A comparison of self-assessment and tutor assessment of undergraduate psychology students

SOCIAL BEHAVIOR AND PERSONALITY, 2012, *40*(4), 591-604

http://dx.doi.org/10.2224/sbp.2012.40.4.591

A COMPARISON OF SELF-ASSESSMENT AND TUTOR ASSESSMENT OF UNDERGRADUATE PSYCHOLOGY STUDENTS

WALLY KARNILOWICZ
Victoria University

The capacity for self-assessment amongst undergraduate psychology students (N = 64) was investigated in comparison to tutor assessments. Students were able to assess their own performance reasonably and accurately. However, higher achieving students underestimated their performance, while the opposite was found for lower achieving students. These results reflect the characteristics of self-enhancement and self-diminishment bias.

Keywords: self-assessment, tutor assessment, psychology students.

Assessment is a vital aspect of learning and is used to evaluate students' knowledge of material, understanding of content, ability to conceptualize, and capacity to think critically. Within this assessment framework, colleges and universities also emphasize self-directed education in order to maintain and improve academic performance. An important component of this form of education is *self-assessment*, defined by Boud (1991) as the "*involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards*" (p. 5).

University faculties, schools, and departments appreciate and reward knowledge of academic content. Students, in turn, tend to be more competent if they have the ability to learn independently, be critical, and meet and confront new and

Wally Karnilowicz, School of Social Sciences and Psychology, Victoria University.
The author thanks research assistants Kirstie Fleetwood and Sarah Rees, Bachelor of Arts students from the Department of Psychology, University of Bath, UK, on placement at Victoria University.
Correspondence concerning this article should be addressed to: Wally Karnilowicz, School of Social Sciences and Psychology, Footscray Park Campus, Victoria University, Ballarat Road, Footscray, P.O. Box 14428, Melbourne, VIC 8001, Australia. Email: **wally.karnilowicz@vu.edu.au**

different learning expectations associated with an independent awareness of learning requirements and careful observation and scrutiny of one's own performance. Self-assessment as a form of self-regulated learning is often used interchangeably with, and referred to as, metacognitive learning (refer to Hofer & Sinatra, 2010). It involves a conscious effort by learners to examine their thoughts and actions and to engage themselves in controlling them (Dinsmore, Alexander, & Loughlin, 2008). In the academic sense, it involves increasing the learner's capacity to know and understand his/her knowledge and ability and to control the cognitive processes associated with a specific learning task.

Ward, Gruppen, and Regehr (2002) contend that although there is value in the theory of self-assessment, methodological issues plague the measurement of self-assessment ability. Nevertheless, self-assessment has been consistently associated with improved student learning. For example, Dochy, Segers, and Sluijsmans (1999), in their review of 63 studies published between 1987 and 1998, concluded that self-assessment improves performance and the quality of student learning. In addition, Lopez and Kossack (2007) reported that a relationship between performance and self-assessment over time resulted in course grades for students who used continuous self-assessment showing a consistent increase across the unit tests. Lopez and Kossack also reported that end-of-course self-assessment correlations with students' actual course grades were more significantly aligned for the continuous self-assessment group. Dochy et al. (1999) further concluded that self-assessment "leads to more reflection on one's own work, a higher standard of outcomes, responsibility for one's own learning, and increasing understanding of problem-solving" (p. 337).

Researchers have also investigated the accuracy of self-assessed performance indicators. Although Eva, Cunnington, Reiter, Keane, and Norman (2004), in a study of 217 medical students, reported low correlations between student self-assessment and student performance, the majority of researchers tend to support a positive correlation between self-assessment and measures of performance. For example, Falchikov and Boud (1989) in their meta-analysis of 57 quantitative studies predominantly involving undergraduate and graduate college students and their instructors, reported a close correspondence between self- and teacher marks. In a related study, Boud and Falchikov (1989) in their meta-analysis of 48 quantitative studies reported that student marks tended to agree with teacher marks. Lindblom-ylänne, Pihlajamäki, and Kotkas (2006), in a study of 15 law students, reported a strong consistency between self- and teacher assessments. Lynn, Holzer, and O'Neill (2006), in a study of 56 residents in general psychiatry programs, concluded that residents were able to accurately assess their strengths and weaknesses as measured using a standard examination of psychiatric knowledge. Fitzgerald, White, and Gruppen (2003) examined self-assessment accuracy from the first year to the third year of medical school with a sample of

medical students (N = 500) and reported moderate to high measures of self-assessment accuracy that were also relatively stable over time. This was particularly the case with more versus less familiar tasks. Fitzgerald, Gruppen, White, and Davis (1997) reported that in a class of 170 first-year medical students, between 137 and 164 who supplied self-evaluations accurately predicted their test scores in a series of course quizzes.

Self-assessment accuracy has also been characterized by a number of other performance factors. For example, Maki, Shields, Wheeler, and Zacchilli (2005), in a study of 159 college student volunteers from general psychology courses, reported metacomprehension measures associated with the reading of difficult texts. They concluded that students with lower verbal abilities were overconfident in predictions of future performance while students with higher verbal abilities had less confidence in judging past performance. In contrast, Dochy et al. (1999), in a review of literature based on the narrative analysis of 63 studies, reported that the accuracy of self-assessment improves with feedback and develops over time. Dochy et al. also reported that advanced students showed greater performance prediction skills in comparison to novice students. Lejk and Wyvill (2001), in a study with 172 participants, concluded that students towards the top of the group in terms of their performance tended to underassess themselves relative to peer assessment while students performing towards the bottom of the group tended to overassess themselves. In addition, Boud and Falchikov (1989) reported that graduates and students in the latter years of a course tended to be both more accurate in their assessments and underestimate their performance, although they also reported that there was an absence of a consistent tendency to overestimate or underestimate. Falchikov and Boud (1989) also reported a general lack of support for a conclusive trend related to the accuracy of self-assessment among senior students and concluded that expertise in a field is more predictive of accuracy in self-assessment than are seniority or duration of enrolment. Eva et al. (2004) also did not find evidence to support an improvement in self-assessment performance over time while Lynn et al. (2006) reported a lack of evidence to support an improvement in self-assessment accuracy related to an advanced status.

A further trend in this research indicated variability in self-estimates of performance according to measures of competence. For example, Cassidy (2007), in a study of 160 first-year undergraduate students, carried out a comparison of their estimated mark for completed coursework with the tutor's mark and concluded that there was a high level of competency among the majority of students in self-assessment skills, but there was a tendency to underestimate their mark. Fitzgerald et al. (1997) reported that students in the lower quartile of the performance distribution tended to overestimate their performance and students in the upper quartile underestimated their performance. Similarly, John

and Robins (1994) studied 106 Master of Business Administration students and reported a self-enhancement bias effect in self-evaluations, with most students overestimating their performance and a lesser but still sizeable minority exhibiting a self-diminishment bias in underestimating their performance. John and Robins (1994) concluded that students whose evaluations were the most highly unrealistic relative to teacher evaluations were highly narcissistic, while those with the lowest self-perception scores also scored extremely low on narcissism.

The accuracy of self-assessment has also been investigated in relation to students' academic ability. For example, Cassidy (2007) concluded that the higher the student's estimated mark, the less accurate the assessment, while the higher the student's actual mark, the greater the degree of estimated accuracy, with the latter indicating that the more academically capable students had superior self-assessment skills. Boud and Falchikov (1989) reported a general tendency for students to either overrate or underrate their self-assessments in contrast to teacher assessments. However, they also concluded that more able students tended to self-assess more accurately and to underrate their performance, while their less able peers tended to overrate their assessments. Kennedy, Lawton, and Plumlee (2002), in their study of 209 undergraduate and graduate college students, concluded that less capable students significantly overestimated their performance, while more capable students underestimated their performance. However, the less capable group became more accurate in self-assessments over time. Lew, Alwis, and Schmidt (2010), in a study of 3,588 first-year university students, also reported that more academically competent students had a greater capacity to accurately self-assess. The tendency for those with low ability levels to overestimate has been supported in psychology theory generally and in some instances (Kruger & Dunning, 1999) considered as illustrative of personal incompetence because it was deemed that these individuals lacked the metacognitive abilities to distinguish accuracy from error. This characteristic of performance is also consistent with self-enhancement (Shrauger, 1975) and self-diminishment theories in which it is posited that low achievers tend to overestimate or self-enhance the value of their work while high achievers tend to underestimate or self-diminish theirs.

A review of the literature leads to two predictions. First, students in general will be able to assess their own performance reasonably and accurately relative to tutor assessment. Second, higher achieving students will tend to underestimate their self-assessment relative to tutor assessment, while lower achieving students will overestimate their performance relative to tutor assessment. The latter hypothesis reflects the characteristics of self-enhancement and self-diminishment bias.

Method

Participants

Of the 64 undergraduate psychology students attending Victoria University who participated in the study, there were 12 males with a mean age of 21.08 years (SD = 1.31) and 52 females with a mean age of 21.75 years (SD = 3.70). The distribution of students by gender was representative of the population of students enrolled in their course. They were enrolled in a compulsory third- and final-year unit of study (History and Theories in Psychology) as part of the major in psychology. The unit was conducted over one semester (12 weeks) and consisted of a one-hour lecture and a one-hour seminar each week. The assessment task used in this study was a short critical review based on readings and associated activities presented in the first four seminar sessions. Participants overall were in attendance in 80% of seminar sessions, with 41 students attending all sessions. All students consented to participate in the study voluntarily.

Procedure

The assessment task was aligned with the requirements of the core unit of study:

History and Theories in Psychology

Content

The place of psychological theories and practices in twentieth century thought is pursued through lecture presentations and seminar discussions on recent philosophies of science, including positivist, constructivist/interpretive, and critical approaches, and utilizing within-psychology case examples such as behaviorism, psychoanalysis, cognitivism, information processing, and critical psychology.

Aims

- to integrate understandings of psychology as a discipline.
- to introduce the key concepts of the history and philosophy of science as they apply to the discipline of psychology.
- to consider the objects and methods of inquiry within particular schools of psychology, and their respective conceptualizations of "truth" and "science".
- to relate the activities and concepts of particular schools to broader historical factors, and consider their relative merits and limitations.
- to encourage reflective process in the practice of psychological activities.
- to develop thinking in the application of psychology to particular research questions.
- to develop each student's understanding of their preferred approach to psychology.

Assessment: Three critical reviews: the first two 500-750 words in length each (30% each) and the third 1250-1500 words (40%).

Participants were required to complete a short critical review, which constituted 30% of their overall assessment requirement. The task was as follows:

Critical review (due week 4)

The reviews provide a space for you to explore. These initial explorations will form the basis of informed class discussions, which may then lead to a successful review. There are recommended readings associated with each seminar, but you may choose other readings. Students are expected to have read at least two readings before each seminar. ***Reviews are to be based on two readings.*** Critical reviews should try to capture key issues in two of the set readings for one particular week's lecture, *or any other pair of readings, so long as the pairing is first discussed with your seminar leader.* Choose readings from the first four weeks.

- establish the central argument(s) in each reading
- find a place within the discipline of psychology and the history of the twentieth century for the ideas (note that some readings do not refer directly to psychology, so you have to make the connection)
- consider the particular models or metaphors implied by the language used and the ideas relating to (for example) science, context, power, change, society, and gender/culture
- be reflective, and include personal responses.

The review must be presented with references and be sourced appropriately (30 marks).

The task was relatively unfamiliar to these students given that the opportunity to critique articles as an assessment task was limited in other units in this course. The criteria for assessment (Appendix I) were outlined to students within their unit guide and provided in hard copy and electronically. The assessment criteria were based on the Structure of the Observed Learning Outcome (SOLO) taxonomy, in which desired learning outcomes are specified in terms of the quality of student response. It is a form of criterion-referenced assessment in which the criteria are predefined for each grade level, with expected learning outcomes becoming more complex at each level. Within a range, the mark is dependent upon the quality of argument and presentation.

The assessment task was largely consistent with level 1 of the framework for self-assessment activities proposed by Adams and King (1995). Students were given an initial tutorial in seminar session one (week one) of the unit in which an outline was given of the characteristics and details related to the assessment criteria associated with the critical review. The students were presented with a hard copy of the assessment criteria. Two research assistants presented a spoken scripted and Microsoft PowerPoint presentation related to the criteria to the students. During the presentation students were provided with, and asked to study, the assessment criteria and afterwards they were provided with two exemplars from the previous year's History and Theories in Psychology critical reviews. The students were not told that the first exemplar had been assessed as a high distinction and the second at the level of a fail. The students were asked to study the exemplars and, using the assessment criteria, place each into the appropriate grade boundary. Participants were then informed of the assessed grade boundaries for each exemplar and the reasons for this were outlined in the presentation.

The tutor taught the subsequent seminars as outlined in the course guide. The students submitted their completed 500-750 word critical reviews by the end of week four. The research assistants subsequently collected the self-assessments from the students in week five. The tutor is a full-time tenured staff member in the School of Social Sciences and Psychology, Victoria University, with four years of experience in the unit teaching and assessing critical reviews using the assessment criteria. The exemplars presented to the students were examples of the tutor's assessment practices. The tutor could not identify the individual assessments as the students were identifiable only through their student numbers as forwarded to the research assistants. Within the next week the tutor independently assessed each student's critical review, which included the assessed mark and grade and a description of the associated assessment criteria. The assessed work was then returned to the students.

Data Analysis

The first prediction examining the assessment of the student's performance related to the tutor assessment was tested using the Pearson correlation coefficient. The second prediction that higher achieving students would tend to underestimate their self-assessment relative to the tutor assessment, and vice versa for lower achieving students was tested using a two-way within-subjects analysis of variance (ANOVA). The high-achieving and low-achieving students constituted the between-subjects levels of the independent variable and the tutor assessments and student self-assessments composed the within-subjects repeated measures dependent variable. The assessments were based on a percentage mark.

Results

Accuracy of Self-evaluations Relative to Tutor Evaluations

Pearson correlation coefficients were used to examine the pairwise correlations between the student's self-assessment and the tutor's assessment of the critical review. The results indicated that students' self-assessments were relatively accurate estimates of the tutor's assessment (r = .420, p = .001, two-tailed). A further examination of the accuracy of self-evaluations relative to tutor evaluations was completed using a subgroup of students according to their level of performance. Participants were divided into a high-achieving group comprising students who scored 70 or above in the tutor's assessment (n = 28) and a low-achieving group comprising students who scored less than 60 in the tutor's assessment (n = 25). A closer examination of these results indicated that the self-assessments of the high-achieving group related to the tutor's assessments (r = .430, p = .025, two-tailed) were more accurate than the self-assessments of the low-achieving group related to the tutor's estimates (r = .266, p = .199, two-tailed).

Self-enhancement Versus Self-diminishment Bias: Student Self-assessments in Contrast to Tutor Assessments

A further examination of the data was used to test for the effect of self-enhancement and self-diminishment according to estimated levels of performance relative to the tutor's assessments. Descriptive statistics associated with the tests of the difference between the low-achieving group ($n = 25$) and the high-achieving group ($n = 28$) in terms of tutor and self-assessments are shown in Table 1.

Table 1. *Means and Standard Deviations of Self-assessments and Tutor Assessments across Levels of Achievement*

	Level of achievement					
	High		Low		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*
Self-assessed mark	72.74	7.36	67.42	6.84	70.18	7.54
Tutor-assessed mark	82.85	6.92	50.68	4.91	67.38	17.30
Overall	77.93	8.70	59.05	10.31	68.94	13.38

The means indicate that the students in the high-achieving group tended to underestimate their self-assessed mark relative to the tutor's mark but low-achieving students tended to overestimate their self-assessments relative to the tutor's mark. A further descriptive analysis of the data indicated that on average the high-achieving group underestimated their mark by 10.11 points ($SD = 7.63$) and the low-achieving group overestimated their mark by an average of 15 points ($SD = 10.56$).

Levene's test of equality of error variances indicated that variances were uniform across all levels of the repeated measures variables because associated significance values were greater than .05. The results of the ANOVA indicated that there was a significant difference in assessments between high-achieving and low-achieving students across self- and tutor assessments of performance $F(1, 50) = 153.918$, $p = .000$, partial $\eta^2 = .755$. In addition, the tests of within-subjects main effects indicated a significant difference between students' self-assessments and tutor assessments $F(1, 50) = 10.221$, $p = .002$, partial $\eta^2 = .170$. Specifically, students' self-assessments were higher than tutor assessments of performance. The most important point central to the study predictions, was that there was a significant interaction effect between tutor's assessments and students' self-assessments $F(1, 50) = 167.709$, $p = .000$, partial $\eta^2 = .770$. High-achieving students' self-assessments were lower than tutor assessments, while low-achieving students' self-assessments were higher than tutor assessments (see Figure 1).

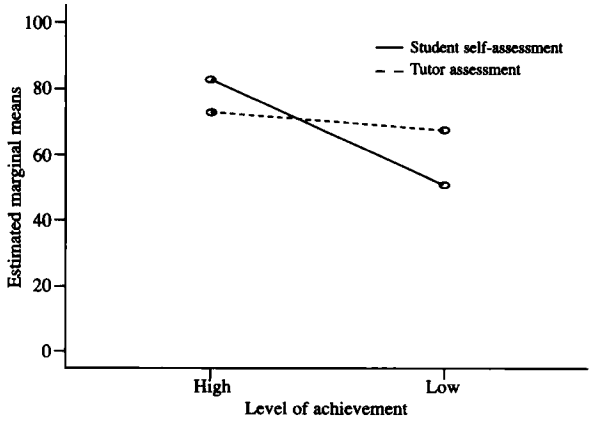

Figure 1. *Mean student self-assessments in contrast to mean tutor assessments.*

Discussion

The results strongly supported the study predictions. First, students in general were able to assess their performance reasonably accurately relative to the tutor assessment. Nonetheless, and consistent with research (see e.g., Moreland, Miller, & Laucka, 1981), low-achieving students were less accurate than high-achieving students when judging their work against the assessment of the tutor. Second, higher achieving students tended to underestimate their self-assessment and lower achieving students tended to overestimate their performance relative to the tutor assessment. The support for the latter prediction corroborated the general theories of self-enhancement and self-diminishment in that low-achieving participants tended to overestimate or self-enhance the value of their work while high-achieving students tended to underestimate or self-diminish the value of their work.

Of particular concern, and consistent with much of the research, were the misconceptions of lower achieving or less able students in their evaluations of performance. In a similar study, Moreland et al. (1981) reported that less able students had difficulty in evaluating their own course work. However, Moreland et al. considered that these students' inaccurate self-evaluations were not because of a lack of knowledge of the grading criteria but rather an inability to apply assessment criteria to their own course work.

A further explanation for the misconceptions and lesser ability of low-achieving students to apply set criteria to assessment may be related to general difficulties with being objective (Lindblom-ylänne et al., 2006). One of the limitations in this study was that the self-assessment of a conceptual question was based on

readings and interpretation of the text. Consistent with the concerns of Eva et al. (2004), is the likelihood of strong assessments in this instance, particularly with a lower achieving student, being restricted. An alternative theory is that if students are unskilled it may cause them to bear a dual burden (Kruger & Dunning, 1999). The first burden is that the unskilled student will reach erroneous conclusions and make unfortunate choices and the second is that this incompetence robs the student of the metacognitive ability to realize the extent of the error. Instead he or she believes he/she is succeeding and performing according to his/her own expectations.

In contrast to the low achievers, high-achieving students were more accurate in their self-assessments in relation to the tutor's assessment but nonetheless they underestimated their performance relative to the tutor's assessment. This result is more difficult to explain. However, in using the self-diminishment paradigm in relation to self-image, students with positive academic self-images may self-diminish in order to be rated more positively by others. Under these circumstances there is a reduced drive to self-enhance. John and Robins (1994) offer a further explanation in contending that high achievers who are low in narcissism underestimate their contributions owing to self-diminishment bias.

In any case, improving the capacity to self-assess is important in a system which encourages self-directed learning. A possible means of improving self-assessment from educational psychology and within metacomprehension is based on Zhao and Linderholm (2008), in their synthesis and review of the literature on metacomprehension related to reading. High self-assessment accuracy may be achieved when the main obstacles to accuracy are overcome. First, test or assessment uncertainty should be reduced by anchoring of judgments less on self-perceptions of ability. Students need to be engaged in discussions on test and assessment formats and should be provided with more information about the content of assessment tasks. Second, techniques should be developed to improve the diagnostic validity of the experiential cues that are the bases of the assessment tasks, for example, engaging in a dialog with peers about a completed assessment piece.

The results also need to be considered in light of the Falchikov and Boud (1989) contention that the practice of using teacher (tutor) marks as the standard against which the reliability or accuracy of peer marks is measured is problematic given that it cannot be assumed that teachers' ratings themselves have satisfactory reliabilities. Tutor assessment ratings are related to the type of assessment task and the characteristics of the tutor. For example, Magin and Helmore (2001) investigated the reliabilities of peer and teacher summative assessments of engineering students' oral presentation skills in a fourth-year communications subject. They concluded that although teacher assessments were more reliable than peer assessments, the use of a single teacher rating of oral presentation

skills was inadequate as a reliable assessment measure. A further limitation in the current study was the absence of measures associated with psychological and personality traits. For example, AlFallay (2004), in a study of 78 students of English as a foreign language, concluded that, in terms of assessing their own performance and those of their peers, learners possessing the positive side of a trait (i.e., motivation types and self-esteem) were more accurate than those possessing a negative side. Furthermore, Brown, Collins, and Schmidt (1988) contended that self-serving or self-enhanced biases are "principally the refuge of those with low self-esteem" (p. 445) and concluded that persons with low or high self-esteem will exhibit different self-enhancement biases under different conditions.

The grades students receive at university can have lasting effects in many areas of life, including their self-concept, relationships, and subsequent careers. As a result, it is not surprising that anxiety levels in students have been found to be strongly related to grades and performance (Moreland et al., 1981). Anxiety may also result from a discrepancy between the expected grade and the actual grade. The attribution made for the reasons for the grade varies according to academic ability. Less capable students are more likely to attribute their grades to external, unstable factors, such as luck, task difficulty, or the marker of their paper. On the other hand, those who achieve higher grades often make internal attributions towards the causes underlying their performance, such as the amount of effort put into the assignment and their ability (Moreland et al., 1981). Differences between more able and less able students have also been seen in the area of self-esteem, as the latter may have lower self-esteem and as a result have a tendency to grade their work unrealistically high in order to enhance and maintain their view of self (John & Robins, 1994).

There is a need for conceptually relevant and psychometrically sound criterion measures against which to compare self-perceptions (John & Robins, 1994). Furthermore, there is a need for well-designed studies incorporating careful attention to the construction and development of reliable and valid independent variables (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000). Fulfilling this need would clearly result in a closer correspondence between student and teacher marking. Nonetheless, the results gained in this study are promising and add to the knowledge about the value of self-assessment as a means of promoting self-directed learning in higher education.

References

Adams, C., & King, K. (1995). Towards a framework for student self-assessment. *Innovations in Education and Training International, 32*, 336-343. **http://doi.org/hnm**

AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System, 32*, 407-425. **http://doi.org/cx272z**

Boud, D. (1991). *Implementing student self-assessment* (2nd ed.). HERDSA green guide no. 5. Kensington, NSW, Australia.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529-549. **http://doi.org/hnn**

Brown, J. D., Collins, R. L., & Schmidt, G. W. (1988). Self-esteem and direct versus indirect forms of self-enhancement. *Journal of Personality and Social Psychology, 55*, 445-453. **http://doi.org/hnp**

Cassidy, S. (2007). Assessing inexperienced students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*, 313-330. **http://doi.org/b7ht5v**

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*, 391-409. **http://doi.org/c77zjd**

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

Eva, K. W., Cunnington, J. P., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Sciences Education, 9*, 211-224. **http://doi.org/hnr**

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430. **http://doi.org/hns**

Falchikov. N., & Goldfinch, F. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287-322. **http://doi.org/hnt**

Fitzgerald, J. T., Gruppen, L. D., White, B. A., & Davis, W. K. (1997). *Medical student self-assessment abilities: Accuracy and calibration.* Presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, April. Retrieved from **http://eric.ed.gov/PDFS/ED410296.pdf**

Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education, 37*, 645-649. **http://doi.org/cz6h6k**

Hofer, B. K., & Sinatra, G. M. (2010). Epistemology, metacognition, and self-regulation: Musings on an emerging field. *Metacognition and Learning, 5*, 113-120. **http://doi.org/bj3h86**

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206-219. **http://doi.org/hnv**

Kennedy, E. J., Lawton, L., & Plumlee, E. L. (2002). Blissful ignorance: The problem of unrecognized incompetence and academic performance. *Journal of Marketing Education, 24*, 243-252. **http://doi.org/bps9j**

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*, 1121-1134. **http://doi.org/dks**

Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of selfassessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education, 26*, 551-561. **http://doi.org/brg2sw**

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35*, 135-156. **http://doi.org/hnw**

Lindblom-ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education, 7*, 51-62. **http://doi.org/b5849w**

Lopez, R., & Kossack, S. (2007). Effects of recurring use of self-assessment in university courses. *International Journal of Learning*, *14*, 203-216.

Lynn, D. J., Holzer, C., & O'Neill, P. (2006). Relationships between self-assessment skills, test performance, and demographic variables in psychiatry residents. *Advances in Health Sciences Education*, *11*, 51-60. **http://doi.org/cxbgkr**

Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education*, *26*, 287-298. **http://doi.org/bmxtj9**

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*, 723-731. **http://doi.org/hnj**

Moreland, R., Miller, J., & Laucka, F. (1981). Academic achievement and self-evaluations of academic performance. *Journal of Educational Psychology*, *73*, 335-344. **http://doi.org/hnk**

Shrauger, J. S. (1975). Responses to evaluation as a function of initial self-perceptions. *Psychological Bulletin*, *82*, 581-596. **http://doi.org/bctz6d**

Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, *7*, 63-80. **http://doi.org/cjhqh2**

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review*, *20*, 191-206. **http://doi.org/bvrsrg**

Appendix I
Guidelines for Assessment Criteria

In History and Theories, the SOLO (Structure of the Observed Learning Outcome) assessment criteria below are used. In this approach to assessment desired learning outcomes are specified in terms of the quality of student response. It is a form of criterion-referenced assessment, in which the criteria are predefined for *each* grade level, with expected learning outcomes becoming more complex at each level. Within a range, the mark will depend on the quality of argument and presentation.

HD (24-30, i.e., 80%+) You must be able to:
- *understand and analyze* arguments presented by both authors;
- *evaluate* the arguments in relation to their place within relevant psychological discourses;
- *reflect* critically on their implications for psychological science and practice, drawing on your own experience so far, and wider reading;
- *demonstrate* capacity for original thinking around the key issues arising from these readings;
- *write* in a manner best suited to expressing all of the above.

D (21-23, i.e., 70-79%) You must be able to:
- *demonstrate clear understanding of* arguments presented by both authors;
- *explain* the implications of these articles/chapters for psychological science and practice, drawing on your own experience so far and/or wider reading;
- *demonstrate* capacity for clear, critical thinking around the key issues arising from these readings;
- *write* in a manner that assists the reader to understand your argument.

C (18-20, i.e., 60-69%) You must be able to:
- *demonstrate understanding* of the main points presented by both authors;
- *make some connection* between these articles/chapters and your own experience and understanding of psychological science and practice so far;
- *demonstrate* capacity for clear thinking around at least one issue arising from these readings;
- *write* in a manner that does not make it difficult for the reader to understand your argument.

P (15-17, i.e. 50-59%) You must be able to:
- *demonstrate reasonable understanding* of the main points presented by at least one of the authors;
- *outline any connection or difference between* these articles/chapters;
- *demonstrate* capacity to think for yourself around at least one issue arising from these readings;
- *write* in a manner that makes it possible for the reader to understand your argument.

N (<15, i.e., <50%)