Working together
www.rcis.ro

Revista de Cercetare si Interventie Sociala

ISSN: 1583-3410 (print), ISSN: 1584-5397 (electronic)

INACCURACY AND OVERCONFIDENCE IN METACOGNITIVE MONITORING OF UNIVERSITY STUDENTS

Florin Vasile FRUMOS, Silviu-Petru GRECU

Revista de cercetare și intervenție socială, 2019, vol. 66, pp. 298-314

https://doi.org/10.33788/rcis.66.16

Published by:
Expert Projects Publishing House

expert projects
publishing

On behalf of:
„Alexandru Ioan Cuza" University,
Department of Sociology and Social Work
and
HoltIS Association

REVISTA DE CERCETARE SI INTERVENTIE SOCIALA
is indexed by Clarivate Analytics (Web of Science) -
Social Sciences Citation Index
(Sociology and Social Work Domains)

Inaccuracy and Overconfidence in Metacognitive Monitoring of University Students

Florin Vasile FRUMOS[1], Silviu-Petru GRECU[2]

Abstract

The aim of the paper is to stress the relationship between individual metacognitive accuracy and academic performance. Moreover, we tested the relationship between the accuracy, items difficulty and bias score and exam results. Metacognitive monitoring calibration of 100 university students was tested in exam settings, using postdated confidence judgments. Absolute local and global accuracy and total bias score were related with test performance, difficulty and types of the items: multiple choice and open ended items. The most important results show local and local inaccuracy or overconfidence, but also an unexpected greater accuracy on low performing subjects compared with there's more performing counterparts. The open ended items low, but positively correlate with metacognitive monitoring inaccuracy, both local and global, a possible illustration of the hard-easy effect. The bias score is globally negatively related with performance, but the more the subjects respond to the difficult open-ended items, the lower the bias in self-appreciation. We conclude that there is a global expected relationship between test results and both accuracy and bias score. Also, particular results show a more nuanced relationship between local and global accuracy, items difficulty and type and bias score. Some theoretical and methodological issues are discussed and futures research direction is proposed.

Keywords: metacognition, accuracy, overconfidence, overestimation, university students.

---

[1] Alexandru Ioan Cuza University of Iasi, ROMANIA. E-mail: frumos@uaic.ro

[2] Alexandru Ioan Cuza University of Iasi, ROMANIA. E-mail: silviu.grecu@uaic.ro

Introduction

Although metacognition is a tempting and promising area of research in learning psychology, is quite difficult to perceive its conceptual complexity. The hidden part of this construct implies the lack of accuracy of metacognitive judgments, namely the overconfidence or the underconfidence of the learners when they make estimates of their learning. Accuracy represents a measure of metacognitive monitoring process, an important metacognitive ability which refers to the awareness in two different cases: comprehension awareness and performance awareness, both during the process of performing a particular cognitive task (Nietfeld, Cao & Osborne, 2005). Tracking own cognitive processes is tough, but important for control processes, i.e. corrective or ameliorative intervention of ongoing cognitive actions. This difficulty is reflected in a high frequency of inaccurate judgments of one's demarche (Hacker, Bol & Horgan & Rakow, 2000). Monitoring of the own cognitive processes seems to have a crucial role into learning improvement, because monitoring operates right to the ongoing cognitive processing or right after they have ended (Grimes, 2002; Hacker, Bol & Horgan & Rakow, 2000).

Schraw, Wise, and Rose (2000) notice that metacognitive monitoring differs from metacognitive control. Metacognitive regulatory processess refers to one's supervising their own thinking processes and applying appropriate regulatory strategies, if and when it is necessary. Therefore, if metacognitive monitoring refers to processes that occur during or after a learning activity and provide information about the effectiveness of those activities, metacognitive control refers to regulatory processes that occur prior to or during a learning activity that direct the course of cognitive activities. Monitoring is however important because it provides self-generated feedback to the control system. Innacurate monitoring may affect the control of one's performance so it may be even impossible (Schraw, Wise & Rose, 2000).

The accuracy of metacognitive judgment is critical for learning and successfully problem solving (Miller & Geraci, 2014). Into some early studies on metacognition (Flavell, 1979), young children were asked to study a list of words until they believed they had memorized them; although they believe memorized the words, actually had not performed very well in the memory test, so they are overconfident on their abilities.

Monitoring is more or less accurate, depending on how one's evaluation is far away from the actual performance (Nietfeld, Cao & Osborne, 2005). Accuracy of metacognitive monitoring affects for example learning of texts (Thiede, Anderson &Therriault, 2003). Strengthening the accuracy means to enhance the calibration of metacognitive monitoring. This is important because, in a general sense, only an enough accurate metacognition may serve as a good starting point for a relaible metacognitive control and, in a broader sens, for understanding and evaluating human cognition (Cheng, 2010). Therefore, metacognitive monitoring accuracy or calibration is a relevant indicator of metacognitive monitoring and is expressed as the overlapping of one's self-assessment and actual performance involved in a cognitive task. The more the subjective self-assessment fit with an objective evaluation, the more calibrated metacognitive monitoring. Monitoring accuracy are critical for learning and memory, offering possibilities for translating laboratory findings in educational real world (Koriat, 2011).

The relationship between metacognitive monitoring and test performance are also important because some particularities of test such as items difficulty or items type – pass-fail or grade tests – may relate with monitoring accuracy (Barenberg & Dutke, 2013). Accuracy of metacognitive judgements is rare also in adults (Pressley, Yokoi & Van Meter & Van Etten & Freebern, 1997), and often poor academic results correlate with overconfidence and underconfidence (Hacker et. al, 2000) and represents a judgements bias. Some prior studies on metacognitive monitoring accuracy shed some light on this metacognitive process. Nietfeld, Cao, and Osborne (2005) noticed global monitoring are more accurate than local monitoring. Also, they reported that local and global monitoring are different qualities of monitoring ability. These results and interpretation are anyway too general and do not explain global and local monitoring processes. Another limit of their research is the small sample of subjects, which may affect statistical power of theirs results.

Hacker *et al.*, (2000) studied metacognitive monitoring ability in students using predictions and postdictions over three exams. However, they report only global monitoring, without pursuing local monitoring ability of the subjects. Schraw's study (1994), although significantly clarify the relationship between local and global monitoring, was conducted in laboratory. That affect the ecologic validity of the results.

Methodology

Our purpose for this study is to examine how local and global monitoring accuracy are related with academic performance into a testing situation, and how the typology of the items and their difficulty are related with monitoring accuracy.

Specifically, we followed three objectives: (1) To investigate in what ways absolute local accuracy and absolute global accuracy are related with exam results, because other studies indicates that achievement level is related with calibration accuracy (Hacker *et al.*, 2000; Nietfeld *et al.*, 2005); To explore if item difficulty and type is related with absolute local accuracy and absolute global accuracy, that because task difficulty may influence calibration accuracy (Bol & Hacker, 2012); (3) To examine the relationship between bias score and academic performance, and relationship between bias score and local and global accuracy. The present study follows one of recommendations of Bol and Hacker (2012) to enhance the research on what extent does calibration accuracy predict achievement. This study may be useful because in the last years only few empirical researches were conducted for exploring relationship between metacognitive accuracy and performance. Accuracy (metacognitive accuracy, metacognitive sensitivity) is one of the key concepts of this study. Accuracy represent the degree of fit or discrepancy between one's judgment of their own future or past performance and their actual cognitive performance.

Metacognitive monitoring accuracy was operationally defined in various ways: as feeling of knowing (FoK), a subjective experience of knowing, or the degree to which one has access to previously learned information in memory; as confidence judgments, a estimation of performance on a future or for an already passed test; ease of learning (judgments of encoding perceived difficulty); judgments of learning (i.e., the degree to which information was learned during the study phase), and performance judgments (i.e., assessments of performance accuracy) (Schraw, Wise & Rose, 2000). It is also expressed as the degree of realism of judgments of confidence that subject makes to their own cognitions (Buratti & Alwood, 2012).

In the present study we operationalized metacognitive accuracy as confidence judgment, which are postdicted estimations of correctness of answers to each items of a test (local or partial accuracy) and to the whole test (global accuracy), respectively. The confidence judgments of participants was compared with their actual test score for determining the accuracy of their metacognitive monitoring (Dinsmore & Parkinson, 2013). We must note that confidence judgments are a somewhat unclear measure of accuracy, because is not evident if subjects make a "guess" of theirs future or past performance – a cognitive operation of estimation – or rather they indicate a level of confidence in their own estimation – a different cognitive operation regarding the estimation itself. This ambiguity was observed by Miller and Geraci (2011) when they try to explain contradictory results of metacognitively unskilled, but somewhat aware subjects. They examined subjective confidence associated with predictions of performance and defined two kind of overconfidence: functional overconfidence (the „guess" described above) and subjective overconfidence: overcertainty: „errors of overestimating one's ability (predicting that one will perform better than one does) as functional overconfidence and errors of overcertainty (being overly certain of one's predictions) as subjective overconfidence" (Miller & Geraci, 2011: 2).

The accuracy of judgements on performance may be splitted in two different measures, namely calibration and resolution. Both contribute to the overall "accuracy" of probability judgments. Calibration or absolute accuracy is defined in terms of whether the predicted value assigned to an item is followed by the occurrence of that value on the criterion test or "the goodness of fit between probability assessments and the corresponding proportion of correct responses" (Fleming & Lau, 2014: 6). In other words, calibration refers to the "correspondence between mean metacognitive judgments and mean actual memory performance, and reflects the extent to which metacognitive judgments are realistic" (Koriat, 2007: 29). A good overlap of estimated and actual performance reflect a calibrated monitoring process, namely a good abolute metacognitive accuracy.

On the other hand, resolution is a measure of the variance of the probability assessments, measuring the extent to which correct and incorrect answers are assigned to different categories (Fleming & Lau, 2014: 6). For example, the relative accuracy or resolution of metamemory is a measure of how sensitive a participant is to the differential recallability between two studied items (Ruiz & Aroyo, 2016). In other words, if calibration (absolute accuracy) reflect the degree of global overlap between the estimated performance and the actual performance, the relative accuracy or resolution reflect the subject's power to discern between his or her good and bad responses. As Fleming and Lau note, resolution is a measure of variance and, accordingly, "a larger variance is better, reflecting the observer's ability to place correct and incorrect judgments in distinct probability categories" (Fleming & Lau, 2014: 14). Redford, Thiede, Wiley, and Griffin (2012) studied resolution as relative accuracy, arguing that concept mapping generate cues for more accurate comprehension judgments.

In this study, we used confidence judgments for measuring only absolute accuracy, summing confidence judgment scores item by item (local or partial accuracy) and for the whole test (global accuracy). Metacognitive monitoring calibration differs according to the moment of cognitive task. If we make estimation before solving the task is about calibration of comprehension (prediction), and if we make after the task is calibration of performance (postdiction). (Glenberg & Epstein, 1987). Calibration of performance seems to be more reliable than calibration of comprehension, probably because the subject already have processed the task (Glenberg & Epstein, 1985). In our study the subjects made confidence judgments after each item and after the test, therefore they will display a calibration of performance.

The second key concept analyzed in this study is the metacognitive bias, e.g. the overall level of confidence expressed by a person into their metacognitive judgment. According with Fleming and Lau (2014), "it is important to distinguish this two aspects, namely sensitivity and bias". If "metacognitive sensitivity is also known as metacognitive accuracy, type 2 sensitivity, discrimination, reliability, […] metacognitive bias is also known as type 2 bias, over- or underconfidence or calibration" (Fleming & Lau, 2014: 14).

Metacognitive bias is the overall level of confidence expressed, independent of whether the trial is correct or incorrect (Fleming & Lau, 2014: 2). That means bias somewhat indirectly reflect the confidence of the subjects in their estimates. Overestimation may reflect overconfidence, and underestimation may reflect underconfidence, but is difficult to interpret a correct estimation without an explicit measure of the confidence per se (see Discutions section). In our study, we calculated metacognitive bias for global confidence judgments, as a measure of marked and signed discrepancy between estimations and results.

Participants

Because some previous research on metacognitive accuracy reported differences between results from laboratory and classroom setings (McCormick, 2003), we choose to conduct our research in „natural" settings, in a real exam conditions. In this research, we explored metacognitive monitoring accuracy to the final/ semester exams on Pedagogy for the student following the teacher's training program. Participants: 100 undergraduates, 14 male and 86 female, in the second year of university studies, from three faculties: Biology (40 sujbects), Economy and Bussiness Administration (11 subjects) and Philosophy and Social-Political Sciences (49 subjects) of an European university. All subjects received a small grade point for participation and they have been assured for the confidentiality of responses.

Instruments and procedure

Data were colected by applying a 14 items test on Pedagogy, that includes ten multiple choice items, each with four alternatives and with just one correct or optimal answer, and four open-ended items, wich require a constructed and written respons. After solving each item, subjets scored postdicted confidence judgments on a scale form 0 to 100, with intervals marked from 10 to 10, reflecting the estimated degree of corectness of their solutions. The same scale was used after they finished the test for a single confidence judgment on the whole test (for computing the global accuracy). The task for the subjects, reflecting theirs confidence judgment, after each item and after the whole test, was: "Mark with X on the scale to what extent you think you gave the correct answer".

We calculated a Alpha Crombach reliability analysis on the whole test and obtained 0.865. Confidence judgments we used are usually measured on Likert scales, and some researchers suggest to use 0-100 scales, considered more sensibles than others scales. (Nietfeld & Schraw, 2002). The procedure we used in the research allowed calculation of the three important indicators: absolute local (or partial) accuracy score (on the 14 items of the test), the global accuracy score (on the whole test) and the bias score. Then, calibration of metacognitive judgments is reflected by the two scores, namely local or partial accuracy score and global accuracy score, deriving from a single global judgment on the test as a whole (Schraw, 1994; Hacker et. al, 2000).

Accuracy, the first key concept in the present research are therefore dual: the couple of indicators of accuracy: absolute local accuracy and global accuracy. Local or partial accuracy reflect the online metacognitive regulation of the subjects (Nietfeld, Cao & Osborne, 2005). We compute this indicator as the the difference between the mean of confidence judgement for the test items and the total score on the test, divided by 100. Scores may range between -1 and 1, where 0 represents perfect accuracy, and -1 and 1 represent total lack of accuracy. The scores near zero indicate high metacognitive accuracy and a good calibration of metacognitive monitoring.

The second indicator, global metacognitive accuracy score was calculated directly as the ratio between the global confidence judgment (final confidence judgment) and the test score. The reference value is 1, and a score near 1 reflect a good global metacognitive calibration. Result above 1 reflect overestimation, and lower than 1 underestimation.

The third indicator of metacognitive accuracy is bias score, the the magnitude of discrepancy between estimation and reality, reflecting overconfidence or underconfidence. The more the score is further on 0, the more biased it is (Nietfeld, Cao & Osborne, 2005).

A significant bias parameter is the bias direction of the results, which refers to the sign of the difference, + or -. Positive biased scores indicate overconfidence, and negative biases indicate underconfidence. Most frequently, the subjects are overconfident, overoptimistic regardind theirs performances (Schraw & Roedel, 1994).

Results

The main purpose of our study was to investigate how type and difficulty of items in an exam situation are related with monitoring accuracy and academic performance.

For the first objective of our study, to investigate how absolute local accuracy and absolute global accuracy are related with exam results we analyse the main variables: (absolute) local accuracy, (absolute) global accuracy and total test score. Supplementary, we examine a repartition of the subjects in quartiles following their test results.

Mean of local accuracy for the subjects is 0,22. This mean is, however, significantly different from perfect accuracy (zero): $t(85) = 11.55$, $p<0.01$, meaning that subjects overestimates theirs performance. When the participant are splitted in quartiles according with theirs results to test, the subjects from the first quartile ($q1=35$) are much more accurate in local confidence judgments: $t(20)=26.87$, $p<0.001$ than the subjects from the last quartile ($q3=51,25$), $t(24)=16.40$, $p<0.001$.

The mean for global accuracy for all the subjects are 1,55. The t test for global accuracy is significantly different from perfect accuracy (value 1): $t(99)=9.38$, $p<0.01$, showing overestimation of performance for all the subjects. The subjects from the first quartile ($q1=1,15$) according with theirs results, less overestimates: $t(21)=15.07$, $p<0.001$ than theirs pairs from the last quartile ($q3=1,89$): $t(27)=6,43$, $p<0.001$. We noticed that the less performing students are, however, more accurate, both locally and globaly, than theirs counterparts.

Table 1: Descriptives of local accuracy, global accuracy and total test score

		total test score	Local/partial accuracy	Global accuracy
N	Valid	100	86	100
	Missing	0	14	0
Mean		42,53	**,2202**	**1,5586**
Median		42,00	,2286	1,5016
Mode		30	,20[a]	1,11
Std. Deviation		11,595	,17666	,66601
Minimum		13	-,38	,16
Maximum		69	,55	4,23
Percentiles	25	35,00	**,1182**	**1,1592**
	50	42,00	,2286	1,5016
	75	51,25	**,3500**	**1,8929**
a. Multiple modes exist. The smallest value is shown				

Table 2: Correlation between absolute local/partial accuracy, global accuracy and total test score

Correlations		Local/partial accuracy	Global accuracy	Total test score
Local/partial accuracy	Pearson Correlation	1	,753**	-,526**
	Sig. (2-tailed)		,000	,000
	N	86	86	86
Global accuracy	Pearson Correlation	,753**	1	-,685**
	Sig. (2-tailed)	,000		,000
	N	86	100	100
Total test score	Pearson Correlation	-,526**	-,685**	1
	Sig. (2-tailed)	,000	,000	
	N	86	100	100
**. Correlation is significant at the 0.01 level (2-tailed).				

Correlation matrix between the main variables local accuracy, global accuracy and total test score show, as expected, a strong positive corelation of global and local accuracy: $r=0.753$, $p<0.01$

Each type of accuracy moderately and negative correlate with total test score, that means the higher the test score, the lower (near to zero) the value of accuracy (local: $r=-0526$, $p<0.01$ and global: $r=-685$, $p<0.01$). We highlited that lower scores of accuracy means greater accuracy and greater scores on accuracy means inaccuracy (miscalibration), respectively.

For testing predictive value of absolute global accuracy for the test results, we used a regression model (Freedman, 2009).

Table 3: Regression model for global accuracy and test results

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
,737	,543	,539	,204
The independent variable is global accuracy.			

Table 4: Analyse of variance between global accuracy and test results

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	4,859	1	4,859	116,630	,000
Residual	4,083	98	,042		
Total	8,942	99			
The independent variable is global accuracy					

We notice an important relation between global accuracy and test results: R^2=0,543, F=116,63, p<0.001. This regression model has a good predictive value, 73,7% of test results variance are predicted by global accuracy score. The relationship between global accuracy and test result can be observed in the exponential function graph: the higher the test scores, the more global accuracy value is lower, near the reference value 1 (perfect accuracy).

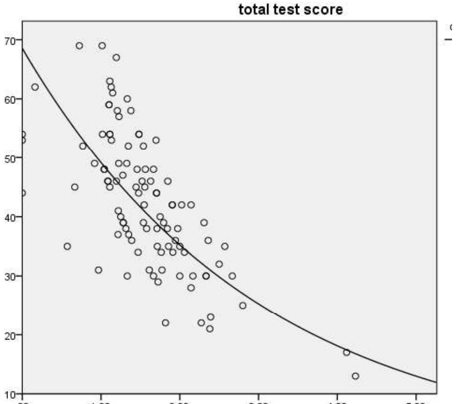

Figure 1: Relationship between total test score and global accuracy

For testing predictive value of local (partial) accuracy for the test results, we used again a regression model (Freedman, 2009).

Table 5: Regression model for local accuracy and test results

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
,526	,276	,268	9,980
The independent variable is local/partial accuracy.			

We notice a weak relation between local/partial accuracy and test results: R^2=0,276, F=32,050, p<0.001. This regression model has a moderate predictive value, 52,6% of test results variance are predicted by local/partial accuracy score.

Table 6: Analyse of variance between local/partial accuracy and test results

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Regression	3192,181	1	3192,181	32,050	,000
Residual	8366,342	84	99,599		
Total	11558,523	85			
The independent variable is local/partial accuracy.					

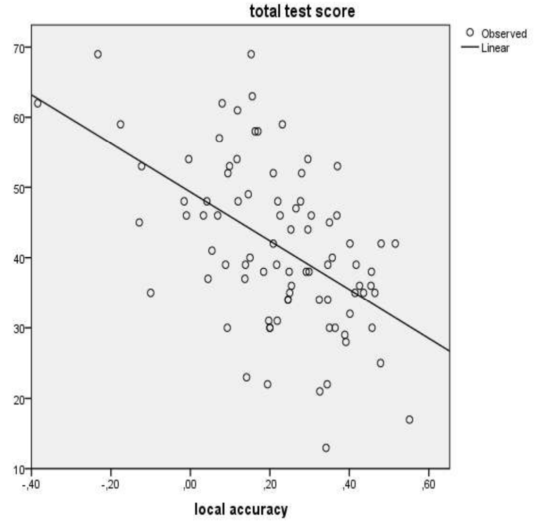

Figure 2: Relationship between total test score and local accuracy

We observe a moderate negative asscociation between local accuracy and total test score: R^2=0,276, F=32,05, p<0.01. In this case, the relation is liniar (graph 2): the higher the test scores, the more the value of absolute local (partial) accuracy is getting closer on value 0 (perfect accuracy). We highlited again that lower scores on global or local accuracy reflect greater accuracy.

Because we was interessted if item difficulty and type are related with absolute local accuracy and absolute global accuracy, we tested how item's difficulty is associated with local and global accuracy; we divided multiple choice items and open ended items into easy and difficult items, respectively.

We determined first the difficulty of the items as follows: we divided all the items into easy or difficult items, according to the frequency of the correct answers provided by the subjects, then we calculated for each subject the total number of easy and difficult solved items.

For the ten multiple choice items we split the items in easy and difficult using mediane method. We obtained six difficult items and four easy items. For the four open ended items, we use mediane and also the skewness. The two difficult open ended items have mediane 4 and skewness 0,392, and median 6 and skewness -0,103 respectively. The two easy open-ended items have mediane 7,72 and skewmess -1,12, and mediane 6,96 and skewness -1,18.

We calculated then the correlation between local and global accuracy and difficult and easy, multiple choice and open-ended items. Relevant results are presented in *Table 7*.

Table 7: Correlation between difficult multiple choice or open ended items and local accuracy or global accuracy

	Local accuracy	Global accuracy	Results
Multiple choice difficult items	r=-0,169, p=0,2	r=-0,208, p=0,053	Poor negative or no correlations
Open ended difficult items	r= 0,251, p<0.01*	r=0,226, p<0.01*	Low positive correlations

For the multiple choice items we obtained only a negative poor and marginal correlation for global accuracy (r=-0,208, p=0,053). When we calculated correlation for open ended difficult items, we obtained low positive correlation both with local and global accuracy; so, greater results to open ended difficult items correlates with increased values of innacuracy, that seems to support the hard-easy effect (Juslin, Winman & Olsson, 2000).

To determine the relationship between the bias score (calculated as the difference between the average confidence and the average performance for the test result) and the exam results, we calculated first, for all subjects and for all test items, the Spearman's rank correlation coefficient and we obtained ρ = -0.532, p <0.01, i.e. the higher the test results, the lower the bias. We tested also the relationship between the bias score and the two forms of accuracy. We obtained a Spearman's rank correlation coefficient ρ=0.844, p<0.001 between bias and local accuracy, ρ=0.960, p<0.001 between bias and global accuracy. That means the innacurated students are overconfident (Kruger & Dunning, 1999) and the accurate subjects seems to be more realist regarding theirs test performance. We also calculated correlations between the bias score and the type and difficulty of the items. We obtained a single weak negative correlation between the bias score and the difficult open-ended items, r = -0.193, p <0.05, i.e. the more the subjects respond to the difficult open-ended items, the lower the bias in self-appreciation. In all other cases we did not obtain any significant correlations.

Discussion

In this study we try to find out on what extent calibration accuracy predicts tests achievement on university students. In the last years only few studies were conducted in this topic, but the question of relationship between metacognitive monitoring and academic results remain under debate.

The first question of study was about the relationship of accuracy and performance. The two forms of accuracy, local (partial) and global accuracy are, both, inversely correlated with performance, the subjects showing, as expected, overconfidence both on local and global confidence judgments. The overconfidence of judgments may relate with the ”weighing the evidence” mechanism described by Griffin and Tversky (1992): people are overconfident when strength of evidence is high and weight is low. In this case, the students may overestimate their results: they respond to many items, but different items have different weight as score, per item and on the whole test.

We find out here another interesting result: the subject from the first quartile, the weakest, are, paradoxically, more accurate in their local and global judgments, comparing with their more performing counterparts. This result may be due to the fact that they are unprepared for the exam, but they are not totally unaware on this. One possible explanation for this phenomena was proposed by Clayson (2005), who suggest that the students are generally aware of their performance, but systematically overestimates their abilities, perhaps due to their previous experiences and or expectations. Similar explanation is proposed by Miller and Geraci (2011), the poorer overconfident subjects yet having low confidence in their predictions. The weakest students are indeed unskilled but that they may have some awareness of their lack of academic ability or preparation. On the other hand, the matrix of correlations (see Table 2) shows that both local and global accuracy are moderately and negative correlate with total test score, therefore the higher the test score, the lower the inaccuracy value (the greater the accuracy).

The second question of this study was if items difficulty is somewhat related with accuracy. Supplementary, we choose to test if item type is related with accuracy. In the case of multiple choice items, items difficulty shows no correlation with global accuracy. This lack of correlation may be due to the lack of subjective clues for confidence judgment, this kind of items favourising rather the guess of the solution, and also to the effect of confidence judgment delayed (at the end of the test). Also, items difficulty could be related with ease with which information is accessed, the fluency of retrieving, the perceived familiarity of items and so on (Koriat, 2007). It is also possible that storage strength, due to massed learning (the day before exam) function as a clue for retrieve strength. This heuristic is erroneous, because the two processes – storage and retrieval seems to be independent (Bjork & Bjork, 1992). This ambiguous results could be clarified by systematic variations of items type series in a future research.

We also find out that open ended items are low, but positively correlated with metacognitive monitoring inaccuracy, both local and global. This results reflect that the task of elaboration of answers, imposed by the open ended difficult items affects the metacognitive accuracy, both locally and globally, the students becoming less aware on their performances. The hard-easy effect revealed in this study (that means students tend to be overconfident on difficult items), must be taken cautiously. The hard–easy effect seems to be confirmed only for local and global confidence judgment related to difficult open ended items. In fact, this effect is questioned by Juslin et. al. (2000) who highlights some methodological biases in others researches.

The third question of the study was how the magnitude of overconfidence, i.e. the bias score, are related with academic performance. As expected, we obtained an average negative correlation, the most performing students being less biased by the overconfidence, and the poorer being more overconfident. This result is consistent with others research (Kruger & Dunning, 1999; Hacker et. al, 2000; Nietfeld et.al, 2005). Bias and accuracy are inversely related, the inaccuracy as overestimation being doubled by the overconfidence. Overconfidence produce underachievement, undermining students' retention and learning (Dunlosky & Rawson, 2001).

Conclusion

Metacognitive monitoring accuracy remains an important topic in the larger area of metacognition. This implies a constant and deep empirical and conceptual exploration. We tried in present study to test some classical findings and questions on the subject of overestimation and overconfidence in metacognitive monitoring. Overconfidence of the subjects in their estimation are the most robust result of this study. Items difficulty and type seems to have impact on accuracy, the subject overestimating their performance in difficult and more processed items. Although the poorer students are, paradoxically, more calibrated than theirs performing counterparts, still academic performance is related with greater metacognitive accuracy, both locally and globally.

One main strength of the present study is the aims to explore metacognitive monitoring processes in academic environment in two main problematic areas: the noticeable overestimation of results and the magnitude of overconfidence displayed by the subjects on theirs outputs. Completing Hacker et al. (2000) work, we have tested global monitoring and also local (or partial) monitoring accuracy.

A second strength of study is testing subjects on exam condition, and this increase validity of the results. The third strength is the association of difficulty and type of items with accuracy of judgments made by the subjects. We believe it's also relevant for the statistical power of results the number of subjects (100), unlike the study of Nietfeld, Cao, and Osborne (2005) although they made repeated measures, have a smaller study group (only 27 subjects).

As expected, a robust finding is the global correlation of greater results to the test with growing accuracy on self-made confidence judgments. Anyway, two distortions seem to appear in this global picture. The first distortion is that the poorer students display a greater local and global confidence comparing with the more performing classmates. They are not prepared for the exam, but they now that (Miller & Geraci, 2001). The second distortion is the weak positive correlation of inaccuracy (both local and global) and the difficult open-ended items. It seems that the effort of resolving this tough items distort the accuracy of self-evaluation performance, probably as expression of the hard-easy effect (Juslin et. al., 2000). Anyway, contradictory findings in monitoring accuracy in other studies was also observed by Efklides (2012), attributable to methodological issues as variation of age subjects, measures of monitoring accuracy used, timing of monitoring, type of the task etc.

Some weaknesses of the study are the correlation design, the confidence judgments as operationalization of accuracy, the using confidence judgment only as postdictions and the lack of metacognitive monitoring exercise of the subjects. Some of this shortcoming may be solved in future researches, by explicitly taking account of others empirical strategy (a quasi-experimental design), use of both prediction and postdiction, and explicit metacognitive training for improving accuracy. Also, we made some theoretical and methodological considerations and issues for future research.

First, we think it is important to carefully separate the estimation of a future or of a past cognitive task from the confidence or trust on this estimation. Confidence judgments are somewhat ambiguous in this matter, because when we made it is not clear if is about metacognitive (prediction or estimation of performance), or affective (subjective feeling, trust in our own prediction). That affects also the interpretation of result: cognitive and affective processing are not separated. For that, we propose an explicit measure of metacognitive estimation: (how much of task you estimate you will solve?) and another for confidence in previous estimation (how confident are you in your previous estimation?).

We expect that future research on accuracy will be focused on resolution (relative accuracy, Dunlosky & Rawson, 2012) and less on calibration (absolute accuracy). If calibration is easier to be tested, resolution remains poor studied. This ability of the subjects to discern between theirs bad and good responses is, however, crucial for practical reasons, such educational interventions. Is not enough to know that student is generally inaccurate and overconfident, it is important to know how and in which cognitive area this inaccuracy and overconfidence appear. Still, an increased number of empirical studies on metacognitive monitoring accuracy are necessary for deeper and clearer understanding of this phenomenon.

References

Barenberg, J., Dutke, S. (2013). Metacognitive monitoring in university classes: anticipating a graded vs. a pass-fail test affects monitoring accuracy. *Metacognition Learning*, *8*, 121–143.

Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In Healy, A. F., Kosslyn S.M., Shiffrin R. M. (coord), *Essays in honor of William K. Estes (Vol. 1: From learning theory to connectionist theory; Vol. 2: From learning processes to cognitive processes)*, Hillsdale, NJ: Erlbaum, pp. 35-67.

Bol, L., Hacker, D. (2012). *Calibration Research: where do we go from here? Frontiers in Psychology*, Hypothesis and Theory Article.

Buratti, S., & Allwood, C., M. (2012). The accuracy of metametacognitive judgments: regulating the realism of confidence. *Cognitive Processing*, *13*, 243–253.

Chao-Ming, C. (2010). Accuracy and stability of metacognitive monitoring: A new measure. *Behavior Research Methods*, *42*(3), 715-732.

Clayson, D. (2005). Performance Overconfidence: Metacognitive Effects or Misplaced Student Expectations? *Journal of Marketing Education*, *27*(2), 122-129.

Dinsmore D.L., & Parkinson, M.M. (2013). What are confidence judgments made of Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, *24*, 4-14.

Dunlosky, J., & Rawson, K.A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271-280.

Efklides, A., (2012). Commentary: How readily can findings from basic cognitive psychology research be applied in the classroom? *Learning and Instruction*. *22*, 290-295.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, *34*, 906-911.

Fleming, S., & Lau, H. (2014). How to measue metacognition. *Frontiers in Human Neuroscience*, *8*: 443. DOI: 10.3389/fnhum.2014.00443

Freedman, D.A. (2009). *Statistical Models, Theory and Practice*. Cambridge: Cambridge University Press

Glenberg, A. M., Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *11*(4), 702-718. DOI: 10.1037/0278-7393.11.1-4.702

Glenberg, A.M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*(1), 84-93. DOI: 10.3758/BF03197714

Griffin, D., & Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology, 24*, 411- 435. DOI: 10.1016/0010-0285(92)90013-R

Grimes, P.W. (2002). The overconfident principles of economics students: an examination of a metacognitive skill. *The Journal of Economic Education, 33*(1), 15-30. 10.1080/00220480209596121

Hacker, D.J., Bol, L., Horgan, D., & Rakow, E.A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*, 160-170. DOI: 10.1037/0022-0663.92.1.160

Juslin P., Winman A., & Olsson H., (2000), Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect, *Psychological Review*, *107*, 2, 384-396. DOI: 10.1037/0033-295X.107.2.384

Koriat, A. (2007). Metacognition and Consciousness. in Zelazo P.D., Moscovitch, M.,Thompson E. (coord.), *Cambridge Handbook of Consciousness*, Cambridge: Cambridge University Press.

Koriat, A. (2011). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*, 296-298. DOI: 10.1016/j.learninstruc.2012.01.002

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. DOI: 10.1037/0022-3514.77.6.1121

McCormick, C.B. (2003). Metacognition and learning. In Reynolds, W. M., Miller, G. E. (Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 79-102). New York, NY: Wiley.

Miller, T.M., & Geraci, L. (2011). Unskilled but Aware: Reinterpreting Overconfidence in Low-Performing Students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 502-506. DOI: 10.1037/a0021802

Miller, T.M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition*. *29*, 131-140. DOI: 10.1016/j.concog.2014.08.008

Nietfeld, J.L., & Schraw, G. (2002). The effect of knowledge and strategy explanation on monitoring accuracy. *Journal of Educational Research*, *95*, 131-142. DOI: 10.1080/00220670209596583

Nietfeld, J.L., Cao, L., & Osborne J.W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. *The Journal of Experimental Education*, *74*(1), 7-28.

Pressley, M., Yokoi, L., Van Meter, P., Van Etten, S., & Freebern, G. (1997). Some of the reasons why preparing for exams is so hard: What can be done to make it easier? *Educational Psychology Review*, *9*, 1-38.

Redford, J.S., Thiede, K.W., Wiley, J., & Griffin, T. (2012). Concept mapping improves metacomprehension accuracy among 7th Graders. *Learning and Instruction*, *22*, 281-289.

Ruiz, M., & Arroyo, C. (2016). JOLer: A Java standalone application for simulating the Weaver & Kelemen's judgment of learning (JOL) model. *Anales de Psicología*, *32*, 3, 893-898.

Schraw, G. (1994). The Effect of Metacognitive Knowledge on Local and Global Monitoring. *Contemporary Educational Psychology, 19*(2), 143-154. DOI: 10.1006/ceps.1994.1013

Schraw, G., Roedel, T.D. (1994). Test difficulty and judgment bias. *Memory & Cognition,* 22 (I), 63-69.

Schraw, G., Wise, S.L., & Rose, L.L. (2000). Metacognition and computer based testing, In: Schraw, G., Impara C.J., *Issues in the measurement of metacognition,* (pp.263-260), Lincoln NE: Buros Institute of Mental Measurement.

Thiede, K.W., Anderson M.C.M., & Therriault, D. (2003). Accuracy of Metacognitive Monitoring Affects Learning of Texts. *Journal of Educational Psychology*, *95*, 1, 66-73. DOI: 10.1037/0022-0663.95.1.66