The effectiveness of terminal feedback delivery to improve the clinical skills of pre-registration health professional students in simulated environments: A systematic review

J. Stanhope[1, 2] *& R. Gill*[1, 3, 4]

Aim: To determine whether terminal feedback is effective in improving the practical, clinical skills of pre-registration health professional students in simulated environments, and how it is best delivered.

Methods: Eight databases were searched systematically to identify studies addressing our aim. Eligible studies were published 2004–2014, in English language, within peer-reviewed journals. Study designs were determined and assigned to the National Health and Medical Research Council hierarchy of evidence, with the PEDro scale used to appraise ≥Level III_1 studies. Data regarding the country, participant characteristics, sample size, task critiqued, prior experience with the task, feedback delivered, follow-up time, outcome measures, methods of statistical analysis and the results were extracted and reported descriptively.

Results: Eight studies (Levels II–V) were included. Most investigated medical students performing laparoscopic and/or knot tying tasks. PEDro scale scores ranged from 18–54%. Results were mixed when comparing terminal verification feedback and no feedback (n=3), and comparing different types of feedback (n=3). Overall, it would appear that, in terms of terminal feedback, elaborative feedback is more effective in improving the practical, clinical skills of pre-registration health professional students (n=4).

Discussion: Whilst there is some indication that elaborative feedback may be more effective than no feedback, the small number of studies and poor methodological quality of the included studies precludes any strong conclusions from being drawn. There

1 School of Health Sciences, University of South Australia
2 School of Biological Sciences, University of Tasmania
3 Country Health South Australia
4 Physiotherapy, School of Health Sciences, Flinders University of South Australia

Correspondence:
Robyn Gill
School of Health Sciences
Health Sciences Building
Flinders University of South Australia
216 Daws Road
Daw Park, SA 5041
Australia
Email: robyn.gill@flinders.edu.au

ISSN 1442-1100

are a number of gaps in the current literature, particularly investigating non-medical professions, that should be addressed in future research. Furthermore, larger-scale, well-designed studies in the area are required to guide clinical teaching.

Keywords: review; education, health; feedback; universities; clinical competence.

Introduction

Feedback has been defined as "specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance" (van de Ridder, Stokking, & McGaghie, 2008, p. 193) and is considered an important component of teaching clinical skills (Branch & Paranjape, 2002), which shapes and maintains the student's appropriate responses (Skinner, 1968).

We were particularly interested in feedback that could be applied to summative assessments as, although they are primarily designed to assess competence, they may also serve as an important clinical teaching opportunity. In practical classes, only limited time can be spent providing students with feedback on their skills, and some students are more likely to request and engage with feedback than others. For instance, in their qualitative study, Bok et al. (2013) found that during their clinical clerkships, medical students tended to seek feedback with the intention of receiving positive reinforcement, whilst avoiding feedback with potentially negative consequences. This same phenomenon is likely to occur within practical classes, with the consequence of students who are most likely in need of feedback being less likely to request, and therefore receive, formative feedback. As such, summative assessments provide a potentially important opportunity for all students to receive feedback on their performance. A recent study reported that 95.7% of the medical students participating in their study viewed their summative feedback (Harrison et al., 2013), indicating that students are interested in the feedback that is delivered following summative assessments. This review investigates terminal feedback delivered following the performance of a clinical task, as this may be applied to summative assessments.

The delivery of feedback is an important consideration, as it may have positive or detrimental impact, depending on the manner in which it is provided (Shute, 2008). Parameters for feedback delivery include the format it is delivered (e.g., written or verbal), group or individual feedback and verification or elaborative. Verification feedback provides the learner with information about the correctness of his or her response (Archer, 2010). Also known as knowledge of results (Jaehnig & Miller, 2007), this type of feedback provides the least information, as the learner is only given information on whether their answer is correct or not, and they are not given the correct answer. In elaborative feedback, additional information or an explanation of the correct answer is provided (Archer, 2010). The potential benefit of elaborative feedback would need to be weighed against the added time, and therefore costs, involved in its provision.

This review sought to answer the following questions:

1. Does the provision of terminal feedback to pre-registration health professional students improve their practical skills in simulated environments?
2. How should feedback be delivered in this situation to optimise learning outcomes?

ISSN 1442-1100

For the purposes of this review, health professions included medical, nursing, midwifery, paramedic, dentistry and allied health. Practical, clinical skills refers only to hands-on skills in simulated environments (e.g., not with real patients). We only considered feedback from tutors or devices, rather than feedback provided by patients or peers.

Methods

A search strategy was developed based upon a PICO framework:

- Population (P): Pre-registration health professional students
- Intervention (I): Terminal feedback from a tutor or scoring device regarding students' practical, clinical skills performed in a simulated environment
- Comparison (C): No feedback or another type of terminal feedback from a tutor or scoring device regarding students' practical, clinical skills performed in a simulated environment
- Outcome (O): A change in the performance of the practical, clinical skill.

Using the terms reported in Table 1, in August 2014, JS conducted a search in eight educational and allied health databases: Education Research Complete, Academic Search Premier, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Health Source—Nursing/Academic Edition (all EbscoHost), Scopus (SciVerse), Education Resources Information Center (ERIC) (ProQuest), Medline (OvidSP) and Informit—Health Collection. The term "feedback" was searched within title only, whereas all other terms were searched within the title, keywords and abstract, with related terms also searched, where permitted by the databases. Where possible, searches were limited to peer-reviewed and English language, published 2004–2014. If the database did not allow this, studies were manually excluded on this basis.

All studies obtained were exported into an Endnote library where duplicates were removed and titles and abstracts screened for inclusion according to the exclusion criteria reported in Table 2. Full texts were then obtained, and excluded, using the same criteria.

Table 1
Search Terms

Feedback*
AND
medic* OR nurs* OR physiotherap* OR therap* OR podiatr* OR dentis* OR midwi* OR surg* OR paramed* OR prosthet* OR orthotist* OR audiol* OR optom* OR "speech path*"
AND
student*
AND
perform* OR skill* OR techni* OR abilit* OR practic* OR clinic*
AND
chang* OR improv* OR effect* OR impact*

ISSN 1442-1100

Table 2
Exclusion Criteria

General	• Published in a language other than English • Published in a non-peer-reviewed journal • Not published in full text (e.g., conference abstract) • Published prior to 2004 • Opinion papers (e.g., editorials, narrative reviews) • Qualitative studies • Pre-post studies with no comparative group
Participants	• Did not investigate pre-registration health professional students
Intervention	• The task was not a practical, clinical task • Feedback was not provided by a tutor/teacher or a scoring device, based on performance of the task • Feedback was provided during the task • If feedback was only provided with another intervention (except where the additional intervention was also provided to the comparison groups).
Comparison	• The comparison feedback was not provided by a tutor or scoring device • The comparison feedback was provided during the task
Outcomes	• Did not investigate a change in the performance of practical skills (e.g., only investigated students views regarding feedback delivery) • Practical skills were assessed by a tutor or device (e.g., not a score given by a patient or the student)

Study designs were determined and allocated to the National Health and Medical Research Council (NHMRC) hierarchy of evidence (NHMRC, 2009). The risk of bias was determined by a trained PEDro rater (JS) using the PEDro scale (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003) for studies of Level III_1 or higher. The PEDro scale was developed for appraising clinical trials for the Physiotherapy Evidence Database (Maher et al., 2003). Although the scale was designed for physiotherapy studies, it consists of generic items that are applicable to other research topics. A common issue with critical appraisal tools is that they lack validity and reliability evidence and do not provide clear instructions regarding their use (Crowe & Sheppard, 2011). The PEDro scale has been shown to be reliable (Maher et al., 2003) and valid (de Morton, 2009; Macedo et al., 2010), making it suitable for use in this review. Lower-level designs were not appraised due to the biases inherent in their study designs.

One reviewer (JS) extracted data regarding: the country (first author's affiliation if where the study was conducted was not stated); participant characteristics; sample size; task critiqued; prior experience with the task; feedback delivered; follow-up time; outcome measures; methods of statistical analysis; significant results. Data are reported descriptively.

Given the broad nature of the review questions, it was expected that there would be heterogeneity in the population, interventions and outcome measures; hence, a meta-analysis was not conducted.

ISSN 1442-1100

Results

The database search yielded 620 studies, eight of which were included in this review (see Figure 1: *Flow chart for study inclusion*). Four studies were from the United States of America (Boehler et al., 2006; Lazarski, Susarla, Bennett, & Seldin, 2007; O'Connor, Schwaitzberg, & Cao, 2008; Van Sickle et al., 2005), three from Canada (Manzone, Tremblay, You-Ten, Desai, D., & Brydges, 2014; Porte, Xeroulis, Reznick, & Dubrowski, 2007; Triano, Scaringe, Bougie, & Rogers, 2006) and one was from Denmark (Strandbygaard et al., 2013).

Six studies were randomised controlled trials (RCTs) (Level II), one was a prospective cohort (III_2) and one was a single-subject shifting baseline design (Level IV) (see Table 3); therefore, most studies were high-level designs. The PEDro score for the RCTs

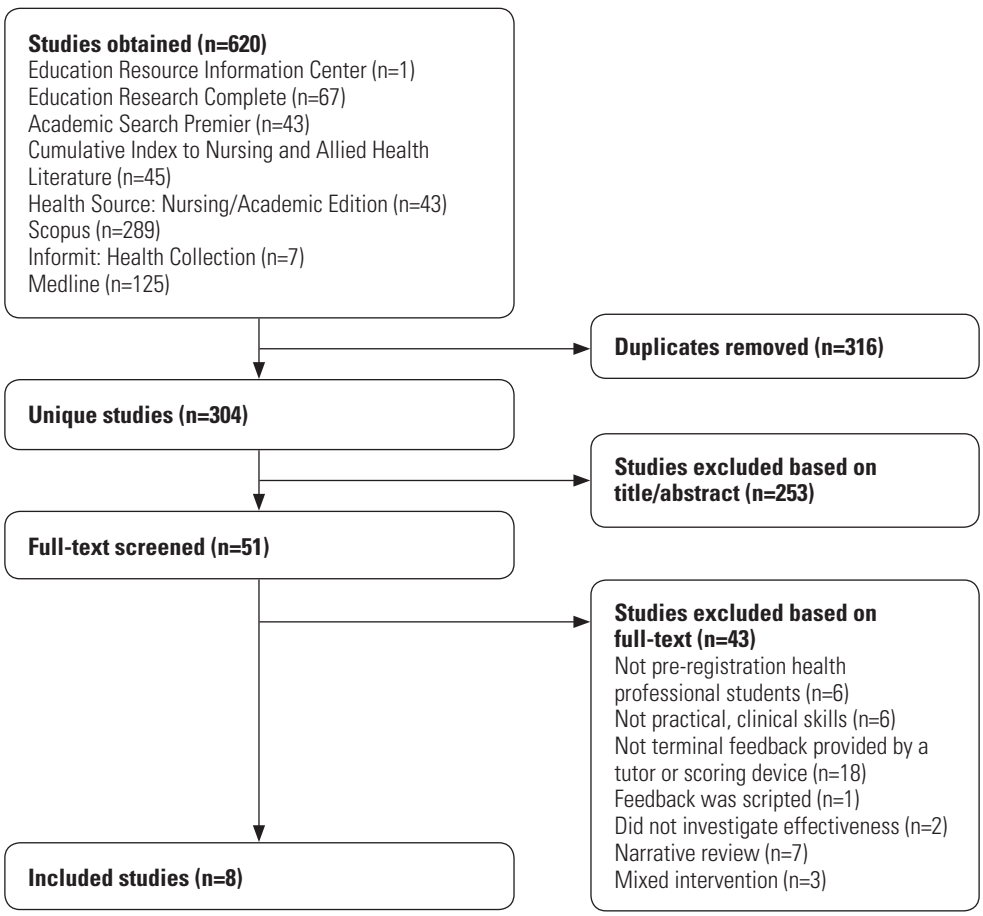

Figure 1. Flow chart for study inclusion.

ISSN 1442-1100

ranged from 18–54% (see Table 4), indicating that despite using high-level designs, these studies are at high risk of bias, and their results should be considered with caution. Most studies investigated medical students, with most tasks being laparoscopic and/or knot tying techniques (see Table 3 and 5). Boehler et al. (2006) did not report the participant's level of experience in the task; however, all other studies included novices only. The outcome measures included those determined by devices and by expert/tutor ratings, and studies varied widely in their follow-up times. Details of these, as well as the statistical analyses, are reported in Table 5.

Comparisons between feedback and no feedback were investigated in six of the studies; two compared verification feedback and no feedback (Lazarski et al., 2007; O'Connor et al., 2008), and five compared elaborative feedback and no feedback (Boehler et al., 2006; O'Connor et al., 2008; Strandbygaard et al., 2013; Triano et al., 2006; Van Sickle et al., 2005).

Mixed results were reported from the three studies comparing verification feedback and no feedback. O'Connor et al.'s (2008) found there to be a significant difference in the learning curves, performance variance and scores for the last trial between the groups, in terms of time, path length and smoothness, in favour of the feedback group. The feedback group had received feedback regarding the time taken, path length and smoothness of tool manipulation, whereas the control group received no feedback.

In contrast, Lazarski et al. (2007) found there to be no significant difference in the performance scores between feedback and control groups. Their study involved four intervention groups: one received verification feedback only; one verification feedback and standardised, written instructions; one standardised, written instructions only; and the last group received no feedback or instructions.

Table 3
Study Design, Health Professions and Sample Sizes

Study	Study design (NHMRC level of evidence)	Health profession (level of training)	Sample size
Manzone et al. (2014)	RCT (II)	Medical students (pre-clerkship)	41
Strandbygaard et al. (2013)	RCT (II)	Medical students (Years 4–6)	99
O'Connor et al. (2008)	RCT (II)	Medical students (Years 1–2)	9
Porte et al. (2007)	RCT (II)	Medical students (Year 1)	45
Lazarski et al. (2007)	Prospective cohort with concurrent controls (III_2)	Pre-clinical medical and dental students (not reported)	61
Boehler et al. (2006)	RCT (II)	Medical students (Years 2–3)	33
Triano et al. (2006)	RCT (II)	Chiropractic students	40
Van Sickle et al. (2005)	Single subject shifting baseline (IV)	Medical students (Years 1–2)	12

Note: NHMRC—National Health and Medical Research Council; RCT—randomised controlled trial

ISSN 1442-1100

Table 4
Task, Follow-up Time, Outcome Measures and Statistical Analysis

Study	Task performed	Follow-up time	Outcome measures	Statistical analysis
Manzone et al. (2014)	Intubation	20 trials before post-test; approximately 2 weeks for retention test	Number of hand movements measured using Polhemus PATRIOT hardware; ratings of surgical skills using the Global Rating Scale (GRS)	Analysis of covariance & analysis of variance
Strandbygaard et al. (2013)	Laparoscopic salpingectomy	Participants could practise up to 3 hours/day for up to 2 months until "expert level" achieved	Number of repetitions and time practised to reach expert level; performance score when expert level reached (determined by virtual reality simulator)	General linear univariate model; Mann-Whitney test
O'Connor et al. (2008)	Laparoscopic knot-tying & suturing	Practiced 1 hour/ day, 6 days/week over 4 weeks	Task performance examined for time, smoothness of instruments and instrument path length using ProMIS optical tracking system; knot errors	ANOVA; Post-hoc Tukey analysis
Porte et al. (2007)	Interrupted suturing and instrument knot-tying	18 trials before post-test; retention test 1 month later with no practice	Expert rating using a modified GRS; computer assessment: total distance, time per movement, number of movements and average speed of each movement	Kruskal-Wallis; Mann-Whitney U test with Bonferroni adjustment; one-way ANOVA
Lazarski et al. (2007)	Fixation of fractured mandible	5 attempts – all measured	Strength of fixation	Tukey honest significant difference procedure
Boehler et al. (2006)	Knot tying	Not reported	Final trial videotaped and reviewed by three individuals	Paired-sample t-tests
Triano et al. (2006)	Spinal manipulation	A 10-minute distraction task was performed between the feedback and post-test	Total force and total movement	Student t-test
Van Sickle et al. (2005)	Knot tying	Not reported	Knot quality	Unpaired t-test; Fisher's exact test; one-way ANOVA

ISSN 1442-1100

Table 5
PEDro Scale Results

	Manzone et al. (2014)	Strandbygaard et al. (2013)	O'Connor et al. (2008)	Porte et al. (2007)	Triano et al. (2006)	Boehler et al. (2006)
Eligibility criteria were specified		✓				
Subjects were randomly allocated to groups	✓	✓	✓	✓	✓~	✓
Allocation was concealed		✓				
Groups were similar at baseline regarding the most important prognostic indicators					✓	✓
There was blind of all subjects						
There was blinding of all therapists who administered the therapy					^	
There was blinding of all assessors who measured at least one key outcome	✓			✓	^	✓
Measure of at least one key outcome were obtained from > 85% of the subjects initially allocated to groups	✓	✓			^	
All subjects for whom outcome measures were available received the treatment or control condition as allocated, or where this was not the case, data for at least one key outcome was analysed by "intention to treat"						
The results of between-group statistical comparisons are reported for at least one key outcome	✓	✓	✓			✓
The study provides both point measures and measures of variability for at least one key outcome	✓	✓		✓		
Total (%)	**5 (45)**	**6 (54)**	**2 (18)**	**3 (27)**	**3 (27)**	**4 (36)**

✓ A point was awarded
~ Participants were paired and then randomised to groups, with no explanation regarding how or why they were paired prior to randomisation. A point has been awarded in the scoring.
^ Study states single-blinded; however, it does not state the type of blinding. One point was awarded.

Triano et al. (2006) compared the provision of visual feedback, in the form of a graph of total force and movement for student performance and that of an expert, with no feedback. They did not perform between-group comparisons for the outcomes of interest in this review; however, they found that those receiving feedback had significant improvements in force duration, duration of moment load, speed of force and moment, and mean peak amplitude force, but not moment. They did not perform the same analysis for those in the control group.

ISSN 1442-1100

Given there are only three studies, of poor methodological quality, it remains unclear whether verification feedback is effective in improving the simulated, practical, clinical skills of pre-registration health professional students. There is a clear need for high-level, well-designed studies investigating the effectiveness of verification feedback in comparison with no feedback in this population.

A comparison of the effectiveness between elaborative feedback and no feedback was investigated in four of the included studies (Boehler et al., 2006; O'Connor et al., 2008; Strandbygaard et al., 2013; Van Sickle et al., 2005). Three of these (Boehler et al., 2006; O'Connor et al., 2008; Strandbygaard et al., 2013) performed between-

Table 6
Details of Studies Comparing Elaborative Feedback and No Feedback

Study	**Feedback intervention**	**Control intervention**	**Findings**
Strandbygaard et al. (2013)	Instructor feedback (1–3 sessions) using a standardized template covering how to hold instruments, optimal use of electric cautery and removal of the fallopian tube. This feedback was not individually tailored.	No feedback	Time and repetitions performed to reach the expert level was significantly lower in the FG in comparison with the CG; the performance score was significantly higher for the CG in comparison with the FG
O'Connor et al. (2008)	Elaborative feedback (as for verification, and instructions on how to improve their performance)	No feedback	Learning curves, performance variance and scores for the final trials (in terms of time, path length and smoothness) were significantly better for the FG in comparison with the CG; the FG had significantly lower overall workload and number of errors than the CG
Boehler et al. (2006)	Elaborative feedback (personal, immediate, constructive feedback)	Demonstration & complements (scripted, pre-arranged)	The FG had significantly better performance ratings than the CG
Van Sickle et al. (2005)	FG 1: Elaborative feedback (how to improve their task performance) FG 2: Elaborative feedback (how to improve their task performance) & observation of their final pre-test	CG 1: final pre-test observed CG 2: all trials observed	The knot quality scores (KQS), completion time and slip percentage did not improve for any participant in CG1. In CG 2, 2 of 3 participants had significant improvements in KQS and slip percentages, with no change in completion time on average, with one taking significantly longer post-feedback. In FG1, 2 of 3 participants had significant improvement in KQS and 1 of 3 in slip percentage. After feedback, those in FG2 on average took longer to complete the task. There were no between group analyses.

ISSN 1442-1100

group comparisons, all in favour of elaborative feedback over no feedback. A summary of the interventions and the results are reported in Table 6. Although this suggests that elaborative feedback may lead to a greater improvement in simulated, practical, clinical skills for pre-registration health professional students, the small number of studies and poor methodological quality limits the strength of this finding.

Three studies (Manzone et al., 2014; O'Connor et al., 2008; Porte et al., 2007) investigated different types of feedback delivery. Manzone et al. (2014) compared the effectiveness of four types of feedback: task-oriented numerical, ego-oriented numerical, task-oriented comment and ego-oriented comment. For task-oriented numerical feedback, participants were given feedback regarding the number of hand movements and the time taken to complete the task and were asked to plot their performance curves on graphs for both time and hand movements based on the feedback provided. Those in the ego-oriented numerical group also received feedback on their time and number of movements, as well as the scores of others (medical students, second-year anaesthesia residents, anaesthesia fellows and anaesthetists). The task-oriented comment group received feedback regarding their strengths and weaknesses from an anaesthesia fellow, and participants were then asked to record the feedback. Finally, for those in the ego-oriented comment group, the anaesthesia fellow made comments as to which "training category" the participant fit (e.g., senior resident, staff member), and the participant then recorded their feedback. In this study, participants were to perform endotracheal intubation in four variations: normal, supine, lateral and ice pick.

With regards to the number of hand movements and scores on the Global Rating Scale, there were no significant differences between the groups for the normal, lateral and ice pick variants for both the post-test and retention test. At the post-test, the numerical groups performed significantly better in the supine variant than the comments groups; however, there were no significant differences at the retention test. With regards to the Global Rating Scale scores in the supine variation, the task-oriented groups were significantly worse at post-test than the ego-oriented groups, but there were no significant differences at the retention test. It therefore appears that, in the longer term, feedback delivery does not impact upon the performance of endotracheal intubation under these conditions.

Three types of feedback were investigated by Porte et al. (2007): motion analysis feedback without criterion, motion analysis feedback with criterion and expert feedback. Motion analysis feedback included the number of movements made, and the criterion provided was the number of movements made by an expert surgeon performing the same task. Expert feedback included ways of improving performance and followed a script. Demonstrations and questions were permitted. Scores on the modified Global Rating Scale improved significantly following feedback for all groups at post-test; however, only those receiving expert feedback retained scores significantly greater than pre-test at the retention test. Similarly, the significantly improved scores from pre-test to post-test were not sustained into the retention test for those who received motion analysis feedback. Those who received expert feedback improved significantly from pre- to post-test, with these scores maintained in the retention test. Between-group analysis was not conducted; however, it would appear that expert feedback had longer lasting impact on skill performance than receiving motion analysis with or without criterion.

ISSN 1442-1100

In addition to investigating verification feedback and no feedback, and elaborative feedback and no feedback (as discussed above), O'Connor et al. (2008) also compared those receiving these two forms of feedback. They found no significant difference in the learning curves or performance variance for time, path length and smoothness between those receiving elaborative or verification feedback. Those receiving elaborative feedback had significantly lower overall workload and number of errors than those receiving verification feedback. Elaborative feedback may, therefore, be favourable to verification feedback; however, it should be noted that although O'Connor et al.'s (2008) study was an RCT, they only had nine participants across three intervention groups and scored poorly on the PEDro scale. As such, these results should only be considered an indication of the potential benefits of elaborative feedback over verification feedback in teaching clinical skills to pre-registration health professional students.

Discussion

This is the first systematic review to investigate the effectiveness of terminal feedback and feedback delivery methods in improving the practical, clinical skills of pre-registration health professional students in simulated environments. There is a paucity of literature in the area, with only eight relevant studies identified. The majority of research was from North America, indicating a need for feedback styles to be examined within the context of health professional education in Australia and New Zealand, particularly given recent changes to university funding in Australia.

The eight included studies (NHMRC Level II to IV) were at a reasonably high risk of bias, with the included RCTs only achieving PEDro scale scores ranging from 18–54%. Key issues for these RCTs were the lack of detail reported regarding eligibility and blinding, the lack of analysis regarding baseline differences between groups and the lack of reporting that the interventions were delivered as intended. Surprisingly, two of these RCTs (Porte et al., 2007; Triano et al., 2006) did not perform between-group comparisons for the outcomes relevant to this review. There is a clear need for high-level, well-designed studies in this area of research to better inform teaching of pre-registration health professional students.

The majority of studies investigated medical students, with one including dental students alongside medical students (Lazarski et al., 2007) and another investigating chiropractic students (Triano et al., 2006). Therefore there are a range of other health professions, such as occupational therapists, nurses, paramedics and physiotherapists, that have not been investigated. These students may differ in their response to different types of feedback, and the clinical skills required of these professions differ from those investigated in the studies included in this review. This highlights potential areas for further research.

Despite the widely accepted view that feedback is required for learning (Archer 2010; Shute, 2008; van de Ridder et al., 2008), this review found little evidence to support the use of terminal feedback in improving the practical, clinical skills exhibited in simulated environments by pre-registration health professional students. The results of the three included studies were mixed; however, studies of elaborative feedback

ISSN 1442-1100

more often indicated that feedback was more effective than no feedback, in contrast to studies focused on verification feedback. This finding, however, should not be over-rated, given few studies investigated this, and these studies were generally at high risk of bias. Similarly, mixed results were obtained when comparing ways in which feedback should be delivered to improve the clinical, practical skills of pre-registration health professional students.

Other factors in feedback delivery should also be considered, such as verbal versus written or group versus individual feedback. Regarding these parameters, no studies that met the inclusion criteria for this review were identified. This highlights potential areas for further research. In giving feedback to higher education students in general, verbal feedback should be used in preference to written feedback (Black & McCormick, 2010). While Cramp (2011) suggested that students prefer individual to group feedback, some studies suggest that there are benefits to providing feedback via a group discussion (Hayes & Devitt, 2008). With large student numbers, individual verbal feedback becomes logistically difficult. Other more efficient methods of delivering feedback, including giving numerical or graded responses only or written feedback as a group summary, lose individual specificity.

There are other feedback parameters that should also be considered. It has been recommended that, for medical students, feedback should be provided within a safe learning environment, commence with a student's self-evaluation, relate to the student's learning goals, be interactive, specific and descriptive, focus on the behaviour observed and highlight only two or three points; in addition, positive feedback should be provided prior to negative and immediately follow the task performance (Bokken, Linssen, Scherpbier, Van Der Vleuten, & Rethans, 2009). These recommendations were, however, not based on research evidence and must, therefore, be considered with this in mind.

Different types of feedback may suit different types of students, and this is a factor that was not explored in any of the studies included in this review. Bok et al. (2009) identified that students are more likely to request feedback if they feel it will be positive. Similarly, Harrison et al. (2013) found that when web-based feedback was accessible to all students, those who performed well were more likely to engage with the feedback than those who just passed and who are, arguably, most in need of using the feedback. This may have implications for the manner in which feedback is delivered to different groups of students and should be investigated in future research.

The conclusions drawn from this review are limited by the underlying research, in addition to the limitations of this review. Pragmatically, it was only possible to include studies published in English; hence, there is potential for studies that may have informed the review questions to have been missed. The date limitation imposed on the search may have also led to the exclusion of otherwise relevant studies. This limitation was, however, selected to ensure the review was relevant to current students with regards to generational differences in learning as well as the advent of increased use of technology in teaching. Another pragmatic limitation of this review was the use of one reviewer

ISSN 1442-1100

to identify studies, appraise and extract data. To reduce the impact of this, any queries were followed up with a second reviewer; however, this may still have had an impact upon the quality of this review.

In conclusion, there is a paucity of research published within the past 10 years, in English, that has investigated whether terminal feedback is effective in improving clinical, practical skills performed by pre-registration health professional students in a simulated environment. Findings are, therefore, based upon a small number of studies with high risk of bias. The current literature suggests that elaborative feedback may be more effective than no feedback, with mixed results regarding the comparison between feedback types, and verification feedback and no feedback. There are significant gaps in the available literature. There is a need to investigate these factors using well-designed studies, with sufficiently powered sample sizes, that address the PEDro criteria, thus minimising potential sources of bias. More specifically, there are opportunities to investigate feedback within other occupational groups, with other tasks, particularly outside of North America. Furthermore, studies should consider the feedback needs of different student groups, especially those with poorer clinical abilities.

Acknowledgement

The authors would like to acknowledge Dr Julie Walters (University of South Australia) for her input into the planning of and feedback on the initial draft of this review.

References

Archer, J. C. (2010). State of the science in health professional education: Effective feedback. *Medical Education*, *44*(1), 101–108. doi:10.1111/j.1365-2923.2009.03546.x

Black, P., & McCormick, R. (2010). Reflections and new directions. *Assessment and Evaluation in Higher Education*, *35*(5), 493–499. doi: 10.1080/02602938.2010.493696

Boehler, M. L., Rogers, D. A., Schwind, C. J., Mayforth, R., Quin, J., Williams, R. G., & Dunnington, G. (2006). An investigation of medical students' reactions to feedback: A randomized controlled trial. *Medical Education*, *40*(8), 746–749. doi:10.1111/j.1365-2929.2006.02503.x

Bok, H. G., Teunissen, P. W., Spruijt, A., Fokkema, J. P., van Beukelen, P., Jaarsma, D. A., & van der Vleuten, C. P. M. (2013). Clarifying students' feedback-seeking behaviour in clinical clerkships. *Medical Education*, *47*(3), 282–291. doi: 10.1111/medu.12054

Bokken, L., Linssen, T., Scherpbier, A., Van Der Vleuten, C., & Rethans, J.-J. (2009). Feedback by simulated patients in undergraduate medical education: A systematic review of the literature. *Medical Education*, *43*(3), 202–210. doi: 10.1111/j.1365-2923.2008.03268.x

Branch, W. T., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine*, *77*(12), 1185–1188.

ISSN 1442-1100

Cramp, A. (2011). Developing first-year engagement with written feedback. *Active Learning in Higher Education*, *12*(2), 113–124. doi: 10.1177/1469787411402484

Crowe, M., & Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *Journal of Clinical Epidemiology*, *64*(1), 79–89.

de Morton, N. A. (2009). The PEDro scale is a valid measure of the methodological quality of clinical trials: A demographic study. *Australian Journal of Physiotherapy*, *55*(2), 129–133.

Harrison, C. J., Könings, K. D., Molyneux, A., Schuwirth, L. W. T., Wass, V., & van der Vleuten, C. P. M. (2013). Web-based feedback after summative assessment: How do students engage? *Medical Education*, *47*(7), 734–744. doi: 10.1111/medu.12209

Hayes, K. D., & Devitt, A. A. (2008). Classroom discussions with student-led feedback: A useful activity to enhance development of critical thinking skills. *Journal of Food Science Education*, *7*(4), 65–68. doi: 10.1111/j.1541-4329.2008.00054.x

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *The Psychological Record*, *57*(2) , 219–232.

Lazarski, M. P., Susarla, S. M., Bennett, N., & Seldin, E. B. (2007). How do feedback and instructions affect the performance of a simulated surgical task? *Journal of Oral and Maxillofacial Surgery*, *65*(6), 1155–1161.

Macedo, L. G., Elkins, M. R., Maher, C. G., Moseley, A. M., Herbert, R. D., & Sherrington, C. (2010). There was evidence of convergent and construct validity of Physiotherapy Evidence Database quality scale for physiotherapy trials. *Journal of Clinical Epidemiology*, *63*(8), 920–925.

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro Scale for rating quality of randomized controlled trials. *Physical Therapy*, *83*(8), 713–721.

Manzone, J., Tremblay, L., You-Ten, K. E., Desai, D., & Brydges, R. (2014). Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. *Medical Education*, *48*(4), 430–440. doi: 10.1111/medu.12397

National Health and Medical Research Council (NHMRC). (2009). *NHMRC additional levels of evidence and grades for recommendations for developers of guidelines.* Canberra, Australia: Author. Retrieved from http://www.nhmrc.gov.au/_files_nhmrc/file/guidelines/stage_2_consultation_levels_and_grades.pdf

O'Connor, A., Schwaitzberg, S. D., & Cao, C. G. L. (2008). How much feedback is necessary for learning to suture? *Surgical Endoscopy and Other Interventional Techniques*, *22*(7), 1614–1619. doi: 10.1007/s00464-007-9645-6

Porte, M. C., Xeroulis, G., Reznick, R. K., & Dubrowski, A. (2007). Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *The American Journal of Surgery*, *193*(1), 105–110.

ISSN 1442-1100

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi: 10.3102/0034654307313795

Skinner, B. F. (1968). *The Technology of Teaching*. New York: Appleton-Century-Crofts.

Strandbygaard, J., Bjerrum, F., Maagaard, M., Winkel, P., Larsen, C. R., Ringsted, C., . . . Sorensen, J. L. (2013). Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: A randomized trial. *Annals of Surgery*, *257*(5), 839–844. doi: 10.1097/SLA.0b013e31827eee6e

Triano, J. J., Scaringe, J., Bougie, J., & Rogers, C. (2006). Effects of visual feedback on manipulation performance and patient ratings. *Journal of Manipulative and Physiological Therapeutics*, *29*(5), 378–385.

van de Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & Ten Cate, O. T. J. (2008). What is feedback in clinical education? *Medical Education*, *42*(2), 189–197. doi: 10.1111/j.1365-2923.2007.02973.x

Van Sickle, K. R., Smith, B., McCluskey, D. A., Baga, I., Smith, C. D., & Gallagher, A. G. (2005). Evaluation of a tensiometer to provide objective feedback in knot-tying performance. *The American Surgeon*, *71*(2), 1018–1023.

ISSN 1442-1100