Do university students know how they perform?*

Piergiorgio BATTISTELLI
University of Bologna, Italy

Alessia CADAMURO
University of Modena and Reggio Emilia, Italy

Alessandra FARNETI
Libera Università of Bolzano, Italy

Annalisa VERSARI**
University of Bologna, Italy

Abstract

The aim of the research is to study the capacity for self-evaluation of University students undergoing tests involving mathematics, linguistic and formal reasoning. Subjects were asked to estimate the number of correct answers and subsequently to compare their performance with that of their peers. We divided the subjects into three groups on the basis of performance: poor, middle and top performers. The results demonstrate that all the subjects in all tests showed good awareness of their level of actual performance. Analyzing comparative assessments, the results reported in literature by Kruger and Dunning were confirmed: poor performers tend to significantly overestimate their own performance whilst top performers tend to underestimate it. This can be interpreted as a demonstration that the accuracy of comparative self-evaluations depends on a number of variables: cognitive and metacognitive factors and aspects associated with self-representation. Our conclusion is that cognitive and metacognitive processes work as "submerged" in highly subjective representations, allowing dynamics related to safeguarding the image one has of oneself to play a role.

Keywords: metacognition, self- evaluation, cognitive performance, university students, self- image.

Introduction

Metacognition

Metacognition is the totality of psychic activities overseeing the cognitive function (Cornoldi 1995). These activities comprise the knowledge an individual has in relation to mental functions and mechanisms of control and self-regulation activated whilst carrying out first level cognitive activities.

Metacognitive knowledge refers to what a subject knows or believes about a number of cognitive processes, such as memory, understanding, studies, etc. It may include ideas about cognitive functioning in general, convictions about one's own skills, the awareness of the existence of cognitive problems and one's ability to solve them, knowledge about the efficacy and use of strategies and personal strengths and weaknesses in this regard. All these elements may derive from personal experience or from the observation of the behaviour of others (De Beni & Moè, 2000).

Control and self-regulating mechanisms, on the other hand, play a guiding and supervisory role over cognitive processes. They include, for example, planning of the task, anticipating the performance, choosing a suitable strategy and verifying the choices made on the basis of the evaluation of results.

The distinction between knowledge and metacognitive control derives from studies carried out in three parallel areas of research and which are the origins of the two leading aspects attributed to metacognition: studies into cognitive development following the developmental theory of Piaget (1974, 1975), the work of Vygotskij (1978) on the social origin of cognitive control and studies based on the *Human Information Processing* (HIP) model (Richard, 1990). Whilst references to developmental psychology and, in particular, to Piaget's theories, have stressed the *awareness* of the subject in relation to the functioning of his/her mental states, studies based on cognitive psychology and the HIP model have pointed to the role of *control* the subject can exercise over his/her cognitive activities. References to Vygotskij have underlined the central role of *regulation mechanisms*, the importance of *cultural transmission* and the educational role of the adult in relation to both metacognitive knowledge and the use of the various strategies.

From the historical point of view, the origin of the metacognitive theory resides in the studies of Flavell at the beginning of the seventies. The term 'metacognition' was used for the first time, in fact, by Flavell in his pioneering work of 1976, mainly in relation to studies on memory.

In his model Flavell (1981) included regulation aspects in his definition of metacognition, meaning by it "the totality of knowledge or cognitive activities which have as object or regulate all the aspects of mental acts" (Flavell, 1981, p. 37): alongside knowledge metacognitive experiences are

introduced, understood as ideas, thoughts, sensations, relating to cognitive activities acting at all levels of the task, before, during and after.

Beginning with these initial models, there was then a proliferation of studies which gradually attributed greater and greater importance to control and monitoring aspects, alongside the aspects linked to knowledge of cognitive processes, going so far as to affirm that metacognition influenced cognitive activities, among other ways, through monitoring, regulation and orchestration (Brown, A. L., & DeLoache, 1978; Campione & Brown, 1978).

The model put forward by Brown (1987) focused specifically on the monitoring activity that accompanies carrying out the task and suggested that there are various types of metacognitive control processes: anticipation of the performance level, planning, monitoring and evaluation.

In 1985 Borkowski put forward a model in which various metacognitive skills of control and regulation can be identified, including: awareness of one's own cognitive function and of this function in general, expectation, planning, monitoring, metacognitive review, evaluation, abstraction and transfer.

Similarly, Pintrich, Wolters and Baxter (as cited in Borkowski, 1996, p.393) distinguished between three correlated aspects of metacognition: Knowledge, Judgement-Monitoring and Self Regulation.

Consequently, the most recent metacognitive models have been enriched by contributions from emotive-motivational theory (Borkowski & Muthukrishna, 1995; De Beni & Pazzaglia, 1991; Hultsch, Herzog, Dixon & Davidson, 1988; Moè & De Beni, 1995), describing metacognition as a complex interactive system with diverse components: variables associated with personal and motivational states (attributive style, motivation to use a strategic form of behaviour), self-esteem and self-efficacy (sense of personal value, knowledge of possible selves, awareness of one's aims), in addition to knowledge of strategies and control processes.

Self-image and causal attributions

Within these variables it seems that an important place is occupied precisely by those personal factors which may act as a driver to activate, maintain and, if necessary, correct one's cognitive activity: the concepts of self-efficacy and the expectation of a result (Bandura, 1986, 2000; Mazzoni, 2000). The first referred to the degree of confidence of an individual in relation to the likelihood of achieving an objective he has set himself. The second referred to the relationship between the way a task is carried out and the result the individual expects to achieve, given the way the task is to be carried out. Evaluations of self-efficacy varied on the basis of three dimensions: difficulty of the task, degree of generality/specificity of the evaluation and the strength of the evaluation. The generality/specificity dimension referred to the awareness an individual has of possessing some or many skills, whilst the intensity of the sense of self-efficacy referred to the

degree of conviction an individual has in relation to his skills. There was a positive correlation between a high degree of conviction and good performance. This is because those with a high sense of self-efficacy persist in tasks where they initially fail (Bandura, 1986).

Moè and De Beni (1995) distinguished between an objective of mastering a task (or learning aims) and the aim to achieve personal success. According to the authors, those who had the aim to achieve mastery wish to improve their culture, believed in co-operating with others and wanted to learn new strategies, applied themselves and thought that understanding is more important than memorizing. On the contrary, those who seeked personal success were motivated by the need to feel superior to others, they believed this was necessary in order to be successful without making much effort (Ames & Archer, 1988). Clearly this model was close to that of Dweck (1999) who distinguished between motivation based on mastery and motivation based on performance.

Petter (1992) distinguished between direct motivations, based on the quality of the activity or prestige, and indirect motivations associated with "projects" or "problems" and extrinsic motivations, represented by marks, rewards and punishments.

Closely linked to motivation is the subject's style of attribution. The process of attribution takes place when an individual, observing an event, attributes to that event a specific cause (Frieze & Bar Tal, 1980). The importance of attributions is given by the fact that they influence cognitive performances and learning at school, persistence, the choice of a task, emotions and expectancies, contributing to produce success and failure.

Heider (1958) was one of the first researchers to propose a classification based on the attribution of inner or outer causes, distinguishing between events attributed to oneself and events attributed to external causes.

Other authors, including Weiner, Frieze, Kukla, Reed, Rest and Rosenbaum (1978), introduced the analysis of stability in relation to the cause, distinguishing between stable causes such as skills and unstable causes such as luck. The dimension of stability influences changes in the expectations of the individual after a success or failure.

Weiner (1986) further enriched these classifications by introducing the idea of the controllability of these causes or lack of it. He pointed out that emotions linked to self-esteem (for example satisfaction, confidence, guilt, etc.) are closely correlated with the attribution locus. The attribution of a success to oneself (inner attribution locus, e.g. skill), generates good self-esteem, whereas the attribution to oneself of a failure causes a lack of self-esteem. If the cause of success/failure is attributed to the task, the result may be a sense of satisfaction (for a success) or sense of guilt (for a failure).

In the situations in which the attribution is for an external attributive locus (e.g. the help of others), the feelings are of gratitude in the event of success and anger in cases of failure.

In the light of all these theories and models it seems clear that, in relation to metacognition, alongside cognitive factors, motivation and processes linked to emotions/affections play an important role.

In this regard, the formulation of Nisbett and Ross (1980) was particularly lucid: the biases of human inference may be attributed to logical errors the subject commits while processing information, or to the interference of motivational or emotional factors which disturb and deform the resulting representations. In the authors' definition, explanations of the first type are "cold" cognitions, those of the second type are "hot" cognitions. Although specifying that there are no scientifically validated reasons for opting for one interpretation or the other, Nisbett and Ross declared their preference for "cold" explanations; and, in fact, it is known that their paper was one of the crucial moments in heuristic research and in cognitive processes "with limited rationality".

Finally, as Rivière (1999) pointed out, these two approaches (hot vs. cold) can also be found in studies on the development of meta-representative thought where they are focussed on computational models based on the processing of information and on models based on the construction of representations of a socio-cognitive nature

Self-evaluation of cognitive performance

An interesting sector within the metacognitive approach, where metacognitive knowledge, control processes and emotional-motivational aspects are intertwined, is the area of metacognitive assessments. Self-evaluation of performance and cognitive skills is considered a fundamental dimension of the control functions carried out by metacognitive monitoring and depends, as we have already seen, on a number of cognitive, metacognitive and emotional-motivational variables (Cadamuro, 2004; Cornoldi, 1995; Flavell, 1981; Izaute & Chambres, 2002; Mazzoni & Nelson, 1998; Schwartz & Perfect, 2002).

Metacognitive assessments are subjective judgements relating to the personal ability to succeed in a given task (De Beni & Moè, 2000). When preparing to carry out the task and in assessing the results, there is a spontaneous anticipation of the likely performance and reflection about the results. This becomes the basis for modifying forecasts of results in subsequent tests.

The awareness of one's own cognitive performance limits was studied in depth by Kruger and Dunning (1999). The authors asked various subject samples to carry out tests involving logical reasoning, to assess examples of humour, to undergo tests involving syntactical skills and then to evaluate their performance and skills in each area. Subjects were asked to provide

these assessments referring to the "average performance and skills of students at their University", using a percentage scale of 0 to 100, whose meaning was self-evident but was also explained. The results showed one phenomenon very clearly and, to some extent, paradoxically: the subjects who obtained the lowest actual performance scores overestimated both their performances and their skills in relation to performance. On the other hand, the subjects with the highest scores tended to underestimate their performance and skills. The explanation of the phenomenon seems to involve a lack of metacognitive skill, accompanied by low skills as shown by the tests. In other words those who do not know how to do things also don't know that they don't know how to do them; they also fail to properly assess others' skills as some of the variations of the experiment of Kruger and Dunning show. For example, some of the subjects who had been tested for syntactical skills were later asked to look at the tests of 5 others with similar scores. The least able in terms of the test were also the least able in assessing others' tests and the most able in terms of the tests were also the best able to assess others' tests.

The underestimation by the most able subjects may be due to the difficulty in assessing the average performance of others, an effect called "false consensus" consisting in over-optimistic assessments of the abilities of others. In order to verify this hypothesis, Kruger and Dunning asked low scorers to undergo first a test of logical reasoning, then training in logic to provide them with the cognitive and metacognitive skills required to correct their overestimations. This training significantly reduced errors in self-evaluation in the lowest scorers, confirming, in the authors' opinion, the hypothesis that poor basic skills are accompanied by low metacognitive awareness. For the high scorers, it was enough to give them some low-scoring tests to correct their optimistic assessments of the average skills of others.

In 2002 Krueger and Mueller joined the debate by objecting that the phenomenon reported by Kruger and Dunning (1999) was in fact due to the joint action of heuristics called *better-than-average* and the statistical effect of regression.

This heuristics consists in the tendency of people to assess themselves as above average: this excess of optimism is a highly irrational bias in that it is logically impossible for everyone to be above average (on the other hand, the assessments are given individually and hence the question does not arise in these terms).

The phenomenon of regression consists in the fact that the average of many repeated measurements tends to nullify the extremes: hence the self-evaluation values of subjects tends towards the average.

Krueger and Mueller (2002) replicated the research of Kruger and Dunning (1999) applying some statistical controls to nullify the regression effect. In this way they highlighted the effect of focussing on oneself and the

degree of confidence in estimates of performance as intermediate variables in the process. To sum up, in their opinion the hypothesis based on statistical regression and the heuristics of *better than average,* provide a more complete explanation of the results in question. In the same edition of the journal, Kruger and Dunning reaffirmed the consistency of the phenomenon even after statistical controls of regression.

Burson, Larrick and Klayman, in a study dated 2006, also supported the hypothesis that the results were due to methodological artificiality: in this case the variable responsible for the observed effect in the research of Kruger and Dunning (1999) were the perceived difficulty of the task. When subjects perceived the task as extremely hard, they believed they will encounter difficulties and their performance will not be very good and, failing to properly account for the degree to which others also experience this difficulty, assessed their performance as worse than average. Burson and colleagues argued that, if everyone produces similar estimates (estimates that are high for tasks perceived to be easy but low for tasks perceived to be difficult) what dictates accuracy is less a matter of greater insight on the part of some participants, more a matter of perceived difficulty. When a test seems easy, everyone believes they have performed well in relation to their peers but only top performers are accurate, leaving bottom performers overconfident. When the test is thought to be hard, however, everyone thinks they have done poorly in relation to their peers and bottom performers will be more accurate than their more competent peers. In short, Burson et al. (2006) argued that whether top or bottom performers are most inaccurate was a result artificially produced by the perceived difficulty of the task.

Burson and colleagues took their results as evidence that the Kruger and Dunning (1999) pattern of over- and underestimation of relative performance was simply a function of using seemingly easy tasks and, as such, did not provide evidence of a relationship between skill level and accuracy in self-assessments.

More recently, Ehrlinger, Johnson, Banner, Kruger and Dunning (2008) examined the relationship between self-insight and level of competence. They considered three explanations for the overconfidence observed among the unskilled: it is a statistical or methodological artefact, stemming from insufficient motivation to be accurate and from a genuine inability to distinguish weak from strong performance. The studies described here are consistent with Kruger and Dunning's (1999) explanation that a lack of skill leads individuals to perform poorly and makes them unable to recognize their poor performances. They found that overestimation among poor performers emerged across a variety of tasks in real world settings too (in which participants had a reasonable amount of prior experience and feedback on the tasks). They asked undergraduates to estimate how well they had performed on course exams and asked members of college debating teams to evaluate their tournament performance. They

provided evidence against the possibility that overestimation among poor performers was a product of insufficient motivation to provide accurate assessments.

They offered incentives (monetary and social) to encourage participants to provide accurate self-assessments and the results demonstrated that not only did incentives failed to improve assessment skills, but actually had the opposite effect: poor performers under incentives became more overconfident. Furthermore, this pattern of overestimation cannot be attributed to a mere statistical artefact, as suggested by Krueger and Mueller (2002), based on notions of statistical reliability and measurement error.

The phenomenon in question, i.e. the overestimation of one's own skills and/or the performance of less skilled subjects, is pervasive and can also be documented in areas which are very different from those of classic cognitive operations. It can be found in the appreciation of practical and professional skills: research carried out on chess players, hunters, doctors and nurses has reported the same phenomenon (Dunning, Johnson, Ehrlinger & Kruger, 2003).

If anywhere, the problem arises in the interpretation of these results and the explanation of the phenomenon: as we have seen, one of the most crucial problems relates to broadening the explanatory model via the inclusion of the variables Nisbett and Ross (1980) call "hot" and Piaget "extra-logical" and which, essentially, are related to one's self-image.

It should also be stated that the phenomenon in question has strong applications significance in any learning process; in fact, as we highlighted in the introduction, the evaluation of the results of a test to a large extent determines the outcome of the process.

Present Study

The aim of the study was to investigate the ability to self-evaluate performance in tests of reasoning of a linguistic, mathematical and formal nature, in a group of University students.

Subjects were asked to provide one objective evaluation (number of correct answers) and two comparative evaluations (comparison with the performance and abilities of a group of peers).

More specifically, following the example of Kruger and Dunning, we intended to verify the hypothesis that subjects less skilled in cognitive tasks tend to overestimate themselves compared to their peers and that more skilled subjects, on the other hand, tend to underestimate themselves.

We expected that, although the subjects can assess their performance quite accurately in objective terms, when asked to make a comparative assessment, they may make errors due to a lack of metacognitive skills and affective components. As Borkowski's model explains (Borkowski, Chan, &

Muthukrishna, 2000), successful information processing results when there is an integration of these metacognitive and affective components.

Instruments

Three cognitive tasks, each with 20 item, were created using item taken from Test di Struttura dell'Intelligenza (Calonghi, Polácek & Ronko, 1974) and from Test di Intelligenza Non Verbale (Pearson & Wiederholt, 1998):

- a task of arithmetic involving the completion of number sequences according to a pattern;
- a task of formal reasoning, taken from the, requiring subjects to complete sequences of geometrical shapes;
- a task of linguistic reasoning asking subjects to identify linguistic analogies, choosing two out of six words linked semantically.

Procedure

Our sample comprised 65 female students at the Faculty of the Science of Primary Education at the University of Reggio Emilia. Mainly female students attend this Faculty, but, as known from the literature, gender does not play a role in self-assessment abilities.

Tests were set in groups and in such a way that upon completion, subjects were asked to estimate:

- how many correct answers they thought they had given (from 0 to 20);
- on a scale of 10, to assess their performance in that specific task "in relation to people who are similar to you";
- on a scale of 10, to assess their general ability in that domain, "in relation to people who are similar to you".

Essentially, with the last two assessments, we asked subjects to give themselves a mark from 1 to 10. To compare these assessments with actual scores (from 0 to 20) in the tests, we converted the scores out of 20 into a score out of 10.

Subjects were divided into three groups, poor, middle and top, each with about a third of the total sample, on the basis of the actual scores (see act.score) obtained in each task.

For each task (arithmetic, formal reasoning and linguistic) a ANOVA, for repeated measures, 3 (groups*: poor, average and top performers*) x 4 *(act. score, est. score, est. perf., est. abil.*) was conducted to verify the effect of the group variable (*between)* on the scores (*within).*

These were as follows:

- actual score *(act. score)* for the test (transformed into a mark out of 10);

- estimated score (*est. score)*, i.e. the number of correct answers the subject thought she had given (also transformed into a mark out of 10);
- comparative assessment of performance (*est. perf.)*, i.e. the score out of 10 attributed to herself by the subject;
- comparative assessment of ability (*est. abil.)*, i.e. the score out of 10 attributed for ability.

We assumed that the data we took a sample from were normally distributed.
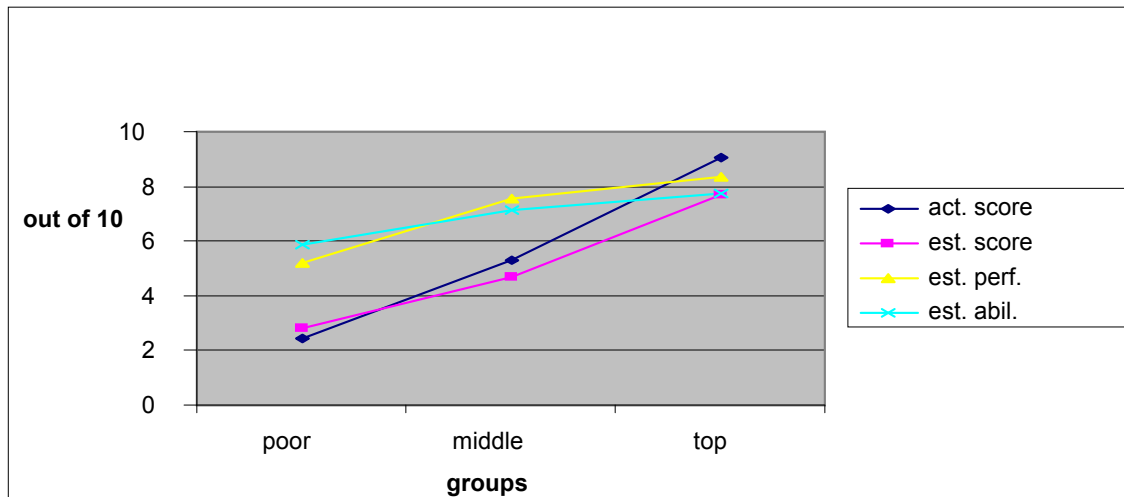
Results

The results of ANOVA [$F(6, 114) = 11.16$; $p < .0000$] showed significant differences among the three groups (poor, average, top performers) for the arithmetic test. (See Table 2). The group of "poor" performers obtained an actual score of $M = 2.42$ ($SD = .60$), out of 10 whilst the self-evaluation score was 5.22 for performance (see Table 1 and Graph. 1) and 5.89 for ability. In the group of "top" performers the actual score was $M = 9.07$ ($SD = .79$) with an average for self-evaluation 8.37 for performance and 7.75 for ability.

Table 1. Average values out of 10 for actual scores, estimated number of correct answers, estimated performance and estimated ability for the "arithmetic task"

	Poor performers M (SD)	Average performers M (SD)	Top performers M (SD)
Actual Score	2.42 (.60)	5.31 (1.33)	9.07 (0.79)
Est. score	2.83 (2.75)	4.71 (2.06)	7.73 (3.09)
Est. perf.	5.22 (2.59)	7.56 (1.21)	8.37 (2.19)
Est. abil.	5.89 (2.52)	7.13 (1.09)	7.75 (1.84)

Table 2. ANOVA: Group (3) x scores (4) for self-assessment of the arithmetic task

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Scores	108,021	3	36,007	24,738	0,000	0,394
Scores*group	97,444	6	16,241	11,158	0,000	0,370
Error (Arithmetic)	165,930	114	1,456			
Intercept	5.797,791	1	5.797,791	628,981	0,000	0,943
Group	407,250	2	203,625	22,090	0,000	0,538
Error	350,275	38	9,218			

Graph 1. Real scores and self-evaluation for the arithmetic task for the three groups.

A second ANOVA was conducted on formal reasoning with group (poor, average, top performers) as independent variable and actual score, estimate score, estimate performance and estimate ability as dependent variables (see Table 4).

The results of ANOVA [$F(6, 123) = 8.42$; $p < .0000$] showed significant differences among the three groups.
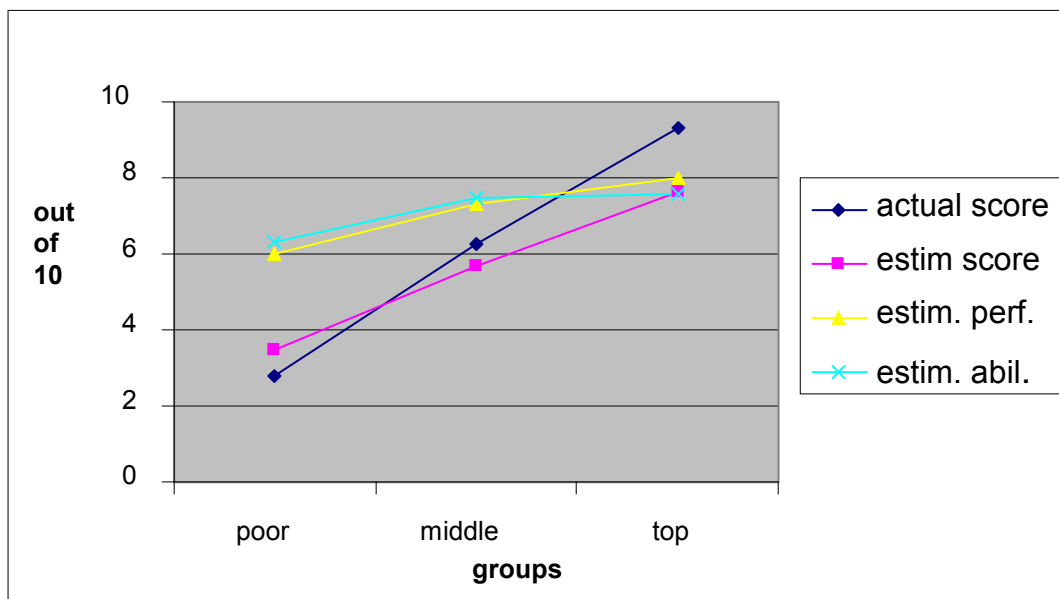
For formal reasoning (see Graph 2), the group of "poor" performers obtained an actual average score, out of 10, of $M = 2.80$ ($SD = .84$), whilst the self-assessment of performance was 6.00 and the self-assessment of ability 6.30. In the "top" performers the average actual score was $M = 9.29$ ($SD = .54$), the average self-assessment of performance 8.00 and the average self-assessment of ability 7.58. (See Table 3 and Graph 2).

Table 3. Average values out of 10 for actual scores, estimated number of correct answers, estimated performance and estimated ability for the "formal task"

	Poor performers M (SD)	Average performers M (SD)	Top performers M (SD)
Actual Score	2.80 (0.84)	6.27 (1.43)	9.29 (0.54)
Est. score	3.50 (2.36)	5.66 (2.90)	7.32 (1.66)
Est. perf.	6.00 (1.33)	7.32(1.25)	8.00 (1.28)
Est. abil.	6.30 (1.49)	7.46 (1.14)	7.58 (1.50)

Table 4. ANOVA: Group (3) x scores (4) for self-assessment of the Formal task

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
scores	66.511	3	22.170	12.113	.000	.228
scores*group	92.509	6	15.418	8.424	.000	.291
Error (Formal)	225.123	123	1.830			
Intercept	6.618.282	1	6.618.282	1,212.136	.000	.967
group	265.117	2	132.558	24.278	.000	.542
Error	223.861	41	5.460			

Graph 2. Actual scores and self-evaluation for the formal task for the three groups

A third ANOVA was significant for the linguistic test [$F(6, 114) = 7.94$; $p <$ .0000] (See Table 6). The group of "poor" performers obtained an actual average score was $M = 2.11$ ($SD = .97$), whilst the self-assessment of performance 5.43 and the self-assessment of ability 6.57. In the "top" performers the average actual score was $M = 8.81$ ($SD = .94$), the average self-assessment of performance was 6.86 and the average self-assessment of ability 7.21 (see Table 3 and Graph 3).

Table 5. Average values out of 10 for actual scores, estimated number of correct answers, estimated performance and estimated ability for the "linguistic task".

	Poor performers *M (SD)*	Average performers *M (SD)*	Top performers *M (SD)*
Actual Score	2.11 (0.97)	5.38 (1.24)	8.81 (0.94)
Est. score	3.59 (3.23)	5.37 (2.83)	6.74 (2.58)
Est. perf.	5.43 (1.90)	6.05 (2.01)	6.86 (1.87)
Est. abil.	6.57 (2.22)	6.15 (1.81)	7.21 (1.89)

Table 6. ANOVA: Group (3) x scores (4) for self-assessment of the Linguistic task

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Partial Eta Squared*
Scores	42,908	3	14,303	6,380	0,000	0,144
Scores*group	106,784	6	17,797	7,938	0,000	0,295
Error (Linguistic)	255,579	114	2,242			
Intercept	4.669,537	1	4.669,537	472,982	0,000	0,926
Group	184,248	2	92,124	9,331	0,001	0,329
Error	375,157	38	9,873			
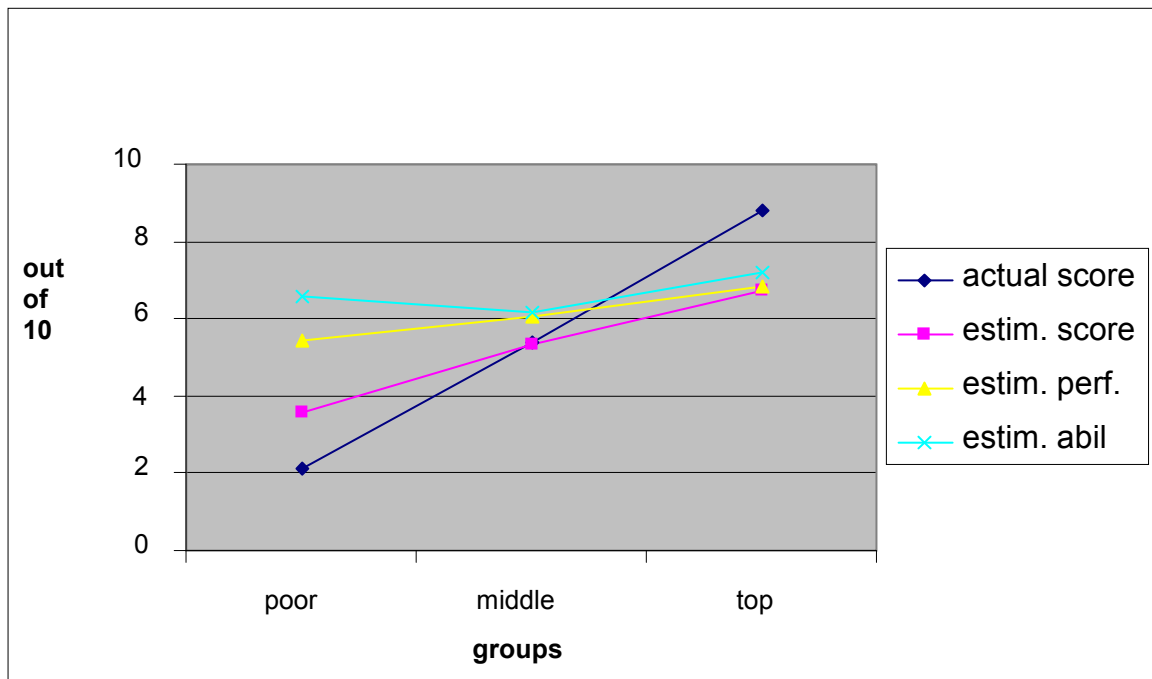
Graph 3. Real scores and self-evaluation for the linguistic task for the three groups.

Finally a post-hoc analysis was conducted using the Tukey method to verify significant differences among groups for the ability to estimate the number of correct answers in the three tasks (see Table 7). Analysis showed that in the highly skilled group the estimated number of correct answers was always less than the actual number of correct answers and this difference was significant in the linguistic task. In this group there is also a significant trend for the arithmetic and formal task.

Table 7. Significance of differences between actual scores and estimated scores (Tukey test).

Test	Poor	Average	Top
Arithmetic	.99	.96	.08
Formal	.99	.94	.11
Linguistic	.79	.99	.01 *

*. Post hoc differences are significant at the 0.05 level

Discussion

In our study we found that self-assessment of the number of correct answers (estimated score) differed between the above average, average and below average performers.

In general there was an increasing numerical difference between the actual score and the average self-evaluated score, which was smallest for the estimate of the number of correct answers and largest for the estimate of ability. This showed that subjects were accurate when assessing the number of correct answers in a test, but they were increasingly unskilled when comparing themselves with their peers.

The group of poor performers, which provided a very low number of correct answers, and were aware of the fact, when asked to provide comparative evaluations of performance and ability, overestimated its own abilities.

Top performers were the opposite, underestimating themselves in relation to others. Their self-evaluation of number of correct answers coincided almost perfectly with the comparative evaluation of performance and ability.

It can therefore be concluded that subjects were fairly accurate self-assessors. However, this accuracy in terms of performance and evaluation was not perfect and it was in the inaccuracy that the phenomenon under investigation was revealed.

Conclusions

In this manuscript we examined the capacity for self-evaluation of University students. We intended to verify the hypothesis that subjects less skilled in cognitive tasks tend to overestimate themselves compared to their peers and that more skilled subjects, on the other hand, tend to underestimate themselves.

The results demonstrated that all the subjects in all tasks showed good awareness of their level of actual performance. Analyzing comparative assessments we found that poor performers tend to significantly overestimate their own performance whilst top performers tend to underestimate it.

We found also an increasing numerical difference between the actual score and the average self-evaluated score, which was smallest for the estimate of the number of correct answers and largest for the estimate of ability.

Even within the comparative evaluations, there was an important difference: the evaluation of performance, in the specific test, was presumably very influenced by the feedback concerning the test: the subject knew if he/she has given the right answer to each question. The more general evaluation of ability for that type of test seems to reflect more self-image, irrespective of the test carried out.

To formulate an explanatory hypothesis, we could begin with one fact (which was also observed in the second study carried out by Kruger and Dunning in 1999): in the poor performers, the estimate of correct answers ("estimated score" in the graphs) was very close to the actual number of correct answers ("actual score" in the graphs).

This means that the poor performers were well aware of how few questions they had got right. The discrepancy between self-evaluation and actual performance emerged only in the comparative evaluations, a metacognitive operation based on an uncertain, and essentially fictional, reference group. Comparative evaluation obliged subjects to refer their self-evaluation to an average level of performance that they did not and could not know, and this lack of any concrete data allowed them to fall back on defence mechanisms to safeguard their self-image; the lack of determination gave them room to use highly subjective criteria of self-evaluation. It's a bit like saying: "I didn't do the test well but I didn't do any worse than most other people". This leads to a kind of optimism in self-evaluation reinforcing one's self-image and seems to be centred more on the person than on the task. What comes to the fore is a *self-focused* defence mechanism which seems to correspond to the heuristic *better than average*, the general tendency to overestimate oneself compared to the average. In reality, in our opinion, it seems more that poor performers assessed average performance on the basis of their own performance, and hence underestimated it.

On the other hand in the top performers group the estimated number of correct answers was always less than the actual number of correct answers and this difference was significant in the linguistic task. In this group there is also a significant trend for the arithmetic and formal task. This might be due to the expression of particularly rigorous and strict epistemic motivations: these subjects performed extremely well but also doubted that they performed so well: a sort of "methical doubt"? This

particular metacognitive style, expressed in the self-assessments of top performers could be correlated with the level that Mason (2001), citing Kitchener (1983) and Kuhn (1999), indicated as the third "epistemic" level, above the cognitive and metacognitive levels.

A further contribution to the interpretation of data may be provided by the motivational theories of Dweck (1988; 1999) and Moè and De Beni (1995). The two motivational styles, focused on "learning - and mastery-oriented" versus "performance-oriented", seem to match to the behaviours we observed in the poor and top performers. Motivation focused on performance involves the need to protect one's self-image from the possibility of failing, which is precisely what happened in the poorly performing group. On the other hand, the top performers, who underestimated their performance and ability, seem to be more focused on the margin of error and hence more interested and motivated by the possibility of improving themselves (De Beni & Moè, 2000).

A more general way of looking at the phenomenon could start with the consideration that cognitive and metacognitive processes are regulated by highly subjective representations of oneself and the world around us.

Nisbett and Ross (1980) dealt with these matters at the crossover of "hot cognition" (in which "errors" are explained by emotional and motivational dispositions) versus "cold cognition" (in which errors are the result of mistakes in processing information), and were led "to confess a prejudice on our part [...] that errors of inference and judgement originate not from motivational factors but from perception and cognitive factors" (Nisbett and Ross, 1980, p. 46).

Examining the phenomenon of "self-overestimation" and "self-underestimation" respectively in poor and top performers, we confess an opposite prejudice. We believe we have found some data supporting the "hot cognition" hypothesis. The evident functional and motivational significance of the phenomenon of overestimation indicate that explanations are to be sought in the safeguarding of the self-image.

It is also clear, however, that the phenomenon requires further extensive investigation of the variables and context to clarify the real forces in play.

First of all a larger and more representative sample would be necessary in order to confirm the results also in the Italian population.

Second, there is a possibility that attributional processes play a role, linked to the nature of the task (easy vs. difficult), as well as personality variables such as those discussed above in relation to motivational systems (performance vs. mastery) and *locus of control* (internal vs. external). Finally, of particular significance, from various points of view including applications, may be evolutionary-genetic research of the phenomenon to study how it begins and develops in children.

• • •

Piergiorgio Battistelli is full professor in General Psychology at the University of Bologna. His research interests are in children's theory of mind, metarepresentative thought, metacognition and various topics of Social and Political Psychology (representation of terrorism, social identity theory).

Alessia Cadamuro, PhD in General and Experimental Psychology, senior lecturer in Developmental Psychology at School of Education, University of Modena and Reggio Emilia. Her research interest is in Educational Psychology, Metacognition, Cognitive and Learning Styles, Communication and Developmental Psychology.

Alessandra Farneti is full professor in Developmental Psychology at School of Education, University of Bolzano. Her research interest is in Educational Psychology, familiar dynamics and grandparenthood, metacognition, social animation by clowning.

Annalisa Versari is a PhD student in General and Experimental Psychology currently working at the Department of Psychology of the University of Bologna. Her research interest is in Theory of Mind, Metacognition and self evaluation processes.

References

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Student's learning strategies and motivation processes. *Journal of Educational Psychology*, *80* (3), 260-267.

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

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: Freman. Borkowski, J. G. (1996). Metacognition: Theory or chapter heading? *Learning and Individual Differences. 8* (4), 391-402.

Borkowski, J. G. (1985). Sign of intelligence: Strategy generalization and metacognition. In S. R. Yussen (Ed.), *The growth of reflection in children*, New York: Academic Press.

Borkowski, J. G., Chan, L. K. S. & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In G. Schraw (Ed.), *Issues in the measurement of metacognition* (pp. 1-41). Lincoln, NB: Univ. of Nebraska Press.

Borkowski, J. G., & Muthukrishna, N. (1995). Learning environments and skill generalization: How contexts facilitate regulatory processes and efficacy beliefs. In F. E. Weinert & W. Schneider (Eds.), Memory performance and competencies: Issues in growth and development. (pp. 283-300). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanism. In F. Weinert & R. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 65-116). Hillsdale, NJ: Erlbaum Associates.

Brown, A. L., & DeLoache, J. (1978). Skills, plans, and self-regulation. In R. S. Siegler (Ed.), Children's thinking: What develops (pp. 3- 35). Hillsdale, NJ: Erlbaum.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. Journal of Personality and Social Psychology, 90, 60-77.

Cadamuro, A. (2004). Stili cognitivi e stili di apprendimento: da quello che pensi a come lo pensi. (Cognitive styles and learning styles: from what you think to how you think). Roma: Carrocci Editore.

Calonghi, L., Polacek, K., & Ronco, A., (1974), Test di Struttura dell'Intelligenza TSI di Rudolf Amthauer Manuale. (Intelligence Structure Test). Firenze: Organizzazioni Speciali.

Campione, J.C., & Brown, A. L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. Intelligence. 2(3), 279-304.

Cornoldi, C. (1995). Metacognizione e apprendimento. (Metacognition and learning). Bologna: Il Mulino.

De Beni, R., & Moè, A. (2000). Motivazione e apprendimento. (Motivation and learning). Bologna: Il Mulino.

De Beni, R., & Pazzaglia, F. (1991). Lettura e metacognizione. (Reading and metacognition). Trento: Erickson.

Dunning, D., Johnson K., Ehrlinger, J. & Kruger, J. (2003). Why people fail to recognize their own incompetence. Current Directions in Psychological Science, 3, 83-87.

Dweck, C. S. (1999). Self theories: Their role in motivation, personality, and development. Philadelphia, PA: Psychology Press/Taylor and Francis.

Dweck, C. S., & Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. Psychological Review, 95, 256–273.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). "Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent". Organizational Behavior and Human Decision Processes 105, 98–121.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In B. Resnichl (Ed.), The nature of intelligence (pp.231-236). Hillsdale, NJ: Lawrence Erlbaum.

Flavell, J. H. (1981). Monitoring social cognitive enterprises: something else that may develop in the area of social cognition. In J. H. Flavell & L. Ross (Eds). Social cognitive development (pp. 272-287). Cambridge: Cambridge University Press.

Frieze, I. H, Bar Tal, D. (1980). Developmental trends in cue utilization for attributional judgments. Journal of Applied Developmental Psychology. 1 (2), 83-94.

Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (1998). TINV. Test di intelligenza non verbale (Non verbal intelligence scale). Trento: Erickson.

Heider, F. (1958). The psychology of interpersonal relations. Oxford, England: John Wiley.

Hultsch, D. F., Herzog, C., Dixon, R. A., & Davidson, H. (1988). Memory self-knowledge and self-efficacy in the aged. In M. L. Howe & C. J. Brainered (Eds.), Cognitive development in adulthood. New York: Springer-Verlag.

Izaute, M., & Chambres, P. (2002). Metacognition: process, function and use. Dodrecht, Netherlands: Kluwer Academic Publisher.

Kitchener, K.S. (1983). Cognition, metacognition, and epistemic cognition. A three-level model of cognitive processing. Human Development, 26 (4), 222-232.

Kuhn, D. (1999). A developmental model of critical thinking. Educational Researcher, 28 (2), 16-26.

Krueger, J., & Mueller, R. A., (2002). Unskilled, unaware or both? The contribution of social-perceptual skills and statistical regression to self-enhancement biases. Journal of Personality and Social Psychology, 82, 180-188.

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

Mason, L.(2001). Verità e Certezze. (Truth and certainties). Roma: Carocci.

Mazzoni, G. (2000). L'apprendimento: comportamento, processi cognitivi, neurobiologia. (Learning: behaviour, cognitive processes, neurobiology). Roma: Carocci Editore.

Mazzoni, G., & Nelson, T. O. (1998). Metacognition and cognitive neuropsychology. Monitoring and control processes. Mahwah, NJ: Lawrence Erlbaum.

Moè A., & De Beni R. (1995). Metodo di studio: dalle strategie ai programmi metacognitivi. In O. Albanese, P. A. Doudin & D. Martin (Eds), Metacognizione ed educazione

(pp.121-140). (Study methods: from strategies to metacognitive programmes. In Metacognition and education). Milano: Franco Angeli.

Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and Shortcomings of Social Judgment. Prentice Hall, NJ: Englewood Cliffs.

Pearson N. & Wiederholt L. (1998). Test di intelligenza non verbale TINV. (Non verbal intelligence scale). Firenze: Organizzazioni Speciali.

Petter, G. (1992). Dall'infanzia alla preadolescenza. (From infancy to pre-adolescence). Firenze: Giunti.

Piaget, J. (1974). Reussir et comprendre. (Succeeding and understanding). Paris : PUF.

Piaget J. (1975). L'équilibration des structures cognitives: problème central du development. Etudes d'épistémologie génétique, vol. 33. (Cognitive structure equilibration: a central problem of development). Paris: Presses Universitaires de France.

Richard, J. F. (1990). Les activitès mentales: comprendre, raisonner, trouver des solutions. Paris : Colin.

Riviére, A. (1999). Teoria della mente e metarappresentazione. In F. Braga Illa (Ed.), Processi di rappresentazione, soggetto e società. (Theory of mind and metarepresentation. Processes of representation, subject and society). Urbino: Edizioni Quattro Venti.

Schwartz, B., & Perfect, T.J (2002). Applied metacognition. New York: Cambridge University Press.

Vygotskij, L. S. (1978). Mind in society. Cambridge, Ma: Harvard University Press.

Weiner, B., Frieze, I.H., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R.M. (1978). Perceiving the causes of success and failure. Morristown NJ: General Learning Press.

Weiner, B. (1986). Attribution, emotion, and action. In: R. M. Sorrentino, & E. T. Higgins (Eds), Handbook of motivation and cognition: Foundations of social behaviour. (pp. 281-312). New York, NY, US: Guilford Press.