Bulletin of the *Transilvania* University of Braşov - Special Issue
Series VII: Social Sciences • Law • Vol. 10 (59) No. 2 - 2017

METACOGNITIVE MONITORING ACCURACY AND MATH COMPETENCES IN PRIMARY SCHOOL

F. VASILE-FRUMOS[1] I. IURCIUC[2]

Abstract: *Knowing the pupils' metacognitive skills represents a challenging task for many compulsory education teachers in different school levels and various disciplines. The main purpose of our study was to investigate how math competences of elementary school subjects are related to their metacognitive monitoring accuracy. 112 second grade elementary pupils were examined with a math test and they also provided an estimation of their own result. The discrepancy between the estimated and the actual score offers a measure of their metacognitive monitoring accuracy. The results show that the lower the score, the optimistic the pupil's estimate is. The best performing students show a good calibration. Some educational implications are discussed.*

Key words: *metacognitive calibration, accuracy, performance, primary school pupils.*

1. Introduction

Metacognition appears as an experience which involves the use of cognitive or metacognitive strategies and also metacognitive regulation (Brown, 1987). These processes help regulate and improve learning, and consist of anticipating and planning future cognitive tasks; monitoring and controlling the ongoing cognitive activities; and finally assessing or checking the outcomes of those activities. Flavell, one of the pioneers of research in the metacognitive area, defined this concept as "the knowledge of an individual about his own cognitive processes, about its products, and about everything related to, for example, the characteristics relevant to learning tasks" (Flavell, 1976, p. 232). Also, metacognition refers to the active monitoring, regulation and organization of cognitive processes (Flavell, 1976).

Metacognitive monitoring is the process of supervising and verifying the quality of learning by reference to the learning goals. For example, during the learning process,

[1] Alexandru Ioan Cuza University of Iasi, frumos@uaic.ro
[2] Secondary School Zmeu, Iasi County, ilie_teacher@yahoo.com

individuals monitor their performance and compare it with the desired level of performance. If there is a discrepancy between the two, individuals will continue to study until this discrepancy is reduced to zero. In other words, individuals will cease to learn when “the performance achieved overlaps or exceeds the desired level of performance” (Thiede, Anderson & Therriault, 2003, p. 69).

We must differentiate between monitoring and control. Monitoring processes reflect the ability to accurately represent one's own mental states and processes, and control is the ability to effectively regulate one's cognitive processes. According to Nelson and Narens’ metacognitive monitoring model (1994), during metacognitive control, the metacognitive level informs the object level so that either a new action is initiated (e.g. if the subjects read a text, they reread the more difficult parts in order to understand) or the actions already started are continued or stopped (e.g., individuals decide to finish reading).

Winne (2001, p. 56) considers metacognitive monitoring to be the “pivot that self-regulated learning is based on” as it involves assessing the level of understanding and progress toward achieving goals that affect how individuals address learning tasks and make adaptive changes in learning.

The accuracy of metacognitive monitoring is “a crucial aspect of self-regulated learning, because the decisions of adjusting the strategies used or the time and effort invested in learning are made on the basis of monitoring results” (Alexander, quoted in Mihalca, 2013, p. 19). For example, if monitoring is inaccurate, the allocated time to learning is shortened, in the sense that individuals decide that it is not necessary to persist in solving tasks for a longer time.

Usually, individuals are considered to be able to monitor and properly control and adjust their own learning process. However, this theoretical assumption has not been confirmed. Some empirical data revealed that “most individuals have low ability to monitor correctly” (Graesser & McNamara, quoted in Mihalca, 2013, p. 19). It has been found that the especially the novices in a knowledge area (individuals with a low background of prior knowledge) have low skills to correctly assess the way they learn, resulting in wrong decisions about what needs to be learned, and for how long, with negative outcomes of the achieved performance (Ross, Phillips, Klein, & Cohn, 2005).

There are two types of metacognitive monitoring accuracy: absolute accuracy and relative accuracy. Despite of the fact that it is useful to measure both types of accuracy, recent studies have focused more on absolute accuracy than on relative accuracy. This phenomenon is probably due to the fact that “absolute accuracy provides useful information about the extent to which items are taught and the need for their study for a while longer term” (Bruin & Van Gog, quoted in Mihalca, 2013, p. 21). Also, absolute accuracy is easier to measure.

The absolute accuracy (or calibration) occurs when we have a measure of the degree of correspondence between the subjects’ metacognitive judgments and their objective performance. The variation of the overlap between the estimation and the actual performance indicates either good calibration quality or poor calibration. More specifically, if individuals estimate their performance as good (metacognitive judgments about their own performance) and indeed the objectively measured performance confirms this estimate, then it can be said that the individuals are well calibrated. A low quality of calibration is indicated by both overestimation of performance (overconfidence), in which case the scores corresponding to metacognitive judgments are higher than the performance obtained, and the underestimation of performance (under-appreciation, under-confidence), in which the reverse is valid (Schraw, quoted in Mihalca, 2013, p. 21).

Relative accuracy refers to the degree of association between metacognitive judgments and the performance obtained in solving multiple tasks (more precisely, it is the intra-subject correlations between the two variables), indicating the accuracy with which individuals discriminate between the better learned subjects and the less learned (Nelson, quoted in Mihalca, 2013, p. 21). In other words, relative accuracy measures the individual's ability to predict the likelihood of achieving superior performance for an item / task compared to another item or task (for example, the prediction that item A will be better learned than item B, as a result, item A will be recovered from memory at a later test, while Item B will not (Meeter & Nelson, quoted in Mihalca, 2013, p. 22).

Absolute accuracy research has indicated that calibration quality is generally low and very few individuals are capable of a near perfect calibration (the lack of bias), in which the judgment scores are close to the actual performance (Bjorkman, quoted in Mihalca, 2013, p. 23). This type of accuracy is influenced by at least three variables, namely: task characteristics (e.g. difficulty level or length of tasks); external constraints imposed on processing (e.g. feedback, explicit training), and characteristics of individuals (e.g. level of expertise or previous knowledge (Nietfeld & Schraw, as cited in Mihalca, 2013). For example, Schraw and Roedel (as cited in Mihalca, 2013) demonstrated, using the bias as a measure of absolute accuracy, that overestimation of own performance increases with the difficulty level of the task. Burson, Larrick, and Klayman (2006) showed that difficult tasks decrease the accuracy of metacognitive judgments in both high and low performance individuals. In addition, these authors indicated that individuals with higher performance overestimated their results (in other words, they were more inaccurate) than those with low performance (Mihalca, 2013, p. 23).

Inaccurate monitoring also seems related to the type of learning judgments (global or local) and the practice and feedback. In the study by Mazzoni and Nelson (as cited in Mihalca, 2013) it was demonstrated that when students provided judgments for assessing their progress in global learning, overestimation was reduced compared to the situation where these judgments were made for each learned item (local JOLs). In other words, in this study, global judgments proved to be more correct than local judgments. Hacker, Bol, Horgan, and Rakow (2000) demonstrated that combining feedback with practice and training provided in courses had beneficial effects on accuracy and performance. More specifically, the results indicated that the absolute accuracy (calibration) of judgments measured both as predictions and as post-dictions increased with time as a result of practice, but only in the case of students with superior performance.

Based on the results from the empirical studies referred to above, it can be concluded that "learning is inversely proportional to calibration bias and directly proportional to the accuracy of calibration" (Winne, 2001, p. 156).

On the other hand, Mihalca (2013) presents a study by Koriat (1997), which showed that practice negatively affects calibration or absolute accuracy, increasing underestimation of performance (the so-called under-confidence-with-practice effect), but it improves relative accuracy, i.e. judgments on which items are better learned that others. According to this author, the improvement with practice in relative accuracy occurs because the clues of learning judgments change from exterior aspects such as perceived items' difficulty to intern ones, such as items' processing fluency.

According to the empirical evidence reviewed in these paragraphs, it can be concluded that improving the accuracy of monitoring is not an easy task, requiring "intensive and explicit interventions that should include the practice and feedback" (Nietfeld Cao, & Osborne, as cited in Mihalca, 2013, p. 26).

2. Objectives

The individuals' ability to correctly monitor their own cognitive processes may result in increased performance through its positive effects on the process of learning self-regulation.

Considering the research results presented above, that demonstrate that a proper monitoring has beneficial effects on self-regulated learning, and that most individuals have low ability to correctly monitor their own cognitive processes, an essential question arise: how the accuracy of metacognitive judgments is related to academic performance in elementary school children.

The answer to this question is very important for the educational practice because the relationship between accurate monitoring and the performance achieved contributes to the design of materials or practical instructional interventions.

The main objective of our study was to test how mathematical competences of elementary school subjects relate to their metacognitive accuracy. A secondary objective of this study was to test if the predicted and post-dicted (immediate and delayed) confidence judgments influenced monitoring accuracy (Hacker, Bol, Horgan, & Rakow, 2000).

3. Participants and Procedure

We tested the relationship between metacognitive accuracy and levels of performance in Mathematics, by applying a evaluation test with exercises of division, one of the most complex arithmetic operation in elementary school.

The participants to this study were 112 children from two secondary urban schools from Târgu Frumos (Iasi County) and from and two secondary rural schools: Zmeu (Iași County) and Pribesti (Vaslui County). All subjects were in second grade elementary schools, approximatively eight years old, and the gender distribution was 60 boys and 52 girls. 80 of the pupils were from urban schools and 32 from rural schools. Ethnic composition of the sample was: 97 Romanian subjects and 15 subjects belonging the Roma ethnic group.

The procedure consisted of assessing children's arithmetic abilities with a five-item or categories of items test. The pupils solved exercises and an arithmetic problem using the operation of division, as follow: 1. The first category item consisted in computing nine divisions: 24:6; 7:7; 50:5 etc. The second set of tasks consisted of six multiplications or divisions for completing the gaps. Example: □ :4=6 or 20: □ =10. The third category of items was a repeated division with two numbers: “Find the numbers two times smaller than: 20, 12, 10, 8”. The fourth category requested to calculate divisions, subtractions, additions and multiplications, observing the order of the operations: 25:5+79; 84–40:5; 2X9+32:4. Ultimately, the fifth item was an arithmetic problem: “Ioana has 63 pencils. Andrei has seven times less than Ioana. How many pencils do they have together?”. For each category of items performance descriptors for three levels of performance and corresponding scores were provided. The students’ papers were evaluated on a 0-10 points scale. The test lasted for twenty minutes and was preceded by an estimate (a confidence judgment) of their own result. Immediately after the test, the children provided again a new estimation of the result. At the end of the school day, a third estimate of the outcome was requested. All three estimates were on a same 0-10 points scale and were compared with the teacher objective evaluation of the tests.

All subjects were tested by the same researcher (the second authors of the study), in standardized conditions: first class of the school day and with the same instructions. The ethics of the research was ensured through the anonymity of the tests and by the honest communication of the research goal.

The confidence judgments (CJs) we used in our study represent a measure of metacognitive monitoring accuracy. This technique consists of statements or numeric estimation of a future or past performance. Confidence judgements are concerned with the degree of trust of individuals in successfully completing learning or evaluating tasks. These types of judgments can be measured either as prospective judgements or predictions (e.g. prospective monitoring - which takes place before solving a learning task, for example, "How sure are you going to resolve the next task correctly?"), or as post-dictions (retrospective monitoring - after solving the learning task) or retrospective judgment of confidence in the given answer, for example, "How certain are you that you have correctly managed this task?" (Dinsmore & Parkinson, 2013).

4. Results

First, we tried to find out if confidence judgments, expressed as numeric estimation in predictions, post-dictions and delayed post-dictions of test performance are statistically significant different from the teacher’s evaluation. Three paired-samples t-tests were conducted to compare the teacher’s evaluation to the pupils’ estimations of their results before, immediately after the test and later after the test. All the differences between the teacher's evaluation and the subject’s self-estimations of results were statistically significant, as shown in Table 1.

We can notice that the mean of children’s self-evaluation is more optimistic (approximatively 1,5 points of overestimation) than their actual results to the test. Another point of interest was whether the prediction, post-diction and delayed post-diction were different one from each other. Paired-samples t-test were conducted to compare the estimation before test - estimation immediately after the test; estimation before the test –

Paired samples t tests for mean difference between teachers’ evaluation and pupils *pretest, posttest and later posttest self-evaluations* Table 1

	M	SD	t	df	p
Pair 1 teacher's evaluation - estimation before test evaluation	1.52	2.78	5.81	111	.001
Pair 2 teacher's evaluation - estimation immediately after the test	1.54	2.53	6.44	111	.001
Pair 3 teacher's evaluation – delayed estimation after the test	1.53	2.43	6.67	111	.001

delayed estimation after the test and estimation immediately after the test - delayed estimation after the test, respectively. There were no significant difference in the scores of the subjects' self-evaluation, as Table 2 shows. If none of the pairs of means compared are significantly different from each other means that the subject basically does the same estimation in predictions, post-dictions and delayed post-dictions.

Because this global approach is not sufficient for testing the research question, we made supplementary calculus. The subjects' results were distributed in four categories or levels of performance, as follows: from 0 to 3 point – level 1; from 4 to 6 points –level 2; from 7 to 8 points –level 3 and from 9 to 10 points –level 4, respectively. Because the prediction, post-diction and delayed post-diction do not differ one from another, for further statistical calculations we choose as reference the subjects' predictions (the first self-evaluation).

Sample t tests were conducted for comparing performances (teachers' evaluation) means for each level, with estimations (subjects' evaluation) means. The result are presented in Table 3.

As we notice in the table 3, there are significant differences between the scores of the pupils' self-evaluation and the teachers' evaluation, for the three of the performance levels. The subjects form lower performance levels: level 1, level 2 and level 3 respectively gradually differ from the teachers' assessment. On the other hand, there are no significant differences between subjects from the highest level of performance (the fourth level) and the teachers' evaluation: the most performing subjects evaluate themselves very close to the teachers' evaluation.

Table 2

Paired samples t tests for mean differences between each pairs of the three self-evaluation

	M	**SD**	**t**	**df**	**p**
Pair 1 estimation before test - estimation immediately after the test	.018	1.35	.13	111	.89
Pair 2 estimated before test evaluation - delayed estimation after the test	.009	1.64	.05	111	.95
Pair 3 estimation immediately after the test - delayed estimation after the test	.009	1.35	.07	111	.94

Table 3

One-Sample t test for mean comparison of each actual performance level and self-evaluation of performance

	M	**SD**	**t**	**p**
teach_eval_level1	.94	1.29	$t_{(16)} = 7.76$	.001
self_eval_level1	6.41	2.87		
teach_eval_level2	5.07	1.91	$t_{(25)} = 7.77$	.001
self_eval_level2	7.92	.79		
teach_eval_level3	7.41	.51	$t_{(11)} = 3.24$	.008
self_eval_level3	8.91	1.62		
teach_eval_level4	9.8	.39	$t_{(56)} = 1.58$	.118
self_eval_level4	9.56	1.13		

The results show that the subjects from the highest level of performance (level 4) made realistic estimation of their performance (mean difference -.23). On the other hand, the subjects that have poorer results made inaccurate estimation of their performance (mean difference 1.51 at level 3; mean difference 2.92 at level 2 and mean difference 5.41 at level 1).

Therefore, inaccuracy grows when performance decreases. This phenomenon is obvious in the Figure 1.

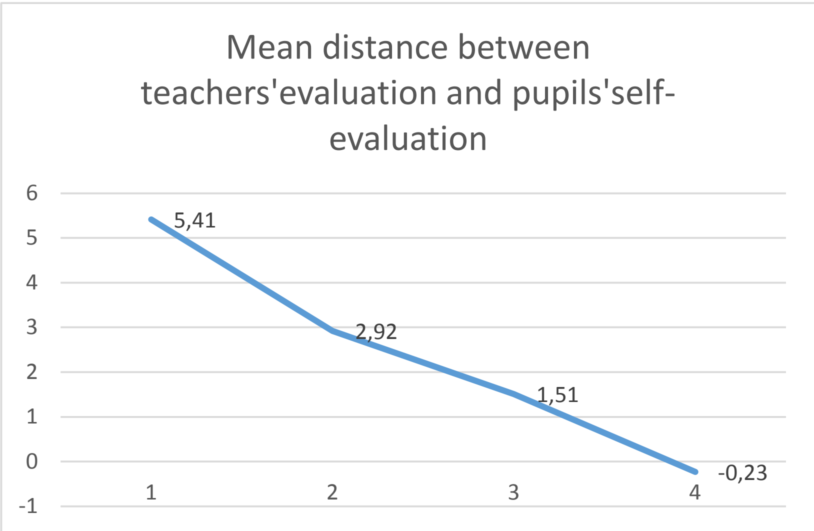

Fig. 1. *Relation between accuracy and performance*

Negative evolution of results and accuracy are also intuitively presented in the next graph (Figure 2), which represents the dynamic of Euclidean distances between teacher’s evaluation and studen’s self-evaluation.

5. Limits of the Study, Conclusions and Discussion

Some limits of the study are: 1) the subjects, elementary school pupils, do not have self-evaluation skills, due to traditional manner of evaluation: the teacher made almost all the assessments and the children do not have a real ability to do it. 2) The study concerns only mathematical skills, whereas others academic abilities (i. e. literacy skills) could dramatically influence the results. 3) We cannot conclusively explain why the subject does the same estimation in predictions, post-diction and delayed post-dictions.

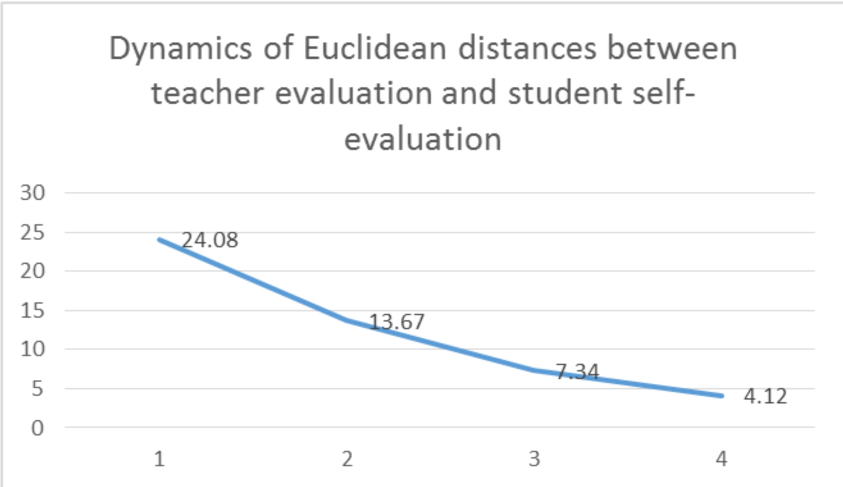

Fig. 2. *Dynamic of Euclidean distances between teachers' and pupils self-evaluation*

The most important results of our study are:

a) the accuracy of metacognitive monitoring is closely related to academic performance, which is confirmed by other studies on this topic (Hacker, Bol, Horgan, and Rakow, 2000); the performing subjects have good metacognitive accuracy, and the less performing are metacognitively de-calibrated. Possible reasons for this poor calibration are: primary schools' pupils in the Romanian educational system have poor metacognitive self-monitoring skills, because this kind of abilities are not practiced systematically.

Another reason could be the assessment scale; in Romanian elementary education an ordinal scale with four levels: Very Good, Good, Sufficient and Insufficient (Failed) is used, which may raise difficulties in self-evaluation on a numerical scale like the one used in the study. A third possible explanation might be the need for a positive self-esteem of the subjects included in the study.

b) As the results decrease in the assessment test, the metacognitive accuracy decreases dramatically. In other studies, it became clear that the higher the difficulty of test items is, the more the metacognitive accuracy decreases. In our study we highlighted a progressive decrease in metacognitive accuracy with lower test performance (Graesser & McNamara, as cited in Mihalca, 2013; Schraw and Roedel, as cited in Mihalca, 2013).

c) the moment of evaluation does not influence metacognitive accuracy. It is possible that the moment of estimation, as well as other variables that we have not considered, will influence the maintenance of the same level of self-evaluation in the three moments. Pieschl (2012, p. 18), for example, believes that we need to explore more “learners’ calibration to internal standards in order to learn more about their awareness of their own cognitive processes of knowing”. On the other hand, Dunlosky and Nelson (1992), referring to the differences between self-assessment moments in learning, show that "the kind of cue that is used for JOLs is critical". However, more other research is needed in this subject.

Other information may be obtained from the address: frumos@uaic.ro

References

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Lawrence 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*(1), 60-77.

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., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOLs) and the delayed-JOL effect. *Memory & Cognition*, *20*, 373–380.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Lawrence Erlbaum.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160-170.

Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349-370.

Mihalca I. L. (2013). *Implementarea unui program de antrenament metacognitiv în scopul sprijinirii învăţării autoreglate în mediile educaţionale computerizate*, [Implementation of a metacognitive training programme in order to support self-regulation of learning in computerized educational environments]. București: Editura Muzeului Literaturii Române. Retrieved from: http://www.cesindcultura. acad.ro /images/fisiere/rezultate/postdoc/rapoarte%20finale%20de%20cercetare%20stiintifica %20ale%20cercetatorilor%20postdoctorat/lucrari/Mihalca_Loredana.pdf

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe, & A. P. Shimamura, (Eds.), *Metacognition: Knowing About Knowing* (pp. 1-25). Cambridge, MA: MIT press.

Pieschl, S., (2009). Metacognitive calibration-an extended conceptualization and potential applications. *Metacognition Learning*, *4*, 3-31.

Ross, K. G., Phillips, J. K., Klein, G., & Cohn, J. (2005). Creating expertise: A framework to guide technology-based training. (Final Technical Report for Contract #M67854-04-C-8035 for the Marine Corps Systems Command/Program Manager for

Training Systems). Fairborn, OH: Klein Associates. Retrieved from: https://issuu.com/cognition/docs/ross__k.g.__phillips__j.k._2005_creating_expertise

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.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153-189). Mahwah, NJ: Lawrence Erlbaum Associates.