ATH05117

SOME ISSUES WITH ADULT SELF-EVALUATIONS IN EDUCATION AND TRAINING

James A Athanasou
University of Technology, Sydney

This paper promotes a social-cognitive framework to categorise those factors that enhance self-evaluation. These factors include social messages, personal factors and situational factors in the self-evaluation context. In addition, some technical issues with the accuracy of self-evaluations are outlined. These related to (a) the adequacy of the criterion, (b) the method of statistical comparison, (c) the quantitative or qualitative basis for self-evaluations and (d) the procedures used in obtaining self-evaluations. It is argued that self-evaluations are inherently economical, reliable and valuable especially when they are geared to idiographic, criterion-referenced and formative assessments of performance rather than ability.

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. Self-evaluation is helpful where “...people need to know their own capacities and limitations” (Wood, 1989, p. 231). The purpose of this paper is to outline a framework for the study of self-evaluations in adult education and training.

At the outset it may be helpful to clarify some aspects of terminology since a number of terms seemingly refer to the same phenomenon viz., ‘self-evaluation’, ‘self-assessment’, ‘self-rating’ or ‘self-estimate’. A self-estimate is typically a person’s qualitative or

---

1 Paper presented at the Australian Association for Research in Education, Annual Conference, University of Western Sydney, November 2005.
quantitative description, such as a self-rating. A self-assessment refers to the broader and over all process of comparison against an implicit or explicit standard and it may or may not result in an overt estimate. Self-evaluation, however, is the judgement of the merit or worth of a self-estimate that has been produced. Self-evaluation can be a formal public process or an informal private perception. Typically, McMillan (2004, p. 55) incorporated all three aspects and defined it as “…students’ reporting on or evaluating themselves”. In reviewing the earlier research and in order to include as many studies as possible, these terms have all been categorised under this broader heading of self-evaluation. The following section outlines a theoretical framework of self-evaluation from a social cognitive perspective.

SOCIAL COGNITIVE THEORY AND SELF-EVALUATION

Theories of self-perception have long proposed that people can often be accurate, participant observers of their own behaviour (e.g., Bem, 1972; Duval & Wicklund, 1972). There is no denying the range and accessibility of the information base on which they may reflect (e.g., Shrauger & Osberg, 1981). Furthermore it has been demonstrated that people have a drive to evaluate themselves, their opinions, attitudes, career success and abilities (Festinger, 1954, p. 117; Heslin, 2005).

The social-cognitive concept of self-efficacy refers to an evaluation of capability that follows a self-assessment. It 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 construct emanates from the work of Bandura (1986) who also noted that people have a sound idea of their talents. Factors that influence self-efficacy include: previous successes; reassurances from others; the observations of the success of others (Ormrod,
Elements of self-regulation that are also linked with self-efficacy are: (a) the performance standards that people set; (b) the observations of our own behaviour; and (c) self-reinforcement. If one were to depict this schematically it would be a recursive model with some of the components outlined in Figure 1. The following sections 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. No claim is made that the model is complete but it is promoted as a heuristic framework for the analysis of self-evaluation in adult learning contexts.

Insert Figure 1 about here

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 feedback that we receive from others and the social and cultural stereotypes that form the background of our perceptions.

Comparisons we make with 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 (i.e., proxies). Indeed, the practice of making self-evaluations that are norm-referenced (i.e., 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.
This reflected Leon Festinger’s theory of social comparison (Festinger, 1954), which proposed that people estimate their abilities (but also opinions, performance and their situation) by using others as the basis for comparisons. Festinger proposed that comparison arises when evaluation is not possible through direct knowledge so people compare themselves with others. Brickman and Janoff-Bulman (1977) first described these forces for social comparison as adaptive (seeking information for improvement) or hedonic (seeking to preserve and enhance self-esteem or well-being). An adaptive motivation for self-improvement may lead people to make upward comparison with those who are superior or a hedonic value may accompany a downward comparison with those who are inferior (Wood, 1989).

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 proxy ability levels 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.

The feedback that we receive from others. 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). Lin-Agler, Moore and Zabrucky (2004) investigated self-evaluations of confidence and understanding prior to testing and estimates of the percentage correct following three in-class tests with 36 undergraduate student in a human development course. They reported that students, especially high self-monitors, adjusted their self-evaluations on later tests. High self-monitors increased their ratings from the second to the third exam.

Social and cultural stereotypes that form the background of our perceptions. 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.
Other enduring features from our socialisation also have an impact on self-evaluation. For example, Betsworth (1999) reported that women continue to underestimate their abilities but Halman and Fletcher (2000) reported that female customer service staff attending a development centre (a type of assessment centre), showed greater accuracy than males. Female assessment of communication skills and planning/organizing were statistically significantly correlated with observer ratings of performance. Marx and Roman (2002) demonstrated that for women who had already been identified as motivated with mathematics, who had obtained a minimum SAT score of 650 and who had enrolled in at least one mathematics course there was a cushion or barrier with self-estimates in the presence of a competent role model. The correlation of their self-estimates with a 15-problem mathematics test was .28.

**Personal factors**

One key personal factor in the ability to self-estimate 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 (e.g. 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 .61 compared with .34 for below average and .33 for average students (Wright, 2000).

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. These
include biases such as the leniency of the rater, which was defined “…as the degree to which raters hand out undeservedly high scores” (Fox, Caspy & Reisler, 1994, p. 45).

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 prior experience with and the relevance of the criterion; (b) whether the self-evaluation is made prior to or following learning; or (c) other measurement conditions (e.g., social desirability associated with the judgement, norm-referenced or criterion-referenced judgements, format and manner in which the self-evaluation is elicited).

*Prior experience with and relevance of the criterion.* While self-evaluation is relevant to the field of post-compulsory 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? Klein and Buckingham (2002) also concluded that ambiguity of one’s own performance leads to bias but this effect was lessened when the ambiguity of the task was reduced and the criterion was clearly defined and verifiable (e.g., typing speed).

*Whether the self-evaluation is made prior to or following learning.* 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 .04 to .24 while post-examination and oral presentation correlations ranged from .25 to .33. In a sample of undergraduate students, correlations
(gamma) between pre-test estimates of reading ability and the number of comprehension questions correct were .14 compared with .28 for the post-test (Lin, Moore & Zabrucky, 2001). An increase in self-rating correlations was found with 111 customer service employees of a telecommunications company after completing an assessment centre. Pre-ratings correlated .07 to .25 while post-ratings correlated .09 to .33 (Halman & Fletcher, 2000, p. 431).

Other measurement conditions. When Westbrook, Buck and Wynne (1994) controlled six measurement conditions for collecting self-ratings, they found that the test-retest reliabilities for self-ratings were in the range .72 to .80 (p. 158) and that validity coefficients ranged from .36 to .66. They assessed ninth grade students on eight aptitudes (verbal reasoning, numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, spelling, language use). They controlled for the following conditions: a match between the criterion ability and the self-rating; self-evaluation of performance rather than ability; self-evaluation of past performance; self-evaluation in relative terms; an expectation of validation; and the provision of experience in self-evaluation.

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

For instance, one view of self-estimates 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. This overconfidence in self-evaluation has also been termed ‘illusion of knowing’ (Glenberg, Wilkinson & Epstein, 1982).

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 can 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 .37 and .35 for a year later; and between perceived competence and grade point average was .52 and .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. Fitzgerald, White and Gruppen (2003) compared the examination performance of medical students across three
years and reported correlations between estimates and achievements of .41 between winter and autumn semesters in Year 1, .37 between autumn and winter semesters in Year 2 and .36 between winter semester Year 2 and a practical clinical examination in Year 3. These were correlations for each person rather than group estimates. Equally importantly, they reported that the test-retest reliability of self-estimates was .81 for Year 1 and .79 for Year 2 compared with .60 and .70 for the actual performance scores (p. 648). It is possible, therefore, that our self-evaluations reflect a more stable assessment than the results of some educational assessments.

The relation of self-evaluation to other evaluations has been studied in two separate meta-analyses. From their review of self-assessment and achievement in a higher education context, Falchikov and Boud (1989) reported a moderate mean effect size of .47 (1989, p. 419); a mean correlation between teacher and student marks of .39 (1989, p. 420); and that 64.1% 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 .29. They reported that 88% of the correlations were greater than zero and ranged from -.026 to .80. It is relevant to compare these correlations with the predictive validity of assessments. Ghiselli (1973) reported that the average predictive validity coefficient of aptitude tests with performance on-the-job in major occupational categories was around .39 for training criteria and .22 for proficiency criteria (p. 475). Accordingly there is some prima facie support for the accuracy of self—estimates across a range of aptitudes and abilities but there is also evidence that the ability to self-assess 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-ratings are compared with criteria. The correlation would increase if there was a correction for attenuation.

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-rating 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 .3 and each had a reliability of only .5 (a low reliability), then in theory the attenuated correlation between the two imperfect measures could be as high as .6.

Furthermore, any correlation from group comparisons automatically standardizes the self-evaluations (i.e., rescales them to have a mean of zero and a standard deviation of one) and also standardizes 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 (e.g., best, second best,
third best and so on) and this rank ordering is the basis for comparison. This procedure is
used in the Career Interest Card Sort where adults rank their seven career interests
(Outdoor, Practical, Scientific, Creative, Business, Office, and People Contact) from first
to last (Athanasou & Hosking, 2003). Fitzgerald et al. (2003) commented, “Intra-
individual analysis enables us to characterize the accuracy of individual students, as
opposed to an inter-individual analysis, which produce group level estimates of accuracy”
(p. 645).

THE EDUCATIONAL VALUE OF SELF-EVALUATIONS 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 (e.g., Fancourt, 2005) it is really little more than the positive use of formative assessment as an instructional or educational tool. A social learning theory framework may be helpful in providing a basis for modifying self-evaluations by varying the social messages that people receive, by training people to become accurate in their self-perceptions and by varying the situations under which self-perceptions are made.

Moreover, there are some benefits in using self-evaluations as a routine accompaniment to formal standardized assessments in education. The accuracy of self-evaluations may have value as an indicator in its own right. The self-evaluation may provide a benchmark for comparison in order to ensure some fairness in evaluations. Furthermore, there may be some value in encouraging learners to think about whether they have the aptitude for success or the extent to which they have achieved learning outcomes. Klenowski (1995) identified additional roles for self-evaluation along a scale from minimal (learning to compare achievement with an external measure) through to the situation where learners
are encouraged to structure their learning around their abilities as well as their weaknesses and finally having learners grade their own work. Stefani (1994) described how science students participating in self-assessment commented that self-evaluation was thought-provoking, that it enhanced learning and was challenging.

While one rationale for an interest in educational self-evaluation 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 ‘Self-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 (e.g., 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-assessment and self-evaluation as a key ingredient of one’s learning or achievement. Viewed in this way self-assessment leads to self-evaluation and builds in a feedback loop.

**Acknowledgements**

Preparation of this paper was supported by a Research Excellence Grant from the University of Technology, Sydney, The cooperation and assistance of Dr Joanne Earl is
gratefully acknowledged. An earlier version of this paper will appear in the *Australian Journal of Adult Learning*, November 2005.

**References**


Figure 1 A recursive social cognitive framework for investigating self-evaluation