Australian Journal of Adult Learning
Volume 45, Number 3, November 2005

Self-evaluations in adult education and training

James A Athanasou
Faculty of Education
University of Technology, Sydney

This paper focuses on two key aspects of self-evaluation in adult education and training through the perspective of (a) a social-cognitive framework which is used to categorise those factors that enhance self-efficacy and self-evaluation, and (b) the accuracy of self-evaluation. The social-cognitive framework categorises the factors that enhance self-evaluation, namely, social messages (e.g. comparison with others, feedback from others, social and cultural stereotypes), personal factors (e.g. the ability level of the rater, the standards and goals of the rater) and situational factors (e.g. the content area being evaluated). The paper reviews the accuracy of self-evaluations and concludes (a) that there is prima facie support from previous meta-analyses for their accuracy, (b) that the accuracy of self-evaluations is likely to be underestimated, and (c) that a focus on individual rather than group comparisons may be more useful for adult education. The educational value of formative self-evaluation for adult education and training contexts is supported.

Introduction

An adult learner who monitors his/her performance or estimates what might be their future achievement engages in a form of self-evaluation. Adult self-evaluation can occur in an evening college class, technical education, higher education, in a workplace situation or whenever someone is engaged in informal learning tasks. People may self-assess in order to determine their past or future response to situations such as their likelihood of success or the value of their investment of time and effort or the extent of their learning. The purpose of this paper is to consider two aspects of the nature and value of self-evaluations in adult education and training. The first of these contexts relates to a theoretical framework for studying self-evaluations and the second deals with the accuracy of such self-evaluations.

At the outset it may be helpful to clarify some aspects of terminology since a number of terms seemingly refer to the same phenomenon, namely, 'self-evaluation', 'self-assessment', 'self-rating' or 'self-estimate'. A self-evaluation is the judgement of the merit or worth of a self-estimate that has been produced. Typically, self-evaluation follows a self-estimate or a process of self-assessment and it can be a formal public process or an informal private perception. Both self-evaluation and the psychological concept of self-efficacy focus on capability, and the following section outlines a theoretical framework of self-evaluation from a social-cognitive perspective.

Social-cognitive theory and self-evaluation

Self-efficacy indicates a person's probability of engaging in a task depending on how capable they believe they are in carrying out the task successfully. This concept emanates from the work of Bandura (1986) who noted that people have a sound idea of their talents. Factors that influence self-efficacy include: previous successes, reassurances from others, and the observations of the success of

others (Ormrod 2005, pp. 144–145). Elements of self-regulation that are also linked with self-efficacy are: the performance standards that people set, the observations of our own behaviour, and self-reinforcement. If one were to depict this schematically it would be a recursive model with some of the components outlined in Figure 1. This framework may offer a helpful basis for studying self-evaluation and the following paragraphs provide an analysis of the components of the social-cognitive process that are linked to self-evaluation and self-efficacy under three broad headings: social messages, personal factors and situational factors.

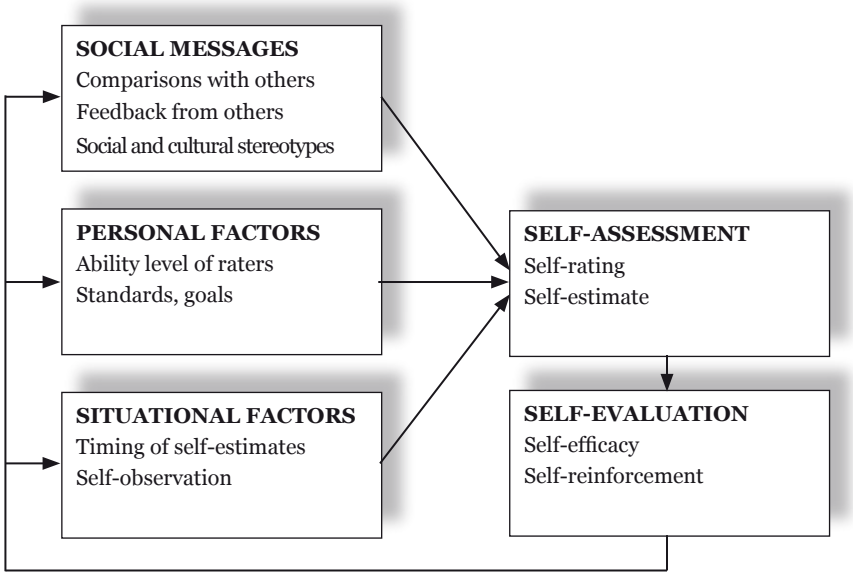

Figure 1 A recursive social cognitive model for self-evaluation

Social messages

Three social messages that have an impact on self-evaluation have been defined in previous research. These are the comparisons that we make with others, the social and cultural stereotypes that form the background of our perceptions and the feedback that we receive from others.

Rather than have people rate themselves on some amorphous concept such as 'average', 'above average' or 'below average', Mabe and West (1982) advocated the use of directions that accentuate comparison with others. Indeed, the practice of making self-evaluations that are norm-referenced (that is, ratings such as high or low, average) implies that the person has an implicit understanding of the normal distribution or a detailed understanding of the likely performance in a comparison group.

Accordingly, there is an increasing emphasis on realistic comparisons with others taking into account the ability levels of these proxies. Recently, Martin, Suls and Wheeler (2002) reported that self-raters' perception of the ability levels of others influenced their self-ratings. People rated themselves lower in relation to superior proxies and higher in relation to inferior proxies. In considering ways to improve self-ratings, they included the use of a competent role model as a basis for comparison in order to overcome gender differences and the use of feedback.

As far back as 1902, Cooley described the 'looking-glass self' in which the feedback provided by others is centrally important to the development of an individual's perceptions of himself or herself. Bergee and Cecconi-Roberts (2002) reported that the use of discussions about performance and group feedback improved the congruence between self and other ratings with music education and music performance majors. Swann, Wenzlaff, Krull and Pelham (1992) described a desire to elicit self-confirmatory feedback

especially amongst people who were clinically depressed or people with negative views about themselves. They wrote mainly about unfavourable appraisals of social interactions and interpersonal relations, but the same phenomenon has been reported amongst children and adolescents in relation to perceived competence in a particular domain, such as athletics, arts and crafts, social acceptance or scholastic competence (Cassidy, Ziv, Mehta & Feeney 2003). The general conclusion was that of a vicious circle in which people with negative self-views tend to seek negative feedback to confirm the original negative self-evaluations (see also Bernichon, Cook & Brown 2003).

Other social characteristics also contribute to the type of self-evaluations that people are likely to make. At a macro-level, culture has also been reported to influence workers' perceptions of their ability. Farh, Dobbins and Cheng (1991) compared the job performance ratings of 982 supervisor and subordinate pairs in Taiwan and the USA and found that Chinese employees displayed a modesty bias. They rated their job performance less than their supervisors whereas US employees were reported to be more lenient with their ratings than their supervisors. Enduring characteristics from our socialisation also have an impact on self-evaluation. For example, Betsworth (1999) reported that women continue to underestimate their abilities. Marx and Roman (2002) demonstrated that there was a limit in the level of self-estimates in the presence of a competent role model for women who had already been identified as motivated with mathematics, who had obtained a minimum SAT score of 650 out of 800 and who had enrolled in at least one mathematics course. The correlation (that is, r – a statistical measure of relationship varying from –1 through 0 to +1) of their self-estimates with a 15-problem mathematics test was low (r = 0.28).

Personal factors

One key personal factor in the ability to self-evaluate appears to be the ability or level of achievement of a person. Although the ability levels of raters have long been recognised as moderating the ability to accurately self-rate, there is recent evidence in some studies (for example, Kruger & Dunning 1999) that under-performers and under-achievers were more likely to overestimate performance than high performers on tasks related to humour, logical reasoning or grammar. Correlations between the grade point average and ratings for above average students were moderate (r = 0.61) compared with low (r = 0.34) for below average and low (r = 0.33) for average students (Wright 2000). Finally, not all individuals have the same training to make accurate self-perceptions. There may be rating errors and biases or they may use inappropriate judgemental heuristics.

A second personal aspect relates to relevant standards and goals. While self-evaluation is relevant to the field of adult education and training, it is not clear that studies of the accuracy of self-evaluations involve contextually relevant comparisons. For instance, what might be a useful basis for comparison with a self-evaluation of adult reading? Tousignant and DesMarchais (2002) demonstrated that prediction of performance was much better than prediction of ability. Klein and Buckingham (2002) also concluded that ambiguity of one's own performance led to bias, but this effect was lessened when the ambiguity of the task was reduced and the criterion was clearly defined and verifiable (for example, typing speed).

Situational factors

Situational factors in an adult context may have a greater impact on the self-evaluation process than first imagined. These factors include *inter alia*: (a) the specific content area; (b) the prior experience with the criterion; (c) whether the self-assessment is made prior to or following learning; (d) whether there is any social desirability

associated with the judgement; (e) whether the criterion is norm-referenced or criterion-referenced; or (f) the format or manner in which the self-assessment is elicited.

An additional situational factor is that people do not apply similar calibrations before and after tests. Tousignant and DesMarchais (2002) compared the degree of certainty in estimating ability to perform before and after an exam. The correlation between pre-examination and oral presentation ranged from no relationship (r = 0.04) to a low relationship (r = 0.24) while post-examination and oral presentation correlations were low (ranging from r = 0.25 to r = 0.33). In other words, there was a slightly higher correlation or relationship between the self-evaluation after an assessment rather than for self-evaluation prior to an assessment. In a sample of undergraduate students, correlations (gamma) between pre-test estimates of reading ability and the number of comprehension questions correct were very low (0.14) compared with low (0.28) for the post-test (Lin, Moore & Zabrucky 2001).

As noted previously, there is scope to use this framework for research and the preceding comments provide some indication of how the social-cognitive perspective of self-efficacy might be linked quite usefully to self-evaluations. More importantly, it highlights a heuristic framework that can be adjusted in order to improve the process of self-evaluation. There are, however, divergent views on the value and accuracy of self-evaluations for adult learners.

The accuracy of self-evaluations

One view of self-evaluations from an adult learning perspective is that they are not valid. Knowles, Holton and Swanson (1998, p. 130) concluded that '…if adult learners rely on proxy measures – *self-assessment* of anticipated outcomes, they will most likely make false conclusions based on invalid data' (p. 130; italics in original). This is consistent with a self-enhancement effect known as the 'above

average effect' (van Lange & Sedikides 1998) where it has been noted that people rate themselves more favourably than they should.

As one would expect, there has been considerable attention on educational aspects of public self-evaluation but this has been mainly in classroom contexts and has involved a number of studies using school and college students. These have examined the ability of people to estimate their performance on formal educational assessments (Lunneborg 1982). For example, Mihal and Graumenz (1984) reported that individuals could accurately rate their performance on more objective and easily measured dimensions.

Longitudinal research in the area is still rare, but one study (Obach 2003) pointed to the predictive value of self-ratings in determining future performance. The correlation between perceived competence and standard achievement tests was 0.37 and 0.35 for a year later; and between perceived competence and grade point average was 0.52 and 0.36 for a year later. Obach (2003) suggested that self-perceptions of ability predicted performance a year later but these results could be interpreted as suggesting either a longer-term stability in self-ratings or potentially a self-fulfilling prophecy in operation.

The relation of self-evaluation to assessment results has been studied in two separate meta-analyses. From their review of self-evaluation and achievement in a higher education context, Falchikov and Boud (1989) reported a moderate mean effect size of 0.47 (1989, p. 419); a mean correlation between teacher and student marks of 0.39 (1989, p. 420); and that 64% of self-assessors had grades that agreed with those of faculty markers (1989, p. 420). In a psychological context, Mabe and West (1982) undertook a substantial meta-analysis of the literature and found that the average correlation between self-ratings and abilities was 0.29. They reported that 88% of the correlations were greater than zero. Accordingly, there is some *prima facie* support for the accuracy of self-evaluations across a range

of aptitudes and abilities but there is also evidence that the ability to self-evaluate may itself vary from person to person. Moreover, any variation in the reported values of the different studies when comparing groups is likely to be a function of sampling error. The important point is that, even with large groups across diverse abilities, the correlations are all positive when self-evaluations are compared with criteria. The correlation would increase if there was a correction for attenuation, in other words, some correction needs to be made for the fact that self-evaluations are typically made across a few points on a rating scale and have a narrow range compared with assessment results which often vary across a wide range.

Understandably, earlier research has focused on the congruence between self-evaluations and formal assessments mainly through correlation coefficients, but this does have some technical problems. Typically a group of people is asked to estimate their performance and the estimate is compared against educational achievement, teacher ratings, supervisor ratings or peer ratings. The correlation or other statistic is produced and any comparison of a self-evaluation with respect to a criterion assumes that the criterion itself can be determined reliably, that is, consistently and in a stable fashion. This resulting correlation is difficult to justify because every educational phenomenon has a degree of unreliability. Since a quantitative or qualitative self-evaluation also has a degree of unreliability, then the comparison is confounded by the interaction of both unreliabilities. Statistically, it is possible to attenuate the correlations so that the unreliability in the criterion is controlled, but this is applicable only to group data and is not of great assistance to an individual learner. Thus, if self-evaluations and a criterion correlated only 0.3 and each had a reliability of only 0.5 (a low reliability for an assessment), then in theory the attenuated correlation between the two imperfect measures could be as high as 0.6.

Furthermore, any correlation from group comparisons automatically standardises the self-assessments (that is, rescales them to have a mean of zero and a standard deviation of one), and also standardises the performances on the criterion in the same way allowing a comparison to be made on the same scale. The first problem is that converting descriptive ratings to numbers is problematic and may not represent true measurement since ratings are not additive units of behaviour. Moreover, this only answers the question of whether the group's relative ordering of self-evaluations is comparable to the group's relative ordering of performance. It does not indicate the accuracy of self-evaluation.

As noted previously, most investigations of self-evaluation focus on nomothetic or group comparisons. If one wished to investigate the accuracy of self-evaluations, then an alternative approach is to focus on ipsative or idiographic approaches. With an ipsative approach, the person is compared within a set of his or her potential achievements. For instance, they may be asked to rank their relative achievements (best, second best, third best and so on) and this rank ordering is the basis for comparison. If these measurement limitations were controlled, then the relationship between self-evaluations and assessment results would be substantially higher.

The educational value of self-assessments for learning

The self-evaluation approach to adult learning involves individuals becoming the direct source of information about themselves. Especially in those contexts where there is no reason for disguise or concealment, Mischel (1977) contended that '…what the person tells us directly turns out to be as valuable an index as any other more direct sign' (p. 248). Writing from a perspective of self-evaluation of personality characteristics, Burisch (1984) concluded:

> … if self-ratings are (a) directly communicable, (b) the ultimate in economy, and (c) also more valid than their questionnaire counterparts, then we will have to face the embarrassing

> question of just why we continue to construct personality inventories at all ... (p. 225)

There would, however, hardly be any substantive community acceptance for a program of research that sought to introduce self-evaluation for summative purposes such as certification or high stakes assessments; but there would in all likelihood be widespread acceptance for self-evaluation as a formative process, as an indicator of learning, or as a benchmark against which a more formal assessment might even be compared. While assessment for learning is now a popular term (for example, Fancourt 2005), it is really little more than the positive use of formative evaluation as an instructional or educational tool. A social learning theory framework may be helpful in improving the basis of self-evaluations.

While one rationale for educators' interest in educational self-evaluations has related to finding alternative approaches to formal assessments, a more important consideration has been the role of self-evaluation as a component of any learning process. Commenting from a higher education perspective, Falchikov and Boud (1989, pp. 426–427) noted '[s]elf-assessment may be regarded as a skill and, as such, needs to be developed... Self-assessment can be a valuable learning activity, even in the absence of significant agreement between student and teacher, and can provide feedback to the student about both learning and educational and professional standards'.

For too long the spotlight in education has been on the intricacies of formal methods of summative assessment (Athanasou & Lamprianou 2002). Adult education, however, that is freely chosen and freely pursued in a non-threatening and non-judgemental context really obtains little value from these advances in educational measurement. Here the emphasis ought to be on the formative uses of self-evaluation as a key ingredient of one's learning or achievement and there is some evidence to support the value of such self-evaluations.

Acknowledgements

Preparation of this paper was supported in part by a Research Excellence Grant from the University of Technology, Sydney. An earlier version of this paper was presented at the 14th National VET Research Conference, Wodonga, July 2005. The cooperation and assistance of Dr Joanne Earl is acknowledged together with the helpful comments of two anonymous reviewers.

References

Athanasou, J.A. & Lamprianou, I. (2002). *A teacher's guide to assessment*, Sydney: Social Science Press.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*, Englewood-Cliffs, NJ: Prentice-Hall.

Bergee, M.J. & Cecconi-Roberts, L. (2002). 'Effects of small-group peer interaction on self-evaluation of music performance', *Journal of Research in Music Education,* 50(3), 256–268.

Bernichon, T., Cook, K.E. & Brown, J.D. (2003). 'Seeking self-evaluative feedback: The interactive role of global self-esteem and specific self-views', *Journal of Personality and Social Psychology,* 84, 194–204.

Betsworth, D.G. (1999). 'Accuracy of self-estimated abilities and the relationship between self-estimated abilities and realism for women', *Journal of Career Assessment*, 7(1), 35–43.

Buckingham, J.T. & Alicke, M.D. (2002). 'The influence of individual versus aggregate social comparison and the presence of others on self-evaluations', *Journal of Personality and Social Psychology*, 83 (5), 1117–1130.

Burisch, M. (1984). 'Approaches to personality inventory construction', *American Psychologist*, 39, 214–217.

Cassidy, J., Ziv, Y., Mehta, T.G. & Feeney, B.C. (2003). 'Feedback seeking in children and adolescents: Associations with self-perceptions, attachment representations and depression', *Child Development*, 74, 612–628.

Cooley, C. (1902). *Human nature and the social order*, New York: Scribner.

Falchikov, N. & Boud, D. (1989). 'Student self-assessment in higher education: A meta-analysis', *Review of Educational Research*, 59, 395–428.

Fancourt, N. (2005). 'Challenges for self-assessment in religious education', *British Journal of Religious Education*, 27(2), 115–125.

Farh, J-L., Dobbins, G.H. &Cheng, B-S. (1991). 'Cultural relativity in action: A comparison of self-ratings made by Chinese and US workers', *Personnel Psychology*, 44(1), 129–148.

Klein, W.M.P., & Buckingham, J.T. (2002). 'Self-other biases in judgements of ambiguous performance and corresponding ability', *Psychologica Belgica,* 42(1/2), 43–64.

Knowles, M.S., Holton III, E.F. & Swanson, R.A. (1998). *The adult learner,* 5th edition, Woburn, MA: Butterworth-Heinemann.

Kruger, J. & Dunning, D. (1999). 'Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments', *Journal of Personality and Social Psychology*, 77 (6), 1121–1134.

Lin, L-M., Moore, D. & Zabrucky, K.M. (2001). 'An assessment of students' calibration of comprehension and calibration of performance using multiple measures', *Reading Psychology*, 22, 111–128.

Lunneborg, C.E. (1982). 'Systematic biases in brief self-ratings of vocational qualifications', *Journal of Vocational Behavior*, 20, 255–272.

Mabe, P.A. & West, S.G. (1982). 'Validity of self-evaluation of ability: A review and meta analysis', *Journal of Applied Psychology*, 67, 280–296.

Martin, R., Suls, J. & Wheeler, L. (2002). 'Ability evaluation by proxy: Role of maximal performance and related attributes in social comparison', *Journal of Personality and Social Psychology*, 82(5), 781–791.

Marx, D.M. & Roman, J.S. (2002). 'Female role models: Protecting women's math test performance', *Personality and Social Psychology Bulletin*, 28 (9), 1183–1193.

Mischel, W. (1977). 'On the future of personality measurement', *American Psychologist*, 32, 246–254.

Mihal, W.L. & Graumenz, J.L. (1984). 'An assessment of the accuracy of self-assessment of career decision making', *Journal of Vocational Behavior*, 25, 245–253.

Obach, M.S. (2003). 'A longitudinal-sequential study of perceived academic competence and motivational beliefs for learning among children in middle school', *Educational Psychology*, 23(3), 323–338.

Ormrod, J.E. (2004). *Human learning*, 4th edition, Upper Saddle River, NJ: Pearson.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L. & Roberts, R. (2002). 'The role of individual differences in the accuracy of confidence judgements', *The Journal of General Psychology*, 129(3), 257–299.

Tousignant, M. & DesMarchais, J.E. (2002). 'Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: A correlation study', *Advances in Health Sciences Education*, 7, 19–27.

Van Lange, P.A.M. & Sedikides, C. (1998). 'Being more honest but not necessarily more intelligent than others: Generality and explanations for the Muhammad Ali effect', *European Journal of Social Psychology*, 28, 675–680.

About the author

Dr James Athanasou *is Associate Professor in the Faculty of Education at the University of Technology, Sydney. His areas of specialisation are the assessment of learning and educational-vocational psychology. He is the editor of the Australian Journal of Career Development published by the Australian Council of Educational Research.*

Contact details

Faculty of Education, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007
Tel: (02) 9514 3712 *Fax: (02) 9514 3939*
Email: Jim.Athanasou@uts.edu.au