



Contents lists available at ScienceDirect

## The American Journal of Surgery

journal homepage: [www.americanjournalofsurgery.com](http://www.americanjournalofsurgery.com)

## The value of psychometric analysis of the advanced trauma life support cognitive test: Outcome of an ACS-Accredited educational institute multisite study



Jenny Guido <sup>a</sup>, Britany Telford <sup>a, c</sup>, Luis E. Llerena <sup>a, b, \*</sup>, Spiros Stergiopoulos <sup>d</sup>,  
Charlotte Huber <sup>e</sup>, Andrew Cross <sup>f</sup>, Michael Brannick <sup>a, c</sup>

<sup>a</sup> Center for Advanced Medical Learning and Simulation (CAMLS), USF Health, University of South Florida, Tampa, FL, USA

<sup>b</sup> Morsani College of Medicine, USF Health, University of South Florida, Tampa, FL, USA

<sup>c</sup> Psychology Department, University of South Florida, Tampa, FL, USA

<sup>d</sup> Medical Physics Department, Athens University Medical School, Athens, Greece

<sup>e</sup> William Maul Measey Institute for Clinical Simulation and Patient Safety, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, USA

<sup>f</sup> Eastern Virginia Medical School, Norfolk, VA, USA

### ARTICLE INFO

#### Article history:

Received 25 April 2018

Received in revised form

25 July 2018

Accepted 30 July 2018

#### Keywords:

Advanced trauma life support

Psychometrics

Educational measurement

### ABSTRACT

**Background:** The Advanced Trauma Life Support® (ATLS®) course provides a standard approach to trauma. Participants must pass the ATLS® post-test. We deployed the test online to allow ongoing psychometric item analysis and potential objective refinement.

**Methods:** A two-phase study was undertaken with the ACS COT permission. In the first phase, ATLS® post-test #2 was computerized and deployed using Qualtrics®. Data were collected from fourteen courses conducted between 2014 and 2015 ( $n = 306$ ) at one ACS AEI site. In the second phase, the same post-test was administered to 238 trainees in 10 courses via secured computers at four ACS AEI sites in 2016.

**Results:** Phase 1 item analyses showed two items with very low percentages correct, and one of these also showed a low discrimination index. Phase 2 item analyses suggested four items as candidates for review and possible revision. We also found differences by learner background and by instructional site.

**Conclusions:** This study demonstrates computerized delivery of the ATLS post-test is feasible, promotes psychometric analysis, and could improve the quality of the test. Further collaboration between the ACS COT and ACS AEI would be beneficial.

© 2018 Elsevier Inc. All rights reserved.

### Introduction

The Advanced Trauma Life Support® (ATLS®) course is designed to teach participants a standardized method of trauma patient evaluation and treatment. This course provides training for all members of the trauma team and was developed by the American College of Surgeons Committee on Trauma (ACS COT).<sup>1,2</sup> The ATLS course is the international standard for training trauma care providers and is mandatory for all United States surgical and emergency medicine residents. Passing the ATLS course has important consequences for the learner; thus the ATLS exam is considered to

be a high-stakes test. As with all high-stakes tests, the ATLS exam should be psychometrically sound.<sup>3</sup> The ACS COT has undertaken rigorous evaluation of ATLS since its inception.<sup>4</sup>

To evaluate the psychometrics of the ATLS post-test, test-taker data are required. Currently, the post-test is administered via paper and pencil, making data analysis difficult. Someone has to score the test by hand, introducing human error into the scores. If item-level data are to be analyzed, someone must input the item responses for each examinee into a database. The evaluation of item quality is a standard practice in high-stakes testing. For example, both the Medical College Admission Test (MCAT)<sup>5</sup> and United States Medical Licensing Examination (USMLE)<sup>6</sup> have been evaluated and redesigned accordingly.

Evaluating the item quality can be done in part by considering each item's discrimination and difficulty. *Item discrimination* is the

\* Corresponding author. USF Health CAMLS, Morsani College of Medicine, University of South Florida, 124 South Franklin Street, Tampa, FL, 33602, USA.

E-mail address: [lllerena@health.usf.edu](mailto:lllerena@health.usf.edu) (L.E. Llerena).

degree to which test takers that score highly on the overall exam also answer that item correctly.<sup>7</sup> This is a point biserial correlation, thus values range approximately from  $-1$  to  $+1$ . Higher values indicate that the item effectively distinguishes between high scorers and low scorers. *Item difficulty* is the mean score on the item and represents the percentage of test takers that correctly answer the item.<sup>7</sup> This index ranges from 0 to 1 with easier items having higher values.

Examining item discrimination and item difficulty values is a critical step in the assessment evaluation process. High-stakes tests should not be stagnant entities but rather undergo a circular use, evaluation, and modification process. Thus, after creation, they should be evaluated and modified appropriately to strive towards validity and fairness. High-stakes tests such as that incorporated in the ATLS should also be modified over time as the critical content (e.g. evidence-based standards) or learner qualities change. In order to evaluate an exam, item-level data for a large number of learners must be collected. The overall scores for each learner are not sufficient to calculate statistics such as item discrimination and difficulty.

An important feature of the ATLS course is that both the instructional curriculum and proficiency assessment are standardized. Standardization is important for consistency, fairness, and evaluation. However, the administration of the course is decentralized, being offered to medical practitioners with different backgrounds by multiple instructors in diverse institutions around the world. The collection of data from the proficiency assessments allows monitoring of the differences in item and total scores by background, location, and instructor. The monitoring of such differences creates the potential to spot problems, and to identify and capitalize upon best practices.

Thus, we created a computerized version of the ATLS post-test using a third-party platform to easily collect and manage item-level data. After receiving permission from ACS COT, we chose to prove the overall concept at one institution followed by a larger multicenter project involving members of the American College of Surgeons Accredited Education Institutes (ACS AEI). The use of a common platform allowed straightforward data analysis for psychometric qualities of the ATLS examination as well as for score differences in participant background and score differences across ACS AEI testing centers.

As the ATLS proceeds to the 10th edition, a new, more Socratic method of teaching is being introduced. The analysis of item content can help pinpoint difficulties in the new teaching model as well as help identify best instructional practice across institutions. As ATLS continues to grow in popularity, its audience continues to expand beyond physicians. Item analysis can identify problem areas for course participants with different backgrounds. Such analysis can also provide different instructors in the same institution with feedback on their effectiveness. In other words, the analysis of the post-test could be useful in quality assurance for the ATLS program. We hypothesized that administration of a computerized ATLS post-test would enable ongoing item analysis that could show differences in item responses by instructional site and by participant background.

## Material and methods

### Participants

*Phase 1.* Data were collected from ten ATLS<sup>®</sup> Provider and four ATLS<sup>®</sup> Refresher courses conducted between 2014 and 2015 at the USF Health Center for Advanced Medical Learning and Simulation (CAMLS). All courses were taught by the same course director. A total of 306 participants completed the computerized post-test. The

majority of the participants were attending physicians (54%), followed by resident/fellows (38%), and advanced practitioners (8%). Phase 1 provided a relatively large sample of participants as the basis of stable estimates of item difficulty and discrimination.

*Phase 2.* Participants in this study were 238 ATLS course participants enrolled in 10 courses at four AEI sites from March–September 2016. The participating sites were the University of South Florida Health Center for Advanced Medical Learning and Simulation (CAMLS;  $n = 135$ , 6 courses), Athens University Medical School (Athens;  $n = 52$ , 2 courses), Lewis Katz School of Medicine at Temple University (Temple;  $n = 31$ , 1 course), and Eastern Virginia Medical School (EVMS;  $n = 20$ , 1 course). Both refresher and two-day course learners were included in the sample. Participants included attending physicians ( $n = 82$ ), advanced practitioner (physician assistants and nurse practitioners;  $n = 21$ ), and residents ( $n = 135$ ).

### Procedure

The American College of Surgeons Accredited Educational Institutes (ACS AEI) sites were invited to participate in the study because of their emphasis on education relevant to ATLS. Since it launched in 2005, ACS AEI has set standards for how surgical education and training should be offered at Accredited Education Institutes. The Institute's ultimate goal is to promote patient safety. AEI members may evaluate the impact of education through long-term follow-up of learners and investigate how best to incorporate aspects of surgical research into surgical education.<sup>8</sup> Therefore, the collaboration of several ACS AEI member Institutes to investigate the psychometric properties of the ATLS examination follows directly from the goals of the organization.

The studies were determined exempt by the USF IRB. Permission to conduct the multi-site study was also received from the ACS Committee on Trauma and the ACS AEI Research and Development Committee. At the 2016 ACS AEI annual meeting, we convened interested AEI sites. For both studies, the ATLS<sup>®</sup> post-test #2 was deployed on secured, monitored computers using Qualtrics<sup>®</sup> software. There was no increased cost as each center had computers currently in use and the software was provided by the researchers. Test data were collected from regular participants in ATLS classes in collaborating sites as an ordinary part of the course (March–September 2016); the only difference from the status quo was that the tests were administered online rather than on paper. Data were subsequently downloaded at CAMLS for analysis. All analyses were conducted using SPSS (study 1 used version 22; study 2 used version 24).

## Results

*Overall.* There were no reports of lost data, compromised post-tests, or course deployment delays for online administration; thus, all enrolled learners' data were usable. The ATLS passing score is 70%. Phase 1 participants' mean score was 32.25 (81%;  $SD = 3.87$ ). Phase 2 participants' mean score was 31.50 (79%;  $SD = 3.65$ ).

*Differences by site.* Phase 1 concerned only a single site. For Phase 2, analysis of variance (ANOVA) revealed that score means differed significantly between sites [ $F(3,237) = 3.25$ ,  $p = .02$ ,  $\eta^2 = 0.04$ ]. A post hoc Tukey test showed score means between two sites, CAMLS ( $M = 31.96$ ) and Athens ( $M = 30.17$ ) differed significantly ( $p = .01$ ). Mean differences were not significantly different for learners at Temple ( $M = 31.94$ ) or EVMS ( $M = 31.25$ ).

*Differences by background.* For Phase 1, ANOVA revealed significant mean differences in post test scores [ $F(2,303) = 15.86$ ,  $p < .001$ ]. Specifically, the mean score for attending physicians [ $M = 32.79$  (82%),  $SD = 3.49$ ,  $n = 175$ ] was not significantly different

from the mean of residents/fellows [ $M = 31.94$  (80%),  $SD = 3.74$ ,  $n = 104$ ], while the mean score for advanced practitioners (PA, ARNP) was significantly lower [ $M = 28.74$  (72%),  $SD = 4.71$ ,  $n = 27$ ].

For Phase 2, significant differences between all professional groups' mean scores were evident [ $F(3,237) = 12.98$ ,  $p < .001$ ;  $\eta^2 = 0.10$ ]. Advanced practitioners scored lower ( $M = 28.43$ ; 71%) than attending physicians ( $M = 32.66$ ; 82%;  $p < .0001$ ) and residents ( $M = 31.28$ , 78%;  $p < .01$ ). In turn, residents on average scored lower than attending physicians ( $p < .05$ ). Thus, the pattern of results for training background is similar for both studies.

Figs. 1 and 2 show scatterplots of item difficulties for each of the 40 items in the test. Fig. 1 shows the relations between item difficulties comparing residents to advanced practitioners; Fig. 2 shows item difficulties from residents and attending physicians. The figures also include reference lines at 0.7, which is the passing proportion for the test as a whole. The reference lines split the figures into four quadrants. In the bottom left quadrant, both groups fail the item on average. In the top right, both groups pass the item. The other two quadrants show items that only one of the groups passes. Both graphs are based on data from Phase 2, so data are aggregated across multiple sites. The numbers next to the circles indicate the item number on ATLS post-test 2. As can be seen in Fig. 2, five items show proportions correct of less than 70% for both residents and attending physicians (those five items also show less than 70% for the advanced practitioners). Fig. 2 shows that there are several items that the attending physicians have mastered, but where the residents show less than 70% correct. Similarly, Fig. 1 shows a series of items that advanced practitioners find problematic, but residents do not. By cross referencing this data with Table 3, we can now focus not only on the underperforming question but also the content theme of these questions.

**Psychometrics.** Difficulty and discrimination indices for the both the Phase 1 and Phase 2 total samples are shown in Table 1. As can be seen in Table 1, item statistics were stable across the two studies. Item difficulty was measured by calculating the item mean and item discrimination was measured using the corrected point-biserial correlation. Difficulty and discrimination indices, by site, for four underperforming items are shown in Tables 2 and 3

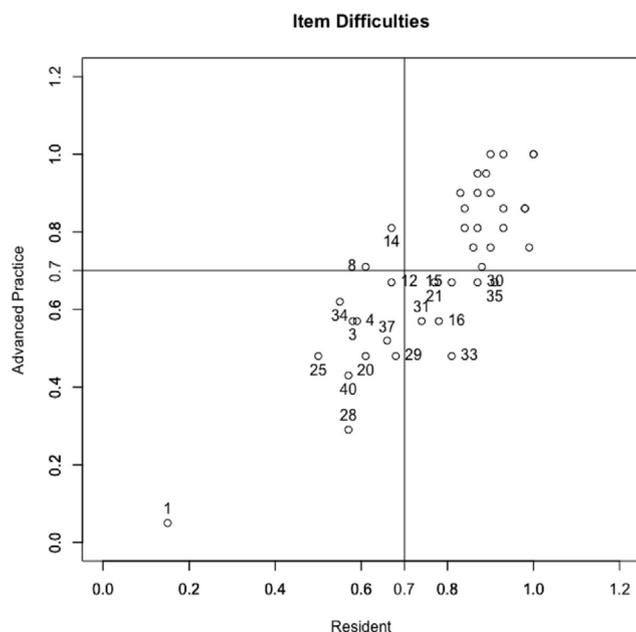


Fig. 1. A scatterplot of item difficulty indices for Phase 2 residents and advanced practice examinees showing passing and failing percentages for each group.

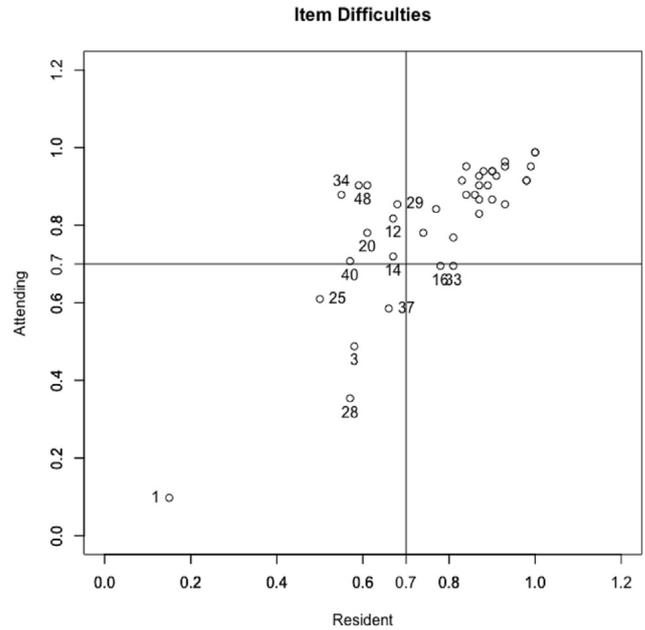


Fig. 2. A scatterplot of item difficulty indices for Phase 2 residents and attending physician examinees indicating which items showed passing and failing percentages for each group.

respectively. Phase 1 also flagged the same four items as unusually difficult, although only Item 1 showed a consistently poor discrimination index (Tables 2 and 3).

**Discussion**

This study illustrated the usefulness of computerized deployment of the ATLS post-test. Our item analysis proved easily applicable across multiple instructional sites and identified four questions as candidates for revision. We also found differences among instructional sites and learner backgrounds consistent with our hypotheses. In what follows, we discuss the main findings plus additional practical details, mention limitations of the study, and provide suggestions for further research. We also relate our findings to those of previous investigators.<sup>9,10</sup>

*Differences in item and total test scores*

One interesting finding concerned score differences between professional groups. In phase one of our study, attending physicians and residents scored higher than other providers. In the second phase, attending physicians also had higher scores than the residents. Although such a finding is not very surprising, it does suggest that some groups may require more preparation than others. Examination of the items that are answered correctly by more than 70% of one group, but less than 70% of another group (e.g., items 4, 8 and 34) suggest that specific on-the-job experiences may explain the differences in performance on the items. Such an analysis shows how data might be used to identify problem areas for different groups of examinees.

Much of our analysis focused on item scores. However, total test scores are also of interest and online data collection facilitates comparisons of total test scores by participant background (as we mentioned) as well as by site, by instructor, and perhaps other characteristics of instructional or participant characteristics. With regard to site, a difference of approximately four percent was seen

**Table 1**  
Analysis of ATLS test items. Difficulty and discrimination indices for each item in post-test#2 (Phase 1, N = 306; Phase 2, N = 238).

| Item | Phase 1         |                     | Phase 2         |                     |
|------|-----------------|---------------------|-----------------|---------------------|
|      | Item Difficulty | Item Discrimination | Item Difficulty | Item Discrimination |
| 1    | .10             | 0.05                | .12             | .04                 |
| 2    | .82             | 0.23                | .86             | .22                 |
| 3    | .61             | 0.13                | .55             | .16                 |
| 4    | .76             | 0.17                | .70             | .16                 |
| 5    | .95             | 0.10                | .92             | .12                 |
| 6    | .97             | 0.10                | .95             | .00                 |
| 7    | .83             | 0.31                | .86             | .19                 |
| 8    | .85             | 0.16                | .72             | .14                 |
| 9    | .89             | 0.11                | .88             | -.01                |
| 10   | .94             | 0.15                | .90             | .04                 |
| 11   | .90             | 0.23                | .89             | .23                 |
| 12   | .79             | 0.23                | .72             | .18                 |
| 13   | .92             | 0.10                | .95             | -.04                |
| 14   | .75             | 0.18                | .70             | .15                 |
| 15   | .79             | 0.21                | .78             | .17                 |
| 16   | .73             | 0.16                | .73             | .15                 |
| 17   | .94             | 0.24                | .95             | .19                 |
| 18   | 1.0             | 0.00                | 1.0             | -.01                |
| 19   | .85             | 0.25                | .89             | .18                 |
| 20   | .62             | 0.15                | .66             | .10                 |
| 21   | .83             | 0.31                | .79             | .24                 |
| 22   | .85             | 0.24                | .87             | .07                 |
| 23   | .91             | 0.07                | .90             | .08                 |
| 24   | .97             | 0.17                | .93             | .24                 |
| 25   | .64             | 0.20                | .54             | .12                 |
| 26   | .94             | 0.24                | .96             | .24                 |
| 27   | .99             | 0.02                | 1.0             | .01                 |
| 28   | .37             | 0.21                | .47             | .15                 |
| 29   | .80             | 0.27                | .72             | .33                 |
| 30   | .81             | 0.22                | .84             | .11                 |
| 31   | .81             | 0.13                | .74             | .13                 |
| 32   | .91             | 0.23                | .88             | .25                 |
| 33   | .75             | 0.27                | .74             | .26                 |
| 34   | .77             | 0.17                | .67             | .17                 |
| 35   | .89             | 0.14                | .89             | .10                 |
| 36   | .91             | 0.14                | .92             | .13                 |
| 37   | .64             | 0.23                | .62             | .24                 |
| 38   | .89             | 0.27                | .85             | .20                 |
| 39   | .92             | 0.14                | .87             | .15                 |
| 40   | .62             | 0.28                | .61             | .13                 |

between mean scores for two sites, CAMLS and Athens. Differences in mean scores across sites do not allow us to conclude why the differences exist; they merely serve as a flag to investigate possible reasons. However, routine data collection allows the development of norms and comparisons across sites that would likely be of interest to the sites as feedback concerning where they stand relative to other sites offering ATLS instruction. Thus, such analysis provides a second example of how computerized data collection can support quality assurance for the ATLS program. Another important aspect of quality assurance is the review of the quality of the test items, which we address next.

*Item analysis*

Item analysis revealed four problematic items which suggest the post-test may need revision. For example, item one had a difficulty index of 0.12 which indicates that only 12% of learners correctly answered this item. ATLS post-test questions contain 5 answer options, therefore the chance of learners randomly guessing the answer is 20%. It is important to note that low item difficulty may be caused by several factors. First, the item could cover difficult material that the learner needs to know. A low item-difficulty on such an item is acceptable on technical grounds. However, such an

**Table 2**  
Difficulty indices for four underperforming items on post-test#2 by site.

| Item | USF <sup>a</sup><br>(N = 135) | Athens <sup>a</sup><br>(N = 52) | Temple <sup>a</sup><br>(N = 31) | EVMS <sup>a</sup><br>(N = 20) | Total <sup>a</sup><br>(N = 238) | USF <sup>b</sup><br>N = 306 |
|------|-------------------------------|---------------------------------|---------------------------------|-------------------------------|---------------------------------|-----------------------------|
| 1    | .02                           | .33                             | .16                             | .20                           | .12                             | .10                         |
| 3    | .60                           | .40                             | .65                             | .40                           | .55                             | .61                         |
| 28   | .37                           | .42                             | .58                             | .35                           | .47                             | .37                         |
| 40   | .61                           | .63                             | .48                             | .65                           | .61                             | .62                         |

Note. USF = University of South Florida Center for Advanced Medical Learning and Simulation; Athens = Athens University Medical School; Temple = Lewis Katz School of Medicine at Temple University; EVMS and Eastern Virginia Medical School.

<sup>a</sup> Phase 1.  
<sup>b</sup> Phase 2.

**Table 3**  
Discrimination indices for four underperforming items on post-test#2 by site.

| Item | USF <sup>a</sup><br>(N = 135) | Athens <sup>a</sup><br>(N = 52) | Temple <sup>a</sup><br>(N = 31) | EVMS <sup>a</sup><br>(N = 20) | Total <sup>a</sup><br>(N = 238) | USF <sup>b</sup><br>N = 306 |
|------|-------------------------------|---------------------------------|---------------------------------|-------------------------------|---------------------------------|-----------------------------|
| 1    | .00                           | .35                             | .04                             | .06                           | .04                             | .05                         |
| 3    | .11                           | .25                             | .17                             | -.02                          | .16                             | .13                         |
| 28   | .22                           | .42                             | -.01                            | .08                           | .15                             | .21                         |
| 40   | .21                           | .00                             | .16                             | .10                           | .13                             | .28                         |

Note. USF = University of South Florida Center for Advanced Medical Learning and Simulation; Athens = Athens University Medical School; Temple = Lewis Katz School of Medicine at Temple University; EVMS and Eastern Virginia Medical School.

<sup>a</sup> Phase 1.  
<sup>b</sup> Phase 2.

outcome means that the instruction has not resulted in participant mastery of the material covered by that item. In other words, such a result indicates that the need for training remains – the participants' knowledge or performance gap has failed to close. Second, low item difficulty might also be caused by a confusing item that learners are answering incorrectly for reasons unrelated to their knowledge. Such items are undesirable and should be rewritten or removed. Finally, low item difficulty might also occur when the test item addresses material that was not covered in the course. If the mastery of such material is critical, then the course content should be modified. If the material is not critical for the learner to master, then the item should be removed. Due to its many antecedents, item difficulty scores should always be interpreted within the context of the assessment and other statistics.

The item discrimination index offers some insight as to whether there may be something wrong with the question or whether learners just failed to remember the pertinent information. Item 1 also demonstrated an extremely low item discrimination (0.05 in the first sample, 0.04 in the second), which indicates that high-performing learners are no more likely to choose the correct answer than low-performing learners. Subsequent investigation of item one reveals that 87% of learners chose the same incorrect answer. Additionally, as shown in [Tables 2 and 3](#), these low indices are replicated across testing sites. These findings suggest item one is a poor item and a good candidate for revision. Additionally, due to low difficulty and discrimination values, items three, twenty-eight, and forty should also be reviewed by subject matter experts. The identification of several problematic items on ATLS post-test #2 suggests that the test is in need of revision, which is a normal, critical process for all high-stakes tests.

On the other hand, if the five items that show difficulty indices less than 0.70 are markers for important information and also technically adequate, then they provide information about where the instruction could be improved, again providing information for quality assurance.

#### *Hardware and software*

Computerized testing requires infrastructure. The contributing sites for this study all had their own computers and access to the online survey software for test delivery. We did not encounter any problems using third-party software (Qualtrics®) to administer the post test and to collect the data from multiple sites. The study demonstrates data collection in multiple locations without compromising the high-stakes exam. However, for sites that do not currently own enough computers to test all course participants, tablet computers with internet browsers are sufficient to the purpose. Such tablets may be purchased relatively inexpensively (at the time of this writing, approximately \$250 per tablet) and each site would require approximately eight to twelve tablets (enough that each participant can use a single tablet). Just as for laptop and desktop computers, it is possible to buy 3-year service agreements for the tablets, but it is up to the institution to decide whether the money spent on the service agreement is worthwhile. It may also prove worthwhile to buy a cart for recharging the tablets assuming the table is the course of action that your organization pursues. Various options are available to enhance both test security (e.g., lockdown browsers) and tablet theft prevention (e.g., GPS locators). We do not allow participants to use their own devices (computers, tablets or phones) because of test security. The cost of Qualtrics (survey software) depends upon how much it is used and whether the package is purchased through an academic institution. Academic institutions may already have established software licensing agreements because the software is used for a variety of research purposes. Estimated cost for the basic software package is five

thousand dollars per year, although adding capabilities to the software can increase the cost.

#### *Results in context*

The ACS COT strives to ensure ATLS is rigorously examined and improved. Before deployment of the ATLS 9th edition, the ATLS subcommittee worked with international trauma experts to assure the multiple-choice questions performed well both in relation to test specification matrix and item analysis. However, all beta-testing involved ATLS experts. They concluded further research using student data would be required.<sup>4</sup> Data collection such as that described in the paper would greatly facilitate such research.

Mobily and colleagues studied student learners across two sites in the US. They found increased failure rate among learners with the following characteristics: age greater than 55, English as a second language, pretest score less than 75, and non-trauma/emergency medicine backgrounds.<sup>9</sup> However, no item analysis was performed on this cohort. Such an analysis could provide insight about whether particular items and their content domains are disproportionately difficult for specific groups of learners.

Not only have the learners and their test results been analyzed but also the ATLS instructors have been researched; Germany reviewed their experience deploying ATLS nationwide from 2008 to 2012. Their results pointed to instructor quality improvement in teaching during the first 100 lessons taught.<sup>11</sup> Computerization of the test makes tracking of instructor performance much less laborious.

The ACS AEI created a consortium of simulation centers with high standards to collaborate and share best practices. There are currently ninety-three accredited institutes worldwide. They hold yearly international meetings for surgical educators to share educational research, collaborate on curriculum development, and foster faculty development.<sup>8</sup> A recent needs assessment performed by the Association of Surgical Education found surgical educators very interested in qualitative methodology and data analysis.<sup>10</sup> This ACS organization is ideal for collaborative research endeavors in this arena.

#### *Study limits and future research*

The current study evaluated the psychometrics of only the ATLS post-test version #2. Alternate versions should be separately examined following the same methodology. Additionally, the current study is also limited by the small number of participants at several sites and the inclusion of only one international site. Going forward, ATLS will continue to provide necessary trauma education; inclusion of sites beyond ACS AEI would allow analyses that could inform local, state, and regional quality assurance programs. Additionally, a larger number of international site participants would allow for comparison of item functioning in different cultures. Thus, future work in this area should include a larger collaborative initiative between the ACS COT and ACS AEI for deployment of this methodology to continue ATLS post-test performance improvement.

ATLS instruction is not limited to ACS AEI sites. Other sites offering ATLS instruction should be included in further research and application of computerized test administration. Research should also address some of the other aspects of quality assurance, such as how best to provide feedback to instructors and to institutions about their instructional outcomes.

#### *Prior presentation*

Portions of this paper were described during poster

presentations during the ACS Accredited Educational Institute National meeting in March 2016 in Chicago, Illinois and in March 2017 in Chicago, Illinois.

### Conflicts of interest

No author on this paper has any financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

### Acknowledgements

The authors wish to thank the following organizations: The American College of Surgeons COT, Monique Drago, EdD; ACS AEI Research and Development Committee, Amy Johnson; USF Health CAMLS, Sue Frye.

### References

1. Baker MS. Advanced trauma life support: is it adequate stand-alone training for military medicine? *Mil Med.* 1994;159:587–590.
2. Van Olden GD, Meeuwis JD, Bolhuis HW. Advanced trauma life support study: quality of diagnostic and therapeutic procedures. *J Trauma.* 2004;57:381–384.
3. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association; 1999.
4. Bustraan J, Henny W, Kortbeek JB, Brasel KJ, Hofman M, Schipper IB. MCQ tests in advanced trauma life support (ATLS): development and revision. *Injury.* 2016;47(3):665–668.
5. Dunleavy DM, Kroopnick MH, Dowd KW, Searcy CA, Zhao X. The predictive validity of the MCAT exam in relation to academic performance through medical school: a national cohort study of 2001–2004 matriculants. *Acad Med.* 2013;88(5):666–671.
6. Swanson DB, Holtzman KZ, Johnson DA. Developing test content for the United States medical licensing examination. *J Med Licens Discip.* 2009;95(2):22–29.
7. Allen MJ, Yen WM. *Introduction to Measurement Theory.* Monterey, CA: Brooks/Cole; 1979.
8. American College of Surgeons. Consortium description. AEI Consortium Available at: <https://www.facs.org/education/accreditation/aei/consortium>. Accessed February 16, 2017.
9. Mobily M, Branco BC, Joseph B, Hernandez N. Predictors of failure in the advanced trauma life support course. *Am J Surg.* 2015;219:942–946.
10. Phitayakorn R, Salles A, Falcone JL, Jensen AR, Steinemann S, Torbeck L. A needs assessment of education research topics among surgical educators in the United States. *Am J Surg.* 2017;213:346–352.
11. Luedi NN, Wolf CC, Weiferich K, Dogjani A, Kauf P, Doll D. Teaching Advanced Trauma Life Support (ATLS): a nationwide retrospective analysis of 8202 lessons taught in Germany. *J Surg Educ.* 2016;74(1):161–166.