ORIGINAL RESEARCH

Assessment of US Paramedic Professionalism: A Psychometric Appraisal

L Michael Bowen
Brett Williams (iD)

Department of Community Emergency
Health and Paramedic Practice, Monash
University, Melbourne, Australia

Introduction: Professionalism is an essential behavior for paramedic students to demonstrate. In the United States, paramedic accreditation standards require educators to evaluate and document summative affective evaluation on each paramedic student before graduation. The 2009 Emergency Medical Services Education Standards identified the affective behaviors as one of the three learning domains and published a grading tool to help educators recognize professional behaviors. However, little attention was given to the validity or reliability of this tool. Therefore, the aim of this study was to evaluate the psychometric properties of the 5-point Paramedic Affective Domain Tool.

Methods: This was a retrospective study with educators that completed evaluations on paramedic students from May 2013 to January 2017. A total of 707 cases met inclusion criteria and 131 unique evaluators from 27 different paramedic programs. A Rasch Partial Credit Model was used to analyze the data.

Results: Almost 97% of the paramedic students received passing scores and 28.1% (n=199) received perfect scores. Only 3.5% (n=25) failed the evaluation. Scores ranged from 11 to 55 (M = 46, SD = 9.02) and α = 0.97. Evidence suggests that the tool is not valid and the clustering of scores suggests minimal information can be gleaned from the results.

Conclusion: Serious consideration should be made in the continued use of this tool and future research should focus on developing a new tool that is both valid and reliable.

Keywords: allied health personnel, paramedic, professionalism

Introduction

Paramedics are invited into brief moments of a patient's life that range from the worst, best, first or last day of their life. A diverse set of professional behaviors is required to manage patient encounters effectively with a judgment-based approach to meet the unique needs of each patient and their situation. These behaviors should enhance the ability to display leadership skills, design treatment plans and communicate decisions in a respectful manner. The situation might be different where care is provided in an unsupervised situation to a vulnerable patient, patient encounters like this require a moral compass.[1] Professional behaviors among paramedics have been shown to improve patient outcomes such as perceptions of empathy,[2] and time management.[3] The paramedic profession continues to evolve and requires self-motivated practitioners to study new treatments, review evidence-based practices, and to self-reflect on previous actions. The affective domain encompasses a variety of professional behaviors which are a critical aspect of patient care, patient outcomes, and standards of care.

In the US paramedic discipline, gaps exist between displaying professional behavior and the occupations expectations. Patients and ambulance agencies voted honesty as the most important quality of a clinician, far above any aspect in the cognitive or

Correspondence: Brett Williams
Department of Community Emergency
Health and Paramedic Practice, Monash
University, Melbourne, Australia
Email brett.williams@monash.edu

psychomotor domain,[4] but still unprofessional actions are concerning issue and were documented at the most frequent complaint.[5] These bad habits and poor behaviors are noticed by colleagues, patients, and bystanders and who then file complaints or even lawsuits against the employee or ambulance service.[6] Fortunately, paramedic students view themselves as the group to break the cycle and desire to be held to a professional standard.[7] The Emergency Medical Service (EMS) community from educators to thought leaders should focus their attention on bridging the gap professional behaviors and expectations for entry-level employees.

The National Highway Traffic Safety Administration (NHTSA) has direct oversight of paramedic education and publishes education standard curricula. The 1998 National Standard Curriculum (NSC) introduced the idea of measuring professional behaviors with paramedic students which included three learning domains; cognitive, psychomotor, and affective.[8] In 2002 NHTSA published the Appendix VI: Rubric Affective Domain Tool (ADT) (Appendix VI - Rubric Affective Domain Tool, 2002) and this provided a grading tool for the affective domain. The ADT has two stated goals. First is to verify competency, and second to identify areas of weakness so that a paramedic student has the opportunity to remediate behavior(s). The ADT outlines 11 traits: Integrity, Empathy, Self-motivation, Appearance and Personal hygiene, Self-confidence, Communications, Time management, Teamwork and diplomacy, Respect, Patient advocacy, and Careful delivery of service (found at https:// one.nhtsa.gov/people/injury/ems/instructor/instructor_ems/ 2002_national_guidelines.htm)[9] The NSC was replaced with the 2009 National Emergency Medical Education Standards: Paramedic Instructional Guidelines (National Emergency Medical Services Education Standards, 2009). Limited evidence exists to prove or disprove the validity and reliability of this scale.[10] The 11 professional traits outlined in the ADT are imprinted into the education process, adopted by accreditation and enforced by the national certification process. The tool used to evaluate these behaviors should be a true and accurate measure of professional ability, but in reality, little is known about its psychometric properties. Therefore, the aim of this study is to gauge the validity and reliability of the ADT published in the 2002 education standards.

Methods
Design
This retrospective study investigated an affective domain tool. A grading tool that was used frequently by EMS educators to measure paramedic student's professional behaviors. The study had three phases. Phase one was data collection and extraction. Data were extracted from Fisdap ™, an online database designed to track and report a comprehensive portfolio of a paramedic students' academic experience. The database had a repository of evaluations completed by EMS educators. The objective of phase one was to ensure data were extracted in an accurate, de-identified manner with a reproducible method. Phase two was data analyses. The objective of phase two was to analyze data in a best practice method. Phase three was to report the results and state the facts.

Instrumentation
A scoping literature review was performed to identify assessments that measure professionalism amongst paramedic students.[10] The National Guidelines for Educating EMS Instructors August 2002, Appendix VI: Rubric Affective Domain Tool (ADT) was selected. It is an openly available grading tool located on the NHTSA website under Emergency Medical Services and instructions guideline. The ADT had educator's rank professional behaviors on a scale of 1–5. Each item has a behavior then provides a description for the corresponding value; 1 = low performance with major infractions, 2 = minor infractions but unacceptable behavior, 3 = acceptable behavior for entry-level provider, 4 = above average consistently, 5 = high performance and role model. The top of the rubric has an instruction section that provides directions on how to interpret the tool. Educators were instructed to focus on patterns of behavior and to avoid judgments from isolated incidents. Overall scores could range from 11 to 55 and to achieve a passing grade a student must earn a score of 33 or above. The instructions outline that most students should receive "3" in each category. The standard setting process used for the evaluation was not reported.

Participants
The participants were a sample of convenience enrolled in paramedic programs throughout the US between May 2013 and January 2017. The inclusion criteria were paramedic students that had agreed to participate in the research through implied consent upon registration, and fully completed evaluations, incomplete evaluations were excluded. Ethical approval was granted by Monash University Humans Ethics Committee.

Data Export

Data were exported in a csv file from Fisdap™ database and de-identified with message-digest (MD5) algorithm with a checksum function output.

Data Analysis

Raw scores were analyzed for descriptive statistics and internal consistency. Data reduction was performed with a Principal Component Analysis (PCA) to measure dimensionality[11] with loading factors (>0.40)[12] and eigenvalues (>1.0).[13] A correlation matrix was utilized for dimensionality in tandem with a scree plot.

Next, the data were transformed with a Rasch-Masters Partial Credit Model (PCM).[14] A mean-square (MSQ) infit and outfit thresholds were used for model fit. Acceptable values range from 0.5 to 1.5 and suggest productive for measurement.[15] Item response-level statistics were count, proportions, logits,[16] and correlation coefficients.

A Mantel differential item function (DIF)[17] analysis was used to gauge how items performed between subgroups of age and gender.[18] Gender subgroups were binned into Male, Female, and unspecified. Age subgroups were grouped to reflect the US Census Age and Sex Composition 2010: 18–24, 25–44, 45–65, and >65. The DIF contrast between two samples should be less than 0.5 logits and with a significant p-value. Dimension reduction was performed with SPSS 24.0 and PCM analysis Winsteps 3.92.

Results

Participants

A total of 1113 forms were completed by 131 unique EMS educators on paramedic students from 27 different paramedic programs. Excluded from the results was non-paramedic students (n=91), and non-research consent (n=315). The remaining 707 cases were included in the study.

Principal Components Analysis

The PCA was used to identify loading factors and dimensionality. The KMO results were 0.969, Bartlett's test of sphericity = 9371.874 (p < 0.001) and suggested an appropriate sample size. The PCA identified a single factor solution with an eigenvalue >1.0 and it explained 77.37% of the variance. The other 10 factors had eigenvalues <0.40. As a result of the PCA and visual inspection of the scree plot, a single-factor solution was identified (see Figure 1).

Unidimensionality

A single factor eigenvalue accounted for 79.42% of all variance. The correlation matrix (Table 1) has the
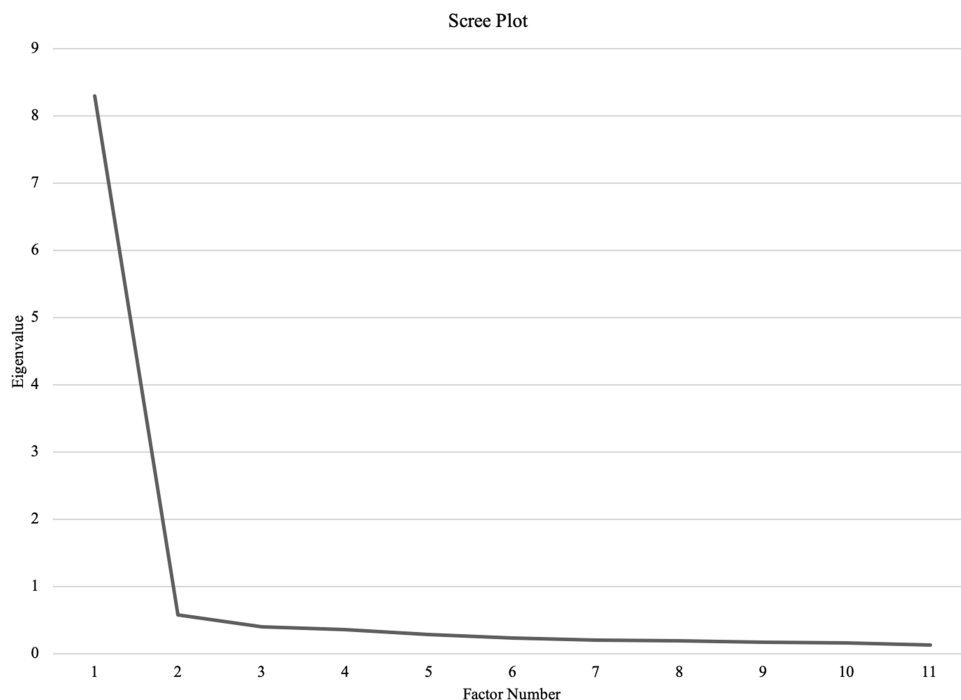

Figure 1 Principal component analysis scree plot.

Table 1 Correlation Matrix (Principal Component Analysis) (n = 707)

Item	Loadings	h^2	r_{it}	Mean	SD
Careful delivery of service	0.91	0.82	0.89	4.13	0.91
Empathy	0.91	0.83	0.90	4.17	0.88
Communication	0.90	0.81	0.88	4.13	0.90
Teamwork	0.90	0.80	0.88	4.20	0.92
Integrity	0.88	0.78	0.87	4.24	0.89
Patient advocacy	0.88	0.77	0.87	4.17	0.91
Respect	0.88	0.78	0.87	4.33	0.90
Self-motivation	0.88	0.78	0.87	4.17	0.95
Self-confidence	0.87	0.75	0.85	4.04	0.95
Time management	0.86	0.74	0.85	4.13	1.00
Appearance	0.81	0.66	0.80	4.30	0.92
Eigenvalues	8.51				
Explained variance	77.37%				

Abbreviations: h^2, communality; *rit*, corrected item-total correlations; SD, standard deviation.

communalities (h^2) account for variance in the variable and high values suggest extracted factors were reliable.

Raw Score

The raw scores were calculated by the sum value of all behavior ratings. The values could range from 11 to 55 with a passing score ≥ 33. Only 25 (3.5%) students received a failing score. Of the 707 cases 96.5% of the students passed and 28.1% (n=199) received perfect scores of 55. The average score was 46, S.D. 9.02, Cronbach alpha = 0.97. Figure 2 is a histogram of score distribution.

Partial Credit Model Analysis

The Rasch-Masters Partial Credit Model (PCM) performed on 707 cases. The logit values were used for ranking item difficulty from hardest to easiest (ranged −0.88–0.77, standard error 0.08–0.09). Table 2 ranks the items on average logit estimates, standard error, and correlation coefficient. Also supporting unidimensionality. The table also includes item Infit/Outfit values. Infit ranged from 0.74 to 1.54 with appearance underfitting the model, and outfit values from 0.67 to 1.47.

Fairness

The population had two demographic subgroups that were used to assess for fairness, Age and Gender. The average age of males was 28 (n=431, SD=6.65) and females 35 (n=165, SD=7.55). Gender was not a required field to report and the unspecified (n=111) cases were excluded from the analysis.

Differential Item Function

A differential item function (DIF) was performed to see if student gender influenced results for each item. Patient advocacy had a moderate to large DIF contrast logit = 0.80 (p < 0.005). Figure 3 displays the contrast between females and males.

A DIF analysis was performed between age subgroups that were grouped 18–24 (n=155), 25–44 (n=251), 45–64 (n=14), ≥ 65 (n=3). A positive DIF contrast exists in the age subgroup 45–64 years of age. The DIF contrast for Self-motivation logits = −0.57 (p < 0.05) and Self-confidence logits = 0.51 (p < 0.05). Figure 4 displays the contrast between age subgroups.

Discussion
Overall

This study was one of the very few to explore the psychometric properties of the ADT used by the US paramedic
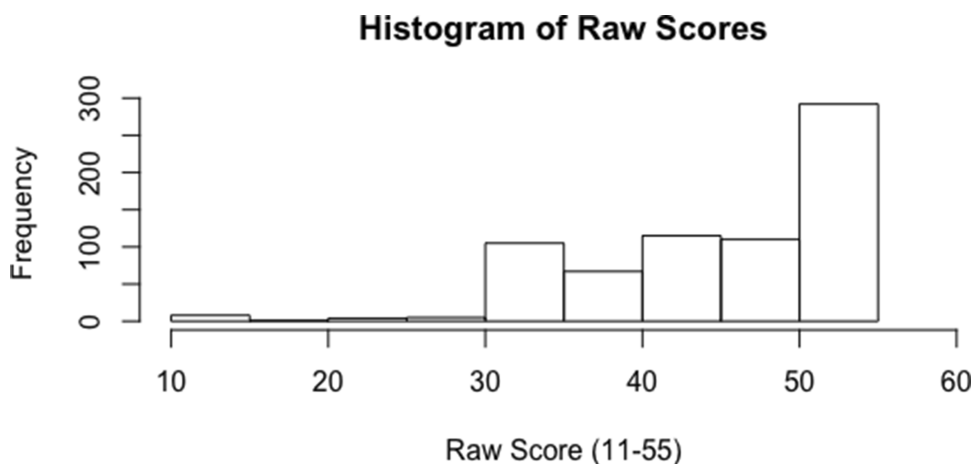

Histogram of Raw Scores

Figure 2 Histogram of raw scores.

Table 2 Partial Credit Model Table

Item	Theta	Point Measure Correlation	S.E.	Infit		Outfit	
				Mean Square	ZSTD	Mean Square	ZSTD
Self-confidence	0.77	0.87	0.08	1.00	0.02	1.09	1.16
Communication	0.33	0.91	0.09	0.80	−2.85	0.81	−2.44
Careful delivery of service	0.32	0.88	0.09	0.76	−3.36	0.76	−3.23
Time management	0.29	0.81	0.09	1.24	2.94	1.20	2.36
Empathy	0.11	0.88	0.09	0.74	−3.83	0.67	−4.44
Self-motivation	0.09	0.90	0.09	1.01	0.10	0.94	−0.68
Patient advocacy	0.09	0.86	0.09	0.98	−0.19	0.97	−0.37
Teamwork	−0.09	0.88	0.09	0.90	−1.33	0.83	−2.10
Integrity	−0.34	0.86	0.09	0.99	−0.18	0.91	−1.12
Appearance	−0.69	0.88	0.09	1.54	6.75	1.47	4.73
Respect	−0.88	0.90	0.09	0.99	−0.19	0.95	−0.53

and EMS sector. An essential part of this study was to understand the significance of this evaluation. Educators completed the evaluation in a standardized process and possibly the only assessment to fulfill an accreditation requirement to confirm a paramedic student had achieved competency in the affective domain. This was the first time ADT was psychometrically analyzed and reported with comprehensive results.

The results confirmed the evaluation tool assess generic types of professional behavior that were vague, imprecise; but nearly all students passed the evaluation. The cluster of perfect or near-perfect scores skewed the results and inflated reliability. The results highlight an issue that raised concerns for how educators had misunderstood the seriousness of ensuring paramedic students were prepared with an entry-level competence of the affective domain. The distribution of scores suggests educators utilized the ADT to document competency at an extreme level and refutes the claim for remedial purpose.
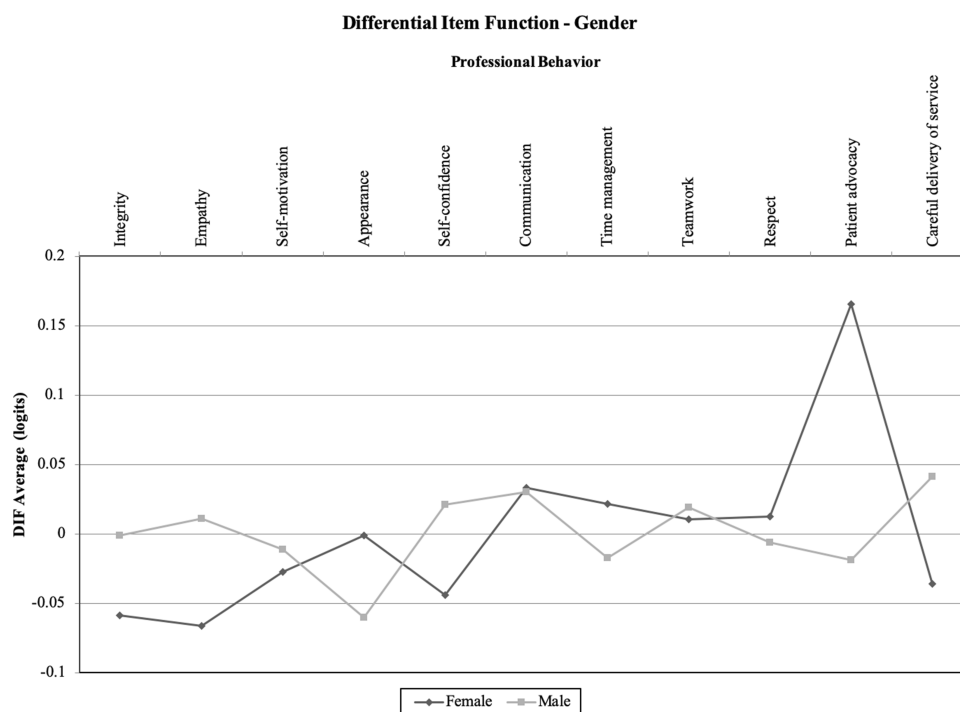

Figure 3 Differential item function by gender.

Differential Item Function - Age
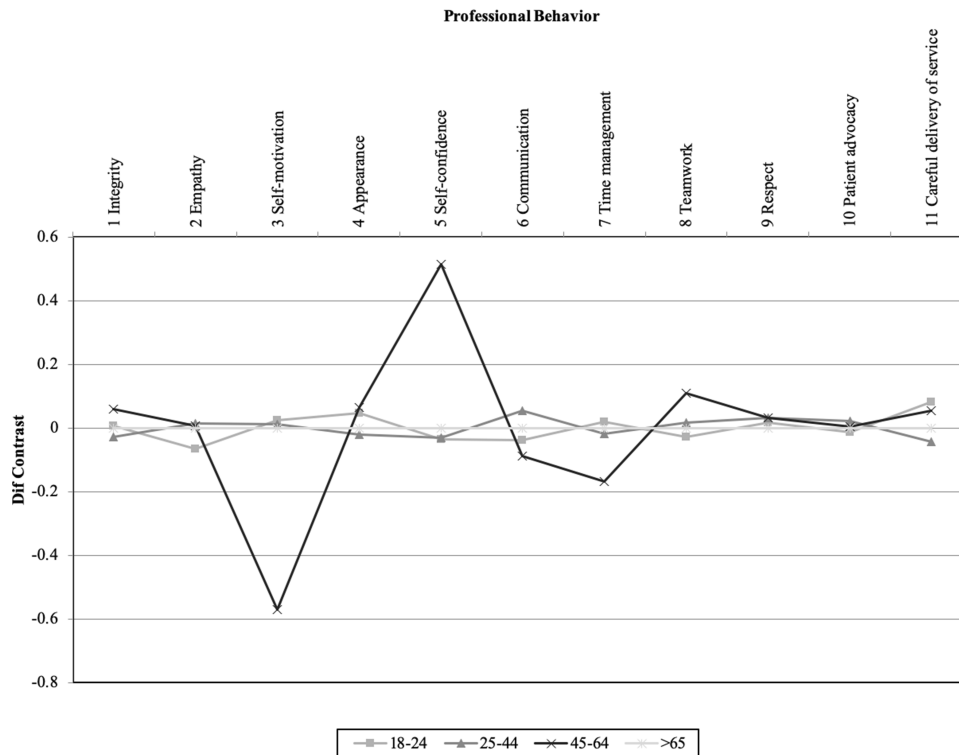
Professional Behavior

Figure 4 Differential item function by age.

Reliability

The reliability statistics for this study were much greater than those published by James in 2004.[19] This evaluation was overtly reliable with a Cronbach's alpha reliability of 0.97 and broke through the upper threshold. Values this high suggest either redundancy within the items or an issue with intra-rater reliability such as "rubber stamping".

The results of the PCM at the item response level support the idea of redundancy between the concepts. The response-level scoring had an increased positive correlation as the options described more professional behavior. This suggests a strong relationship as item difficulty increases high versus low performers were classified appropriately. All response options but the highest value (5) had negative response level correlations.

Cut Score

Essential elements of the standard setting process were not recorded. The ADT pass/fail point was provided in the instructions but how and why those values were agreed to were undocumented. The instructions stated an overall raw score of 33 equates to an acceptable score for an entry-level candidate for work force entry competency. If the ADT was utilized immediately before graduation from a paramedic program, then it is reasonable to think only 25 (3.5%) students failed and students that would have failed were dismissed earlier in the program. Even in that scenario, this evaluation should have been used to capture the behaviors that lead to the dismissal. In 2013 program, dismissal rate was greater than 30% for 69 (13%) of paramedic programs.[20] While the specific of dismissal were not published, these statistics provide evidence that nearly a third of all students failed to complete a US paramedic program. Greater attention needs to be given on documenting professional behaviors for not only students that complete a paramedic program but also those students that fail to complete programs. Identifying the differences in high and low performing students can help bridge the gap of student dismissal rates.

It is unclear why the results were so consistent, but a common-sense approach points to educators misunderstood the importance of the evaluation or the construction of the items introduces construct irrelevant variance. A possible cause for the outcome was that the evaluations were "pencil whipped" by educators. A total of 36% of

the overall raw scores ≥ 44 with 28% having perfect scores (55) and defined to always display role-model type behavior.

DIF Discussion

Fairness is an essential element for validity (Standards for Educational and Psychological Testing, 2014) but the ADT assessed student's ability unfairly. The item for "Patient Advocacy" functioned differently when compared with gender subgroups and it had a negative bias towards Males. Put simply, the behaviors described in this item do not accurately evaluate a male paramedic student's ability to be patient advocates.

Limitations

This study had a number of limitations. Firstly, some caution is required with the study results given a convenience sample technique was used. Secondly, there was no documentation if an educator was trained on how to use the evaluation tool, educator's intentions on the purpose of the evaluation, or their level of engagement with the student. This point requires further exploration given the importance educators play in student evaluation "sign-offs". Additionally, and finally, the sample is limited to a commercial database. A lack of research on this topic in this occupation created the inability to accurately compare the findings.

Conclusion

The ADT is not a valid and reliable assessment and the systemic errors with ratings and reduces the confidence in results. Evidence failed to support that the affective domain tool published by NHTSA is a valid and reliable measure of professional behaviors among paramedic students. The content and descriptions in the ADT were developed 20 years ago and fails to meet today's challenges. Serious consideration should be made before using this assessment. The item design was repetitive, confusing and introduce bias. Furthermore, the instructions failed to describe how to elicit accurate results that reflect a student's true ability. Future research should be aimed at replacing this evaluation that would include multiple raters in diverse settings and captured at predetermined milestones in the program that can be used to show progression.

Abbreviations

ADT, Affective Domain Tool; DIF, differential item functioning; EMS, Emergency Medical Services; MSQ, mean square; NHTSA, National Highway Traffic Safety Administration; PCA, Principal Components Analysis; PCM, Partial Credit Model.

Ethics Approval and Consent to Participate

Monash University Human Ethics Committee provided ethical approval. Participants included in this study consented to research and data were anonymized.

Author Contributions

LMB and BW conceived the idea for research. LMB provided statistical analysis. Both authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. O'Gara PE, Fairhurst W. Therapeutic communication part 2: strategies that can enhance the quality of the emergency care consultation. *Int Emerg Nurs*. 2004;12:201–207.
2. Williams B, Boyle M, Earl T. Measurement of empathy levels in undergraduate paramedic students. *Prehospital Disaster Med*. 2013;28(2):145–149. doi:10.1017/S1049023X1300006X
3. Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;130(25):e344–e426. doi:10.1161/CIR.0000000000000134
4. Kilner T. Educating the ambulance technician, paramedic, and clinical supervisor: using factor analysis to inform the curriculum. *Emergency Med J*. 2004;21(3):379–385. doi:10.1136/emj.2003.009605
5. Colwell C, Pons P, Pi R. Complaints against an EMS system. *J Emerg Med*. 2003;25(4):403–408. doi:10.1016/j.jemermed.2003.02.004
6. Gartell N San Pablo EMT in hot water over Instagram post of biker's mangled leg; 2016. Available from: http://www.eastbaytimes.com/2016/10/27/san-pablo-emt-in-hot-water-over-instagram-post-of-mangled-leg/. Accessed May 28, 2017.
7. Williams B, Fielder C, Strong G, Acker J, Thompson S. Are paramedic students ready to be professional? An international comparison study. *Int Emerg Nurs*. 2017;Volume 8(2):120–126. doi:10.1016/j.ienj.2014.07.004
8. Stoy WM, Paris G, Roth R. EMT-paramedic: National Standard Curriculum. The National Highway Traffic Safety Administration; 1998. Available from: www.ems.gov. Accessed February 6, 2019.
9. National Association of EMS Educators. 2002 National guidelines for educating EMS instructors; 2002. Available from: https://one.nhtsa.gov/people/injury/ems/instructor/instructor_ems/2002_national_guidelines.htm. Accessed January 8, 2020.
10. Bowen LM, Williams B, Stanke L. Professionalism among paramedic students: achieving the measure or missing the mark? *Adv Med Educ Pract*. 2017;8:711–719. doi:10.2147/AMEP.S137455
11. Pett M, Lackey N, Sullivan J. *Making Sense of Factor Analysis*. SAGE Publications, Inc; 2003.

12. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas*. 1960;20:141–151. doi:10.1177/001316446002000116

13. Henson R, Roberts K. Use of exploratory factor analysis in published research: common errors and some comment on improved practice. *Educ Psychol Meas*. 2006;66:393. doi:10.1177/0013164405282485

14. Masters G. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47(2):149–174. doi:10.1007/BF02296272

15. Ames A, Penfield R. An NCME instructional module on item-fit statistics for item response theory models. *Educ Meas*. 2015;34:39–48. doi:10.1111/emip.12067

16. Streiner DN, Cairney G. *Health Measurement Scales. A Practical Guide to Their Development and Use 5th Ed*. New York: NY Oxford University Press.; 2015.

17. Mantel N. Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure. *J Am Stat Assoc*. 1963;58: 690–700.

18. Tennant A, Pallant JF. DIF Matters: A Practical Approach to Test if Differential Item Functioning Makes a Difference. *Rasch Meas Trans*. 2007;20: 42007.

19. James B, Lindstrom J. Construct validity of the professional behavior evaluation instrument from the National Standard Paramedic Curriculum. *Prehospital Emergency Care*. 2004;8(4):434–435. doi:10.1016/j.prehos.2004.06.002

20. Tritt P Outcomes Threshold Report Summary 2015; 2016. Committee on Accreditation of Educational Programs for the Emergency Medical Services Professions. Available from: www.CoAEMSP.org. Accessed January 8, 2020.