

An Investigation of Agreement between Clinical Educators and Peers in Rating Speech Pathology Students' Interview Skills

Hannah Reece
The University of Queensland, Australia

Anne E. Hill
The University of Queensland, Australia

Adriana Penman
The University of Queensland, Australia

Abstract

Peer assessment and feedback is being used more frequently in health science education as it has been shown to enhance self-directed learning. This study investigated the level of agreement between clinical educators and speech pathology student peers when rating students' performance during standardised patient interviews. Participating in this study were 104 undergraduate speech pathology students and six clinical educators who were required to rate students' foundation clinical skills on the Standardised Patient Interview Rating Scale (SPIRS). Students' skills, including communication, interviewing and professional practice, were rated by a clinical educator and a peer. Data from two separate interviews in weeks 4 and 8 of a clinical placement were analysed to determine the agreement between clinical educators and peers in rating a student on individual items on the SPIRS. Results indicated that there were unacceptable agreement levels between clinical educators and peers in both opportunities of rating. Recommendations for improving agreement between peers and clinical educators were made including increasing explicit training with the rating tool, increased collaboration between clinical educators and student raters, and using peers with more clinical placement experience as raters. Further research is required to investigate the use of peer assessment for both formative and summative purposes in speech pathology student education.

Keywords: agreement; assessment; clinical education; peer feedback; speech pathology

*Corresponding Author: Anne E. Hill PhD, The University of Queensland, School of Health and Rehabilitation Sciences, Division of Speech Pathology, St Lucia QLD 4072, Australia

Email: ae.hill@uq.edu.au

Journal URL: <http://e-learning.coventry.ac.uk/ojs/index.php/pblh>

Reece, H., Hill, A.E., and Penman, A. (2018) 'An investigation of agreement between clinical educators and peers in rating speech pathology students' interview skills'. *International Journal of Practice-based Learning in Health and Social Care*, 6 (2), 64–79



© 2018 Hannah Reece, Anne E. Hill, and Adriana Penman. This Open Access article is distributed under the terms of the Creative Commons Attribution Attribution-Non-Commercial No Derivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and is unaltered.

Introduction

The purpose of tertiary health science programs is to provide students with the skills, knowledge and attitudes required for competency on entry to the workforce ([Ladyshevsky 2002](#)). Therefore, it is important that students have access to 'real-world situations' (clinical placements) to apply coursework learning to workplace contexts in preparation for their graduate practice ([McAllister et al. 2010](#)). Clinical placements also facilitate the development of discipline and cross-discipline skills, interpersonal skills and clinical reasoning ([Sheepway, Lincoln, and Togher 2011](#)).

In speech pathology student education in Australia, academic and clinical placement experiences are guided by the Competency-based Occupational Standards (CBOS) for Speech Pathologists – Entry Level ([Speech Pathology Association of Australia 2011](#)). CBOS outlines minimum skills, knowledge base and professional standards that are required for entry level practice in the profession. University curricula, coursework assessments and clinical programs are designed and implemented to ensure that students meet these standards. The teaching of CBOS elements is integrated into clinical placements from the beginning of the program, with foundation clinical skills as an initial priority ([McAllister et al. 2010](#)).

The teaching of foundation clinical skills is critical in the education of speech pathology students ([Hill et al. 2014](#)). Foundation skills include non-verbal communication such as eye-contact and body language, verbal communication skills including appropriate language use, and interpersonal skills, for example building rapport and maintaining a focus on the client. Foundation skills are an important component of speech pathology student education as they provide a platform for developing more discipline-specific clinical skills ([McAllister et al. 2010](#)). [Becker et al. \(2006\)](#) also reported that students beginning their clinical placements with little training in foundation skills can have reduced learning outcomes in these placements. It is important that students are not only aware of the importance of these foundation skills, but that they are assessed on these specific behaviours ([McAllister et al. 2010](#)).

The role of assessment within health sciences education programs has been well researched ([Boud 2000](#), [Boud and Soler 2016](#), [Rush et al. 2012](#), [van der Vleuten et al. 2010](#)). It is recognised that students often dedicate focus and time to learning in those areas in which they know they will be assessed ([van der Vleuten et al. 2010](#)). Therefore, the creation and implementation of meaningful assessment requires significant attention. Assessment should be objective, standardised and based on an in-depth understanding of the task and criteria ([Rush et al. 2012](#)). In addition to yielding the final mark or grade, assessment should offer the opportunity to provide targeted, formative feedback to students to enhance their clinical learning ([van der Vleuten et al. 2010](#)). Such feedback can equip students for life-long learning ([Boud 2000](#)). In keeping with the notion of life-long learning, it has been suggested that a shift from 'assessment of learning' to 'assessment for learning' allows students to continually learn from their environment and achieve success beyond immediate course-related goals ([Boud and Soler 2016](#)). The role of assessing and providing feedback to students on clinical placements is typically undertaken by clinical educators.

Despite the central role of the clinical educator in the assessment process and as a facilitator of student learning, the role of the student in their own learning requires consideration. Students who are passive in their approach to learning, for example, listening but not actively involved in the learning process, may neglect opportunities to take control of their own learning ([Boud 2000](#)). This is likely to limit their ability to build new knowledge without educator guidance ([Boud 2000](#), [Hodges 2011](#)). [Boud and Soler \(2016\)](#) recognise 'active learners' as those who draw on metacognitive skills to learn independently, approach problems from multiple perspectives, and reflect on personal performance. During clinical placements, active learning is encouraged through, for example, reflection activities. These can provide students with an opportunity to reflect on their learning and apply this learning to future performance, thereby enhancing clinical reasoning skills which are essential for speech pathology practice ([Boud and Soler 2016](#), [Hulsman, Harmsen, and Fabriek 2009](#), [McAllister et al. 2010](#)).

In addition to self-reflection, student engagement in active learning may also be facilitated through peer assessment which provides students with the opportunity to evaluate, and be evaluated by, peers of a similar academic level and expertise who share a common knowledge and skill base (Boud and Soler 2016, Finn and Garner 2011). A range of studies in health professional education have explored the notion of peer assessment (Hodgson, Chan, and Liu 2014, Perera, Mohamadou, and Kaur 2010, Rush *et al.* 2012, Speyer *et al.* 2011). Rush and colleagues (2012), in their study of first-year nursing students, highlighted the positive effects of peer assessment and feedback. This included students learning how to give and receive constructive criticism and students gaining insights into how others approach similar clinical problems. Receiving and providing quality feedback has been reported to enhance self-directed learning and improve professional performance (Hodgson, Chan, and Liu 2014, Perera, Mohamadou, and Kaur 2010, Rush *et al.* 2012, Speyer *et al.* 2011).

Despite literature outlining the benefits of peer assessment and feedback, there are some concerns regarding the equity of the assessment process and the quality and accuracy of feedback provided by peers (Kaufman and Schunn 2011). Friendships, collusion among students to allocate the same marks and reluctance to provide negative feedback to peers may influence the outcomes of peer assessment (Hodgson, Chan, and Liu 2014). Given these concerns, if peer assessment is to be utilised and valued by students, it is essential that peer ratings are valid and justified. One method of investigating this is through a comparison of students' ratings with the ratings of their peers or more experienced raters. Dannefer and colleagues (2005), in their study of second-year medical students, investigated the agreement between peer raters in assessing an individual student. Each student was evaluated on a rating scale by 15 peers during a standardised patient interview. Results indicated that there was no statistically significant correlation between each peer's rating of an individual's 'interpersonal habits', including demonstration of respect and understanding. However, the study reported moderate correlation between peers when rating 'work habits', including explaining results and being well-prepared (Dannefer *et al.* 2005). This study corroborates existing research that suggests that peer assessment may be better suited to technical skill evaluation rather than the assessment of interpersonal skills (Falchikov and Goldfinch 2000, Finn and Garner 2011, Kogan *et al.* 2010).

Another avenue for exploring the worth of peer assessment is to compare a rating of students' skills by their peers with ratings by an experienced clinical educator. Peer and educator ratings were explored by Falchikov and Goldfinch (2000) in a meta-analysis. Overall, the meta-analysis concluded that there was strong evidence of agreement between peer and educator ratings (a mean correlation of 0.69 across the 45 studies). In addition to critiquing the reliability and validity of peer assessment, the meta-analysis highlighted certain variables that influenced the level of agreement between peer and educator assessment. Firstly, there was a higher level of agreement between peers and educators when academic products, that is, exams or assignments were marked compared to the rating of professional practice skills. Additionally, Falchikov and Goldfinch (2000) found that peer assessments were more valid when there were predetermined criteria detailing exactly what skills and knowledge students were required to demonstrate. The authors also acknowledged the benefits of peer assessments such as improving student learning. While the results of this meta-analysis are encouraging, and peer assessment in student education and training was generally supported by the cited literature, it is important to note that the review only included studies from 1959 to 1999. The studies analysed were also not specific to health professional education, therefore further research is required to determine their applicability to the current study's discipline of speech pathology. In addition, meta-analyses have limitations including reviewer bias, heterogeneity of studies, and problems with combining studies of different methods and sample sizes (Falchikov and Goldfinch 2000). Therefore, they need to be interpreted with caution.

More recent studies have explored the concept of peer and clinical educator ratings in health science education and found that there were overall acceptable levels of agreement between raters (Chenot *et al.* 2007, Evans, Leeson, and Petrie 2007, Machado *et al.* 2008). Chenot and colleagues (2007) compared student tutor and clinical educator assessment of medical students on four Objective Structured Clinical Evaluation (OSCE) stations. The student tutors were in

their fourth year of a medical degree and the students being assessed were in their third. The rating tool included a checklist of individual skills and a global rating scale (5-point Likert scale) for overall performance on each station. Overall kappa statistics showed moderate to good agreement levels between clinical educators and student tutors. It is important to consider that in this study, the peer assessors were student tutors in these OSCE stations, were one year further into their degree than the students, had been trained as raters and were familiar with the task and assessment criteria. Therefore, the results of this study cannot be readily generalised to other studies involving peer assessors of the same university year level with less explicit training.

In contrast, [Machado et al. \(2008\)](#) found that peers of the same year level were significantly more lenient than clinical educators when assessing medical students on clinical skills using pre-established criteria (5-point Likert scale). Peer assessment contributed to students' final grades. The authors found no statistically significant correlation between peer and educator ratings. However, these results must be interpreted with caution as the authors highlighted that peers may have rated fellow students more highly in an attempt to inflate their overall grade ([Machado et al. 2008](#)). [Evans, Leeson, and Petrie \(2007\)](#), in their study of postgraduate dentistry students performing a common procedure, reported an excellent level of agreement between peer and clinical educator ratings of performance. Unlike [Machado et al.'s \(2008\)](#) study, peer ratings in the [Evans, Leeson, and Petrie \(2007\)](#) study did not contribute to students' grades but compared peer and student ratings on an assessment checklist in which the steps of the dental procedure were objectively detailed for rating, unlike the studies by [Chenot et al. \(2007\)](#) and [Machado et al. \(2008\)](#) whereby Likert scales were used to assess clinical skills

Taken together, results of the above studies have demonstrated the value of peer assessment and feedback for student learning. To the authors' knowledge there is currently no research focusing on the use of peer assessment in speech pathology programs. Given the importance of active engagement with learning in promoting life-long learning it can be postulated that peer assessment, including the rating of foundation clinical skills, would be a worthwhile addition to speech pathology clinical education programs ([Boud and Soler 2016](#), [McAllister et al. 2010](#)). However, conflicting outcomes from available medical and health education literature comparing peer and educator assessment suggest that further research on the reliability of peer ratings is required. Therefore, the aim of the current study was to determine the level of agreement between clinical educators and peers when rating the performance of speech pathology students in a client interview conducted within a clinical placement.

Methods

This quantitative comparative study had ethical clearance from the health and research ethics committee at an Australian university.

Context of the study

This study was conducted during the first clinical placement opportunity undertaken by second-year speech pathology students enrolled in a four-year undergraduate program at an Australian university. Students worked in groups of six with one clinical educator for one half-day placement per week for 12 weeks. The clinic consisted of a series of clinical workshops discussing foundation clinic skills and clinical processes involved in supporting children with speech sound disorders. Additionally, students participated in two interviews with a standardised patient (SP) and a screening assessment of a paediatric client. SPs are frequently used within simulation clinics in medicine, nursing and allied health and are reported to encourage safe skill acquisition, student-centred learning and uniform case experiences ([Hill 2012](#)). Simulation clinics using SPs provide a gradual transition from academic boundaries of lectures and tutorials to actual clinical practice ([Hill, Davidson, and Theodoros 2010](#)). The introduction of SP interaction before 'real' client experiences has been shown to increase student confidence and decrease anxiety ([Hill, Davidson, and Theodoros 2013](#)). Simulation clinics have also been shown to assist students in understanding their professional role within a

low-risk environment and preparing them for future practice ([Becker et al. 2006](#)). The SPs in this study portrayed the role of a parent or grandparent of a child with a speech sound disorder and had received nine hours of training in how to portray this role and provide feedback to students

Participants

The participants of this study were 104 undergraduate speech pathology students in their second year of a four-year program and eight clinical educators. The students were aged from 18 to 36 with a mean age of 20 (standard deviation = 2). The clinical educators had a range of 6–31 years' clinical experience with a mean of 17 years of experience (standard deviation = 8). All participants were female. The participants were fully informed of the research project and provided informed consent prior to the commencement of the clinic. Students were informed that peer assessment was an important learning activity within the clinical placement but that their consent to participate in this study was independent of the course content and that failure to participate would have no impact on their grade for the course.

Procedure

SP interviews were conducted in pairs and the remaining students in the clinic group were required to observe all interviews and rate the performance of another student in their interview; this rating did not contribute to students' overall grades. The first interview occurred in the fourth week of the clinical placement and involved conducting a case history with the SP parent/grandparent. The second interview was held in week 8 and involved providing feedback to the client, including an explanation and interpretation of assessment results and discussion of a management plan. Peers and clinical educators both rated students' performance in both SP interviews using the Standardised Patient Interview Rating Scale (SPIRS) ([Hill, Davidson, and Theodoros 2015](#)) (see [Appendix 1: Week 4: Case History Interview](#), and [Appendix 2: Week 8: Follow-up Interview](#)). Students were introduced to the SPIRS in a pre-clinical lecture of one hour's duration and then practised using the tool to rate two recorded case history interviews of an experienced clinician interviewing a client. Clinical educators also received three hours' training in the use of the tool for rating students' performance through group discussion and practice interview rating. For both the case history and feedback interviews, students were rated by the clinical educator. Clinical educator ratings on the SPIRS contributed to students' final grade for the clinic, providing evidence of the students' competency at the end of the placement. For the current study, the SPIRS forms were collated and de-identified, and peer and clinical educator ratings were stored separately. Clinical educators, students and peers were assigned a numerical code and the data was entered into Excel.

Interview Assessment Tool

The SPIRS is a purpose-built assessment tool designed for clinical educators to evaluate student performance on specific foundation skills within a simulation environment. The SPIRS includes six individual items (non-verbal communication, verbal communication, interpersonal skills, interviewing skills, professional practice skills and clinical skills) which are rated on a 5-point Likert scale. In addition, an overall rating of interview performance is provided (also a 5-point scale). It is a validated tool with high internal consistency ($\alpha = 0.743$) and acceptable levels of inter-rater reliability ([Hill, Davidson, and Theodoros 2015](#)). This tool was ideal for use in the current study as it had been validated with clinical educators undertaking assessment of students in the same clinical context.

Data Analysis

Percent exact agreement (PEA) and weighted kappas were used to analyse the inter-rater agreement between peer and clinical educator ratings in this study. Inter-rater agreement is the degree to which two or more evaluators assign the same rating to a performance in an identical situation using the same rating scale ([Shweta, Bajpai, and Chaturvedi 2015](#)). PEA is defined as the number of agreements divided by the total number of agreements plus disagreements, while the kappa statistic measures the difference between observed and expected agreement

(Shweta, Bajpai, and Chaturvedi 2015). PEA and kappa values were calculated for the ratings of each item on the SPIRS; for example, the clinical educators' ratings of students' performance on SPIRS item 1.1 (non-verbal communication) were compared to the ratings of those same students on SPIRS item 1.1 by peers (total of 7 kappa values for rating on each of weeks 4 and 8 interviews). Clinical educators rated all students in their allocated clinic groups while each observing peer rated one student only out of another pair. Landis and Koch (1977) outlined guidelines to determine the strength of inter-rater agreement. Kappa values ≤ 0 indicated poor agreement, 0.01–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement and 0.81–1.00 indicated almost perfect agreement. Shweta, Bajpai, and Chaturvedi (2015) suggested that PEA should be at least 70% to be considered acceptable when there are between 5 and 7 rating levels; the SPIRS has 5 levels. Weighted kappas and PEA values were calculated using Statistical Package for Social Sciences (SPSS) software (2016). The mean PEA and range of PEA was also calculated for the data for both weeks.

Results

The results for week 4 and 8 are detailed separately below (Table 1).

Week 4 Ratings

Complete matched clinical educator-peer week 4 data was available for 69 students. Percent exact agreement (PEA) between clinical educator and peer ratings across the 7 items of the SPIRS ranged from 35.29% to 48.53% with a mean of 42.57%. This represents an unacceptable level of agreement (Shweta, Bajpai, and Chaturvedi 2015). Kappa values ranged from -0.053 to 0.161 which, according to Landis and Koch (1977), reflects poor to slight agreement. Individual PEA and kappa values for each item are detailed in Table 1. The SPIRS item 1.2 (verbal communication) and SPIRS item 2 (professional practice skills) yielded the lowest PEA and kappa values. PEA was highest for SPIRS item 4 (overall score), although similarly high PEA was achieved on SPIRS item 1.4 (interviewing skills) and SPIRS item 3 (clinical skills).

Table 1. Agreement between peers and clinical educators in rating students' interview skills.

Item*	Skill	Kappa values		Percent Exact Agreement (PEA)	
		Week 4	Week 8	Week 4	Week 8
1.1	Non-verbal communication	-0.033 (none**)	0.148 (slight)	42.03%	58.23%
1.2	Verbal communication	-0.053 (none)	0.205 (fair)	35.29%	55.70%
1.3	Interpersonal	0.045 (slight)	0.229 (fair)	41.54%	57.69%
1.4	Interviewing	0.161 (slight)	0.132 (slight)	48.53%	52.56%
2	Professional practice	-0.103 (none)	0.076 (slight)	35.38%	52.56%
3	Clinical	-0.008 (none)	0.156 (slight)	46.67%	56.58%
4	Overall	0.043 (slight)	0.209 (fair)	48.57%	58.02%

* = rated on the SPIRS (Hill, Davidson, and Theodoros 2015)

** = level of agreement reported by Landis and Koch (1977).

Week 8 Ratings

Complete matched clinical educator-peer week 8 data was available for 79 students. In week 8 PEA between clinical educators and peers across the seven items of the SPIRS ranged from 52.56% to 58.23% with a mean of 55.19% which is also considered unacceptable agreement (Shweta, Bajpai, and Chaturvedi 2015). In week 8, kappa values ranged from 0.076 to 0.229, indicating slight to fair agreement. PEA was highest for SPIRS item 1.1 (non-verbal communication). Similar to week 4, a high PEA was also achieved on SPIRS item 4 (overall score). The lowest PEA and kappa values were found for SPIRS item 1.4 (interviewing skills) and SPIRS item 2 (professional practice skills).

Rating trends between clinical educators and peers

A visual inspection of the data revealed that no students were rated as a 1 (unacceptable) or 2 (poor) on any item of the SPIRS in week 4 or 8 by clinical educators or peers. Both peers and clinical educators rated students primarily at a 4 (good) or 5 (excellent) level. In week 4, 87.7% of clinical educator and peer ratings were at a 4 or 5 level and in week 8, 96.2% of ratings were at a 4 or 5 level. In terms of PEA, there was an increase in the agreement between clinical educators and peers between weeks 4 and 8 ratings.

Discussion

The aim of this study was to investigate the extent of agreement between clinical educators and peers when rating speech pathology students' clinical performance in a standardised patient interview using a validated tool. The results of this study indicate low levels of agreement between clinical educators and peers on two occasions of rating (weeks 4 and 8 of the clinical placement). These results are inconsistent with the outcomes of limited available research that has reported adequate levels of overall peer and clinical educator agreement (Chenot *et al.* 2007, Evans, Leeson, and Petrie 2007, Falchikov and Goldfinch 2000). Although the findings of the current study suggest that peer assessment may not be reliable, the study adds to the growing body of research and can inform future health education practices. The benefits of peer assessment, including increasing active learning, have been well documented and it is likely that peer assessment will continue to be used. In light of this, factors that may have influenced the level of agreement between clinical educators and peers in this study and potential considerations to improve agreement are discussed below.

Overall, PEA and kappa results suggested inadequate agreement between peers and clinical educators in both weeks 4 and 8. In contrast to the results of Machado and colleagues (2008), peers did not consistently rate students consistently more highly than clinical educators, or vice versa. The relationship between PEA and kappa values was sometimes unbalanced. For example in week 4, SPIRS item 1.1 (non-verbal communication) had a PEA of 42.03% and a kappa value of -0.033 (no agreement) while SPIRS item 1.3 (interpersonal skills) had a PEA of 41.54% and a kappa value of 0.045 (slight agreement). Sertdemir *et al.* (2013) acknowledged that using the kappa statistic to calculate agreement between raters on a 5-point ordinal scale (like the SPIRS) may underestimate agreement, particularly if the majority of ratings are at the highest or lowest level. Shweta, Bajpai, and Chaturvedi (2015) also reported that kappa values may be low if the results are concentrated at one rating level. This is the case for the current study in which clinical educators and peers rated students predominantly at a 4 or 5 on the 5-point scale. Given this limitation of kappa statistics and the rating patterns of both clinical educators and peers, PEA results may represent study outcomes more realistically.

Variable levels of peer and clinical educator agreement were achieved across different items of the SPIRS in weeks 4 and 8. The highest level of percent exact agreement was achieved for SPIRS item 4 (overall score) in week 4 and SPIRS item 1.1 (non-verbal communication) in week 8. Higher agreement for SPIRS item 4 may suggest that both clinical educators and peers are relatively comparable in 'summing up' overall performance in a rating. This finding is consistent with other studies which have reported higher accuracy in rating overall scores

compared with rating on discrete checklist items, perhaps suggesting that overall ratings may actually be more reflective of clinicians' understanding of clinical competency than checklist items as demonstrated by students in their performance ([Boursicot, Roberts, and Pell 2006](#), [Norman and Feightner 1981](#), [Regehr et al. 1998](#)). The inclusion of an overall rating is justified.

The finding that PEA between peers and clinical educators was higher for SPIRS item 1.1 (non-verbal communication) has similarities to the findings of [Hill, Davidson, and Theodoros \(2015\)](#) who reported the highest level of agreement in SPIRS item 1.1. Rating of non-verbal communication may have had higher agreement as eye-contact, gesture and body language are observable, do not require interpretation and can be objectively assessed ([Hill, Davidson, and Theodoros 2015](#)). In contrast, agreement between peers and clinical educators was low on SPIRS item 2 (professional practice) in both weeks 4 and 8. Rating of students' performance on the professional practice item requires the evaluation and integration of several aspects of the case history and interview process and draws on broader knowledge of course objectives. It is likely that experienced clinical educators are more familiar with skills required by students in demonstrating this item than the peers who were in their first clinical placement. This difference in experience may explain the lack of agreement on this item. An additional SPIRS item with low rater agreement in week 4 and 8 was item 1.2 (verbal communication). Rating of verbal communication requires evaluation of speech volume, rate and intonation as well as appropriateness of language use in context. Objective and reproducible evaluation of these skills has been found to be problematic ([Makoul and Curry 2007](#), [O'Sullivan et al. 2008](#)). When considering the variability in agreement levels across SPIRS items, it is evident that peers may require more information on the process of rating verbal communication skills in order for peer assessments to be validly used.

One way to increase the knowledge and experience of raters is to provide additional training. [Hill, Davidson, and Theodoros \(2015\)](#) reported that the clinical educators participating in their study received group training in use of the SPIRS with practical application. Students in the current study received an introduction to the SPIRS and brief training only. It is possible that increasing the amount of explicit training that student raters receive may increase the level of agreement between peers and clinical educators. This recommendation concurs with research conducted by [Chenot and colleagues \(2007\)](#) who found that, when provided with standardised rater training (using videos and explicit instructions), peers and clinical educators had moderate to good agreement in individual item and global/overall rating on a 1–5 scale, as was used in the SPIRS. This notion is reinforced by results in the current study which indicated that peer and clinical educator agreement increased on the second occasion of rating (none to slight agreement in week 4 compared to slight to fair agreement in week 8). This improvement between week 4 and week 8 lends support to the view that increasing experience with the rating tool and rating procedure may increase agreement between raters.

The differences in the nature of the SP interviews conducted by students in week 4 and week 8 may also be a potential factor in the improvement in levels of agreement between week 4 and week 8. In week 4, students were required to take a case history and in week 8 students were providing feedback from a speech assessment. Other studies have explored the notion of assessment type affecting the level of agreement between raters ([Chenot et al. 2007](#), [Evans, Leeson, and Petrie 2007](#), [Machado et al. 2007](#)). It is possible that the task differences in week 4 and week 8 in the current study contributed to disparate agreement outcomes. Further analysis of task components and students' perception of their relative difficulty was beyond the scope of this study, but future research could address this question.

The findings of the current study are of interest when compared with those of [Hill, Davidson, and Theodoros \(2015\)](#), who found acceptable levels of rater agreement between an 'expert rater' (a rater with extensive experience in student competency rating) and 10 other clinical educators (mean PEA of 82.06%). The level of expertise and experience of raters may have contributed to the discrepancy between the [Hill, Davidson, and Theodoros \(2015\)](#) study and the current study. This experience may also have been a factor which contributed to the acceptable levels of agreement between experienced student tutors and clinical educators reported by [Chenot et al. \(2007\)](#). In contrast, the current study compared experienced clinical educator

raters with novice student raters (in their first speech pathology clinical placement). This may suggest that peer rating may be more reliable with more experienced speech pathology students.

In light of the above factors there are a number of actions that could be implemented which may increase the level of agreement between clinical educators and peers in rating students' performance. Increasing the amount of specific training in the SPIRS for students, which mirrors that given to clinical educators would be beneficial. The training could be complemented with examples of the behavioural descriptors as well as video examples of students performing at a variety of levels on the 5-point scale. A practice rating using the form may also help to improve the reliability of students' ratings, thereby increasing overall agreement between clinical educators (CEs) and peers. In the current study this was demonstrated by observed improvement between week 4 and week 8 ratings. It may also be beneficial for the CEs and peers to discuss the process of rating students so that agreement is reached on the criteria for allocating specific ratings. This collaboration between CEs and peers has been shown to increase clinical reasoning and active learning (Springer, Stanne, and Donovan 1999).

Limitations and Future Directions

There are some limitations to this study, and these highlight opportunities for further research. Firstly, peer and clinical educator rating agreement was only investigated with the SPIRS, therefore these results may not be transferable to other rating forms and tools used in speech pathology student education. Additionally, the participants in this study were a part of one cohort at one university so the results may not be readily generalised to students at other year levels and in other contexts. Students were in their first clinical placement in their second year of a four-year undergraduate program and had varying backgrounds, work and life experiences. Therefore, it would also be valuable to compare the level of agreement between the ratings of clinical educators and peers with more clinical placement experience and knowledge, for example, using students further into their speech pathology program. There is promising evidence to suggest that agreement between peers and clinical educators could be improved by including more training for peers, particularly for the evaluation of less objective skills such as communication and professional practice.

It may also have been valuable to obtain the perceptions of peers and clinical educators about their experience of rating to provide qualitative information about their understanding of the tool, rater bias and why raters rated as they did. Similarly, the evaluation of student learning outcomes as a result of providing and receiving peer feedback would be a valuable addition to research in peer assessment and feedback.

Conclusion

Results of this study indicate that there were low levels of agreement between clinical educators and peers when rating students' foundation clinical skills including communication and interview skills, using a validated single-interview rating form. The results of the current study suggest that caution should be exercised in utilising peer assessment to contribute to students' grades. However, increasing explicit student training in the use of the rating tool and course objectives may help to improve agreement between peer and clinical educator ratings. To the authors' knowledge this is the first study investigating the validity of peer assessment in speech pathology clinical education and therefore adds to the growing body of research in peer assessment and feedback. By taking the opportunity to become assessors and to receive feedback from their peers, students are developing skills in life-long learning, reflection and clinical reasoning which are essential for speech pathology practice.

Acknowledgements

Ethical Approval By: The University of Queensland Medical Research Ethics Committee: clearance number 200901668.

References

- Becker, K.L., Rose, L.E., Berg, J.B., Park, H., and Shatzer, J.H. (2006) 'The teaching effectiveness of standardized patients'. *Journal of Nursing Education*, 45 (4), 103-111
- Boud, D. (2000) 'Sustainable assessment: Rethinking assessment for the learning society'. *Studies in Continuing Education*, 22 (2), 151–167 <https://doi.org/10.1080/713695728>
- Boud, D. and Soler, R. (2016) 'Sustainable assessment revisited'. *Assessment & Evaluation in Higher Education*, 41, 400-413 <https://doi.org/10.1080/02602938.2015.1018133>
- Boursicot, K.A.M., Roberts, T.E., and Pell, G. (2006) 'Standard setting for clinical competence at graduation from medical school: A comparison of passing scores across five medical schools'. *Advances in Health Sciences Education*, 11 (2), 173–183 <https://doi.org/10.1007/s10459-005-5291-8>
- Chenot, J.-F., Simmenroth-Nayda, A., Koch, A., Fischer, T., Scherer, M., Emmert, B., ... Himmel, W. (2007) 'Can student tutors act as examiners in an objective structured clinical examination?'. *Medical Education*, 41 (11), 1032–1038 <https://doi.org/10.1111/j.1365-2923.2007.02895.x>
- Dannefer, E.F., Henson, L.C., Bierer, S.B., Grady-Weliky, T.A., Meldrum, S., Nofziger, A.C., ... Epstein, R.M. (2005) 'Peer assessment of professional competence'. *Medical Education*, 39 (7), 713–722 <https://doi.org/10.1111/j.1365-2929.2005.02193.x>
- Evans, A.W., Leeson, R.M.A., and Petrie, A. (2007) 'Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery'. *Medical Education*, 41 (9), 866–872 <https://doi.org/10.1111/j.1365-2923.2007.02819.x>
- Falchikov, N. and Goldfinch, J. (2000) 'Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks'. *Review of Educational Research*, 70 (3), 287–322 <https://doi.org/10.3102/00346543070003287>
- Finn, G.M. and Garner, J. (2011) 'Twelve tips for implementing a successful peer assessment'. *Medical Teacher*, 33 (6), 443–446 <https://doi.org/10.3109/0142159X.2010.546909>
- Hill, A.E. (2012) 'The use of standardised patients in the clinical education of speech pathology students'. Unpublished PhD thesis. St Lucia, Australia: The University of Queensland, School of Health & Rehabilitation Sciences
- Hill, A.E., Davidson, B.J., McAllister, S., Wright, J., and Theodoros, D.G. (2014) 'Assessment of student competency in a simulated speech-language pathology clinical placement'. *International Journal of Speech-Language Pathology*, 16 (5), 464–475 <https://doi.org/10.3109/17549507.2013.809603>
- Hill, A.E., Davidson, B.J., and Theodoros, D.G. (2010) 'A review of standardized patients in clinical education: Implications for speech-language pathology programs'. *International Journal of Speech-Language Pathology*, 12 (3), 259–270 <https://doi.org/10.3109/17549500903082445>
- Hill, A.E., Davidson, B.J., and Theodoros, D.G. (2013) 'Speech pathology students' perceptions of a standardised patient clinic'. *Journal of Allied Health*, 42 (2), 84–91

- Hill, A.E., Davidson, B.J., and Theodoros, D.G. (2015) 'An investigation of the Standardised Patient Interview Rating Scale (SPIRS) for the assessment of speech pathology students in a simulation clinic'. *International Journal of Practice-based Learning in Health and Social Care*, 3 (1) 58–76 <https://doi.org/10.18552/ijpbhlhsc.v3i1.213>
- Hodges, H.F. (2011) 'Preparing new nurses with complexity science and problem-based learning'. *Journal of Nursing Education*, 50 (1), 7–13 <https://doi.org/10.3928/01484834-20101029-01>
- Hodgson, P., Chan, K., and Liu, J. (2014) 'Outcomes of synergetic peer assessment: First-year experience'. *Assessment & Evaluation in Higher Education*, 39 (2), 168–178 <https://doi.org/10.1080/02602938.2013.803027>
- Hulsman, R.L., Harmsen, A.B., and Fabriek, M. (2009) 'Reflective teaching of medical communication skills with DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients'. *Patient Education and Counseling*, 74 (2), 142–149 <https://doi.org/10.1016/j.pec.2008.10.009>
- Kaufman, J.H. and Schunn, C.D. (2011) 'Students' perceptions about peer assessment for writing: Their origin and impact on revision work'. *Instructional Science: An International Journal of the Learning Sciences*, 39 (3), 387–406 <https://doi.org/10.1007/s11251-010-9133-6>
- Kogan, J.R., Hess, B.J., Conforti, L.N., and Holmboe, E.S. (2010) 'What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills'. *Academic Medicine*, 85 (10 Suppl, RIME: Proceedings of the Forty-Ninth Annual Conference November 7–November 10, 2010), S25–S28 <https://doi.org/10.1097/ACM.0b013e3181ed1aa3>
- Ladyshevsky, R.K. (2002) 'A quasi-experimental study of the differences in performance and clinical reasoning using individual learning versus reciprocal peer coaching'. *Physiotherapy Theory and Practice*, 18 (1), 17–31 <https://doi.org/10.1080/095939802753570666>
- Landis, J.R. and Koch, G.G. (1977) 'The measurement of observer agreement for categorical data'. *Biometrics*, 33 (1), 159–174 <https://doi.org/10.2307/2529310>
- Machado, J.L.M., Machado, V.M.P., Grec, W., Bollela, V.R., and Vieira, J.E. (2008) 'Self- and peer assessment may not be an accurate measure of PBL tutorial process'. *BMC Medical Education* [online], 8 (55) <https://doi.org/10.1186/1472-6920-8-55>
- Makoul, G. and Curry, R.H. (2007) 'The value of assessing and addressing communication skills'. *JAMA*, 298 (9), 1057–1059 <https://doi.org/10.1001/jama.298.9.1057>
- McAllister, S., Lincoln, M., Ferguson, A., and McAllister, L. (2010) 'Issues in developing valid assessments of speech pathology students' performance in the workplace'. *International Journal of Language & Communication Disorders*, 45 (1), 1–14. <https://doi.org/10.3109/13682820902745461>
- Norman, G.R. and Feightner, J.W. (1981) 'A comparison of behaviour on simulated patients and patient management problems'. *Medical Education*, 15 (1), 26–32 <https://doi.org/10.1111/j.1365-2923.1981.tb02311.x>
- O'Sullivan, P., Chao, S., Russel, M., Levine, S., and Fabiny, A. (2008) 'Development and implementation of an objective structured clinical evaluation to provide formative feedback on communication and interpersonal skills in geriatric training'. *Journal of the*

American Geriatrics Society, 56 (9), 1730–1735 <https://doi.org/10.1111/j.1532-5415.2008.01860.x>

- Perera, J., Mohamadou, G., and Kaur, S. (2010) 'The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: Evaluation using a controlled trial'. *Advances in Health Sciences Education*, 15 (2), 185–193 <https://doi.org/10.1007/s10459-009-9191-1>
- Regehr, K.G., MacRae, K.H., Reznick, K.R., and Szalay, K.D. (1998) 'Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination'. *Academic Medicine*, 73 (9), 993–997 <https://doi.org/10.1097/00001888-199809000-00020>
- Rush, S., Firth, T., Burke, L., and Marks-Maran, D. (2012) 'Implementation and evaluation of peer assessment of clinical skills for first-year student nurses'. *Nurse Education in Practice*, 12 (4), 219–226 <https://doi.org/10.1016/j.nepr.2012.01.014>
- Sertdemir, Y., Burgut, H.R., Alparslan, Z.N., Unal, I., and Gunasti, S. (2013) 'Comparing the methods of measuring multi-rater agreement on an ordinal rating scale: A simulation study with an application to real data'. *Journal of Applied Statistics*, 40 (7), 1506-1519 <https://doi.org/10.1080/02664763.2013.788617>
- Sheepway, L., Lincoln, M., and Togher, L. (2011) 'An international study of clinical education practices in speech-language pathology'. *International Journal of Speech-Language Pathology*, 13 (2), 174–185 <https://doi.org/10.3109/17549507.2011.491129>
- Shweta, Bajpai, R.C., and Chaturvedi, H.K. (2015) 'Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods.' *Journal of the Indian Academy of Applied Psychology*, 41 (3), 20–27
- Speech Pathology Association of Australia (SPAA). (2011). *Competency-Based Occupational Standards (CBOS) for Speech Pathologists: Entry-Level (revised)*. Melbourne: The Speech Pathology Association of Australia
- Speyer, R., Pilz, W., Van Der Kruis, J., and Brunings, J.W. (2011) 'Reliability and validity of student peer assessment in medical education: A systematic review'. *Medical Teacher*, 33 (11), e572–e585 <https://doi.org/10.3109/0142159X.2011.610835>
- Springer, L., Stanne, M.E., and Donovan, S. S. (1999) 'Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis'. *Review of Educational Research*, 69 (1), 21–51 <https://doi.org/10.3102/00346543069001021>
- van der Vleuten, C.P.M., Schuwirth, L.W.T., Scheele, F., Driessen, E.W., and Hodges, B. (2010) 'The assessment of professional competence: Building blocks for theory development'. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 24 (6), 703–719 <https://doi.org/10.1016/j.bpobgyn.2010.04.001>

Appendix 1

Standardised Patient Interview Rating Scale (SPIRS)

Week 4: Case History Interview

Student Name: _____ Year level: _____ Date: _____

Following the interview you have observed, please rate the student's performance by circling **ONLY** the appropriate number on the scale below. Student **must** be rated at **one** of the 5 points only (not in between) using the performance descriptors below as a guide. Listed below each skill area are some specific aspects to consider to help you rate the student's performance and formulate your feedback comments. You do not need to limit your comments to the specific aspects listed.

Unacceptable – Demonstrates many behaviours in specified skill area(s) that are inappropriate or have negative outcomes or consequences (make the situation worse). The desired outcome is not achieved.

Average – Demonstrates a sufficient range of expected behaviours in specified skill area(s) to achieve the desired outcome. Some deficiencies exist in the skill area(s) assessed but none are of major concern.

Excellent – Consistently demonstrates the full range of expected behaviours in specified skill area(s) to achieve the desired outcome. An outstanding level of performance is maintained. No deficiencies exist in the skill area(s) assessed.

1. COMMUNICATION / INTERPERSONAL SKILLS		
The student demonstrated behaviours at the following performance levels in these skill areas:		
Skills	Performance	Comments
1.1 Non-verbal Communication	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • eye contact • use of facial expression • body language • use of gesture 	
	Performance	Comments
1.2 Verbal Communication	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • use of language and terminology • use of appropriate level of formality • speech volume • use of intonation • speech rate 	
	Performance	Comments
1.3 Interpersonal Skills	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • building of rapport • response to client's feelings and needs • greeting of client (e.g. stood up, welcomed, introduced self, directed to seat) • allowing client to complete statements without interruption • maintenance of focus on client (e.g. note taking does not disrupt flow of interview) 	
Skills	Performance	Comments

1.4 Interviewing Skills	unacceptable 1	poor 2	average 3	good 4	excellent 5	
	<ul style="list-style-type: none"> • use of open / closed questions / forced choice questions to gain specific information • use of verbal cues to indicate active listening • encouraging client to ask further questions • logical and systematic sequencing of questions • verification / clarification of information 					
2.0 PROFESSIONAL PRACTICE SKILLS						
The student demonstrated the following performance on stated behaviours:						
Skills	Performance					Comments
2.1 Professional Practice	unacceptable 1	poor 2	average 3	good 4	excellent 5	
	<ul style="list-style-type: none"> • Explanation of professional role • Identification and understanding of the reason for visit • Summary of interview for parent • Discussion of follow-up plan • Interview conducted in a professional manner • Integration of knowledge, evidenced by what, when and how information is elicited during the interview • Respecting and maintaining professional boundaries 					
3. CLINICAL SKILLS						
The student gained specific information in each history area below at the following performance level:						
Information	Performance					Comments
	unacceptable 1	poor 2	average 3	good 4	excellent 5	
<i>(Please tick which history area this student completed)</i>	<ul style="list-style-type: none"> • Birth/ developmental history <input type="checkbox"/> • Speech and language history <input type="checkbox"/> • Medical/family history <input type="checkbox"/> • History of previous support <input type="checkbox"/> • Educational history <input type="checkbox"/> • Interaction and socialisation <input type="checkbox"/> 					
4. OVERALL RATING OF INTERVIEW PERFORMANCE						
The student demonstrated communication, interview and interpersonal skills, professional and clinical practice skills at the following level during this interview:						
	unacceptable 1	poor 2	average 3	good 4	excellent 5	

Source: Hill, Davidson, and Theodoros. (2015: Appendix 1).

Appendix 2

Standardised Patient Interview Rating Scale (SPIRS)

Week 8: Follow-up Interview

Student Name: _____ Year level: _____ Date: _____

Following the interview you have observed, please rate the student's performance by circling **ONLY** the appropriate number on the scale below. Student **must** be rated at **one** of the 5 points only (not in between) using the performance descriptors below as a guide. Listed below each skill area are some specific aspects to consider to help you rate the student's performance and formulate your feedback comments. You do not need to limit your comments to the specific aspects listed.

Unacceptable – Demonstrates many behaviours in specified skill area(s) that are inappropriate or have negative outcomes or consequences (make the situation worse). The desired outcome is not achieved.

Average – Demonstrates a sufficient range of expected behaviours in specified skill area(s) to achieve the desired outcome. Some deficiencies exist in the skill area(s) assessed but none are of major concern.

Excellent – Consistently demonstrates the full range of expected behaviours in specified skill area(s) to achieve the desired outcome. An outstanding level of performance is maintained. No deficiencies exist in the skill area(s) assessed.

1. COMMUNICATION / INTERPERSONAL SKILLS		
The student demonstrated behaviours at the following performance levels in these skill areas:		
Skills	Performance	Comments
1.1 Non-verbal Communication	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • eye contact • use of facial expression • body language • use of gesture 	
	Performance	Comments
1.2 Verbal Communication	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • use of language and terminology • use of appropriate level of formality • speech volume • use of intonation • speech rate 	
	Performance	Comments
1.3 Interpersonal Skills	unacceptable 1 poor 2 average 3 good 4 excellent 5	
	<ul style="list-style-type: none"> • building of rapport • response to client's feelings and needs • greeting of client (e.g. stood up, welcomed, introduced self, directed to seat) • allowing client to complete statements without interruption • maintenance of focus on client (e.g. note taking does not disrupt flow of interview) 	
Skills	Performance	Comments

1.4 Interviewing Skills	unacceptable 1	poor 2	average 3	good 4	excellent 5	
	<ul style="list-style-type: none"> • use of open / closed questions / forced choice questions to gain specific information • use of verbal cues to indicate active listening • encouraging client to ask further questions • logical and systematic sequencing of questions • verification / clarification of information 					
2.0 PROFESSIONAL PRACTICE SKILLS						
The student demonstrated the following performance on stated behaviours:						
Skills	Performance					Comments
2.1 Professional Practice	unacceptable 1	poor 2	average 3	good 4	excellent 5	
	<ul style="list-style-type: none"> • Explanation of professional role • Identification and understanding of the reason for visit • Summary of interview for parent • Discussion of follow-up plan • Interview conducted in a professional manner • Integration of knowledge, evidenced by what, when and how information is elicited during the interview • Respecting and maintaining professional boundaries 					
3. CLINICAL SKILLS						
The student gained specific information in each history area below at the following performance level:						
Information	Performance					Comments
	unacceptable 1	poor 2	average 3	good 4	excellent 5	
<i>(Please tick which history area this student completed)</i>	<ul style="list-style-type: none"> • Discussion Assessment results <input type="checkbox"/> • Treatment plan <input type="checkbox"/> • Organisation of the clinic <input type="checkbox"/> • Willingness to participate in therapy program <input type="checkbox"/> 					
4. OVERALL RATING OF INTERVIEW PERFORMANCE						
The student demonstrated communication, interview and interpersonal skills, professional and clinical practice skills at the following level during this interview:						
	unacceptable 1	poor 2	average 3	good 4	excellent 5	

Source: Hill, Davidson, and Theodoros. (2015: Appendix 1).