table 1), and trainers must decide on whether to invest in one of these or use a physical model. Advantages of using a physical model include the use of real instruments and irrigation to resect the prostate, therefore teaching the trainee how to use the equipment. Furthermore, using real equipment and a model means that the visuospatial environment is realistic and does not rely on a computer to try and simulate instrument movement within a virtual 3-D environment. However, virtual reality simulators are becoming more realistic, they are able to simulate bleeding, can automatically measure performance, and can simulate prostates of different sizes and shapes. Several VR simulators have been evaluated in the surgical literature although only a few have assessed construct validity (Table 1). When deciding on a simulator, trainers must take into account the available evidence as well as considering how a given simulator fits into a training curriculum.

Costs must also be considered when deciding on a simulation program. VR simulators are very expensive to purchase but have minimal running costs. The Bristol TURP trainer is much cheaper than currently available VR simulators ($1161 at the time of writing, http://limbsandthings.com/uk/products/bristol-turp-trainer/), but the individual prostate models are single use and cost $117 each. In our study, trainees got between 45 and 75 minutes of use from each prostate model. Using a VR simulator, Kallstrom et al. found that trainees required a mean simulator training time of 254 min to reach proficiency, and this would roughly translate into the use of 4.2 models per trainee at a cost of $495.

**Limitations**

Although this study evaluated face, content and construct validities, it has not assessed whether simulator training results in improved OR performance, and further research is needed. If the simulator is to be considered as a tool for high-stakes assessment, further research with more participants would be needed to evaluate the reliability of the performance measures and compare this with real-life performance.

**CONCLUSION**

This study provides evidence to support the continuing use of the Bristol TURP simulator in urology training as it clearly demonstrates face and construct validities. Although some aspects of the model were not very realistic, it was considered a good training tool. Several metrics of evaluation including the OSATS, the GRS, and resection efficiency appear to be good measures of performance, which
could be used to provide feedback and develop performance-based training curricular.

ACKNOWLEDGMENTS

We would like to thank Olympus, The School of Surgery, and SIMULATE for providing the simulation facilities that made this study possible. Prokar Dasgupta received financial support from the Department of Health via the National Institute for Health Research comprehensive Biomedical Research Centre award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with KCL and KCH NHS Foundation Trust. He also acknowledges the support of the MRC Centre for Transplantation.

REFERENCES


2. Cury J, Coelho RF, Bruschiini H, Srougi M. Is the ability to perform transurethral resection of the prostate influenced by the surgeon’s previous experience? *Clinics (Sao Paulo)*. 2008;63(3):315-320.


**SUPPORTING INFORMATION**

Supplementary material cited in this article is available online at doi:10.1016/j.jsurg.2014.01.013.