The development of 'SATLAB': A tool designed to limit bias and improve feedback in simulation-based assessment

Andrew Makkink*, Craig Vincent-Lambert

Department of Emergency Medical Care, Faculty of Health Sciences, University of Johannesburg

*Corresponding author: amakkink@uj.ac.za, P O Box 524, Auckland Park, 2006

ARTICLE INFORMATION

Received: 22 June 2018
Revision received: 23 July 2018
Accepted: 21 August 2018

Keywords:
assessment
assessor bias
feedback
medical simulation

ABSTRACT

Context and Setting: The development of clinical competence remains a challenge in healthcare education. The reliable assessment of competence requires new approaches to address the perceived limitations of current assessment practices. The Simulation Assessment Tool Limiting Assessment Bias (SATLAB) seeks to provide a novel assessment strategy within the simulation environment.
Concept: The SATLAB is a simulation assessment tool conceptualised by Andrew Makkink as a means to address many of the perceived issues prevalent in simulation assessment.
Implementation: The SATLAB has been in use at the University of Johannesburg since its inception and is currently also used at several other institutions offering simulation-based learning and assessment in emergency medical care.
Impact: The SATLAB was conceptualised as a potential solution to some of the inherent challenges in the assessment of simulation. The current use and ongoing evaluation of the SATLAB is providing exciting insights into the future of simulation assessment. Further research is required to determine the reliability and validity of the tool within the simulation assessment domain.

Makkink A, Vincent-Lambert C. The development of 'SATLAB': A tool designed to limit bias and improve feedback in simulation-based assessment. South African Journal of Pre-hospital Emergency Care. 2020; 1(1):26-34. doi:10.24213/1-1-3024

BACKGROUND

The development of clinical competence in healthcare education remains a challenge for educators. The competent healthcare provider is required to integrate theoretical principles, psychomotor skills and clinical reasoning to diagnose and treat patients. However, a relative disorder exists in the clinical environment, which compounds the teaching and assessment of clinical competency. The Simulation Assessment Tool Limiting Assessor Bias (SATLAB) seeks to address both the assessment of and assessment for learning paradigms. The generation of a percentage provides assessment of learning and the feedback strategies provide the assessment for future learning.

High-fidelity patient simulation is used within the healthcare environment to better facilitate teaching, learning and assessment.[1] High-fidelity patient simulations relate to the use of a person, device or set of conditions with the aim of presenting authentic problems within a simulated environment and is used to emulate anatomical areas, clinical tasks, or sick or injured patients.[2] Some advantages that have been linked to the use of human patient simulation are listed below.

Advantages of Human Patient Simulation

- The ability to provide real-life experiences that may only occur rarely or not be possible to recreate in the real-world.[2]

- Decreased risks for patients.[3]
- Increased incidence of specific patient availability[3] and decreased expense.[4]
- The ability for the educator to control and direct learning experiences towards specific cognitive outcomes and competencies.[2]
- The ability to standardise assessments, resulting in consistency in clinical case replication.[3]

High-fidelity patient simulation is also associated with some disadvantages or limitations, as listed below. Within the context of assessment in simulation, there is a lack of objective and valid measurement tools for assessing clinical competence. This has the potential to negatively impact patient safety if assessment systems lack reliability and validity.

Disadvantages and Limitations of Human Patient Simulation

- There is a lack of realism. This includes areas related to the feel, colour and temperature of the skin, as well as the environment in which the simulation takes place.[3]
- There is a fear that (simulation) technology will dehumanise health care.[3]
- There is a risk of overreliance on simulation as a replacement for actual clinical experience.[3]
- Mechanical breakdown of manikins is a constant risk.[3]
- Cost remains a significant area of concern. Manikins, equipment, staff training and maintenance remain cost concerns.[3]
- A perceived inability to establish congruent findings in simulation research.[3]
- The limited pool of appropriately trained academic staff.[3]
- Students perceive a higher anxiety level during simulation than they would in the real clinical environment.[3]

The terms 'examiner' and 'assessor' are often used interchangeably, and the specific roles are difficult to separate. To clarify the roles specified within this concept article, we define each role in Table 1. The terms 'marker' and 'expert panel member' are also introduced to provide a clear differentiation of the assessor/examiner conundrum.

Table 1: Definitions

Term	**Definition**
Assessor	The person responsible for setting the assessment, determining competency outcomes and compiling the competency outcome descriptors.
Expert Panel Member	The person(s) responsible for determining the individual competency outcome weightings, reviewing the simulation for validity and assessing the relevance of competency outcome descriptors prior to the simulation being administered.
Marker	The person using the SATLAB to determine student competency levels by assigning scores as determined by the assessment rubric.

The developers of the SATLAB, both members of the Department of Emergency Medical Care at the University of Johannesburg, evaluated generic simulation assessment practice by performing a literature search as described in the concept design. In addition, we reviewed our own practices and identified the following potential shortcomings in ourselves, along with traditional methods of assessment in high-fidelity patient simulations. We critically analysed our practices to determine possible areas of weakness. The combination of these processes led us to conclude that the following were the most pertinent issues:

- Marker bias - a clinician's personal experiences affect the way in which they view the criteria for competence. This has the potential to negatively affect the inter-rater reliability of marker ratings related to the scoring of a student's performance.
- A perceived focus on negative feedback - the defensibility of a fail has often been the motivator for a marker to write copious notes related to what a student did wrong. The opposite has generally not been the case in a student who has done well.
- Disparity between assessor and marker - the assessor is often involved in teaching the students being assessed. Conversely, the marker

is usually unfamiliar with what has been taught and uses their own frame of reference to determine a measure of competency.

- Lack of specificity when using Global Rating Scores (GRS) - the GRS asks the marker to provide a global impression of the student's overall performance. This is usually depicted as a percentage and, in the case of the GRS, the final percentage or result fails to specify in which specific areas a student has excelled or performed poorly. In addition, differences in marker scores are difficult to contextualise where a significant deviation exists between markers.
- Lack of feedback or debriefing information - the GRS used in isolation and its perceived focus on negative comments generally failed to identify specific areas of excellence or areas requiring improvement within the simulation. This limited feedback and hampered appropriate debriefing on a student's performance.
- There has traditionally been a lack of detail related to how competency levels were determined because simulation assessment criteria were often driven by a 'tick box list of things to do'. This list generally failed to provide descriptions of specific outcome competency levels. This meant that students were unable to correct their mistakes as there was no detailed marking guide against which to compare their performance.

The above factors had the potential to negatively affect the identification of transfer of learning from the classroom to the simulation, and subsequently to practice. Transfer of learning has been described as "the learning process involved when a person learns to use previously acquired knowledge/skills/ competence/expertise in a new situation".[5] Simulation assessment seeks to determine whether or not transfer of learning has taken place by simulating a real-world problem that the student is required to manage. Traditionally, the problem has been that simulation assessment did not adequately identify where transfer of learning had, or had not, occurred.

The importance of feedback from simulation, especially simulation assessment, was an area we identified as a particular one for improvement. For the student to remediate areas where they had underperformed, they would need to know where and why their performance was not adequate. In other words, the identification of what had been learned and what had not formed the basis for remediation. If there was inadequate information related to what the student had done right, or how to appropriately fix what they had done wrong, the process of remediation would be hampered. The student would not know what knowledge, skills or competencies to use in the future and what to discard as incorrectly learned. This would negatively affect the concept of assessment for learning.

The use of human patient simulation within assessment is an area that requires investigation, and solutions need to be sought for the perceived lack of reliability and validity in simulation assessment within emergency medical care. The SATLAB seeks to address many of the issues associated with simulation assessment. This article seeks to explain how SATLAB functions and how the potential exists to improve simulation assessment.

SATLAB DESIGN AND METHODS

The two authors began by critically reflecting on their own experiences of simulation assessment. This included reviewing current simulation practice within our own Department as well as that of other institutions and faculties where we had additional exposure. We performed a desktop study which focused on the characteristics of assessment processes, assessment criteria, assessor bias, feedback and debriefing. We used the UJoogle search engine (© Innovative Interfaces, Inc. Emeryville, CA) as our primary search tool. UJoogle is a search engine that uses federated search technology and performs simultaneous searching of multiple library sources and databases. We used a number of terms to guide our search, including 'medical simulation', 'simulation assessment', 'assessor bias', and various combinations of these and similar terms. Article titles identified by the search were read and classified for relevance. Where articles seemed appropriate, these were downloaded and read. We consolidated our readings and this served to guide our development of SATLAB.

DEVELOPMENT OF THE SATLAB

Our desktop study yielded several interesting results. We identified specific areas within simulation assessment that we believed we could address using the SATLAB, which employs a sequential process to compile, weight, calculate and provide feedback within simulation assessment. These areas/issues were:

- Reliability
- Fairness
- Validity
- Feedback and Debriefing
- Ease of Use

The SATLAB takes as its starting point the premise that there are a number of competency outcomes that make up the framework of the total simulation experience. The SATLAB seeks to provide detailed descriptors for each competency outcome for both students' and markers' use. Students use the descriptors to analyse the standards for best practice and to compare these to their performance post-assessment. Markers use the descriptors to compare student performance with a clear descriptor of what the assessor requires. These descriptors should be closely linked to the learning outcomes to which the students have been exposed. The SATLAB model assigns weightings to each competency outcome to reflect its individual importance within the context of the specific simulation. The SATLAB seeks to mitigate individual preferences by using a panel of experts to determine the appropriate weighting of each competency outcome. This panel of experts also serves to scrutinise the simulation and its specified outcomes; this sequential process involves six basic steps:

1. Design the simulation.
2. Determine competency outcomes and descriptors.
3. Distribute to panel of experts for review and weighting of competence criteria.
4. Conduct the simulation assessment.
5. Collate and calculate the results.
6. Provide appropriate feedback.

Design the Simulation

The assessor designs a simulation of appropriate complexity with the aim of determining student competence in managing a specific patient. The specific outcomes required for competence are determined and the simulation is designed around meeting these outcomes. The assessor is often required to evaluate the fidelity limitations that may exist for the specific simulation. These limitations include manikin and environmental fidelity, as well as factors related to the degree of exactness when compared to the real-life events that the simulation aims to recreate. The assessor carefully aligns the simulation outcomes to those within the curriculum.

Determine Competency Outcomes and Descriptors

The design phase of the simulation includes establishing the appropriate number of competency outcomes to be assessed. These competency outcomes are simulation dependent and focus on the core areas of patient assessment and management for the specific condition or injury. The number of outcomes is determined by the competency outcomes being assessed and the need to distinguish between specific areas of competence. While we have not yet been able to determine an ideal number, our current experience suggests that the appropriate number of competency outcomes appears to be between five and seven. The question of too many vs too few must be evaluated on a case-for-case basis. Too many competency outcomes may result in markers being overwhelmed by the number of outcomes, whereas too few competency outcomes may not produce a broad enough assessment of overall competence.

The assessor compiles a list of descriptors for each competency outcome. These descriptors are aligned to a six-point competency scale and are designed to guide markers towards using consistent criteria to determine competence levels. This six-point scale is summarised in Table 2.

Table 2: Six-point Competency Scoring Scale

Score	**Description**
3 (100%); Best Practice	The competency outcome is performed perfectly.
2 (67%); Competent	The competency outcome is performed at a level where minimum competency has been demonstrated.
1 (33%); Not Yet Competent	The competency outcome is performed at a level that fails to demonstrate adequate competence
0 (0%); Omitted	The competency outcome is omitted. It is possible that omission may be considered harmful.
-1 (-33%); Some Harm Caused	The way in which the competency outcome is performed results in some harm to the patient.
-2 (-66%); Significant Harm Caused	The way in which the competency outcome is performed results in significant, or life-threatening, harm to the patient.

The six-point scale was chosen due to its simplicity and the association that we are able to make with relation to linking each outcome to a competency descriptor. The process of dividing the simulation into several specified descriptors means that a student can be appropriately rewarded for competency outcomes that they perform correctly. This is not dissimilar to a written assessment where marks are awarded per question. In fact, each competency outcome is actually asking a specific competency question of the student.

The negative scoring aims to appropriately penalise actions that result in patient harm. The negative scoring ultimately results in marks being subtracted for harmful practice; this process is discussed in more detail under the calculation of the results section. In the rare event that a result of less than zero is achieved, the final result remains zero. It is not possible to achieve less than zero as a final result for an assessment that uses SATLAB.

Distribute to Panel of Experts for Review and Weighting of Competence Criteria

The simulation and competency outcome descriptors are sent to the panel of experts. The members of the expert panel should be carefully selected based on criteria such as their experience in simulation assessment, their clinical currency and their familiarity with the SATLAB system. Our current experience indicates that a panel of five experts is adequate. This ensures an adequate spread of views and opinions related to the simulation. Each member of the panel reviews the simulation and provides comments, if necessary, to the assessor. These comments may include aspects related to fidelity, simulation progression, realism or clarity of the competency descriptors. In addition, the panel scrutinises the competency outcome descriptors to ensure they contain sufficient detail and are not ambiguous.

The members of the panel of experts are also tasked with determining the percentage weightings of each competency descriptor. Each competency descriptor contributes a certain percentage to the final result. The assumption is that more critical competency outcomes will carry a higher weighting. These weightings are based on currently accepted evidence-based protocols, guidelines and professional standards. Each expert weights blindly, without any influence from other members of the panel, to ensure that each expert's perceived percentage is not unduly influenced by personal interactions with other members of the expert panel. This contributes to the overall reliability and validity of the simulation result as it limits the effect that one person can have on the final outcome. Final weightings are returned to the assessor who calculates the mean of the submitted weightings for each competency outcome. This calculated percentage represents the weighting of each competency outcome within the context of the simulation. Where significant differences are noted, the individual members of the expert panel can be contacted to determine the motivations for their specific weightings.

Conduct Simulation Assessment

Post review by the panel of experts, the simulation assessment is conducted in an appropriate setting. The importance of conducting a reliable simulation process cannot be overemphasised and should include adequate equipment and venue preparation, control of student movement, time allocation and adherence to set time limits.[6] It is imperative that markers use the competency outcome descriptors as the basis for awarding results for each competency outcome. This creates a synthesis between real-life, the classroom and the assessment. In other words, the markers are comparing student performance with competency outcome descriptors supplied by the assessor. These competency outcome descriptors used in the assessment are therefore directly linked to real-life, to what students were taught in the classroom, and to the simulation itself. In other words, competence is not measured by the marker's personal opinions of what they believe competency to be but rather by structured and consistent criteria.

Collate and Calculate Results

Results from the assessment are collected, and marker scores are used to calculate the student results. An example of this is depicted in Table 3. Each competency outcome makes up a proportion of the total mark for the simulation. The results for each competency outcome are calculated by converting the result (on the six-point scale) to a percentage (by dividing by 3) and multiplying this by the weighting for that specific competency descriptor. The final result is the sum of the individual competency outcomes. Where a negative result has been recorded, the same calculation principles apply, yet a negative result will be produced. This negative mark is used in the final result calculation as is and serves to provide a realistic representation of the effects of a harmful action. In other words, where a student does significant harm, the result should indicate this by reflecting a fail mark.

The process of addressing inter-rater reliability involves comparing the student's results between markers. Where the two markers differ significantly in the final percentage result, an inter-rater comparison of individual competency outcome scores is conducted to determine the possible source of the difference. If the results are such that there are significant differences for a specific competency outcome, this outcome can be reviewed through a moderation process. This moderation involves a review of that specific outcome and a careful comparison between the competency outcome descriptor and the student's performance. It is also possible that despite the detailed descriptors, one marker may be stricter or more lenient than the other. Carefully constructed descriptors limit interpretation and significantly reduce inter-rater differences. Comparative analysis of inter-rater score identify outcome descriptors that may require revision; we consider small, consistent differences as normal. In the event of a significant difference, the individual competency outcome scores and final result are reviewed as a whole. This may or may not result in the need for post-assessment moderation and should be assessed on a case-for-case basis.

Table 3: Final Result Calculation Example

Competency Outcome	**Weighting**	**Mark Out of 6 (%)**	**Final Results**
A	25%	2 (66%)	16.7%
B	30%	3 (100%)	30%
C	45%	1 (33%)	15%
Total	100%		61.7%

Provide Appropriate Feedback

Feedback is an essential component of any assessment. The challenge with assessment is that there are multiple possible permutations of student performance. The focus of simulation feedback should be on guiding students towards best practice. The SATLAB provides students with the same outcome competency descriptors that were used by markers to rate their performance. This creates congruency between assessment expectations, student performance scoring and feedback. This feedback allows students to carefully analyse their own performance and to compare this with the requirements for each competency outcome.

Where appropriate, the marker can comment on specific areas of a student's performance; we do not encourage extensive note-taking as this may distract the marker from watching the student perform. The student is then able to integrate the actions described in their future practice and, in doing so, improve their performance. The facilitator and assessor are able to use the feedback generated from the SATLAB to formatively identify specific areas that may require review, and use these in debriefing sessions. The SATLAB results would guide the facilitator to specific areas where group performance was poor and would assist in identifying any concerning trends.

Individualised feedback and debriefing may be considered impractical due to both lecturer and student time constraints. The SATLAB provides feedback to students by way of the mean, unweighted percentage results for each outcome assessed, depicted in Table 3. This percentage can be seen as a measure of competency where 100% represents best practice (perfect competence) for that specific competency outcome. Any other result is a measure of how close to perfect competency the student's performance was. Students can then identify specific areas of excellence as well as those requiring improvement.

It is essential that feedback is comprehensive enough for students to develop improvement strategies aimed at achieving improved results. Students are given the identical outcome competency descriptors that were used by the markers to determine their result. This means that students are able to compare their performance with the specified outcomes and identify areas of improvement. The student is also able to analyse the criteria required for best practice and adapt their practice to meet these criteria. The descriptors therefore serve as a model answer of sorts, something previously lacking in simulation assessment feedback.

Debriefing of simulation results should emphasise the relevant aspects within the best practice descriptors as a means of guiding student learning and remediation. This process should involve a detailed debriefing with either individual students or the entire student grouping who were involved in the simulation. The format for debriefing would be dependent on the results of the assessment and the preferred practices of the subject lecturer.

The SATLAB seeks to address a number of issues associated with reliability and validity in simulation assessment. We discuss a number of these issues next, and explain how we believe the SATLAB system addresses each one.

RELIABILITY

Reliability relates to the ability of an assessment instrument to produce consistent measurement of achievement.[7] Equivalence, also termed 'inter-rater reliability', refers to the consistency in scoring between markers. One of the criticisms of the global rating score is that significant differences in inter-rater reliability do not have context. The SATLAB requires markers to score each assessment outcome separately. It is therefore possible to analyse inter-marker scorings for each competency outcome and determine levels of inter-marker consistency. Where significant differences exist, the video footage can be reviewed for that particular competency outcome by an external marker or moderator, and the markers' score compared to the observed student performance. The SATLAB also requires the marker to provide a Global Rating Scale for the simulation. This can be compared to the final weighted result and serves as an additional measure of reliability.

The panel of expert members is required to determine the weighting for each specified outcome independently. This is done in an environment free from the stresses associated with simulation assessment. Through the use of a panel of experts who are responsible for independent weightings, the SATLAB seeks to flatten out the potential skewing effect that individual preferences may have on the result.

The SATLAB seeks to provide a well-defined and standardised measurement instrument and procedure for use in simulation assessment. Marker error is a constant factor that threatens the reliability of simulation assessment results. Marker error may be caused by marker stress, interpretation of simulation assessment outcomes, and the non-linear nature of simulation.[2] The SATLAB seeks to reduce marker stress by removing the need for the marker to make global rating score decisions. The marker is blinded to the weightings of each assessed outcome and is therefore unable to calculate the effect that any specific result may have on the final mark. The SATLAB competency descriptors guide the marker's interpretation of the relevant importance of simulation assessment outcomes, reducing the need for the marker to do this themselves.

FAIRNESS

Fairness implies that a test is free of bias, that there is equal opportunity to show proficiency, that in a test of knowledge and skill there is equal opportunity to learn, and that score distributions are as equal as possible across different groups.[2] The SATLAB seeks to limit bias by using a panel of experts to determine the weightings of each simulation outcome. The use of SATLAB competency descriptors aims to reduce marker bias related to personal competency classification scales. The SATLAB competency descriptors allow students equal opportunity to show proficiency in that the descriptors are based on what the students would have been taught in the classroom. Feedback provided by SATLAB is presented by way of competency descriptors. Students are able to classify their performance and compare this to the best practice criteria; this enables the student the same learning opportunity as their classmates. The use of SATLAB competency descriptors promotes score consistency across groups that use the same competency descriptors.

VALIDITY

Validity refers to the interpretation of results and what decision is made based on these results. Validity consists of various sub-headings, namely; internal validity, face validity, content validity and criterion validity.

Internal validity relates to the degree to which the results are related to the intervention of interest and no extraneous variables.[7] In other words, do the results of the assessment accurately reflect the student's performance against the competency outcomes? Another way of saying this would be to ask whether the markers have used the correct criteria to award the mark. The SATLAB uses assessor-generated descriptors for each competency outcome specified for assessment. These ensure that markers use only criteria linked to specific learning outcomes when determining their measures of competence. This links the classroom where the student was taught, the immersive environment in which they are assessed, and the real-world of the marker. Providing the markers with the descriptors encourages them to link the teaching described with their clinical experience and allows them to use a combination of the two to generate a result. Using only one or the other would result in potential bias that could affect the validity of the assessment result.

The perceived validity of an instrument at face value is termed 'face validity'.[8] Face validity has also been termed 'logical validity' and refers simply to whether or not the test measures what it claims to. A simulation measures whether or not a student can appropriately manage a patient. Should the student's management of the patient result in harm being done, it would stand to reason that the logical result would be a fail for the assessment. By the same token, a student who adequately manages the patient should achieve a pass result. Students assessed using the SATLAB should have confidence that the tool measures what it claims to measure, since students' perception of the SATLAB is critical to its success as an assessment tool. Students need to believe that the SATLAB accurately measures the competencies it claims to be measuring.

Hancock researched student perceptions of the quality of feedback produced by the SATLAB.[9] Data were collected using a purpose-designed questionnaire that was completed by 34 respondents. The results suggest that students generally believed the SATLAB feedback had face validity, provided understandable feedback, guided reflection on their performance in the simulation, and was consistent with simulation outcomes. In addition, students indicated that they used SATLAB feedback to improve their simulation preparation for future assessments and to change their approach to simulations, that they preferred written feedback over verbal feedback, and were unanimous that it would be helpful to discuss improvement goals with an assessor while receiving feedback.[9]

Content validity relates to the level to which an instrument measures the designated concept or competency.[7] In other words, does the instrument adequately assess each of the domains under investigation and does it adequately identify areas requiring improvement. Criteria that can affect content validity include the fidelity of the simulated interaction, the accuracy of diagnostic cues provided, as well as the extent to which clinical interventions will affect physiological responses. The SATLAB requires the simulation to be sent to a panel of experts for review. These experts provide feedback on multiple areas of the simulation that are then incorporated into the final assessment. This is a recommended method of ensuring content validity.[7]

Criterion validity refers to the predictability of an outcome based on a specific set of variables.[8] In other words, it is the ability of an instrument to distinguish between performance variations. Garcia et al have referred to this as the ability of a tool to differentiate the expert from the intermediate and the novice.[10] The SATLAB provides descriptors for each category of assessment and requires competency levels to be clearly stated and differentiated. Competency outcome criteria are specified for Best Practice (expert), Competent (intermediate) and Not Yet Competent (novice), and descriptions are provided for actions omitted or considered harmful. This negates the need for the marker to make this decision based on his or her own potentially biased opinions. In addition, the weightings determine the importance of each specific competency outcome, negating the need for the marker to try to determine which outcomes are more or less critical than others. The marker is then able to better focus on assessing individual competencies as opposed to trying to make a complicated global rating decision based on their opinion.

EASE OF USE

We realised that an overly complex assessment tool would not provide an appropriate solution to the challenges being experienced in simulation assessment. If the tool is too complex, markers would not be able to use it as intended and students would simply find it intimidating and resist its implementation. The SATLAB combines a number of logical processes into one process that we believe is reasonably easy to implement, user-friendly and easily understood by both students and markers. The fact that the SATLAB consists of six stages means that it is easier for users to break up into 'bite-sized chunks' and combine these to understand and conceptualise the SATLAB as a whole. Students can receive detailed feedback by means of a simple feedback sheet which can be made available using an appropriate software package such as Microsoft Excel®.

CONCLUSION

The SATLAB was conceptualised as a potential solution to some of the inherent challenges in assessment of simulation. This concept article has explained the rationale behind the tool and has highlighted how it aims to resolve some of the identified challenges. The current use and ongoing evaluation of the SATLAB is providing exciting insights into the future of simulation assessment. Further research is required to determine the reliability and validity of the tool within the simulation assessment domain.

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

AUTHOR CONTRIBUTIONS

AM conceptualised the SATLAB. The SATLAB was implemented with the assistance of CV-L. AM drafted the original manuscript that was critically reviewed by CV-L.

REFERENCES

1. McKenna KD, Carhart E, Bercher D, Spain A, Todaro J, Freel J. Simulation use in paramedic education research (SUPER): A descriptive study. Prehosp Emerg Care 2015;19:432–40.

2. Bewley WL, O'Neil HF. Evaluation of medical simulations. Mil Med. 2013;178:64–75.

3. High fidelity patient simulation: a descriptive white paper report. Healthcare Simulation Strategies, 2007. (`http://sim-strategies.com/downloads/Simulation%20White%20Paper2.pdf`) (visited on 04/18/2018).

4. Bradley P. The history of simulation in medical education and possible future directions. Med Educ. 2006;40:254–62.

5. Malloch M, Cairns L, Evans K, O'Connor BN, eds. The sage handbook of workplace learning. London: SAGE Publications Ltd., 2011:185.

6. Vincent-Lambert C, Bogossian F, eds. A guide for the assessment of clinical competence using simulation. Section 5: conducting the simulation. Johannesburg: Universitas 21 Health Sciences Group, 2017:31–6.

7. Shelestak D, Voshall B. Examining validity, fidelity, and reliability of human patient simulation. Clin Simul Nurs. 5 2014;10:e257–60.

8. Dumestre D, Yeung JK, Temple-Oberle C. Evidence-based microsurgical skills acquisition series part 2: validated assessment instruments-a systematic review. J Surg Educ. 2014;72:80–9.

9. Hancock JJ. Student opinions and experiences related to the quality of feedback of the simulation assessment tool limiting assessor bias (SATLAB). Research rep. University of Johannesburg, 2015.

10. Garcia J, Coste A, Tavares W, Nuño N, Lachapelle K. Assessment of competency during orotracheal intubation in medical simulation. Br J Anaesth. 2015;115:302–7.