

Assessing simulation-based clinical training: comparing the concurrent validity of students' self-reported satisfaction and confidence measures against objective clinical examinations

Owen B J Carter,¹ Brennen W Mills,¹ Nathan P Ross,² Alecka K Miles,² Jonathan M Mould,³ Robert P O'Brien⁴

¹Office of the Pro Vice-Chancellor (Health Advancement), Edith Cowan University, Joondalup Western Australia, Australia

²School of Medicine, Deakin University, Geelong, Victoria, Australia

³Division of Health Sciences, School of Nursing, Midwifery and Paramedicine, Curtin University, Bentley, Western Australia, Australia

⁴School of Medicine, The University of Melbourne, Parkville, Victoria, Australia

Correspondence to Brennen Mills, School of Medical and Health Sciences, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia; b.mills@ecu.edu.au

Accepted 2 March 2016
Published Online First
30 March 2016

ABSTRACT

Introduction Simulation-based education (SBE) literature is replete with student satisfaction and confidence measures to infer educational outcomes. This research aims to test how well students' satisfaction and confidence measures correlate with expert assessments of students' improvements in competence following SBE activities.

Methods N=85 paramedic students (mean age 23.7 years, SD=6.5; 48.2% female) undertook a 3-day SBE workshop. Students' baseline competence was assessed via practical scenario simulation assessments (PSSAs) administered by expert paramedics and confidence via a questionnaire. Postworkshop competence and confidence plus self-reported students' satisfaction were remeasured.

Results PSSA scores increased significantly between baseline and post workshop (35.7%→53.4%, $p<0.001$), as did students' confidence (55.7%→60.5%, $p<0.001$), and their workshop satisfaction was high (71.0%). Satisfaction and postworkshop confidence measures were moderately correlated ($r=0.377$, $p=0.001$). However, competence improvements were not significantly correlated with either satisfaction ($r=-0.107$, $p=0.344$) or change in confidence ($r=-0.187$, $p=0.102$).

Discussion Students' self-reported satisfaction and confidence measures bore little relation to expert paramedics' judgements of their educational improvements. Satisfaction and confidence measures appear to be dubious indicators of SBE learning outcomes.

INTRODUCTION

A worldwide trend identified in the undergraduate training of at least 18 health disciplines has been the adoption of simulation-based education (SBE) to complement the experiential learning of traditional clinical practicums.^{1–2} There is a large body of literature investigating the merit of SBE as an educational modality. However, a consistent criticism of most SBE research from 14 systematic reviews spanning 2005–2015 is a lack of methodological rigour that limits the ability to draw any strong conclusions from the literature.^{1–15} Not all agree; at least one review, drawing on 14 SBE studies, describes 'clear and unequivocal' evidence that SBE outcomes are 'powerful, consistent, and without exception' (p.709).¹⁶ Our own examination of these 14 studies reveals that, without exception, each suffered confounding from clear

dosage effects: groups received SBE in addition to their regular training and were compared to others receiving regular training alone. Norman states better than we might: 'Just as we need not prove that something is bigger than nothing, we also do not need to prove that something+something else is greater than something alone' (p.2).¹⁷ Such methodological shortcomings are emblematic of a lack of high-quality comparative studies within the literature.⁵

Another consistent criticism by reviewers is the heavy reliance on indirect measures, such as student satisfaction and confidence measures, to infer acquisition of skills.³ It is not surprising that satisfaction and confidence measures are popular within the SBE literature; self-reported questionnaires are convenient measures that readily allow quantification and statistical comparison. Furthermore, student satisfaction and improvements in confidence are highly desirable educational outcomes from a pedagogical perspective. However, there is good reason to suspect that indirect measures such as these—based as they are on students' subjective, self-reported data—are unreliable indicators of SBE learning outcomes. It is known that measurements of satisfaction are prone to inflation owing to socially desirable response bias, whereby participants tend to respond favourably out of politeness, apathy or ingratiation.¹⁸ Similarly, students are noted to significantly inflate self-evaluations of their clinical competencies compared to their instructors¹⁹ suggesting there is every chance that self-reported confidence measures are also inflated. For reasons such as these, many reviewers implore researchers to eschew self-reported, indirect measures and rather rely on the 'gold standard' of objective, independent judgements made by clinical supervisors in a standardised manner.^{3 17 20 21} However, as far as we are aware, there is little within the SBE literature that has investigated the criterion validity of self-reported satisfaction and confidence measures to infer SBE-related learning outcomes. Thus, before dismissing SBE research based on self-reported satisfaction and confidence measures, we sought to test the assumptions, formulated as falsifiable hypotheses, that after participating in SBE, independently assessed changes in students' competencies will be linear and positively related to their self-reported satisfaction (H_1) and change in confidence (H_2).



To cite: Carter OBJ, Mills BW, Ross NP, et al. *BMJ Stel* 2016;**2**:30–34.

METHODS

Participants

N=86 first-year undergraduate paramedic science students at Edith Cowan University (ECU), undertaking the unit 'Clinical Skills for Paramedic Practice', served as the participant pool for the present study. This unit was selected because it incorporates three consecutive days of intensive SBE during the mid-semester break to reinforce a step-by-step approach to the application of clinical skills, documentation and therapeutic communication. Students were guided in simulation-based practice by two clinical instructors who held bachelor degrees in paramedical science with 3 and 6 years' experience, respectively, as tertiary educators and 7 and 12 years' respective operational field experience. Students enrolled in the course were under no obligation to participate in the study and provided informed consent to take part. The ECU Human Ethics Committee provided clearance for the study. The final sample of consenting students comprised n=85 students, representing a participation rate of 98.8%. Their average age was 23.7 years (SD=6.5) and 48.2% were female.

Materials

The three measures selected for the study all directly related to the two falsifiable hypotheses. These were SBE learning outcome-based competency to be used as a measure of concurrent comparison with self-reported student satisfaction and confidence.

Competency

Participants' SBE learning outcome-based competency was assessed via practical scenario simulated assessments (PSSAs). After being provided with call-out information by an instructor, participants entered a room carrying a standardised medical kit to find a human patient actor awaiting diagnosis, stabilisation and to be made ready for transport. A standardised PSSA checklist marking instrument was adopted from the ECU School of Medical Sciences. The PSSA instrument was developed in two stages. First, a series of core competency items were derived from those featured in the 2011 second revision of the Australasian Paramedic Professional Competency Standards.²² In the second stage, a panel of content experts—teaching staff from two paramedicine programs within Australia—checked each item to ensure its content validity and that it would accurately reflect the learning outcomes of the 3-day SBE workshop. The final tool featured a list of 24 items against which instructors could score a participant's level of competency: 0, 'requires development'; 1, 'requires supervision' or 2, 'competent'. Scores for each item were aggregated to provide a final score of 48, which was then converted to a percentage.

Satisfaction

To measure participants' level of satisfaction, we adopted the simulation component of the Clinical Learning Environment Comparison Survey (CLECS) from the National Simulation Study validated at 10 nursing colleges across the USA.²³ Two paramedic experts reviewed each item in the questionnaire to confirm its direct relevance, content and face validity for the core competencies and learning outcomes of the SBE workshop. The resultant questionnaire featured 29 items providing participants with four response categories: 1, 'not met'; 2, 'partially met'; 3, 'met' or 4, 'well met'. Scores for each item were aggregated to provide a final score of 116, which was then converted to a percentage.

Confidence

The CLECS items were then modified to incorporate measures of student confidence with the same core competencies outlined in the satisfaction questionnaire. Response categories included 1, 'not confident'; 2, 'partially confident'; 3, 'confident' or 4, 'very confident'. To establish content validity, the same panel of expert paramedic staff reviewed the original 29 items and recommended the removal of 6 items, resulting in a confidence questionnaire of 23 items. The confidence scores for each item were then aggregated to provide a final score out of 92, also converted to a percentage.

Procedure

The day before the 3-day SBE workshop, all participants self-reported their current confidence in their core skills. They then undertook a PSSA requiring them to treat an actor-patient simulating an acute and severe asthmatic episode (preworkshop competence). By measuring the two variables in this order (confidence followed by competence), participants' confidence measures were not confounded by their subsequent performances. The participants then participated in the intensive, 3-day SBE workshop. The day after the workshop, they were asked to complete the self-reported confidence questionnaire a second time (postworkshop confidence) and then the SBE satisfaction survey, immediately followed by a second PSSA (postworkshop competence) where they were required to treat a patient-actor presenting with hypoglycaemia and an altered conscious state. We contemplated randomising the order of students' preworkshop and postworkshop PSSA tasks to eliminate the possibility of improvements being confounded by the second PSSA being easier than the first. However, our previous experiences suggested the risk would be too great of students conferring with one another after their preworkshop PSSAs and quickly divining their postworkshop PSSA task. Thus, we settled on standardised preworkshop and postworkshop scenarios selected by the panel of experts who judged each to be of equal difficulty.

Analysis

Paired samples t tests comparing changes in students' measures of confidence and competence before and after the workshop provided an indication of within-subject changes for these two measures. Within-subject differences between preworkshop and postworkshop measures were also calculated for each participant to provide a single score to indicate change for confidence and competence. Pearson product-moment correlation coefficients were then calculated between satisfaction, and preworkshop and postworkshop change in confidence and competence measures. A power analysis using G*Power (V.3.1.9.2) suggested our sample size (n=85) would provide an 87% chance of detecting a statistically significant ($\alpha=0.05$) bivariate correlation for variables with a 'moderate' relationship using Cohen's conventions to interpret effect size, which we determined as the minimum meaningful relationship ($r=0.30$, $R^2=9\%$). A multiple linear regression using the generalised least squares method was also conducted to examine unique variance between satisfaction and change in confidence as predictors of change in competence.

RESULTS

The entire sample completed all study aspects in their entirety (preworkshop competence and confidence; postworkshop competence, confidence and satisfaction), meaning there were no missing data.

Internal reliability

An internal reliability analysis of all three instruments was conducted using Cronbach's α . An α of ≥ 0.9 is considered a reflection of 'excellent' internal consistency.²⁴ An examination of postworkshop measures suggested all three met this criterion: PSSA $\alpha=0.940$, satisfaction $\alpha=0.947$ and confidence $\alpha=0.967$. This suggests each instrument had excellent internal consistency and measured a single, theoretical construct, justifying the use of aggregated scores for each measure.

Face validity

It was expected that the competency and confidence measures would reflect improvements as a direct result of participating in 3 days of intensive SBE, and if so, then a majority of students would be satisfied with the workshop. As both were repeated measures, participants' preworkshop scores were subtracted from their postworkshop scores to infer improvements in confidence and competency. The within-subject differences in competence before and after the workshop suggested significant and meaningful improvements (+17.7%) ($t(84)=7.336$, $p<0.001$). Similarly, participants enjoyed a significant and meaningful confidence boost (+4.8%) ($t(72)=5.121$, $p<0.001$). Consistent with these results, a large majority of students (70.4%) suggested that the SBE either 'met' (51.4%) or 'well met' (19.0%) their learning needs (figure 1). Thus, we considered all our measures to have face validity as they appeared to accurately reflect each construct as might be reasonably expected.

Concurrent validity

The three measures were then compared using Pearson correlation coefficient. As can be seen in table 1, significant repeated-measure correlations were observed for competence and confidence between participants' respective preworkshop, postworkshop and change measures. Modest correlations were observed between satisfaction and the confidence preworkshop and postworkshop measures. A modest correlation was also observed between the preworkshop measures of competence and confidence but not the postworkshop measures. Neither change

measures of competence nor confidence were significantly correlated with any other measure.

Predictive validity

The linear regression analysis suggested that changes in competency could not be predicted by participants' self-reported measures of satisfaction or changes in confidence; the linear regression model was non-significant ($f(2,74)=1.594$, $p=0.210$) and yielded a very modest adjusted $R^2=1.6\%$.

DISCUSSION

All three measures used in the present study demonstrated psychometrically sound properties, including good content and face validity and excellent internal reliability. Furthermore, participation rate was excellent, thereby minimising the likelihood of response bias, and our study was sufficiently powered to statistically detect even modest relationships between variables. The measures suggested participants had high levels of satisfaction with the 3-day workshop and demonstrated statistically significant improvements in their preworkshop versus postworkshop measures for confidence and competence. These results were in the anticipated direction and suggest our measures were sufficiently sensitive to detect the manipulations imposed on our participants, confirming the appropriateness of our experimental paradigm to test the study hypotheses.

H_1 predicted there would be a significant, positive linear correlation between measures of satisfaction and change in competence. This hypothesis was not supported; the correlation was negative, very small and non-significant (figure 2). H_2 predicted a significant, positive linear correlation between changes in confidence and competence measures. Likewise, this hypothesis was not supported; the correlation was non-significant and so small that it could be considered virtually non-existent (figure 3). The only results of significance were the autocorrelations of the repeated measures and their differences for competence and confidence. These correlations are of no surprise given their repeated-measure nature and little should be made of them; it would have been of far greater concern if they were not correlated. The intermeasure correlation between preworkshop

Figure 1 Mean preworkshop versus postworkshop comparisons of instructor-rated clinical competence, students' self-rated confidence and postworkshop student satisfaction ratings with the SBE (with 95% CI error bars). SBE, simulation-based education.

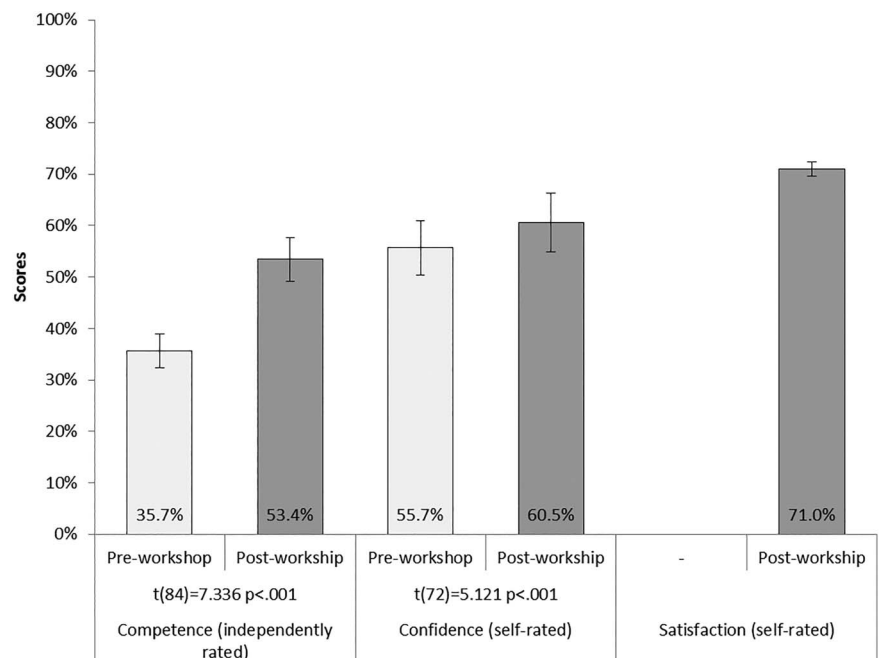


Table 1 Pearson's correlations between students' self-rated satisfaction (post) and confidence (pre, post and change) and instructor-rated competence (pre, post and change)

	Confidence			Competence		
	Pre	Post	Change	Pre	Post	Change
Satisfaction						
Post						
r	0.244	0.365	0.145	0.094	-0.012	-0.104
Sig.	0.030*	0.001**	0.212	0.385	0.915	0.338
Confidence						
Pre						
r	-	0.802	-0.242	0.267	-0.077	-0.200
Sig.	-	0.000**	0.028*	0.011*	0.481	0.065
Post						
r	-	-	0.386	0.206	-0.004	-0.167
Sig.	-	-	0.000**	0.055	0.972	0.122
Change						
r	-	-	-	0.102	0.077	-0.007
Sig.	-	-	-	0.362	0.489	0.950
Competence						
Pre						
r	-	-	-	-	0.347	-0.473
Sig.	-	-	-	-	0.000**	0.000**
Post						
r	-	-	-	-	-	0.662
Sig.	-	-	-	-	-	0.000**

Statistically significant (sig.) correlations denoted at the *0.05 level (two-tailed test) and **0.01 level (two-tailed test).

competence and confidence ($r_{pre}=0.267$) is of some interest, as it infers that instructors' assessments of competence at baseline may have been swayed in some way by the confident deportment of participants, regardless of actual skill. However, correlations do not impart causation, so it is just as likely to be indicating that participants who were objectively deemed more competent were appropriately more confident. Either way, this correlation was small and no correspondingly significant correlation between the respective postworkshop measures was

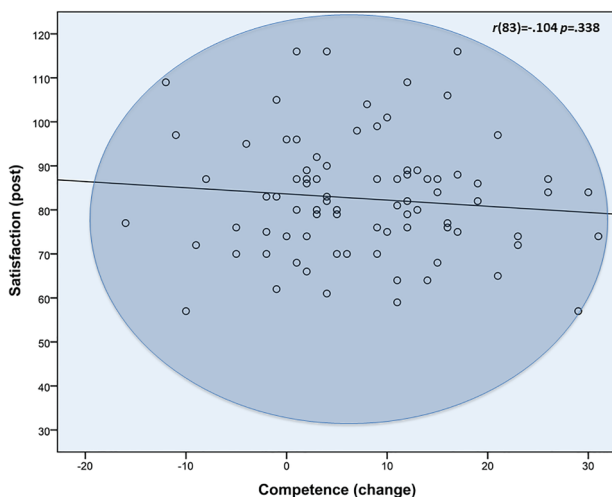


Figure 2 Relationship between self-reported satisfaction and instructor-assessed change in competency (H_1).

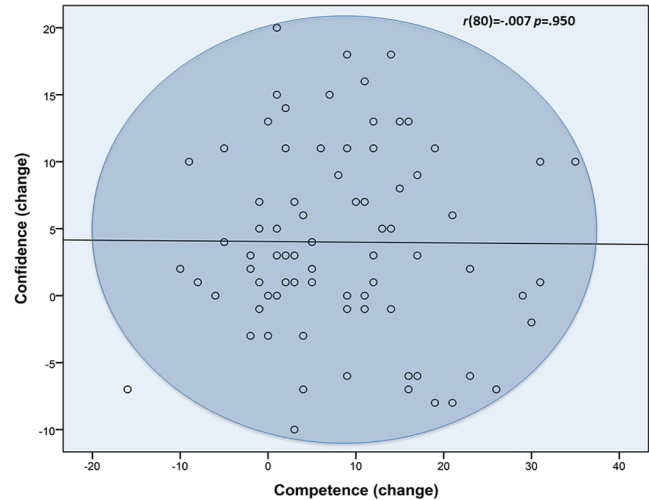


Figure 3 Relationship between change in self-reported confidence and change in competency (H_2).

observed, suggesting the influence of confidence on competence assessments (or vice versa) was, at best, inconsistent and unreliable.

The only other significant correlations were the satisfaction measure with preworkshop and postworkshop measures of confidence ($r_{pre}=0.244$ and $r_{post}=0.365$). These correlations appear reflective of both variables being related to a third, unknown subjective construct; the virtually non-existent relationship between these measures and the objectively assessed change in participants' competence suggest whatever the nature of the unknown construct, it was unlikely related to learning outcomes. Besides that, with such modest shared variance between satisfaction and confidence measures ($R^2_{pre}=5.9\%$ and $R^2_{post}=13.1\%$), the argument for measurement of a third construct is not strong. We deem it more likely that these significant correlations are an artefact of the shared origins of the measures, both originating from the CLECS. However, rather than viewing this as a weakness of our design, we consider it a strength that these measures shared so little variance, suggesting they were largely measuring different constructs despite their common geneses.

Overall, our data suggested that the students' self-rated satisfaction and their self-reported change in confidence due to participation in the SBE workshop bore remarkably little relation to independent, expert assessment of changes in their competency. This is consistent with previous research suggesting students' self-assessments of their clinical skill tend to be inflated compared to their instructors' assessments.¹⁹ It also goes some way to vindicate those calling for abandonment of indirect, self-reported measures, such as satisfaction and confidence, to infer merit in education modalities such as SBE.¹⁷ However, it is beyond the scope of the present study to suggest self-reported measures are without any merit. For instance, it may well be relevant to assess confidence as one dimension of attitude change in relation to SBE, especially in the workplace where skills acquisition is already achieved and SBE is used for other learning outcomes such as interprofessional practice.

In terms of study limitations, owing to the within-subject nature of our study design, it remains possible that the improvements in students' competence scores could be entirely explained by the postworkshop scenario simply being easier than its preworkshop comparator. We doubt this was the case as

the panel of experts deemed the PSSA scenarios to be of equivalent difficulty, and the modest but meaningful improvement in students' mean competency scores is a logical outcome of attending the 3-day workshop. However, we acknowledge the possibility remains. The reach of the present results is also undetermined. It is beyond the scope of this study to suggest that our findings are generalisable beyond paramedic students, or indeed our immediate sample of second-year paramedic students at ECU. Both potential limitations of our study can best be demonstrated through replication of our method with different cohorts and different disciplines. To this end, we encourage criticisms of our findings and would welcome contradiction of our results through empirical replication. In the interim, our data suggest that students' self-reported satisfaction and confidence measures are poor predictors of actual improvements in competency related to SBE. We therefore caution future studies of SBE against the use of student self-reported measures of confidence and satisfaction as proxy measures for skill acquisition.

Twitter Follow Owen Carter at @owencarter

Contributors OBJC and BWM conceived the study and analysed the data. NPR, AKM, JMM and RPO provided expertise in development of materials and data collection. All helped draft the final manuscript.

Competing interests None declared.

Ethics approval Edith Cowan University Human Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Raw data are available on request by contacting OC.

REFERENCES

- Hallenbeck VJ. Use of high-fidelity simulation for staff education/development: a systematic review of the literature. *J Nurs Staff Dev* 2012;28:260–9.
- Health Workforce Australia. *Use of Simulated Learning Environments (SLE) in professional entry level curricula of selected professions in Australia*. Adelaide: Health Workforce Australia, 2010. <http://www.hwa.gov.au/sites/uploads/simulated-learning-environments-2010-12.pdf> (accessed 26 Aug 2015).
- Cant RP, Cooper SJ. Simulation-based Learning in nurse education: systematic review. *J Adv Nurs* 2010;66:3–15.
- Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;27:10–28.
- Jayasekara R, Schultz T, McCutcheon H. A comprehensive systematic review of evidence on the effectiveness and appropriateness of undergraduate nursing curricula. *Int J Evid Based Healthc* 2006;4:191–207.
- Laschinger S, Medves J, Pulling C, et al. Effectiveness of simulation on health profession students' knowledge, skills, confidence and satisfaction. *Int J Evid Based Healthc* 2008;6:278–302.
- Norman J. Systematic review of the literature on simulation in nursing education. *ABNF J* 2012;23:24–8.
- Rudd C, Freeman K, Swift A, et al. *Use of simulated learning environments in nursing curricula*. Adelaide: Health Workforce Australia, 2010. <http://www.hwa.gov.au/sites/default/files/sles-in-nursing-curricula-201108.pdf> (accessed 26 Oct 2015).
- Cheng A, Lockett A, Bhanji F, et al. The use of high-fidelity manikins for advanced life support training—a systematic review and meta-analysis. *Resuscitation* 2015;93:142–9.
- Weaver A. High-fidelity patient simulation in nursing education: an integrative review. *Nurs Educ Perspect* 2011;32:37–40.
- Cook DA, Brydges R, Zendejas B, et al. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med* 2013;88:1178–86.
- Cook DA, Hamstra SJ, Brydges R, et al. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med Teach* 2013;35:e867–98.
- Cheng A, Lang TR, Starr SR, et al. Technology-enhanced simulation and pediatric education: a meta-analysis. *Pediatrics* 2014;133:e1313–23.
- Ilgen JS, Sherbino J, Cook DA. Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. *Acad Emerg Med* 2013;20:117–27.
- Mundell WC, Kennedy CC, Szostek JH, et al. Simulation technology for resuscitation training: a systematic review and meta-analysis. *Resuscitation* 2013;84:1174–83.
- McGaghie W, Issenberg S, Cohen E, et al. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med* 2011;86:706–11.
- Norman G. Data dredging, salami-slicing, and other successful strategies to ensure rejection: twelve tips on how to not get your paper published. *Adv Health Sci Educ* 2014;19:1–5.
- Parmac Kovacic M, Galic Z, Jerneic Z. Social desirability scales as indicators of self-enhancement and impression management. *J Pers Assess* 2014;96:532–43.
- Lee-Hsieh J, Kao C, Kuo C, et al. Clinical nursing competence of RN-to-BSN students in a nursing concept-based curriculum in Taiwan. *J Nurs Educ* 2003;42:536–45.
- Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004;38:199–203.
- Sloan DA, Donnelly MB, Schwartz RW, et al. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995;222:735–42.
- Council of Ambulance Authorities. *Paramedic professional competency standards: Version 2.2*. Paramedics Australasia, 2011. http://caa.net.au/~caanet/images/documents/accreditation_resources/Paramedic_Professional_Competency_Standards_V2.2_February_2013_PEPAS.pdf (accessed 10 Nov 2015).
- Hayden JK, Smiley RA, Alexander M, et al. The NCSBN National Simulation Study: a longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *J Nurs Regul* 2014;5:S1–S64.
- Kline P. *The handbook of psychological testing*. 2nd edn. London: Routledge, 1999. ISBN 0415211581.