2013

Self-assessment of employability skill outcomes among undergraduates and alignment with academic ratings

Denise Jackson
Edith Cowan University, d.jackson@ecu.edu.au

10.1080/02602938.2013.792107

Self-assessment of employability skill outcomes among undergraduates and alignment with academic ratings

Abstract

Despite acknowledgement of the benefits of self-assessment in higher education, disparity between student and academic assessments, with associated trends in over- and underrating, plagues its meaningful use, particularly as a tool for formal assessment. This study examines self-assessment of capabilities in certain employability skills in more than 1000 Australian business undergraduates. It evaluates the extent to which student self-assessments differ from academics, in what ways and the influence of certain individual and background characteristics - such as stage of degree, gender and academic ability - on rating accuracy. Explanations for documented disparities are presented, in addition to implications and strategies for educators.

Keywords

Self-assessment; evaluation; performance; reflection; skill assessment.

There has been considerable focus on the purpose, value and implementation of self-assessment in higher education in recent years. Self-assessment is defined by Boud and Falchikov (1989) as 'the involvement of learners in making judgements about their achievements and the outcomes of their learning' (529). It requires academics and students to engage in a dialogue to specify standards which apply to their work and make judgements on the degree to which they have been met (Boud 1995). Certain principles of self-assessment are discussed later in the paper but there is significant emphasis on learners actively considering, negotiating and evaluating the criteria for self-assessment (see Brew 1999). In her review of literature, Leach (2012) acknowledges variations in the understanding and conception of self-assessment. Self-assessment is typically formative (Leach 2012) and, in its purest form (Andrade and Valtcheva 2009), may offer students opportunities to revise and resubmit their work based on identified strengths and weaknesses. It may also be summative as a one-time assessment reflecting on specific criteria and which results in a mark contributing to the student's final grade, described by Andrade and Du (2007) as self-evaluation.

This paper explores the use of formative self-assessment to complement and enhance employability skill development in undergraduates within an Australian Business Faculty. Employability skills, otherwise referred to as generic, core, key or professional skills, are those which enable new graduates to effectively apply their technical knowledge in the workplace and typically comprise communication, team working, self-management and problem-solving skills (Jackson and Chapman 2012a). Related to this pursuit of the 'rounded' graduate is a shift in pedagogy towards student-centred learning. The fostering of autonomy and self-direction within a climate of academic support and feedback is widely acknowledged as enhancing undergraduate learning (see Johan and Clarke 2012); self-regulation vital as

students monitor and direct their own learning to achieve goals (Andrade and Valtcheva 2009). Lew et al. (2010) argue self-regulation can be enhanced through self-assessment as students decipher expected standards, reflect on their progress and direct their behaviour accordingly to achieve learning goals more effectively. Here, students develop their ability to become 'reflective practitioners' who are able to critically evaluate their own practices (Boud and Falchikov 1989, 530). Certain studies have confirmed students become more responsible learners when actively self-assessing their work (Dochy, Segers and Sluijsmans 1999; Lopez and Kossack 2007); in addition to improvements in other employability skills such as problem solving (Dochy et al. 1999) and critical thinking (Thompson, Pilgrim and Oliver 2005), among others (see Falchikov 2005).

A review of relevant literature reveals our understanding of the dimensions to and influences on self-assessment are far from conclusive. The need to incorporate self-assessment in a way which enhances student learning, rather than simply providing a distraction for students or lessening academic marking load, prompts further investigation. Research questions for this study are i) to what extent do student self-assessments of capabilities in employability skills differ from academic assessments; ii) do students of different abilities vary in their tendency to over or underrate themselves compared to academics; iii) which individual characteristics, if any, influence the self-assessment process; and iv) does the stage of degree impact student ability to rate their performance.

The research questions will be addressed through the self-assessment of capabilities in certain employability skills by more than 1000 undergraduates in the Business faculty of an Australian university. The study will improve our understanding of the accuracy of student self-assessments of their employability skills and the influence of certain factors on their

ratings. The paper first reviews relevant background literature on self-assessment within the context of the skills agenda, problems associated with self-assessment and perceived influences on the process. This is followed by an outline of methodology and a discussion of the findings and implications for stakeholders in undergraduate education.

Background

Self-assessment and the skills agenda

Higher education is no longer solely focused on developing disciplinary expertise through critical inquiry. Industry calls for work-ready graduates, and associated government funding and national skills initiatives, have caused a shift towards producing graduates who are technical experts and appropriately equipped with a range of employability skills considered essential for applying disciplinary knowledge. These highly regarded skills are defined in national skills frameworks which now permeate school, vocational and tertiary education sectors. In Australian universities, the national framework is typically used to produce an institutional framework which defines essential employability skills, or graduate attributes, which students are expected to master upon graduation (Department of Education, Science and Training [DEST] 2002). The use of self-assessment in higher education may complement, and possibly augment, the development of these employability skills in a number of ways.

First, included in Australia's national framework is 'self-management' which encompasses 'evaluating and monitoring own performance'. Developing meta-cognition, defined from an employability perspective as the process of self-regulation, reflection and learning how to learn, is widely acknowledged as vital for graduate employability (Dacre Pool and Sewell 2007; Yorke and Knight 2004), and highly regarded by industry (see

Jackson and Chapman 2012a). Self-assessment is considered a valuable tool for enhancing learner awareness of one's own thinking and performance (see Lew, Alwis and Schmidt 2010) and nurturing meta-cognition.

Second, as the goals of higher education shift, Dochy et al. (1999) acknowledge that different forms of assessment are required to effectively measure skill outcomes. Authentic assessment encourages students to continually monitor and reflect on their performance – as they would in the workplace - not only in regard to achieved outcomes but the process of learning and completing tasks. Self-assessment tasks at university resemble assessment mechanisms in the workplace (see Dochy et al. 1999) and may enhance learning transfer. Student ability to easily transfer the process of self-assessment across contexts should, however, not be taken for granted (Andrade and Valtcheva 2009).

Further, self-assessment is considered critical for developing a propensity for lifelong learning beyond university years (Boud 1989) and may nurture increased engagement and empowerment with the learning process, as well as enhance student motivation (see MacDonald 2011). Instilling the cycle of monitoring, assessing and evaluating one's practices in completing tasks to achieve learning goals will assist graduates in recognising continual improvement and identifying approaches to best achieve it. Embedding self-assessment into undergraduate curricula has therefore become an important part of higher education's efforts to improve work-readiness by producing graduates which are appropriately skilled and can effectively apply and further cultivate their learning in the workplace.

Further perceived benefits of self-assessment are a reduction in student-teacher conflict through the discussion and unpacking of the grading process and improved learner

performance (see Leach 2012). Increased effectiveness of learning material and the identification of the strengths and weaknesses of pieces of work, in addition to decreased anxiety, are additional benefits (Andrade and Valtcheva 2009). The benefits of reduced academic workload are questioned by Cowan (1988) as students must be taught to self-assess and engaging with criteria requires time.

Problems with self-assessment

Leach (2012) cites a number of problems related to self-assessment, the first being student reluctance to self-assess due to perceived inability, a lack of confidence, inclination to avoid responsibility, and/or preference for expert opinion and feedback. Her study revealed that when the given the option, the majority of students did not self-assess. She notes the "dark side" to self-assessment (140), referencing the influential role of self-concept, cultural issues, and social control and accountability.

Problems with self-assessment typically focus on concerns with student ability to accurately self-assess. There has been considerable effort to ascertain to what degree student self-assessments align with academic assessments and in what ways they differ and for what reasons (see Falchikov and Boud 1989; Falchikov and Goldfinch 2000). Further, the impact of certain factors on student ability to self-assess – such as gender, familiarity with the process and stage of study – are considered important. There is considerable evidence to suggest that students' self-assessments are similar to academics (Leach 2012; Stefani 1994); rationalising the use of self-evaluation, where students actively contribute to their overall grade, in undergraduate programmes. Boud and Falchikov's (1989) critical analysis of quantitative studies relating to student self-assessment and teacher grades indicated that in most studies more student marks agreed rather than disagreed with teacher grades.

Conversely, there are many studies which found enough disparity between student and academic assessments to raise concern for incorporating self-assessment into undergraduate curricula in a meaningful way (Strong, Davis and Hawkes 2004; Thompson, Pilgrim and Oliver 2005). Lew et al. (2010) also found weak to moderate accuracy of student ability in self-assessment. Typically, these studies found that self-assessment leads to grade inflation, some arguing against self-grading as students are always motivated to grade higher (Andrade and Du 2007; Evans, McKenna and Oliver 2002; Kirby and Downs 2007). Sullivan and Hall (1997) found 39% of students overestimated their grades and Matsuno (2009) identified several studies with low correlations between student and academic grades.

Porter (2013) reviews mixed evidence of student ability to self-report on learning gains; some studies indicating self-reported data is a valid measure of student learning and others highlighting a lack of cognitive ability in students to accurately self-assess. He highlights the problem of large samples inflating the statistical significance of factors determining student self-assessment and the lack of investigation into exactly why student self-report data is often not correlated with objective measures of learning.

Leach (2012) acknowledges the foundation of these comparisons is the assumption that academic assessments are themselves accurate, despite evidence of unreliability in facilitator grading (Falchikov 2005; Kirby and Downs 2007). Extant literature discusses the difficulties of embedding self-assessment in higher education and the importance role of principles for effective design, now broadly acknowledged in the field of education (see Andrade and Valtcheva 2009). Despite the considerable number of studies in this area, variations in assessment type, establishment of learning criteria and student engagement with the assessments render research design problematic.

Influences on student self-assessment

A number of common trends in the differences between student and academic assessment have been identified. Leach (2012) found higher achieving students tend to underrate whereas low achievers overrate in comparison with academics, supporting similar findings (Papinczak et al. 2007). Aronson et al. (1989) suggest the inflation of grades by less able students may occur for two reasons: reporting higher gains in denial of their own lack of learning and/or students being susceptible to research aims and inflating grades to achieve desired results when investigations concerns their own growth and development. Boud and Falchikov's (1989) meta-analysis of quantitative studies of disparities in student and academic assessments noted different trends in student ratings under different circumstances. They found, however, there was a general trend by student ability in their rating approaches. High achieving students tend to be more realistic and perhaps underestimate their performance while low achievers overestimate and probably to a greater extent than those underestimating. Lew et al. (2010) also found more academically competent students were able to self-assess with greater accuracy than less competent peers.

Boud and Falchicov's (1989) review identified studies suggesting participant seniority impacts the accuracy of self-rating; seniority encompassing stage of degree, age and/or experience. Their meta-analysis, however, indicated that it was expertise in a given field which improved accuracy, not simply age or stage of enrolment. They could not conclude whether student ability at estimating grades improves with time and practice. Lopez and Kossack (2007), however, found student self-assessments became more realistic with practice while Lew et al. (2010) found accuracy did not improve over time and there was no relationship between student's belief in the value of self-assessment for their learning and their accuracy. Nulty (2011) reviewed self-assessment among first year students whose

reflective skills may be systematically less developed and emphasises the need to embrace self-assessment as a vehicle for developing judgment and critical evaluation. A further influence on self-assessment is gender, although Boud and Falchicov's (1989) found differences to be inconclusive with several studies citing differences and others finding none. More recently, Prince et al. (2008) reported females generally give higher estimations yet Langan et al. (2008) reported lower scores by females.

There are several influential factors within the domain of principles for effective self-assessment design; space allowing only a brief review. Falchikov and Boud found student familiarity with rating criteria enhances accuracy and alignment of ratings with academics. Tensions surrounding compulsory self-assessment for students who are not actively engaged with the learning criteria or the process, and are therefore not empowered or motivated but simply 'going through the motions', are discussed by Leach (2012). Further, the different tasks for rating will impact on the accuracy of ratings assigned. Falchikov and Boud (1989) found better prediction in science than social science and attributed this to task content. They also found better alignment in teacher and student assessments for traditional academic tasks (product oriented) rather than tasks involving professional practice. Finally, the form and complexity of the measuring instrument will impact self-assessment accuracy.

Method

Participants

Students enrolled during 2011 in an employability skills development programme, core to the business undergraduate degree, were invited to participate in a Skills Audit. Table 1 summarises the participants' demographic and background characteristics. The programme comprises four units; Units One and Two for first years, Unit Three for second years and Unit

Four for final year students. The response rate for the sample exceeded 90% but was reduced to 77%, of the 1232 enrolled students, once incomplete responses and those with inaccurate student identifiers were removed. [**Insert Table 1**]

Procedures

Data for the Skills Audit was gathered using an online survey in the latter half of the semester, more specifically October 2011. The Audit formed a learning activity for on and off campus students in all four units; students were actively encouraged to participate via lecturers and/or the unit's learning management system. The Audit was undertaken for two reasons: first, to encourage learners to self-reflect on their performance and further engage with developing the defined skills; second, to evaluate the programme's effectiveness in employability skill provision.

Measures

The Audit captured students' demographic and work background characteristics before asking them to self-assess their capabilities in each of the behaviours defined in the programme's employability skills framework (see Table 2). The framework comprises ten skills and forty constituent behaviours and was adapted from an established framework of industry-relevant competencies (Jackson and Chapman 2012b). Jackson and Chapman's framework originally derived from an international review of literature on industry expectations of new graduates (Jackson 2010). [**Insert Table 2**]

Students were asked to rate their current capabilities in performing each of the behaviours in the workplace, self-assessing directly against the behaviour descriptors in Table 2. A familiar scale of one to 10 was used, as recommended by Falchikov and Boud (1989);

one meaning students considered themselves unable to perform the behaviour in the workplace and 10 meaning they were an expert and able to teach others. Each of the four units has three to five core skills which form the basis of its learning activities and assessments (see Table 2). A composite measure was generated for student self-rating against the core skills and also the "communicating effectively" skill set; the latter spanning across assessments in all four units. The academic assessment is a weighted percentage score based on judgement of student performance in the unit's formative assessments. These assessments address the unit's assigned core skills and range from three to six per unit. For analysis purposes, the academic assessment was transformed to a score out of 10. This will subsequently be referred to as 'achieved mark'. The composite student self-rating is compared with the achieved mark to address the first research question. Variations in student ratings by individual characteristics, ability and stage of degree are examined using the composite measure, achieved mark and data on background/demographic characteristics.

A composite measure of student performance in the core skills for each unit is considered most appropriate given the interwoven nature of employability skills. These do not exist in a vacuum and must draw on others for their own effective demonstration (Gibson 2003; Rausch, Sherman and Washbush 2002). Communication is interrelated with many other skills (Casner-Lotto and Barrington 2006; Lowden et al. 2011) and the separation of skills for assessment purposes is difficult (Barrie 2005). Comparing a composite measure of student self-ratings in the core skills and communication skill set with the achieved mark for all assessments was therefore deemed most appropriate.

Validity and reliability

Cronbach alpha values for each skill set in the framework ranged between .866 and .925; ensuring internal consistency among items (behaviours). Further, correlations between behaviours (items) and skill sets (scale) ranged from .608 to .818; confirming the behaviours within each skill set are measuring the same construct. There have been extensive mapping exercises to ensure constructive alignment between each unit's learning outcomes, assessments and the core skills within the framework. Learning outcomes typically derive from the core skill behaviour descriptors and are then cross-referenced to each of the unit's assessments. This provides further assurance that comparisons between the composite student rating and achieved mark are valid and both are measuring student performance in the unit's core skills.

There is significant moderation and training in academic grading within each unit to ensure academics are consistent when assigning marks to students. Processes include: providing facilitators access to banks of moderated assessments for initial marking purposes; moderation of a 10% sample of marked assessments by the unit coordinator and provision of feedback to facilitators to ensure convergence with the norm; and reviewing and feeding back on facilitators' use of standard grading rubrics within a unit's electronic grading system. The training and moderation processes, in combination with shared perception among academics of the core behaviour descriptors, provide confidence of inter-rater reliability in the study.

Self-assessment design

The study considers the research design and analysis principles recommended by Boud and Falchikov (1989), particularly the use of scales. Falchikov and Boud's state familiarity with assessment criteria enhances the accuracy and alignment of student ratings.

Students were accustomed to the skills framework upon which the Audit was based with the ten skills, and their constituent behaviours, incorporated into the units' learning materials, assessments and marking rubrics. There is an ongoing dialogue between academics and students on the meaning and importance of the skills framework throughout the semester although students do not negotiate, critique or amend the criteria as recommended by Brew (1999).

Goodrich's (1996) principles of effective self-assessment design highlight the importance of students understanding the value of self-assessment. Here, reflection and the development of meta-cognitive skills are firmly embedded in the programme with students regularly using blogs and reflective journals, in addition to formally reflecting on their skill development in written and oral assessments. Goodrich also emphasises the need to instruct and assist students with completing self-assessment, as was the case with the Audit; as well as giving opportunities to review performance and identified areas of strengths and weaknesses. Students keep a copy of their completed Audit and are encouraged to discuss their ratings and use them, for example, when creating skills portfolios, career management plans and personal SWOT analyses. By adhering to these principles and using an established framework for measuring employability skills, the study hopes to overcome the challenges of measuring skill outcomes (Barrie 2005) and the risk of generating inaccurate ratings by both parties.

Results

Alignment of student and academic assessments

Table 3 summarises the mean ratings for each unit, and the overall sample, and their associated standard deviation. The achieved mark for each student was transformed from a

percentage score to a rating out of 10 and the mean for each unit and overall sample, with the associated standard deviation, is also given. The effect size (*d*), the difference between the means while taking into account the standard deviation, is calculated for each unit and the overall sample and can be used to indicate the degree of difference between students and academic assessments (see Falchikov and Boud 1989). A positive effect indicates higher grading by students and a negative value the opposite. The larger the effect size, the greater the disparity between the two groups. The effect sizes show that students in each of the four units overrate their performance in comparison to academics, ranging from 0.52 to 0.94 which are considered medium to large in size (see Falchikov and Boud 1989). Paired sample t-tests were conducted to investigate whether the academic and student mean scores differed significantly (p=.05). The results indicate significantly different means for each of the four units and the overall sample (see Table 3). [**Insert Table 3**]

To investigate further the degree to which students rate their performance the same as academics, the product moment correlation coefficient (r) was calculated for each unit, see Table 3. The correlation coefficients for the first two units are less than 0.1 and therefore considered small (see Falchikov and Boud 1989). Unit Three is marginally higher at 0.18 and significant (p<.05). Given this unit's very high effect size, which suggests considerable overrating among students, this degree of correlation between academics and students is surprising. It is important to remember that effect size measures students and academics assigning the same rating *on average* whereas correlation requires *agreement on a scale*, with very different means being quite possible. Again, Unit Four's coefficient indicates a significant correlation (p<.05) between academics and students despite a positive effect size. The Pearson correlation coefficient for the overall sample is positive and significant (r=.18).

Importantly, when calculating effect sizes we assume that academics are a reliable benchmark for comparison, deemed problematic by some (see Falchikov and Boud 1989). Prince et al. (2008) acknowledge that most studies comparing self-report and direct measures data use correlation to draw comparisons. This, however, is limited as it measures the strength of the relationship and not the level of agreement between the two sets of data (see Bland and Altman 1986). They suggest, instead, calculating limits of agreement between the data, in addition to reviewing mean differences. The percentage of student ratings agreeing to within +/-1.0 on the scale of one to ten, equivalent to plus/minus 10%, with academics' assigned mark was therefore calculated. For Unit One, 53% of students agreed with academics; 48% in Unit Two; 40% in Unit Three, and 46% in Unit Four. These results indicate a consistently medium proportion of agreement (see Falchikov and Boud 1989) among students across the units.

Variations in rating accuracy by student ability

Table 4 summarises the number and percentage of students in each grade category which overrated, underrated or assigned an equal rating in comparison to academics. Across the entire sample, 70% of students overrated in comparison to academics and the remaining 30% underrated. There was some minor variation in these proportions across the four units but these appear random. There is substantial evidence supporting the notion that less able students are more likely to overrate their ability and extremely able students, classed here as those achieving 80% and above, will underrate (Boud and Falchikov 1989; Leach 2012). This trend was apparent in all four units and more pronounced in the first unit where the majority of students achieving a Distinction, rather than just a Higher Distinction as in the other three units, also underrated. [**Insert Table 4**]

To investigate further, the difference in student ratings (out of 10) and achieved mark (expressed out of 10) was calculated. The difference therefore represents the extent to which each student's assessment of their own performance agreed with the academic's. Students were placed into one of five classifications: equal (if the academic and student ratings were equal); minor overrate (for positive values up to 2); major overrate (positive values exceeding 2.01); minor underrate (negative values to 2), and major underrate (negative values from 2.01 and below). Figure 1 indicates a pattern in student ratings by ability. Those students who overrate by 20% or more are predominantly failing or achieving pass grades. Conversely those who are significantly underrating their ability, by 20% or more, are predominantly Higher Distinction students. Those underrating by a lesser proportion are still mainly high achievers; those achieving Distinctions and above. Interestingly, those overrating by less than 20% are more evenly spread across students achieving Passes, Credits and Distinctions. [**Insert Figure 1**]

Further examination of the degree of accuracy by ability shows only 11% of Higher Distinction and 1% of Distinction students underrated by a major amount and 64% and 38% underrated, respectively, by a minor amount. In comparison, 18% of Fail and 47% of Pass students overrated by a major amount and 2% and 19% overrated, respectively, by a minor amount. This indicates the more capable students underrate to a lesser degree than which their less capable counterparts overrate.

Influences on self-assessment

A series of univariate ANOVAs was conducted on the absolute difference between academic and student assessments to investigate the potential influence of demographic/background variables defined in Table 1. Values were converted to positives

thus accounting only for magnitude of difference in ratings, not direction. Results are summarised in Table 5; a significance level of .05 was maintained to support the exploratory nature of the study. Findings indicate there were no significant variations for sex, age, degree type, student status, first language or employment status. Tukey post-hoc tests (α=.05) revealed the significant result for continent of birth was due to African students assigning consistently inflated ratings in comparison to their Australasian, Asian and European peers (*p*=.000). There were no other significant differences by origin among the other student groups. There was also a significant result for variations by unit (*p*=.000) which showed the second year students had more significantly inflated ratings than third year students. As this is an isolated result, it appears to be simply a function of the sample. [**Insert Table 5**]

Discussion and implications

Alignment of student and academic assessments

The study indicates a fairly substantial disparity between academic and student assessments of performance in employability skills. Findings based on average student and academic ratings indicate considerable overrating among students, the disparity extending from first through to final year students. Correlation coefficients, however, indicate there is a degree of alignment between academic and student assessments in the second and third year samples, and for the overall sample of students. This aligns with some studies which found student ability to rate accurately improves with stage of degree. The proportion of agreement between students and academics across the units is mediocre and there is no evidence to suggest the level of agreement improves as stage of degree advances.

The overall lack of evidence of improvement in students as they progress through their degree, particularly given reflective exercises are heavily embedded in the employability

skills programme, raises concern. It prompts the question of precisely how aligned graduates' perceptions of their own capabilities are aligned with employers when they enter the workforce. Self-assessment is commonplace in the workplace and pertinent to effective performance management processes. Sadly, inflated perceptions of capabilities, unrealistically high expectations and feelings of self-entitlement in the workplace are frequently associated with Generation Y graduates (see Jackson 2012). This finding urges better articulation to students of precisely how skills are used in the workplace, and to what expected standard, in a range of different professions. Ideally this should assist in revising students' inflated perceptions and encourage more accurate assessments of their own capabilities. Clarification of what constitutes a highly skilled graduate may be better achieved through student interaction with local industry. Strategies may include work-integrated learning – such as placements, volunteering and service learning; direct industry involvement in the teaching and learning of targeted skills in the university setting and/or a greater focus on authentic learning with students participating in role plays and simulated workplace scenarios which encourage them to apply targeted skills and critically reflect on their performance.

Influences on self-assessment

Students judged by academics as more academically competent – those achieving Higher Distinctions and Distinctions – were able to self-assess more accurately than their less competent peers – those achieving fails and passes. This is consistent with other studies (Boud and Falchikov 1989; Lew et al. 2010) which found student ability to self-assess is closely related to academic competence. Lew et al. attribute enhanced accuracy in more capable students to being 'better at self-monitoring, judging their own performance and processes of learning and at identifying their own learning strengths and weaknesses' (147).

The trends in students under or overrating by ability also aligned with existing literature (Boud and Falchikov 1989; Leach 2012). High achieving students tended to underestimate their performance while low achievers overestimated and to a greater extent than those underestimating. This was apparent across all four units. Leach argues a students' propensity for overrating or underrating will depend on their position, as determined by academics, on the grading scale. It is important to note the Audit was conducted before final assessments and the release of certain marks. Students therefore only had partial idea of what their final mark might be at the time of completion.

The influence of gender on self-assessment accuracy is not detected in this study, supporting Falchikov and Boud's (1989) findings. In regard to their broader focus on the influence of 'seniority' (418), there was no evidence of variations for age or work experience. Further, there appear to be little demographic/background influences on a students' ability to self-assess accurately other than African students having more inflated perceptions of their performance levels than other student groups. This lack of variation in agreement scores across different characteristics contravenes certain studies yet aligns with others (see Boud and Falchikov 1989).

Conclusions and future research

This study supports concerns for the use of self-assessment in higher education (see Sitzmann et al. 2010) with further evidence of significant differences between self and academic assessments. The study contributes to extant literature by examining disparities and trends specific to the rating of employability skill outcomes. Inflated self-perceptions and

consistent patterns of under and over rating among students indicate self-assessment is problematic in employability skill development and assessment.

The pronounced disparity between students and academics may be explained in a number of ways. First, it may be that students are simply a poor judge of their own abilities, although this would counteract a number of studies which found strong agreement between academic and student assessments. Second, students' lack of experience in self-assessment impacts on accuracy although there is mixed evidence to support this in both our study and existing literature. Third, students may be inadequately trained in self-assessment as feedback and practice alone appear insufficient for improvement (Lew et al. 2010). Final year students on the employability skills programme have typically had more exposure to reflective practices through the use of blogs and personal journals with extensive academic feedback. They did not, however, demonstrate the ability to self-assess more accurately than their less experienced peers. For self-assessment to be integrated into the employability skills programme in a meaningful way, training and development in the process is required. Educators must also be mindful of their international cohorts as certain groups' demonstrated higher levels of inaccuracy in their self-assessments than others. Otherwise, there is a generic need for development across the undergraduate cohort as individual and background characteristics - including gender, work experience and stage of degree – appear to make little difference in student ability to self-assess accurately.

This raises further questions on developmental approaches for nurturing accurate self-assessment in students. As practice alone does not guarantee improved accuracy, perhaps there should be more guidance on the standards expected for a particular assessment (or in this case skill). This reiterates the importance of academic consideration of the principles of

good self-assessment design and an ongoing dialogue with students on the learning/assessment criteria. Practising the application of assessment criteria to exemplars of work and follow-up moderation and remediation to reach agreement on awarded marks among students and academics may assist (Boud 1989). For the skills framework pertinent to this particular study, a set of rubrics for each skill are currently being developed. These provide a breakdown of the expected standard for a passing effort in each of the defined behaviours for first, second and final year students. With coaching, it is hoped these will improve student ability to self-assess more accurately in future semesters to capitalise on the documented benefits of self-assessment.

Fourth, students may not be engaged with the self-assessment process. Student interest in and motivation for the self-assessment process should enhance accuracy (Longhurst and Norton 1997). If students do not believe self-assessment will contribute to their learning, one might expect them to be less accurate although Lew et al. (2010) found no empirical evidence to support this. Time taken to complete self-assessments could indicate the degree of student engagement for future studies. Leach (2012) argues it is the responsibility of academics to engage students in self-assessment through actively promoting its benefits; encouraging the negotiation of assessment criteria; and nurturing confidence and understanding of the process in reluctant students. Student buy-in, however, does not necessarily guarantee a high level of agreement with academic assessments; further reiterating the importance of development.

Finally, disparity may be due to poor practices in self-assessment design and implementation. In regard to this particular study and the conclusions drawn, Falchikov and Boud's (1989) assertion that it is easier to predict grades – with therefore less disparity

between academic and self assessments – in science than social science may be important. They argued that accuracy in ratings may depend on task content. Given employability skills are notoriously difficult to measure (Halfhill and Nielsen 2007), this may amplify the degree of disparity. Further, correlations between students and academics may only be moderate because students are assessing only the core skills for a particular unit whereas, in reality, the academic's awarded mark may span other areas of the skills framework. Problems raised by any mismatch between academic and student criteria may be investigated in a follow-up study. Interestingly, Hansford and Hattie's (1982) meta-analysis found higher quality studies reported lower correlations between self-measures and performance achievement.

As argued by Boud (1989), if an acceptable point of agreement cannot be reached between academics and students, self-assessments should be restricted to a learning activity rather than formal assessment. Alternatively, measures for controlling the documented biases should be introduced. Boud discusses a number of strategies for incorporating student self-assessment, which contributes to awarded marks, which attempt to combat problems with accuracy yet capitalise on the benefits of the process.

For employers, the problems with graduates' inflated perceptions of personal performance are well-documented (see Jackson 2012). These may extend beyond graduate positions to the increasingly popular Work Integrated Learning (WIL) opportunities which include vacation programmes, internships and placements during undergraduate degrees. These are acknowledged as vital tools in enhancing student learning and their subsequent transfer of acquired skill and knowledge to the workplace upon graduation (Billett 2011). Over-confident and self-important undergraduates which lack humility, however, may

jeopardise their own achievements as well as future relations between local industry and higher education providers.

There are certain limitations to the study. First, the sample derives from a single source. A cross-disciplinary study in multiple institutions may provide a better understanding of the impact of demographic and background characteristics on self-assessment accuracy due to greater variations in social and cultural background, academic ability and demographic profile. However, Porter (2013) notes the problem of sizeable samples, generated by national and institution-wide studies wishing to gauge student learning, producing statistically significant results for factors considered to influence self-reporting tendencies, such as academic discipline. To overcome this, he recommends focusing more on effect sizes to ascertain a true growth in learning attributable to these variables. Further, despite this study adhering to recommended design principles by Boud and Falchikov (1989), Porter recommends students should gauge - on a pre-defined scale - their capabilities in each skill area upon entering university. He also argues they should have access to their previous data when assessing perceived capabilities repeatedly over different time points. Neither process was incorporated into the Audit; providing some points for consideration in future research design.

The study also lacks consideration of precisely why undergraduates are unable to accurately self-assess their abilities in certain employability skills, highlighting a valuable area for future research. Porter (2013) hypothesises there are common factors, other than lack of cognitive ability, which drive student responses on reported learning gains. Applying his proposed belief-sampling approach to self-reporting on employability skills may add significant value to current research and advance our understanding of why disparities exist

between student and academic assessments. The model would investigate the role of certain ‘considerations’ on student self-reporting, perhaps including the degree to which a student’s entry pathway was vocational - thus facilitating a stronger perception of being ‘skilled’; exposure to the workplace through paid employment, volunteering, work-integrated or service learning; their experience in extra-curricular activities; and the strength of their academic background. Given self-assessed data is widely used to gauge student learning and to better understand the impact of certain conditions, infrastructure and pedagogy on learning outcomes and graduate employability, investigating not only whether but also why inaccuracies exist is critical to the future of higher education.

References

Andrade, H., and Y. Du. 2007. Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education* 32, no. 2: 159–81.

Andrade, H., and A. Valtcheva. 2009. Promoting learning and achievement through self-assessment. *Theory Into Practice,* 48, no. 1: 12-19.

Aronson, E., P. Ellsworth, J. Carlsmith, and M. Gonzales. 1989. *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.

Barrie, S. 2005. Rethinking Generic Graduate Attributes. *HERDSA News*, 5 March 2005.

Billett, S. 2011. *Curriculum and pedagogical bases for effectively integrating practice-based experiences – final report*. Strawberry Hills, NSW: ALTC.

Bland, J., and D. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, no. 8476: 307-310.

Boud, D. 1989. The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education* 14, no. 1: 20-30.

Boud, D. 1995. *Implementing student self-assessment.* 2nd ed. Canberra: Higher Education Research and Development Society of Australasia.

Boud, D., and N. Falchikov. 1989. Quantitative studies of student self-assessment and peer assessment. *Higher Education* 18, no. 5: 529–49.

Brew, A. 1999. Towards autonomous assessment: Using self-assessment and peer assessment. In *Assessment matters in higher education: Choosing and using diverse approaches*, ed. S. Brown and A. Glasner, 159–71. Buckingham: The Society for Research into Higher Education/Open University Press.

Casner-Lotto, J., and L. Barrington. 2006. *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st Century U.S. workforce*. USA: The Conference Board, Inc.

Cowan, J. 1988. Struggling with student assessment in Boud D (Ed), *Developing Student autonomy in Learning.* 2nd ed., London: Kogan Page.

Dacre Pool, L., and P. Sewell. 2007. The key to employability: Developing a practical model of graduate employability. *Education + Training* 49, no. 4: 277-289.

Department of Education, Science and Training [DEST] (2002) *Employability skills for the future*. Canberra: DEST.

Dochy, M., D. Segers, and D. Sluijsmans. 1999. The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education* 24, no. 3: 331-350.

Evans, A., C. McKenna, and M. Oliver. 2002. Self-assessment in medical practice. *Journal of the Royal Society of Medicine* 95, no. 10, 511–513.

Falchikov, N. 2005. *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: RoutledgeFalmer.

Falchikov, N., and D. Boud. 1989. Student self-assessment in higher education: A meta-analysis. *Review of Educational Research* 59, no. 4: 395–430.

Falchikov, N., and J. Goldfinch. 2000. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research* 70, no. 3: 287-322.

Gibson, D. 2003. A framework for developing and assessing generic capabilities in QUT law graduates. *B-HERT News* 16, 4-5.

Goodrich, H. 1996. *Student self-assessment: At the intersection of metacognition and authentic assessment.* Unpublished doctoral dissertation. Cambridge, MA: Harvard University.

Halfhill, T., and T. Nielsen. 2007. Quantifying the 'softer side' of management education: An example using teamwork competencies. *Journal of Management Education* 31, no. 1: 64-80.
Hansford, B. and J. Hattie. 1982. The relationship between self and achievement/performance measures. *Review of Educational Research* 52, no. 1: 123-142.
Jackson, D. (2010) An international profile of industry relevant competencies and skill gaps in modern graduates. *International Journal of Management Education* 8, no. 3: 29-58.
Jackson, D., and E. Chapman. 2012a. Non-technical skill gaps in business graduates. *Education + Training* 54, no. 2/3: 95-113.
Jackson, D., and E. Chapman. 2012b. Non-technical competencies in undergraduate Business degree programs: Australian and UK perspectives. *Studies in Higher Education* 37, no. 5: 541-567.
Jackson, D. 2012. Business undergraduate perceptions of their capabilities in employability skills and implications for industry and higher education. *Industry and Higher Education* 26, no. 5: 345-356.
Joham, C., and M. Clarke. 2012. Teaching critical management skills: The role of problem-based learning. *Teaching in Higher Education* 17, no. 1: 75-88.
Kirby, N., and C. Downs. 2007. Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? *Assessment & Evaluation in Higher Education* 32, no. 4: 475–94.
Langan, A., D. Shuker, W. Cullen, D. Penney, R. Preziosi, and C. Wheater. 2008. Relationships between student characteristics and self, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education* 33, no. 2: 179-190.
Leach, L. 2012. Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education* 37, no. 2: 137-147.
Lew, M., W. Alwis, and G. Schmidt. 2010. Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education* 35, no. 2: 135-156.
Longhurst, N., and L. Norton. 1997. Self-assessment in coursework essays. *Studies in Educational Evaluation* 23, no. 4: 319-330.
Lopez, R., and S. Kossack. 2007. Effects of recurring use of self-assessment in university courses. *International Journal of Learning* 14, no. 4: 203–16.
Lowden, K., S. Hall, D. Elliot, and J. Lewin. 2011. *Employers' perceptions of the employability skills of new graduates.* London: University of Glasgow SCRE Centre and Edge Foundation.
MacDonald, K. 2011. A Reflection on the Introduction of a Peer and Self Assessment Initiative. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education* 6, no. 1: 27-42.
Matsuno, S. 2009. Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26, no. 1: 75-100.
Nulty, D. 2011. Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education* 36, no. 5: 493-507.
Papinczak, T., L. Young, M. Groves, and M. Haynes. 2007. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher*, 29, no. 5: 122-132.
Porter, S. 2013. Self-reported learning gains: A theory and test of college student survey response. *Research in Higher Education* 54, no. 2: 201-226.
Prince, S., K. Adamo, M. Hamel, J. Hardt, G. Connor, and M. Tremblay. 2008. A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity* 5, no. 56: 1-24.
Rausch, E., H. Sherman and J. Washbush. 2002. Defining and assessing competencies for competency-based, outcome-focused management development. *Journal of Management Development* 21, no. 3, 184 – 201.

Sitzmann, T., K. Ely, K. Brown, and K. Bauer. 2010. Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education* 9, no. 2: 169-191.
Stefani, L. 1994. Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education* 19, no. 1: 69–75.
Strong, B., Davis, M. and V. Hawks. 2004. Self-grading in large general education classes. *College Teaching* 52, no. 2: 52–7.
Sullivan, K., and C. Hall. 1997. Introducing students to self-assessment. *Assessment & Evaluation in Higher Education* 22, no. 3: 289–306.
Thompson, G., A. Pilgrim, and K. Oliver. 2005. Self-assessment and reflective learning for first-year geography students: A simple guide or simply misguided? *Journal of Geography in Higher Education* 29, no. 3: 403–20.
Yorke, M., and P. Knight 2004. *Embedding employability into the curriculum*. Learning and Employability Series 1. York, England: Higher Education Academy.

Table 1 Sample demographic/background characteristics

Characteristic	Sub-group	Sample	
		n	%
Unit	Unit One	205	22
	Unit Two	327	34
	Unit Three	197	21
	Unit Four	220	23
Sex	Male	426	45
	Female	523	55
Age	16-20 years	215	23
	21-25 years	529	56
	26-30 years	123	13
	31-40 years	51	5
	41+ years	31	3
Degree type	Bachelor of Business	813	86
	Other	136	14
Student status	International	411	43
	Domestic	536	57
Continent of birth	Asia	394	42
	Africa	93	10
	Europe	76	8
	Australasia	379	40
First language	English	503	50
	Other	516	50
Weekly paid employment status	0 hours	231	24
	1 – 9 hours	105	11
	10 – 19 hours	294	31
	20 – 29 hours	211	22
	30 – 37 hours	38	4
	38 hours or more	70	8

Table 2 Employability skills framework (adapted from Jackson and Chapman 2012b)

Employability Skill	Behaviour	Behaviour descriptor
Working effectively with others	Task collaboration	Complete group tasks through collaborative communication, problem solving, discussion and planning.
	Team working	Operate within, and contribute to, a respectful, supportive and cooperative group climate.
Core to Units One, Two and Three	Social intelligence	Acknowledge the complex emotions and viewpoints of others and respond sensitively and appropriately.
	Cultural and diversity awareness	Work productively with people from diverse cultures, races, ages, gender, religions and lifestyles.
	Influencing others	Defend and assert their rights, interests and needs and convince others of the validity of one's point of view.
	Conflict resolution	Address and resolve contentious issues with key stakeholders.
Communicating effectively	Verbal communication	Communicate orally in a clear and sensitive manner which is appropriately varied according to different audiences and seniority levels.
Core to Unit One	Giving and receiving feedback	Give and receive feedback appropriately and constructively.
	Public speaking	Speak publicly and adjust their style according to the nature of the audience.
	Meeting participation	Participate constructively in meetings.
	Written communication	Present knowledge, in a range of written formats, in a professional, structured and clear manner.
Self-awareness	Meta-cognition	Reflect on and evaluate personal practices, strengths and weaknesses in the workplace.
Core to Units One and Four	Lifelong learning	Actively seek, monitor and manage knowledge and sustainable opportunities for learning in the context of employment and life.
	Career management	Develop meaningful and realistic career goals and pathways for achieving them in light of labour market conditions.
Thinking critically	Conceptualisation	Recognise patterns in detailed documents and scenarios to understand the 'bigger' picture.
Core to Unit Two	Evaluation	Recognise, evaluate and retain key points in a range of documents and scenarios.
Analysing data and using technology	Numeracy	Analyse and use numbers and data accurately and manipulate into relevant information.
	Technology	Select and use appropriate technology to address diverse tasks and problems.
Core to Unit Two	Information management	Retrieve, interpret, evaluate and interactively use information in a range of different formats.
Problem Solving	Reasoning	Use rational and logical reasoning to deduce appropriate and well-reasoned conclusions.
Core to Unit Three	Analysing and diagnosing	Analyse facts and circumstances and ask the right questions to diagnose problems.
	Decision making	Make appropriate and timely decisions, in light of available information, in sensitive and complex situations.

Developing initiative and enterprise *Core to Unit Two and Three*	Entrepreneurship/ Intrapreneurship	Initiate change and add value by embracing new ideas and showing ingenuity and creativity in addressing challenges and problems.
	Lateral thinking / creativity	Develop a range of solutions using lateral and creative thinking.
	Initiative	Take action unprompted to achieve agreed goals.
	Change management	Manage change and demonstrate flexibility in their approach to all aspects of work.
Self-management *Core to Unit Three*	Self-efficacy	Be self-confident in dealing with the challenges that employment and life present.
	Stress tolerance	Persevere and retain effectiveness under pressure or when things go wrong.
	Work / life balance	Demonstrate the importance of well being and strive to maintain a productive balance of work and life.
	Self-regulation	Reflect on and regulate their emotions and demonstrate self-control.
Social responsibility and accountability *Core to Units Three and Four*	Social responsibility	Behave in a manner which is sustainable and socially responsible (e.g., consistent with company policy and/or broader community values).
	Accountability	Accept responsibility for own decisions, actions and work outcomes.
	Personal ethics	Remain consistently committed to and guided by core values and beliefs such as honesty and integrity.
	Organisational awareness	Recognise organisational structure, operations, culture and systems and adapt their behaviour and attitudes accordingly.
Developing professionalism *Core to Unit Four*	Efficiency	Achieve prescribed goals and outcomes in a timely and resourceful manner.
	Multi-tasking	Perform more than one task at the same time.
	Autonomy	Complete tasks in a self-directed manner in the absence of supervision.
	Time management	Manage their time to achieve agreed goals.
	Drive	Go beyond the call of duty by pitching in, including undertaking menial tasks, as required by the business.
	Goal and task management	Set, maintain and consistently act upon achievable goals, prioritised tasks, plans and realistic schedules.

Table 3 Mean ratings, effect sizes and correlations by unit and overall sample

	Student rating		**Achieved mark**		**Effect size**	**Pearson correlation**		**Paired samples t-test**		
	Mean	SD	Mean	SD	*d*	*r*	*p*	*t*	*df*	*p*
Unit One	7.00	1.07	6.38	1.20	.52	.09	.18	5.89	204	.00
Unit Two	7.32	1.02	6.59	1.02	.72	.07	.21	9.53	326	.00
Unit Three	7.37	1.16	6.38	1.05	.94	.18	.01	9.88	196	.00
Unit Four	7.59	1.13	7.10	0.78	.63	.32	.00	6.29	219	.00
Total	7.33	1.10	6.62	1.06	.67	.18	.00	15.76	948	.00

Table 4 Patterns in student ratings across units

	Fail (0 – 49%)			**Pass (50 – 59%)**			**Credit (60 – 69%)**			**Distinction (70 – 79%)**			**Higher Distinction (80 – 89%)**			**Total**		
	U	O	E	U	O	E	U	O	E	U	O	E	U	O	E	U	O	E
Unit One	0	13 100%	0	8 14%	47 86%	0	24 34%	47 66%	0	27 60%	18 40%	0	16 76%	5 24%	0	75 37%	130 63%	0
Unit Two	0	16 100%	0	4 7%	50 93%	0	19 17%	93 82%	1 1%	53 44%	67 55%	1 1%	21 91%	2 9%	0	97 29%	228 70%	2 1%
Unit Three	0	9 100%	0	7 10%	63 90%	0	7 12%	52 88%	0	12 27%	32 73%	0	13 87%	2 13%	0	39 20%	158 80%	0
Unit Four	0	0	0	0	7 100%	0	24 28%	62 71%	1 1%	25 28%	66 72%	0	20 57%	14 40%	1 3%	69 31%	149 68%	2 1%
Total	0	38 100%	0	19 10%	167 90%	0	74 22%	254 77%	2 1%	117 39%	183 61%	1 <1%	70 75%	23 24%	1 1%	280 30%	665 70%	4 0%

Table 5 Variations in rating disparity across demographic/background variables

Variable	***df***	***MS***	***F***	***p*-value**	η^2
Unit	3	5.458	5.836	.001	.018
Sex	1	0.578	0.608	.436	.001
Age	4	.852	.897	.465	.004
Degree type	1	.032	.034	.855	.000
Student status	1	1.480	1.560	.212	.002
Continent of birth	4	6.490	7.010	.000	.029
First language	1	.150	.158	.691	.000
Employment status	5	1.165	1.229	.293	.006

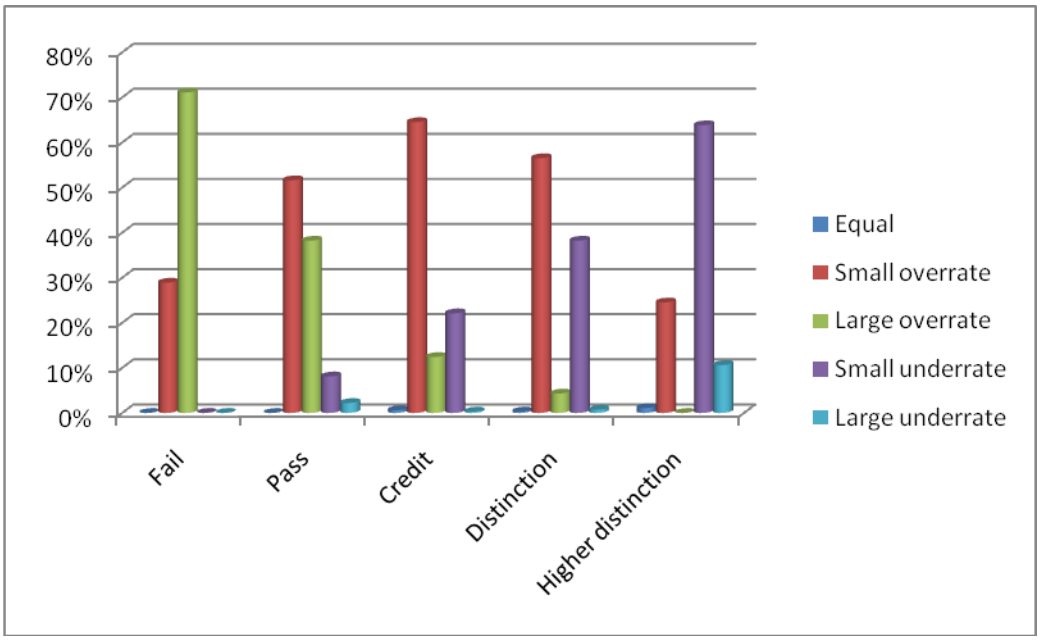

Figure 1 Pattern in ratings by student ability