Students as harsh critics: A comparison of student and industry assessment of placement competencies

SHAREN NISBET
JOHN HAW
Southern Cross University, Gold Coast, Australia
JULIA CALDICOTT[1]
Southern Cross University, Coffs Harbour, Australia

Reflecting on one's performance is a critical aspect of work integrated learning and professional life in general. This project compared 282 business and tourism students' self-assessment of their workplace performance against their industry supervisor's evaluation. The analysis included possible moderator variables such as gender, residential status, level of study and placement type. Overall, supervisors rated the importance of the competencies to their organization, and the level of attainment of the students on these, higher than the students themselves. Close to half of these ratings, on both measures, were found to be significantly different with no significant interactions revealed. These findings will be used to improve the evaluation form by combining some items, including descriptors, and supporting supervisors and students in its use. This will increase the capacity of students and employers to prepare for, engage in, and extract meaning from their work integrated learning (WIL) experiences.

Keywords: Work integrated learning, assessment, self-assessment, internship, supervisor evaluation

Assessment has long been recognized as a powerful tool for learning (Boud & Associates, 2010; Dochy et al., 1999; Hodges et al., 2014). At its core, assessment requires judgement about how a student's work meets certain criteria (Boud & Associates, 2010). Typically, this judgement is made by academics but increasingly, the value of other assessors including peers, workplace or practicum supervisors and the students themselves is recognized. Success is, however, highly contingent upon appropriate design, especially the constructive alignment of learning outcomes, teaching practices and assessment techniques (Biggs & Tang, 2007). In the capstone Work-Integrated Learning (WIL) context good-practice assessment is constructively aligned and systematically organized in a way that builds skills and experiences (Jaekel et al., 2011). Furthermore, capstone WIL: involves students (so as to develop their ability to evaluate quality and accuracy); facilitates lifelong learning; delivers high quality and useful feedback; and provides evidence of overall achievement of learning outcomes (Bailey et al., 2012).

Placements have a long history in certain disciplines such as teaching and nursing and are increasingly being utilized by other disciplines, such as business and tourism management, in an effort to improve graduate employability rates and address the calls of industry groups for graduates to be work ready. Despite the proliferation of other forms of WIL, internships, practicums and professional placements remain the most common (Jackson, 2017). Assessments involved in placement-based WIL often include learning goals, critical reflections and performance evaluations (Patrick et al., 2008). Performance evaluations, a common assessment and feedback tool in WIL, are typically completed by a student's workplace supervisor, and consider attainment on a range of technical and soft competencies (Ferns & Moore, 2012; Jackson, 2018; 2019; McNamara, 2013; Milne & Caldicott, 2016). These evaluations can be used as an assurance of learning tool, allow mapping of graduate outcomes, and contribute to program evaluation and improvement (Whelan, 2017). While it should be easy to assess functioning knowledge,

[1] Corresponding author: Julia Caldicott, Julia.caldicott@scu.edu.au

as placement competencies are often aligned with "real-life professional problems" (Biggs & Tang, 2007, p. 218), the socially constructed nature of the workplace will influence perceptions of performance. That is, while placement evaluations are highly suited to capstone WIL experiences it can, in practice, be difficult to measure attainment due to the influence of environmental and social factors (Biggs & Tang, 2007; Hodges et al., 2014; Smith, 2014).

Furthermore, capturing the student experience during WIL is complex given the number of stakeholders involved, and that learning in the workplace can be informal and incidental (Venville et al., 2018). In addition, stakeholders can have different expectations around the purpose, measures and outcomes associated with student evaluations (Ferns, 2012; Smith, 2014) and the role of the supervisor delivering the feedback is often not well understood (Peach et al., 2014). Whilst involving host supervisors in the assessment of a WIL student's performance is advocated (Bernard & Goodyear, 2013; Gonsalves & Freestone, 2007), associated challenges have been noted such as supervisor resistance to evaluating the student (McNamara, 2013), lack of skill and/or experience in student evaluation (Trede & Smith, 2014) and leniency of evaluation (Jackson, 2018; Milne & Caldicott, 2016; Wolf, 2015). In contrast to a leniency bias, the impact of asymmetric student-supervisor relationships on WIL assessment has been noted when supervisors enact their power through negative end of placement evaluations (Rees et al., 2020) and students conform to the supervisors' expectations rather than exercising initiative for fear of a poor grade (Aprile & Knight, 2020). When poorly executed, supervisor evaluation has little value-add (Peach et al., 2014).

Involvement of the student in the process of evaluating their workplace performance constitutes good assessment practice. It allows the student to develop, or demonstrate, their ability to evaluate the quality, completeness and/or accuracy of their work (Bailey et al., 2012). Self-assessment is particularly valued for its ability to develop lifelong learning skills (Bailey et al., 2012; Biggs & Tang, 2007; Hodges et al., 2014; Richardson et al., 2013), and is viewed as an essential part of professional competence (Boud & Falchikov, 2007). Self-assessment fosters the students' ability to select good evidence when judging their professional performance and make improvements as required in real time (Biggs & Tang, 2007; Montgomery, 2000).

There are, however, several key moderators of self-assessment or self-evaluation effectiveness as compared to academic evaluation including student ability, the timing of the evaluation, and inter-rater reliability, being the level of agreement between two independent assessors (Dochy et al., 1999). Student ability in this context refers to a student's academic performance (Dochy et al., 1999). There is evidence that good students under-rate themselves and weaker students over-rate themselves relative to academic assessments of ability (Biggs & Tang, 2007; Dochy et al., 1999; Jackson, 2014; Leach, 2012). There is limited evidence, however, regarding the applicability of this effect to the work placement setting. One study, (Prozesky et al., 2019) investigating the preparedness of interns, found the Dunning-Kruger effect (Kruger & Dunning, 1999), where students lacking in skill over-rated their abilities, was a plausible explanation for differences between student and supervisor evaluations. Ways to address this inequity in regular assessment settings include providing students an opportunity to discuss and therefore develop their understanding and to align self-assessment with how they will use such skills and knowledge in the future (Biggs & Tang, 2007; Boud & Associates, 2010).

Time effect considers a student's level of learning although again, the evidence is inconsistent as to the influence on self-rating ability. A 1995 study by Griffee (cited in Dochy et al., 1999) found that the year of study did not influence self-assessment but the timing of the assessment within the academic year did. That is, students rated themselves lower at the beginning of the year or session and higher as the

session progressed. This was thought to be related to student's increasing confidence, feedback received during the session, and an actual increase in skill. More recently, Jackson (2014) found that despite a lack of agreement between student and academic performance ratings, final year students self-rated closer to an academic score than their junior peers.

Inter-rater reliability is also a factor to consider in student self-assessment. A strong correlation has been found between student self-assessment and teacher assessment on an essay task, although students, in the study by Longhurst and Norton (1997) were deemed better at grading than at assessing their own deep processing, and weaker students were worse at understanding the criteria. Another study found that while students were capable of self-assessing knowledge and understanding, there was a discrepancy between the self and teacher assessments explained by a lack of alignment between the teaching and assessment practices of the time (Zoller & Ben-Chaim, 1997). These results support the use of constructive alignment and the use of clear descriptors for each criteria to improve the accuracy of student self-assessment and enhance their learning (Biggs & Tang, 2007).

COMPARING SELF AND SUPERVISOR PERFORMANCE RATINGS IN WORK PLACEMENTS

While the evidence regarding the positive impact of student self-assessment on learning is robust, fewer studies have considered differences between student self-assessments and workplace assessor ratings of WIL performance. These include studies comparing the ratings of students versus graduates (Rainsbury et al., 2002), student versus academics (Jackson, 2014), students and their workplace supervisors (Jackson, 2019) and graduates and their supervisors (Stark & Greggerson, 2016). These studies focused on which skills are deemed important in the given workplace. Other studies have sought to identify variation in how different groups evaluate student performance, or skill level in the workplace setting (Jackson, 2014; 2019; Stark & Greggerson, 2016).

Rainsbury et al. (2002) required students and graduates to rank the relative importance of 24 competencies for graduates entering the workforce. These authors predicted that there would be little difference between the two groups in their rankings of cognitive or hard skills and behavioral or soft skills. This was partially proven, as students and graduates similarly ranked five of the 24 measures: computer literacy, customer service orientation, teamwork and co-operation, self-confidence, and willingness to learn. Further, the graduates did place a greater importance on most of the competencies, resulting in a statistically significant difference between the graduates and students' ranking of both skillsets. This seems consistent with the Hodges et al. (2014) time effect discussed earlier. At a minimum, these findings suggest that experience situated in the workplace, such as placement based WIL, may help develop business students' awareness of the importance of graduate competencies in the workplace (Rainsbury et al., 2002).

Some studies have found that students and their academic supervisors rate the student's performance during their WIL placement differently. Jackson (2014) compared undergraduate business student's self-assessment on a range of competencies against a composite measure of academic-assessed student performance. Findings from the sample of approximately 1000 students concluded that students at all levels of learning overrated their abilities relative to academics ($p = .001$), although as above, this difference was more pronounced in the first-year cohort as compared to the final year students. Demographic variables were found to have no effect on the differences in student ratings although there was a weak effect for gender. Jackson (2014) concluded that the students are probably poor judges of their own ability, and likely require further instruction in the use of self-assessment. They further

asserted that the academic moderation processes controlled for differences that might have otherwise occurred, and therefore "provide confidence of interrater reliability"(p. 61).

More recently, Jackson (2019) considered differences in student and workplace supervisor ratings of 17 capabilities using a sample of 212 paired business student-supervisor evaluations. The students and supervisors ranked the same three capabilities highest, while the six lowest ranked capabilities were also the same between both groups. When comparing means of these items, however, Jackson (2019) found that workplace supervisors rated the students significantly higher than the students' self-assessments on 14 of the 17 capabilities. Jackson explains this as a consequence of the investment made by the supervisor in the student, rather than the student's actual performance.

The role of individual characteristics on the student self-ratings, namely gender, age, residency (international or domestic), specialization, and whether the organization was for-profit or not-for-profit was also considered by Jackson (2019). While none of these factors explained variation in ratings, an almost-significant ($p = .052$) effect was reported for international students, explained by the scales of effective communication and team work. Jackson's (2019) explanation for higher supervisor than student self-ratings, an emerging and consistent theme in the literature, is supported by the findings of a 2016 study of counselling interns and their supervisors (Stark & Greggerson, 2016). These researchers surveyed 118 post-masters counsellor-interns and 157 supervisors. These non-paired participants were asked to rate how often particular behaviors, drawn from a validated instrument, were displayed. Comparison of the mean ratings of each of the 51 items yielded nine of significance ($p < 0.05$). In each of these cases, the supervisors rated the frequency of the behavior higher than the student sample. Examples of these nine behaviors included goals focus, collaboration and contribution to the supervision process, tolerance of ambiguity and willingness to attempt new behaviors.

Research into the moderating factors for variation in student-supervisor ratings is also limited. Supervisor anonymity was found to decrease the variation between supervisor and student evaluation (Vinton & Wilke, 2009). Gender is another variable shown to have an impact on evaluation variation. Whilst studies have shown that females tend to receive slightly higher performance appraisals from students and supervisors (Chopra et al., 2020; Jackson, 2014), no significant difference between student and supervisor ratings for gender was found in Jackson's study (2019). Similarly, ratings of student performance has been reported as higher for domestic students than international students by both students and workplace supervisors (Jackson, 2019). Other than Jackson's (2019) study, literature regarding other factors impacting student evaluations during WIL such as remuneration of placement, age of student, and discipline of study, is lacking.

The consequences of variation of student and workplace supervisor rating of skill importance is not clear however ambiguous standards and a lack of expected skills have been cited as common reasons for negative internship experiences (Hartman et al., 2005; Sosland & Lowenthal, 2017). Equipping students with knowledge of which skills are deemed important by employers, and how to access this information from the workplace, may assist in managing expectations of students and workplace supervisors and decrease the disillusion associated with internship placements (Sweitzer & King, 2009). Furthermore, insights into the moderating factors of why students and supervisors rate internship students' skill level could influence WIL program design.

RESEARCH QUESTIONS

Two questions guided the research approach:

1. Is there a significant difference in mean importance ratings of skills between students and supervisors?
2. Is the difference between students and supervisors in rating the intern skill level moderated by the student characteristics of gender, residential status, level of study, and type of placement (paid/unpaid)?

METHOD

Design

To investigate both research questions a cross-sectional survey methodology was employed. The study utilized a paired-samples design (student-supervisor) in a similar approach to Jackson (2018; 2019). This provides greater power and economy over between-subjects design (Howell, 2013) and required two surveys with slight variations in wording depending upon the participant (student or supervisor).

The internship program was a credit bearing, core capstone unit which students undertook for 14 weeks. The placements were predominantly sourced by the students and typically undertaken in a tourism or hospitality organization. The supervisors were experienced managers within these organizations.

Participants

Initially, there were 292 students and 319 supervisors who completed the surveys over the years 2017 and 2018. The removal of the 27 supervisor scores that were not paired with student data left both samples at 292. There were 10 students who had been placed in two internships and completed two surveys and one of these was deleted to ensure each student observation was independent (the survey with the most missing data was deleted). This left a final sample of 282.

The student sample comprised around 70% female, 30% male, 82% undergraduate, 18% postgraduate, 55% international on-shore, and 45% domestic students. Their mean degree Grade Point Average was 4.6 (SD = 0.91) which is between a Pass (50-64%) and Credit (65-74%) in the grading system of most Australian universities. The placements were largely unpaid (54%), and 92% of the students would recommend the organization to future students seeking placement. Of the supervisor sample, 50% of organizations had hosted interns through the WIL program prior to the study. Most (95%) indicated they would consider hosting an intern again and over 93% of organizations indicated that they would be prepared to offer the intern paid employment following the internship if a position was available. Finally, just over half of all students (55%) were offered ongoing employment.

Procedure

An online questionnaire was constructed with Qualtrics© software that requested some descriptive information from the student and supervisor about their placement (hours worked, paid/unpaid) and was followed by 15 skill items for supervisors and students to independently assess. Both survey instruments included a set of 15 identically worded competencies rated on a five-point Likert scale for both their perceived importance to the organization and the level of student attainment. These competencies were originally developed according to the competing values framework developed by

Quinn et al. (2003) and had been progressively modified according to course learning outcomes and industry feedback. Students and supervisors were also asked to reflect upon their strengths and qualities, to offer suggestions for improvement or growth, and discuss the benefits of participation in the WIL program.

Supervisors completed the evaluation at the conclusion of the student's placement. This was an ungraded, summative assessment and, where consent was given, made available to the students. Students self-assess a few weeks earlier, submitting their evaluation with their placement report, which was graded. Data were collected across the 2017 and 2018 academic years and collated, cleaned and de-identified. The student ratings were then paired with and compared to the supervisor ratings. Ethics approval for this project was granted by the Southern Cross University Ethics Committee (ECN-17-026).

RESULTS

Student and Supervisor Ratings of Skill Importance

To examine the differences between students and supervisors in the importance of the competency item to the organization, a one-way repeated measures MANOVA was conducted. The independent variable was the rater, which had two levels (student or supervisor) and the 15 competency items served as dependent variables. The repeated measures MANOVA was conducted as students were matched with their supervisor to control for the different organizations that were being rated. This was necessary as different organizations may place a different level of importance on the competency. Hence, both supervisor and student were paired to ensure they were rating the importance of the skill for the same organization.

Missing data for most dependent variables was detected, with the highest being 30 cases for the 'Computer skills' variable when rated by supervisors. Utilizing listwise deletion across all variables, which removed pairs of cases with any missing data, resulted in the deletion of 73 cases, leaving a final sample size of 209. This still provided a sample size that was almost 14 times the number of dependent variables, meeting the sample size assumption for the planned MANOVA and providing adequate power (.80) to detect small effect sizes (n^2 = .03) at $p < .05$ (Tabachnick & Fidell, 2014). Most of the dependent variables were negatively skewed to varying degrees, however, with the large sample size and a robust statistical procedure, this was not a concern as the few outliers would have little effect on the means and variance (Allen & Bennet, 2008; Tabachnick & Fidell, 2014). It is also noteworthy that correlations and their scatterplots were generated between all dependent variables and met the assumptions of linearity and collinearity.

Most items rated by students covaried in strength from weak to moderate which is ideal for MANOVA (Tabachnick & Fidell, 2014). The lowest was the correlation between 'Computer skills' and 'Capacity for cooperation and teamwork' (r = .13) and the highest was between 'Information Literacy' and 'Analytical and problem solving skills'. It was a similar result for the supervisor data, although with slightly stronger correlations. The weakest was again the relationship between 'Computer skills' and 'Capacity for cooperation and teamwork' (r = .20) and the strongest was between 'Information literacy' and 'Analytical and problem solving skills'. This suggests that organizations that value computer skills do not value teamwork as highly (and vice-versa) and organizations that value information literacy do tend to value problem solving skills. It can also be concluded that each item was measuring something different which is a necessary theoretical consideration for MANOVA (Tabachnick & Fidell, 2014).

The MANOVA result was significant ($F(15, 194) = 4.175$, $p < .001$, $n^2 = .244$) indicating that there was at least one competency item where the mean for students was significantly different to the mean for supervisors. Inspection of univariate tests revealed that there were seven items with significant differences, as displayed in Table 1. As indicated by the mean scores in Table 1, the general trend was for supervisors to rate the importance of each competency higher than the student. This occurred for all items except 'Knowledge of industry' where supervisors rated that slightly lower in importance. It can also be seen that, on average, students and supervisors rated 'Positive work attitude and work ethic' as the most important skill to the organization. However, the least important skill for the students was 'Computer skills' whereas the supervisors rated 'Knowledge of industry' the lowest.

The MANOVA revealed significant differences in seven of the 15 items, as indicated in the F ratio column in Table 1. In each of these, the supervisor had rated the importance of the skill significantly higher than the student had. The largest difference was for the item assessing the importance of 'Appropriate personal presentation and grooming'. The supervisors mean score was 4.41 and the students mean score was 4.07. Overall, the descriptive and inferential statistics show that there is a dissonance between the student intern and their supervisors in rating the importance of the competency items to the organization.

The Effect of Student Characteristics on Ratings of Skill Level

The next analysis focused on the difference between the student and the supervisor in their rating of student performance. Of particular importance was the inclusion of moderating variables to test for any observed differences between the student and the supervisor. These included students' gender, residential status, level of study, and the type of placement. These were selected as interaction effects have been explored in other, similar studies, as documented in the above literature review.

Consideration was given to doing one five-way mixed MANOVA with all variables, but a preliminary cross-tab analysis revealed some very small cell sizes (e.g., there were only two domestic, postgraduate students in the sample and only eight male students who were domestic and in paid positions). This is problematic for MANOVA (Tabachnick & Fidell, 2014). Also, the 2 x 2 x 2 x 2 x (2) design creates a highly complex model that is not well supported by the literature. To best answer the research question it was optimal to perform four two-way mixed MANOVA's and use an adjusted alpha again ($p < .0125$) (Tabachnick & Fidell, 2014). Each of these included the student/supervisor rating of the student's performance as the within-group variable and the between groups variable would be gender, residential status, level of study and placement type.

Gender

Assumptions for the 2 x (2) MANOVA were met (Allen & Bennett, 2008; Tabachnick & Fidell, 2014), with the exception of the cell imbalance for gender. There was almost 2.5 times as many women (n = 150) as men (n = 61) and this imbalance was addressed with the estimation of marginal means (Tabachnick & Fidell, 2014). Results of the MANOVA indicated no significant interaction between gender and rater. That is, the rating of the student performance by students and supervisor was not dependent upon the gender of the student. There was, however, significant main effects for both gender $F(15, 195) = 2.571$, $p = .002$, $n^2 = .165$ and the rater, $F(15, 195) = 3.451$, $p < .001$, $n^2 = .210$.

TABLE 1: Mean scores for students and supervisors on the importance of competency item to organization (N = 209). ***p< .001, **p < .01, *p< .05

	Student		Supervisor		*F*	n^2
	Mean	SD	Mean	SD		
Capacity to learn new skills and procedures	4.22	.68	4.41	.65	9.90**	.05
Appropriate personal presentation and grooming	4.07	.77	4.41	.71	30.43***	.13
Punctuality and attendance	4.51	.68	4.64	.55	5.97*	.03
Computer skills	3.90	.93	4.09	.83	6.21*	.03
Knowledge of industry/sector within both local and international contexts	4.00	.81	3.97	.73	.26	.00
Awareness of cultural, environmental & economic impacts of business management	3.91	.88	4.06	.76	3.79	.02
Written communication skills	4.05	.91	4.21	.74	4.82*	.02
Interpersonal communication skills (verbal)	4.50	.64	4.49	.56	.07	.00
Capacity to appreciate different cultural perspectives/intercultural competence	4.22	.79	4.25	.72	.26	.00
Commitment to ethical standards and professionalism	4.44	.73	4.46	.66	.15	.00
Proactive, responsive and adaptable to change	4.34	.73	4.52	.61	8.39**	.04
Ability to manage large volumes of information from various sources (information literacy)	4.04	.89	4.23	.77	7.04**	.03
Have the analytical and problem-solving skills needed to work effectively	4.36	.73	4.29	.70	1.27	.01
Capacity for co-operation and teamwork	4.51	.73	4.62	.55	3.26	.02
Positive work attitude and work ethic	4.59	.61	4.66	.51	1.84	.01

With regard to gender, the only competency items to achieve significance were teamwork and positive attitude, however these were at $p = .044$ and $p = .025$ and failed to achieve significance with the adjusted $\alpha < .0125$. Supporting this, and the power of the analysis undertaken, the estimated marginal means showed that the difference in the rating of men and women interns was very small. For teamwork the mean for men was 4.40, women = 4.54 and for positive attitude it was men = 4.50 and women = 4.65.

With the observed main effect for the rater variable, there were nine competency items that achieved significance ($p < .0125$). These are indicated in Table 2.

For each item that achieved significance, the supervisor had rated the student's skill level higher than the students rating of themselves and significantly so for nine of these. The item with the largest

difference was 'Appropriate personal presentation and grooming' with the supervisors mean evaluation of the student being 4.42 and the students mean evaluation = 4.15. Overall, the conclusion from this data subset is that supervisors tended to rate the intern higher than the intern themselves did, and this was not dependent upon the gender of the intern and there was also no significant main effect for gender in the competency ratings.

TABLE 2: Student and supervisor ratings of intern skill level (N = 211). *p < .0125

	Student		Supervisor		F	n^2
	Mean	SE	Mean	SE		
Capacity to learn new skills and procedures	4.05	.05	4.28	.06	12.28*	.06
Appropriate personal presentation and grooming	4.15	.06	4.43	.05	18.05*	.08
Punctuality and attendance	4.57	.05	4.56	.05	.05	.00
Computer skills	3.98	.06	4.26	.05	17.38*	.08
Knowledge of industry/sector within both local and international contexts	3.57	.05	3.82	.06	12.32*	.06
Awareness of cultural, environmental & economic impacts of business management	3.64	.05	3.95	.06	16.61*	.07
Written communication skills	3.96	.06	4.15	.06	6.42*	.03
Interpersonal communication skills (verbal)	3.91	.06	4.16	.06	10.79*	.05
Capacity to appreciate different cultural perspectives/intercultural competence	4.07	.05	4.27	.05	7.60*	.04
Commitment to ethical standards and professionalism	4.29	.05	4.39	.05	2.27	.01
Proactive, responsive and adaptable to change	4.18	.05	4.32	.06	3.55	.02
Ability to manage large volumes of information from various sources (information literacy)	3.76	.06	4.06	.06	15.43*	.07
Have the analytical and problem-solving skills needed to work effectively	3.91	.05	4.04	.06	2.98	.01
Capacity for co-operation and teamwork	4.44	.05	4.49	.05	.54	.00
Positive work attitude and work ethic	4.55	.04	4.60	.04	.76	.00

Residential status

The residential status variable comprised 97 domestic students and 114 international students. The assumptions for the 2 x (2) MANOVA were met and the results revealed no significant interaction effect ($p > .0125$). With the same variable as in the gender analysis (i.e. rater), the results suggest that the tendency for supervisors to rate interns higher than the interns themselves was not dependent upon

the residential status of the intern. There was a significant main effect for residential status $F(15, 195) = 6.625$, $p < .001$, $n^2 = .338$. Focusing on the main effect for residential status it was revealed that there were 10 significant competency items. The means and F ratios are presented in Table 3.

TABLE 3: Combined student and supervisor ratings of skill levels for domestic and international students. *$p < .0125$

	Domestic		International		F	n^2
	Mean	SE	Mean	SE		
Capacity to learn new skills and procedures	4.33	.05	4.04	.05	18.03*	.08
Appropriate personal presentation and grooming	4.56	.05	4.09	.05	38.65*	.16
Punctuality and attendance	4.67	.05	4.53	.05	2.79	.01
Computer skills	4.20	.06	4.03	.05	5.12	.02
Knowledge of industry/sector within both local and international contexts	3.61	.06	3.71	.05	1.74	.01
Awareness of cultural, environmental & economic impacts of business management	3.69	.05	3.83	.05	3.49	.02
Written communication skills	4.25	.05	3.90	.05	22.75*	.10
Interpersonal communication skills (verbal)	4.19	.06	3.91	.05	11.72*	.05
Capacity to appreciate different cultural perspectives/intercultural competence	4.22	.05	4.16	.05	.69	.00
Commitment to ethical standards and professionalism	4.52	.05	4.24	.05	17.62*	.08
Proactive, responsive and adaptable to change	4.37	.05	4.17	.05	7.10*	.03
Ability to manage large volumes of information from various sources (information literacy)	4.07	.06	3.83	.05	9.92*	.05
Have the analytical and problem-solving skills needed to work effectively	4.11	.05	3.86	.05	12.56*	.06
Capacity for co-operation and teamwork	4.62	.05	4.40	.04	12.03*	.05
Positive work attitude and work ethic	4.71	.04	4.51	.04	11.58*	.05

As depicted in Table 3 it is evident that international students were rated lower on all but two of the 15 items. The two items where international students performed better than domestic students were 'Knowledge of industry/sector' within both local and international contexts and 'Cultural awareness'. The ten that achieved significance were all in the direction of international students scoring lower on the competency item. Overall, the student and supervisor ratings of competency were not dependent upon the residential status of the intern. That is, supervisors tended to rate the intern higher than the

students regardless of residential status. There was a main effect for residential status however, indicating that, on most items, the domestic students were considered more competent than the international students.

Level of study and placement type

Two separate 2 x (2) MANOVA's were completed for these independent variables however the results were the same for both, and are presented together here. For both, there was no significant interaction with the rater variable (level of study $F(15, 195) = .917$, $p = .546$; placement type $F(15, 195) = .1.556$, $p = .089$) and no significant main effect for either level of study $F(15, 195) = .543$, $p = .914$ or placement type $F(15, 195) = .802$, $p = .674$. These results indicated that whilst supervisors tended to rate students higher than the students themselves, this was not dependent upon the student's level of study or if their placement was paid or unpaid.

Overall, the results for second research question indicated that supervisors tended to rate the competency of the student higher than the students did themselves. This pattern was not dependent upon the gender of the student, their residential status, their level of study or their placement type. There were significant effects for gender and residential status without considering the different raters. That is, both supervisors and students tended to rate women more highly on some competency items than men and also tended to rate domestic students more highly on some competency items than international.

DISCUSSION

This study sought to highlight the similarities and differences between student and industry supervisor ratings of WIL performance with a view to testing the reliability and validity of these assessments. Specifically, it sought to understand significant differences in mean importance ratings between students and supervisors and consider whether differences between skill level ratings are moderated by selected student characteristics. Developing our understanding of differences in the importance ratings is key to our study because it demonstrates whether students are able to evaluate "quality and completeness" (Bailey et al., 2012, p. 44). That is, it indicates student's ability to reflect upon their capacity and capability as industry professionals. Developing such professional acumen is a common learning outcome of WIL (Bowen, 2020).

The first key finding is the general trend for workplace supervisors to rate the importance of the competencies higher than the students. This finding is consistent with other studies (Jackson, 2019; Rainsbury et al., 2002). Of the 15 items, seven were rated significantly higher by the workplace supervisors and students. The largest difference was for the item assessing the importance of 'Appropriate personal presentation and grooming', which was valued much more by the supervisors than students. This was followed by 'Capacity to learn new skills and procedures' and 'Proactive, responsive and adaptable to change'. Many (54, representing 18% of the sample) students undertook placements in hotels where there is an emphasis on professional appearance/grooming. Supervisors attributing higher importance for presentation and grooming is consistent with findings of the importance of aesthetic labour management in the hospitality industry (King & Grace, 2010; Wu et al., 2020). Higher importance of the 'Capacity to learn new skills' and 'Proactive, responsive and adaptable to change' potentially reflects industry supervisors assuming that interns have limited workplace experience.

There were nine competency items where the supervisor had rated the student's skill level significantly higher than the students rating ($p < .0125$), although supervisors rated the students higher on all 15

items. This finding is contrary to studies which report students rating their performance higher than their workplace supervisor (Bailey-Residori, 2017; Jones et al., 2001; Langendyk, 2006; Prozesky et al., 2019). The item with the largest difference was 'Appropriate personal presentation and grooming' with the supervisors mean evaluation being 4.42 and the students mean evaluation for this item being 4.15. This finding indicates that whilst students did not value presentation and grooming as important as workplace supervisors, they did meet the industry expectations in this regard.

This study's findings suggest that evaluations can be subjective and prone to bias. Consistent with other WIL research, such subjectivity challenges the validity and reliability of this process as an indicator of performance (Billett, 2015; Ferns & Moore, 2012; Gonsalvez & Freestone, 2007; Jiang et al., 2015; McNamara, 2013). According to Smith (2014, p. 210), "validity in assessment is a matter of whether and to what degree a protocol […] measures the thing it purports to measure". Ways to address validity issues are to include consideration of multiple dimensions from a variety of sources, and to carefully consider what to assess (Smith, 2014). The goal, then, is to increase the usefulness of the evaluation by providing clear and helpful feedback that evidences overall achievement (Bailey et al., 2012).

The differences between student and supervisor ratings in our study appear to suggest that the instrument and its items lack validity and reliability. Closer examination of the data, however, shows that the actual differences are small, as evidenced by the difference of .27 between student and supervisor ratings of the item with the largest (significant) difference. The item with the next largest difference, of .25, is 'Knowledge of industry/sector within both local and international contexts'. The high number of items with significant differences yet small actual ones could be explained by the very large sample size.

Other explanations considered as per our analysis included gender, residential status, level of study and placement type (paid versus paid). None of these explained the higher overall supervisor skill level ratings although, there were significant effects for gender and residential status without considering the different raters. This echoes some of the findings of an earlier, comparable study by Jackson (2019) but is contrary to recent findings by Arsenis and Flores (2020) of economic undergraduates undertaking a professional training year, who found that females rated themselves lower than males on competencies to perform a job and there was limited evidence of differences in workplace supervisor's assessment by gender. With regard to the effect of residential status on ratings, it is likely that international students (having English as a second language) and their supervisors measure their performance in comparison to domestic students/employees and this would be influenced by their less well-developed English language skills. This finding affirms the ethnocentric attitude of Australian internship hosts and the deficit perception of international students (Andrew, 2020).

LIMITATIONS

Several limitations are associated with this study. The regional location of most of the students and, typically, their WIL workplaces will limit generalizability to other student populations, and the disciplines are limited to business and tourism. Additionally, we could not test for time effect as the sample did not contain multiple year cohorts. While the students were most likely final session students, given the design of the degrees and the placement of subjects within them, this could not be guaranteed. It is possible, and there is some evidence to suggest, that students in the later stages of

their degree have gained the skills and knowledge to be able to self-assess more accurately than first- or second-year students (Jackson, 2014).

Our assumption that there should be parity between student and supervisor ratings of performance, that is, that objectivity is both desirable and possible, is also possibly flawed. It is reasonable to question whether this a fair goal given what is known about the influence of environmental and individual factors in external settings. If validity and reliability could be shown, however, our thesis is that this would indicate good assessment practice. If they are not, then there is scope to provide better descriptors and/or training for supervisors. This goal is supported by the evidence that says parity is possible for academically advanced cohorts of students where good descriptors of the competency exist (Biggs & Tang, 2007).

PRACTICAL IMPLICATIONS

This study has several practical implications. The results will, inform curriculum (re)design at Southern Cross University, and highlight any needed changes to support services for all WIL partners. Specifically, differences in student and supervisor evaluations of workplace performance and competencies have been highlighted. While there are some areas of parity, and this suggests that the instruments are useful and are measuring attainment that can provide students with feedback for future development, there are several areas that require attention. Despite our analysis revealing that each item was measuring something different, there was a high correlation between some items including 'Information literacy' and 'Analytical and problem-solving skills' for both the student and supervisor groups. There is scope to remove or combine these items in the evaluation form and therefore reduce the number overall.

Improved descriptors, to help stakeholders discriminate between items and to better rate student attainment on those items would also likely lead to greater parity and reduce bias, particularly that bias that manifests as lower ratings for international students. Well-designed evaluation forms, or rubrics, ensure fair and consistent marking and promote validity and reliability in the assessment of student performance (Chan & Ho, 2019). This change would also more effectively support workplace supervisors in the completion of the evaluation. Supervisors who have more confidence in their ability to accurately and adequately evaluate student performance will become more effective and confident workplace mentors and this will have flow on effects for employability of all cohorts.

Students could also be supported to develop their understanding of the items and the utility of the process overall. The results of this analysis can be incorporated into the design of the employability /internship preparation unit. This will allow for discussion of the factors valued by internship hosts, and encourage students to consider their own level of attainment for example, in relation to personal presentation/grooming and industry knowledge. Studies have found that careful instruction around the use of evaluation tools such as rubrics decreases the inequities in assessment practices and resultant student performance (Greenberg, 2015; Jones et al., 2017).

During the placement student reflection on any differences in the ratings could be encouraged, via the addition of a short, final assessment item such as a debrief. This might require a change to the timing of the supervisor evaluation, which currently comes at the conclusion of the 14 weeks, after the student has submitted all of their assessments. These changes, accompanied by ongoing evaluation, should lead to a revitalized evaluation form and associated process that will, ultimately, ensure the ongoing viability and utility of the Southern Cross University WIL program while enhancing its value for all

stakeholders. This new form need not be discipline specific, as it measures generic employability skills like those that form the basis of most institution's graduate attributes.

CONCLUSION

In conclusion, this project's comparison of business and tourism students' self-assessment of their workplace performance against their industry supervisor's evaluation has found that supervisors rate the importance and attainment of the competencies higher than the students themselves. These findings will be used to improve the evaluation form, resulting in clearer and more targeted descriptors and a refined assessment regime. As a consequence, the students will extract more meaning from their work integrated learning (WIL) experiences and the institution will be able to assure student learning against key employability measures.

REFERENCES

Allen, P. J., & Bennett, K. (2008). *SPSS for the health & behavioural sciences*. Thomson Learning.

Andrew, L. (2020). Ensuring equitable work-integrated learning opportunities for international students. *Journal of University Teaching & Learning Practice*, *17*(4), Article 13.

Aprile, K. T., & Knight, B. A. (2020). The WIL to learn: Students' perspectives on the impact of work-integrated learning placements on their professional readiness. *Higher Education Research & Development*, *39*(5), 869-882. https://doi.org/10.1080/07294360.2019.1695754

Arsenis, P., & Flores, M. (2020). Confidence and gender differences within a work-integrated learning programme: Evidence from a UK higher education institution. *Higher Education Research & Development*, *40*(5), 1-17. https://doi.org/10.1080/07294360.2020.1798888

Bailey, J., van Acker, E., & Fyffe, J. (2012). *Capstone subjects in undergraduate business degrees: A good practice guide.* Griffith University.

Bailey-Residori, A. L. (2017). *Are students competent? Methods of assessing Bachelor of Social Work student competence* [Ed.D. Dissertation, Olivet Nazarene University]. Olivet Digital Commons. https://digitalcommons.olivet.edu/edd_diss/109

Bernard, J. M., & Goodyear, R. K. (2013). *Fundamentals of clinical supervision.* Merrill.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university: What the student does* (3rd ed.). McGraw-Hill Education.

Billett, S. (2015). *Integrating practice-based experiences into higher education*. Springer.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council.

Boud, D., & Falchikov, N. (2007). Developing assessment for informing judgement. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 181-197). Routledge.

Bowen, T. (2020). Examining students' perspectives on gender bias in their work-integrated learning placements. *Higher Education Research & Development*, *39*(3), 411-424. https://doi.org/10.1080/07294360.2019.1677568

Chan, Z., & Ho, S. (2019). Good and bad practices in rubrics: The perspectives of students and educators. *Assessment & Evaluation in Higher Education*, *44*(4), 533-545. https://doi.org/10.1080/02602938.2018.1522528

Chopra, S., Khan, A., Mirsafian, M., & Golab, L. (2020) Gender differences in work-integrated learning experiences of STEM students: From applications to evaluations. *International Journal of Work-Integrated Learning, 21*(3), 253-274.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331-350. https://doi.org/10.1080/03075079912331379935

Ferns, S. (2012, October 29 - November 2). *Graduate employability: Teaching staff, employer and graduate perceptions*. [Paper presentation]. Australian Collaborative Education Network National Conference, Geelong, Vic, Australia.

Ferns, S., & Moore, K. (2012). Assessing student outcomes in field work placements: An overview of current practice. *Asia-Pacific Journal of Cooperative Education, 13*(4), 207-224.

Greenberg, K. P. (2015). Rubric use in formative assessment: A detailed behavioural rubric helps students improve their scientific writing skills. *Teaching of Psychology, 42*(3), 211–217.

Gonsalvez, C. J., & Freestone, J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist, 42*(1), 23-32. https://doi.org/10.1080/00050060600827615

Hartman, D., Bentley, J., Richards, K. & Krebs, C. (2005). Administrative tasks and skills needed for today's office: The employees' perspective. *Journal of Education for Business, 80*(6), 347–357. https://doi.org/10.3200/JOEB.80.6.347-357

Hodges, D., Eames, C., & Coll, R. K. (2014). Theoretical perspectives on assessment in cooperative education placements. *Asia Pacific Journal of Education, 15*(3), 189-207.

Howell, D. (2013). *Statistical methods for psychology*. Wadsworth Cengage Learning.

Jackson, D. (2014). Self-assessment of employability skill outcomes among undergraduates and alignment with academic ratings. *Assessment & Evaluation in Higher Education, 39*(1), 53-72. https://doi.org/10.1080/02602938.2013.792107

Jackson, D. (2017). Developing pre-professional identity in undergraduates through work-integrated learning. *Higher Education*, *74*(5), 833-853. https://doi.org/10.1007/s10734-016-0080-2

Jackson, D. (2018). Challenges and strategies for assessing student workplace performance during work-integrated learning. *Assessment & Evaluation in Higher Education, 43*(4), 555-570. https://doi.org/10.1080/02602938.2017.1378618

Jackson, D. (2019). Students' and their supervisors' evaluations on professional identity in work placements. *Vocations and Learning*, *12*, 245-266. https://doi.org/10.1007/s12186-018-9207-1

Jaekel, A., Hector, S., Northwood, D., Benzinger, K., Salinitri, G., Johrendt, J., & Watters, M. (2011). Development of learning outcomes assessment methods for co-operative education programs. *Journal of Cooperative Education and Internships, 45*(1), 11-33.

Jiang, Y. H., Lee, S. W. Y., & Golab, L. (2015). Analyzing student and employer satisfaction with cooperative education through multiple data sources. *Asia-Pacific Journal of Cooperative Education, 16*(4), 225-240.

Jones, A., McArdle, P. J., & O'Neill, P. A. (2001). How well prepared are graduates for the role of pre-registration house officer? A comparison of the perceptions of new graduates and educational supervisors. *Medical Education*, *35*, 578-584. https://doi.org/10.1046/j.1365-2923.2001.00955.x

Jones, L., Allen, B., Dunn, P., & Brooker, L. (2017). Demystifying the rubric: A five-step pedagogy to improve student understanding and utilisation of marking criteria. *Higher Education Research & Development*, *36*(1), 129-142. https://doi.org/10.1080/07294360.2016.1177000

King, C., & Grace, D. (2010). Building and measuring employee-based brand equity. *European Journal of Marketing*, *44*(7/8), 938–971. https://doi.org/10.1108/03090561011047472

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognising one's incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Langendyk, V. (2006). Not knowing that they do not know: Self assessment accuracy of third-year medical students. *Medical Education*, *40*(2), 173–179. https://doi.org/10.1111/j.1365-2929.2005.02372.x

Leach, L. (2012). Optional self-assessment: Some tensions and dilemmas. *Assessment & Evaluation in Higher Education, 37*(2), 137-147. https://doi.org/10.1080/02602938.2010.515013

Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation*, *23*(4), 319-30.

McNamara, J. (2013). The challenge of assessing professional competence in work integrated learning. *Assessment and Evaluation in Higher Education, 38*(2), 183-197. https://doi.org/10.1080/02602938.2011.618878

Milne, L., & Caldicott, J. (2016). Exploring differences in industry supervisors' ratings of student performance on WIL placements and the relative importance of skills: Does remuneration matter? *Asia-Pacific Journal of Cooperative Education, 17*(2), 175-186.

Montgomery, K. (2000). Classroom rubrics: Systematizing what teachers do naturally. *Clearing House, 73*(6), 324-328. https://doi.org/10.1080/00098650009599436

Patrick, C.-J., Peach, D., Pocknee, C., Webb, F., Fletcher, M., &. Pretto, G. (2008). *The WIL (work integrated learning) report: A national scoping study.* Australian Learning and Teaching Council. https://eprints.qut.edu.au/44065/1/WIL-Report-grants-project-jan09.pdf

Peach, D., Ruinard, E., & Webb, F. (2014). Feedback on student performance in the workplace. *Asia Pacific Journal of Education, 15*(3), 241-252.

Prozesky, D. R., Molwantwa, M. C., Nkomazana, O., & Kebaetse, M. B. (2019). Intern preparedness for the CanMEDS roles and the Dunning-Kruger effect: A survey. *BMC Medical Education*, *19*(1), Article 422. https://doi.org/10.1186/s12909-019-1836-z

Quinn, R., Faerman, S., Thompson, M., & McGrath, M. (2003). *Becoming a master manager: A competency framework* (2nd ed.). John Wiley and Sons.

Rainsbury, E., Hodges, D., & Burchell, N. (2002). Ranking workplace competencies: Student and graduate perceptions. *Asia-Pacific Journal of Cooperative Education, 3*(2), 8-18.

Rees, C. E., Davis, C., King, O. A., Clemans, A., Crampton, P. E., Jacobs, N., McKeown, T., Morphet, J., & Seear, K. (2020). Power and resistance in feedback during work-integrated learning: Contesting traditional student-supervisor asymmetries. *Assessment & Evaluation in Higher Education*, *45*(8), 1136-1154. https://doi.org/10.1080/02602938.2019.1704682

Richardson, J., Jackling, B., Henschke, K., & Tempone, I. (2013). Developing a collaborative model of industry feedback for work placement business students. *Asia-Pacific Journal of Cooperative Education, 14*(1), 28-43.

Smith, C. (2014). Assessment of student outcomes from work-integrated learning: Validity and reliability. *Asia-Pacific Journal of Cooperative Education, 15*(3), 209-223.

Sosland, J. K., & Lowenthal, D. J. (2017). The forgotten educator: Experiential learning's internship supervisor. *Journal of Political Science Education*, *13*(1), 1-14. https://doi.org/10.1080/15512169.2016.1165106

Stark, M., & Greggerson, K. (2016). Differences in perceptions of supervisee contribution: Supervisors' vs. supervisees' evaluations. *Journal of Counselor Preparation and Supervision, 8*(1), Article 1.

Sweitzer, H. F., & King, M. A. (2009). *The successful internship: Personal, professional, and civic development*. Cengage Learning.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson Education.

Trede, F., & Smith, M. (2014). Workplace educators' interpretations of their assessment practices: A view through a critical practice lens. *Assessment & Evaluation in Higher Education*, *39*(2), 154-167. https://doi.org/10.1080/02602938.2013.803026

Venville, A., Lynch, B., & Santhanam, E. (2018). A systematic approach to the evaluation of the student experience in work-integrated learning. *International Journal of Work-Integrated Learning*, *19*(1), 13-21.

Vinton, L., & Wilke, D. J. (2011). Leniency bias in evaluating clinical social work student interns. *Clinical Social Work Journal*, *39*(3), 288-295. https://doi.org/10.1007/s10615-009-0221-5

Whelan, M. (2017). Road testing graduate attributes and course learning outcomes of an environmental science degree via a work-integrated learning placement. *Asia-Pacific Journal of Cooperative Education, 18*(1), 1-13.

Wolf, K. (2015). Leniency and halo bias in industry-based assessments of student competencies: A critical, sector-based analysis. *Higher Education Research & Development*, *34*(5), 1045-1059. https://doi.org/10.1080/07294360.2015.1011096

Wu, L., King, C. A., Lu, L., & Guchait, P. (2020). Hospitality aesthetic labor management: Consumers' and prospective employees' perspectives of hospitality brands. *International Journal of Hospitality Management*, *87*, Article 102373. https://doi.org/10.1016/j.ijhm.2019.102373

Zoller, Z., & Ben-Chaim, D. (1997, August 26-30). *Student self-assessment in HOCS Science Examinations: Is it compatible with that of teachers?* [Paper presentation]. European Association for Research on Learning and Instruction 7th Biennial Conference, Athens,