

## ARTÍCULO DE INVESTIGACIÓN

# RELATION BETWEEN STUDENTS' EXPECTATIONS ABOUT THEIR GRADE AND METACOGNITIVE MONITORING AND A DEEPER UNDERSTANDING OF METACOGNITIVE JUDGMENTS

## RELACIÓN ENTRE LAS EXPECTATIVAS DE LOS ESTUDIANTES SOBRE SU NOTA Y EL MONITOREO METACOGNITIVO Y UNA COMPRENSIÓN MÁS PROFUNDA DE LOS JUICIOS METACOGNITIVOS

ANTONIO P. GUTIERREZ DE BLUME<sup>1</sup>,  
DIANA MARCELA MONTOYA LONDOÑO<sup>2</sup>

FECHA DE RECEPCIÓN 26/05/2021 • FECHA DE ACEPTACIÓN 02/09/2021

Para citar este artículo: Gutierrez de Blume, A. P., & Montoya Londoño, D. M. (2021). **Relation between students' expectations about their grade and metacognitive monitoring and a deeper understanding o metacognitive judgments.** *Psychología. Avances de la Disciplina*, 15(2) 13-31. <https://doi:10.21500/19002386.5425>

### Abstract

Metacognition is an important higher-order thinking process for successful learning. The present study investigated the relation between students' ( $N = 65$ ) expectations about their grade (expressed as difference scores between expected grade and actual grade) and their metacognitive monitoring accuracy and bias and the extent to which these difference scores in expected grade versus actual grade predicted accuracy and bias, employing an *explanatory sequential quantitative* → *QUALITATIVE* mixed method research design. The study also explored how students develop and refine metacognitive judgments and the types of strategies they employ during this process. Results revealed that there were significant relations between difference scores in expected grade versus actual grade and accuracy and bias ( $r = .02$  to  $r = .89$  in absolute value), and that difference scores significantly predicted both accuracy ( $R^2 = .52$ ) and bias ( $R^2 = .69$ ). Further, qualitative findings revealed that there were differences in how students developed and refined metacognitive judgments as a function of four aspects of learning: effort/preparation, strategy selection/implementation,

1 Associate Professor in the Department of Curriculum, Foundations, and Reading, Georgia Southern University, United States. ORCID: <https://orcid.org/0000-0001-6809-1728> P.O. Box 8144, Statesboro, GA 30460-8144, United States. Doctor of Philosophy in Educational Psychology from the University of Nevada, Las Vegas (United States). Email: [agutierrez@georgiasouthern.edu](mailto:agutierrez@georgiasouthern.edu). Phone: +1-912-478-7831.

2 Faculty in the Department of Educational Studies, Universidad de Caldas, and Faculty in the Psychology Program, Universidad de Manizales. ORCID: <https://orcid.org/0000-0001-8007-0102>

planning, and evaluation. Educators should explicitly teach metacognitive monitoring skills to improve students' self-regulated learning.

**Key words.** Metacognition; Absolute accuracy; Absolute bias; Mixed method (Source: PsycINFO Thesaurus).

## RELACIÓN ENTRE LAS EXPECTATIVAS DE LOS ESTUDIANTES SOBRE SU NOTA Y EL MONITOREO METACOGNITIVO Y UNA COMPRENSIÓN MÁS PROFUNDA DE LOS JUICIOS METACOGNITIVOS

### Resumen

La metacognición es un proceso importante de pensamiento de orden superior para un aprendizaje exitoso. El presente estudio investigó la relación entre las expectativas de los estudiantes sobre su nota (expresadas como puntuaciones de diferencia entre la nota esperada y la nota real) ( $N = 65$ ) y su precisión y sesgo de monitoreo metacognitivo y el grado en que estas diferencias en la nota esperada versus la nota real predijeron la precisión y el sesgo, empleando un diseño de investigación *secuencial explicativo cuantitativo-CUALITATIVO* de método mixto. El estudio también exploró cómo los estudiantes desarrollan y refinan juicios metacognitivos y los tipos de estrategias que emplean durante este proceso. Los resultados revelaron que había relaciones significativas entre las diferencia de puntajes en la nota esperada versus la nota real y la precisión y el sesgo ( $r = .02$  to  $r = .89$ , en valor absoluto), y que estas diferencia de puntajes predijo significativamente tanto la precisión ( $R^2 = .52$ ) como el sesgo ( $R^2 = .69$ ). Además, los hallazgos cualitativos revelaron que había diferencias en la forma en que los estudiantes desarrollaban y refinaban juicios metacognitivos en función de cuatro aspectos del aprendizaje: esfuerzo / preparación, selección / implementación de estrategias, planificación y evaluación. Los docentes deben enseñar explícitamente habilidades de monitoreo metacognitivo para mejorar el aprendizaje autorregulado de los estudiantes.

**Palabras clave:** metacognición; precisión absoluta; sesgo absoluto; método mixto; desempeño (Fuente: PsycINFO Tesaurus).

### Introduction

Research on metacognition has involved two main trends. Some studies focus on the two classic components of metacognition, metacognitive knowledge and regulation. Others shifted to a new paradigm that recognizes individual differences in metacognitive behavior and the development of metacognitive profiles, which has allowed researchers to better understand the importance of relatively obscure aspects such as metacognition's relation to personality, self-concept, types and levels of processing, rhythms in learning, and even locus of control, among other aspects (Gutierrez de Blume & Montoya, 2020; Gutierrez de Blume et al., in press). However, both trends help advance researchers' understanding of how metacognition operates in the learning, problem

solving, and reasoning skills in students of all ages, domains, tasks, and contexts (Azevedo, 2020).

The latest research on metacognition is oriented to the development of research more focused on not only explaining, but also understanding the different mechanisms regarding how students learn from a metacognitive perspective. This orientation examines monitoring as a complex, multilevel process with different layers predicated on a theoretical model that explains the nuanced role of metacognitive monitoring accuracy and error in learning-judgment development (Gutierrez de Blume, 2020; Gutierrez et al., 2016; Gutierrez de Blume et al., 2021). Thus, research underscores the relevance of specifying the underlying aspects of monitoring and control processes in the development of first- and second-order metacognitive judgments such as predictive, concurrent,

and postdictive. Further, research exists that links “warmer” aspects of cognition, including variables such as motivation, attributional style (Gutierrez & Price, 2017), and affect and personality (Gutierrez de Blume & Montoya, 2020) that seem to guide the level of metacognitive awareness during learning. Increasing understanding of potentially generalizable metacognitive skills, applicable in any domain, promises to benefit students in many areas of knowledge acquisition and in everyday life. This is the case because being able to accurately monitor one’s progress towards a learning goal and clearly understanding task demands, as a form of formative-continuous evaluation, can, presumably, improve the effectiveness of later learning episodes (Dunlosky & Rawson, 2019).

Interestingly, many studies regarding metacognitive monitoring and the development of metacognitive judgments have employed a quantitative approach focused on investigating relative or absolute monitoring judgments. These types of monitoring judgments describe the relation between performance in an evaluation task and learners’ confidence in performance judgments (Schraw et al., 2013; Gutierrez de Blume et al., 2021). These concepts are explored next.

Research on metacognitive monitoring has focused on estimating the level of performance, accuracy, and confidence with various measures. Metacognitive judgments can be understood, for instance, in terms of absolute and relative accuracy (Schraw, 2009a, 2009b; Schraw et al., 2013), as well as how accuracy is determined such as the Goodman-Kruskal gamma correlation (Nelson, 1996). Regardless of how researchers measure monitoring or how accuracy is determined, measurement follows a typical format in which learners answer a test item and provide a confidence rating on performance on that item (local), or they provide confidence ratings holistically for an entire assessment to compare against overall performance on the assessment (global), which can also be done prior to (predictions) and/or after (postdictions) the assessment itself (Follmer & Clariana, 2020; Schraw, 2009a, 2009b). Monitoring accuracy is subsequently calculated based on different computational formulas using frequencies of two or more of the four mutually exclusive cells in a 2x2 data matrix, where cell *a* corresponds to correct performance that is judged to be correct; cell *b* corresponds to incorrect performance that is judged to be correct; cell *c*

corresponds to correct performance that is judged to be incorrect; and cell *d* corresponds to incorrect performance that is judged to be incorrect (Gutierrez et al., 2016; Gutierrez de Blume et al., 2021; Schraw et al., 2013). Thus, cells *a* and *d* in this framework correspond to accurate monitoring whereas cells *b* (referred to variously as overconfidence or an *illusion of knowing* [Serra & Metcalfe, 2009]) and *c* (referred variously as underconfidence or an *illusion of not knowing* [Serra & Metcalfe, 2009]) correspond to erroneous monitoring.

Schraw and his colleagues (Gutierrez et al., 2016; Schraw et al., 2013; Schraw et al., 2014) examined different statistical measures and how they relate. These included parametric monitoring indices like sensitivity, specificity, *d*’ (“d prime”), and the G-index, and non-parametric ones such as the odds-ratio, gamma, kappa, and Sokal distance. The main objective of this series of studies was to determine not only different latent dimensions of the metacognitive monitoring process, but also to uncover which of these measures explain more variation in the data, and if the use of multiple measures provides a more complete result. Results of these studies provided a greater understanding of the underlying mechanisms of metacognitive monitoring. Nevertheless, a major shortcoming of these works is that they were all quantitative in nature, thereby prohibiting a deeper, richer understanding of metacognitive monitoring due to a lack of process-oriented data. This necessitates additional research employing qualitative and mixed method research designs.

However, relatively few studies exist that employ alternate research designs. Exceptions to this include an investigation carried out in South Africa with a group of students of a basic chemistry course (Mathabathe, 2019; Mathabathe & Potgieter, 2014). Mathabathe and Potgieter (2014) sought to establish whether students’ overconfidence before instruction was adjusted after instruction. Results revealed that most of the students were too confident in their judgments of performance. In both the pre- and post-tests, the quantitative results showed that students with little preparation were slow to develop accurate metacognitive monitoring skills within the classroom environment that did not include instruction focused on the development of such skills. Along a similar vein, Mathabathe (2019) explored the justifications that

students expressed for their perceived performance in an objective test. Findings indicated that despite teaching, students still overestimated their performance, relying more on feelings when answering test questions than on their lack of information or level of mastery to guide the choice of their performance judgments.

In another study with undergraduate students, researchers examined both the quantitative measures of monitoring judgments and performance, as well as the open responses provided by the students (Dinsmore & Parkinson, 2013). Results showed no significant findings in the quantitative portion. On the other hand, the open-ended responses demonstrated that students based their confidence scores on prior knowledge, text characteristics, item characteristics, riddles, and combinations of these aspects (Dinsmore & Parkinson, 2013). Finally, a purely qualitative research study, which served as the foundation for the present study, found that students base their metacognitive judgments on four aspects: 1) effort/preparation; 2) strategy selection; and two aspects of the regulation component of metacognition, 3) planning and 4) evaluation (Gutierrez de Blume et al., 2017).

Given the dearth of research on metacognitive monitoring incorporating a qualitative or mixed research design, the present study sought to establish the relation between students' expectations about their grade, monitoring accuracy, and monitoring bias (predictions and postdictions) (from a quantitative approach) and to understand how students develop their metacognitive judgments (from a qualitative approach).

## Conceptual and Theoretical Frameworks

This study is situated in the work of Serra and Metcalfe's (2009) discussion regarding the implications of learners' *illusion of knowing* and *illusion of not knowing*. In their discussion, they describe these processes as "feelings of knowing" (p. 292), which learners experience as they prepare for a learning episode. Thus, this study re-conceptualizes the conventional linear approach to metacognitive monitoring as a *feeling of knowing* (FOK) that allows researchers to connect with participants in a meaningful way and tap into their understanding of their own learning process (Avhustiuk et al., 2018).

In addition to Serra and Metcalfe's (2009) work, the present study is framed using self-regulated learning theory (SRL). According to the tenets of SRL, to capture learners' processing more completely, researchers need to consider cognitive, metacognitive, and affective/dispositional characteristics of the learner. Though there are several ways to approach SRL theory (see Panadero, 2017, for a review), this study employs Winne and Hadwin's (2008) Metacognitive Perspective Model (MPM) and Efklides' (2011) Metacognitive and Affective Model of Self-Regulated Learning (MASRL) as theoretical principles to guide substantive interpretation of findings. Metacognitive processes play a central role in both models. According to the tenets of the MPM, learners are perceived as being active, involved self-regulated individuals who control their own learning through the implementation of metacognitive monitoring and strategy use, which are central to the goals of the present study. Along a similar vein, Efklides' MASRL stipulates that metacognitive and motivational processes are also key, centered on task, person, and a combination of both (Efklides et al., 2018). Self-regulated learning theory provided a framework with which to explore the FOKs within our participants expressed as *absolute* monitoring accuracy and bias.

## The Present Study

Based on the literature surveyed, the present investigation had three objectives. The first was to examine the relation between students' expectations about their grade (expressed as difference scores between expected grade and actual grade) and metacognitive monitoring accuracy and bias on four separate assessments. The second was to evaluate whether the difference between students' expected grade and actual grade on the four assessments predicted monitoring accuracy and bias. The final objective was to explore how students develop and refine their metacognitive judgments and the types of strategies they employ in that process as a function of their monitoring accuracy using qualitative data. Hence, the present study was guided by the following research questions.

1. What is the relation between students' expectations regarding their grade on four assessments

(expressed as difference scores between expected grade and actual grade) and metacognitive monitoring accuracy and bias (predictions and postdictions)?

2. Does the difference between students' expected overall grade and actual grade (expressed as difference scores) predict their metacognitive monitoring accuracy and bias?
3. Are there differences in how students develop and refine their metacognitive judgments and the types of strategies they employ in that process as a function of their monitoring accuracy (very high, very low)?

The first two research questions were quantitative in nature, and hence, they necessitate a priori hypotheses, as follows.

*Hypothesis 1:* Students' grade difference on four assessments and metacognitive monitoring were expected to be associated, but this relation was expected to be stronger for postdictions than predictions (1a). Further, students' grade difference on four assessments and metacognitive monitoring accuracy were expected to be positively related to one another such that smaller differences between students' expected grade and actual grade should coincide with increased monitoring accuracy. Conversely, students' grade difference on four assessments was hypothesized to be negatively associated with monitoring bias (error) such that smaller differences between students' expected grade and actual grade should lead to decreased bias (1b). However, grade differences on four assessments and metacognitive monitoring indices were expected to be more strongly related within each assessment than across assessments (1c).

*Hypothesis 2:* The difference between students' expected grade and actual grade was expected to positively predict monitoring accuracy (2a), but negatively predict monitoring bias (2b).

## Method

### Research Design

The present study employed a non-random convenience sampling approach for its quantitative component and a purposive (extreme case) sampling approach for its qualitative component. The research design selec-

ted was an *explanatory sequential quantitative* → *QUALITATIVE* (quan → QUAL = explain significant factors) *mixed method* research design (Creswell & Plano Clark, 2017). In this specific mixed method design quantitative data are collected first and they help inform the qualitative component of the study; the qualitative data also help *explain* the quantitative data. The present study first collected quantitative data on students' expectations about their grade, confidence in performance judgments, and employed these to calculate *absolute* monitoring accuracy and bias indices for predictions and postdictions. Next, the five students with the highest monitoring bias (error) and the five students with the highest monitoring accuracy, based on the quantitative data, were selected as extreme cases, and the depth and quality of their responses to the eight qualitative open-ended questions regarding the development and refinement of their metacognitive judgments was compared across the two extreme subsets. Thus, these 10 students represented the extreme ends of the two sides of the metacognitive monitoring continuum—those with the highest accuracy and those with the greatest error.

## Participants

The study recruited 65 participants from a private university in Colombia. Of these, 11 identified as male (54 as female), and their age ranged from 17-41 ( $M = 20.55$ ;  $SD = 3.42$ ). All study participants were in their fourth or fifth semesters of an undergraduate degree in psychology at the time of the research. To be part of the research, participating students had to be enrolled in the University during the first and second semesters of 2020. It was considered important in the selection of the sample that the students who were part of the sample did not have a school report of major disturbance at the psychiatric level or a significant history of school repetition or long breaks between semesters.

Inclusion criteria were as follows: 1) participants had to be enrolled in the University during the first and second semesters of 2020; 2) students were required to be enrolled in the two courses from which participants were recruited (cognitive neuropsychology and child developmental neuropsychology); 3) none of the students were diagnosed with a neurological or psychiatric con-

dition according to the student's record on the comprehensive monitoring process implemented by the University; and 4) sign an informed consent form, indicating voluntary participation and permission for their data to be used for research purposes. The only exclusion criterion was that data were discarded for those who did not sign an informed consent form.

## Materials and Instruments

### Quantitative

#### Students' Expectations about their Grade.

Students' expectations about their grade were measured via four declarative knowledge tests prepared specifically for the class from which participants were recruited. These tests were of moderate difficulty and length. Each of the tests included 10 multiple-choice items, with four responses per item (one correct response and three distractors). The tests covered topics related to cognitive neuropsychology and child developmental neuropsychology. The items included in the tests were evaluated by independent experts prior to administration. Students' perceptions about their grade were calculated as the sum of correct responses across the 10 items for each test. The difference between students' expected score and their actual score (grade) were calculated by subtracting the former score from the latter for each test. This was done for both prediction and postdiction scores.

**Metacognitive Judgments and Absolute Monitoring Indices.** First, participants were asked to make judgments of future grade (predictions) and retrospective confidence judgments about their grade (postdictions), which were estimated on a continuous scale of 0-100 points (confidence from 0% to 100%). This measurement approach guaranteed a ratio scale rather than a matrix of correct and incorrect responses (e.g., Gamma coefficient), which only allows dichotomous ratings of confidence as low or high.

Next, feeling of knowing judgments (FOKs) (prediction judgments), were administered before each test, in which students indicated their level of confidence regarding the test they were going to take (on a scale of 0 to 100 points). Finally, retrospective confidence judgments (RCJs) (postdiction judgments) were provided after each test and were estimated on a scale of 0 to 100 points.

**Metacognitive Monitoring Accuracy and Bias.** To compute *absolute* monitoring accuracy, difference scores between students' expected grade and actual grade were subtracted from the students' confidence in performance judgments. Therefore, accuracy was evaluated by calculating the absolute value of the continuous difference score between students' confidence judgments and actual grade, such that zero corresponded to perfect monitoring accuracy whereas a higher non-zero score corresponded to lower monitoring accuracy because the difference between confidence and difference scores was greater (e.g.,  $75-75 = 0$  would indicate perfect accuracy whereas  $75-60 = 15$  indicates miscalibration, with higher values indicating poorer monitoring accuracy). *Absolute* monitoring bias, an index of error in judgments, was computed as the signed difference of students' confidence in judgments about their grade and actual grade, such that negative values correspond with underconfidence, or *illusion of not knowing*, and positive values correspond with overconfidence, or *illusion of knowing*.

### Qualitative

Qualitative data were collected by employing a questionnaire with eight open-ended semi-structured questions related to the process by which students develop and refine their metacognitive judgments. Sample questions included, "Explain how you arrived, or what aspects you considered, when making your metacognitive judgments?"; "Could you describe the process you underwent to develop and/or refine your metacognitive judgments regarding your expected grade with the actual grade in the partials?"; "What are some of the specific strategies you used while developing and/or refining your metacognitive judgments related to your expected grade for the exam?"; and "How do you know if your metacognitive judgments are accurate? In other words, what internal criterion (or criteria) do you use to evaluate your judgments?". The 10 participants for the qualitative portion of the study were purposefully selected based on the extreme case and maximal variation principle. This yielded five students with very high monitoring accuracy and five students with very low monitoring accuracy.

### Procedure

Data collection occurred throughout the year 2020 with four different groups from the cognitive neuropsychology and child neuropsychology class that were

in the fourth and fifth semester. During data collection, students learned about the objectives of the research, and, once they agreed to participate in the study, signed the informed consent form. The study adhered to the ethical guidelines provided by Resolution 0084330 of October 4, 1993, for studies considered to be of minimal risk to human beings from the country in which data were collected (Ministry of Health, 1993). Further, participants did not receive credit or any other form of incentive for participating in the study, and they were informed that they could withdraw from the study at any time without penalty.

The administration of the tasks related to the collection of monitoring accuracy and bias and difference scores between expected grade and actual grade was completed via the application of the four assessments during the semester, one for each unit of the topics addressed in the classes. All the assessments had a format that integrated four sections within each test: 1) estimation of confidence judgments about grades and expected grade in the prediction judgments (before the test); 2) application of the 10-question test along with estimation confidence judgements about grades and the expected grade in the postdiction judgments (after the test); and 3) the completion of the eight semi-structured, open-ended metacognitive questions from which the qualitative findings are based. Data collection for both the quantitative and qualitative phases was conducted collectively by all participants as part of the dynamics of the class meetings.

## Data Analysis

### Quantitative Analysis

Quantitative data were first screened for univariate outliers and tested for requisite statistical assumptions prior to data analysis. No extreme outliers were detected in the data, and hence, all 65 cases were retained for quantitative analyses. Failure to account for outliers in data analyses subjects the data to potential biases because of the undue influence these atypical scores exert on measures of central tendency and dispersion in inferential statistics (Tabachnick & Fidell, 2013). Quantitative data met all requisite statistical assumptions, including linearity, homoscedasticity, univariate normality, and lack

of collinearity, and thus, quantitative analyses proceeded without making a statistical adjustment to the data.

The first research question was answered by conducting bivariate, zero-order correlations, Pearson's  $r$ , for each of the four tests, including prediction and postdiction metacognitive judgments. The second research question was answered by conducting a series of ordinary least squares (OLS; standard) regressions. In each of the standard regressions, difference scores between students' expected grade and their actual grade across the four tests served as predictors and their composite absolute monitoring accuracy and bias (across the four tests) served as the criterion in each regression analysis, respectively. The squared multiple correlation coefficient,  $R^2$ , served as the measure of practical significance, or effect size estimate, of the findings. Cohen (1988) provided the following interpretive guidelines for the effect size,  $R^2$ : .010-.499 as small; .500-.799 as medium, and  $\geq .800$  as large.

### Qualitative Analysis

Qualitative data analysis to answer the third research question began with an initial read of participants' open-ended responses to the eight semi-structured questions to become familiarized with the data and to note those sections of the data that were most interesting to the objectives of the present study. Engagement in open reading of qualitative data permitted the individual coding of the data descriptively and to tease out individual meaning units in the data (Saldaña, 2013). Next, codes were developed based on individual meaning units to move deeper into the data and proceed with thematic analysis. More specifically, the analytical process included: 1) repeated readings of the data; 2) the combining of similar codes into categories; 3) identifying broad patterns across the data, resulting in themes; and 4) selection of representative quotations from participants to enrich and support substantive interpretations and meaning. Throughout this process the research team remained transparent and reflexive and continually returned to one of the present study's primary purposes, to more deeply understand the awareness participants had regarding their own metacognition, more specifically how they developed and refined their metacognitive judgments during learning. The research team constantly reflected on how assumptions shaped the interpretive process. Quali-

tative analysis reached an acceptable level of data saturation within the two groups of participants—that is, those with very low monitoring accuracy and those with very high monitoring accuracy. Even though they expressed it in slightly different wording, there was an overlap in the fundamental meaning of participants' experiences regarding how they developed and refined their metacognitive judgments. The differences in how participants with very low monitoring accuracy and very high monitoring accuracy experienced this metacognitive process was also evident in the data, as outlined below. The two authors of the present study independently analyzed the qualitative data to triangulate findings and obviate researcher bias. Inter-rater agreement was exceptionally high, Cohen's  $\kappa = .94$ . The minor disagreements were related to the labeling of the themes that emerged from the data, and these were resolved through a conference between the two raters, thereby reaching total agreement.

## Results

The reporting of results begins with the quantitative findings, first with an explanation of the general descriptive trends in the data, followed by a reporting of the findings of the two quantitative research questions. As the intent of the *explanatory sequential quantitative* → *QUALITATIVE mixed method* research design is for the qualitative findings to help support and explain the quantitative findings, the reporting of results continues with the reporting of the qualitative findings within- and between-groups to answer the third research question and concludes with a brief integration of the quantitative and qualitative findings.

### Quantitative

#### General Trends in Descriptive Data

Descriptive statistics for *absolute* monitoring accuracy and bias (predictions and postdictions) are displayed in Table 1 for each test and those for the difference between expected grade and actual grade are presented in Table 2. Tables 3 and 4 present the zero-order, bivariate correlation coefficients, Pearson's  $r$ , both within and across tests for monitoring accuracy and bias and the difference between expected grade and actual grade, respectively.

Descriptive statistics in Table 1 revealed that students were quite consistent in their monitoring accuracy not only within and across tests, but also between predictions and postdictions. The only exception is Test 4, in which students demonstrated the highest accuracy overall in their predictions and postdictions compared to the other three tests. Regarding bias, or errors in judgment, students tended to exhibit underconfidence (i.e., *illusions of not knowing*) not only within and across exams, but also between predictions and postdictions. Again, the only exception was Test 4, in which students tended to show slight overconfidence in their predictions (i.e., *illusions of knowing*), albeit they appeared to show errors in their postdictions. With respect to differences in expected grade versus actual grade across tests, descriptive statistics in Table 2 demonstrate that students' expectations about their grade were most well calibrated in Test 3 predictions and postdictions, but the least calibrated in Test 4, in which differences were greatest in both predictions and postdictions.

Table 1. *Descriptive Statistics for Absolute Metacognitive Monitoring Accuracy and Bias by Test*

| Variable | Predictions |      | Postdictions |      |
|----------|-------------|------|--------------|------|
|          | M           | SD   | M            | SD   |
| Accuracy |             |      |              |      |
| Test 1   | 1.01        | 0.64 | 1.15         | 0.76 |
| Test 2   | 1.26        | 0.78 | 1.16         | 0.75 |
| Test 3   | 1.12        | 0.60 | 1.10         | 0.71 |
| Test 4   | 0.16        | 0.37 | 0.00         | 0.00 |
| Bias     |             |      |              |      |
| Test 1   | -0.88       | 0.80 | -1.10        | 0.83 |
| Test 2   | -1.05       | 1.06 | -0.97        | 0.99 |
| Test 3   | -0.99       | 0.80 | -1.01        | 0.84 |
| Test 4   | 0.09        | 0.39 | 0.00         | 0.00 |

$N = 65$

*Note.* This table reports students' absolute monitoring accuracy, expressed as the absolute difference between their confidence judgments about their grade and actual grade. Absolute bias scores represent the signed difference of students' error, in which positive scores



indicate overconfidence (*illusion of knowing*) and negative values represent underconfidence (*illusion of not knowing*).

Table 2. Descriptive Statistics for the Difference between Expected Grade and Actual Grade by Test

| Variable                      | Predictions |      | Postdictions |      |
|-------------------------------|-------------|------|--------------|------|
|                               | M           | SD   | M            | SD   |
| Expected Grade – Actual Grade |             |      |              |      |
| Test 1                        | 0.15        | 0.79 | 0.10         | 0.71 |
| Test 2                        | 0.11        | 0.95 | 0.20         | 0.91 |
| Test 3                        | 0.01        | 0.80 | 0.02         | 0.82 |
| Test 4                        | 1.13        | 0.45 | 1.40         | 0.76 |

N = 65

Note. This table displays the difference scores between students’ expected grade and their actual grade on each of the tests.

As is evident, correlation coefficients were generally stronger within each exam and weaker between exams, a pattern that was also evident regarding predictions and postdictions within and across exams. This was consistent for monitoring accuracy and bias and differences between expected grade and actual grade (see Tables 3 and 4).

Table 3. Zero-Order Correlation Matrix for Metacognitive Monitoring Accuracy and Bias by Test

| Variable       | 1 | 2      | 3      | 4      | 5     | 6      | 7      | 8      | 9    | 10     | 11     | 12     |
|----------------|---|--------|--------|--------|-------|--------|--------|--------|------|--------|--------|--------|
| 1. Accuracy1+  | - | -.89** | .54**  | -.53** | .21*  | -.02   | .04    | .10    | .13  | -.17   | .09    | -.11   |
| 2. Bias1+      |   | -      | -.59** | .64**  | -.27* | .06    | -.09   | -.07   | -.13 | .21*   | -.07   | .13    |
| 3. Accuracy1-  |   |        | -      | -.96** | .21*  | -.17   | .14    | -.12   | .11  | -.06   | .11    | -.09   |
| 4. Bias1-      |   |        |        | -      | -.21* | .16    | -.12   | .09    | -.11 | .13    | -.11   | .15    |
| 5. Accuracy2+  |   |        |        |        | -     | -.58** | .74**  | -.37** | .28* | -.34** | .32**  | -.35** |
| 6. Bias2+      |   |        |        |        |       | -      | -.36** | .85**  | -.20 | .21*   | -.23*  | .22*   |
| 7. Accuracy2-  |   |        |        |        |       |        | -      | -.53** | .25* | -.19   | .30**  | -.27*  |
| 8. Bias2-      |   |        |        |        |       |        |        | -      | -.17 | .10    | -.21*  | .16    |
| 9. Accuracy3+  |   |        |        |        |       |        |        |        | -    | -.82** | .71**  | -.65** |
| 10. Bias3+     |   |        |        |        |       |        |        |        |      | -      | -.65** | .81**  |
| 11. Accuracy3- |   |        |        |        |       |        |        |        |      |        | -      | -.89** |
| 12. Bias3-     |   |        |        |        |       |        |        |        |      |        |        | -      |

\*\* p < .01      \* p < .05

Note. The number after each metacognitive monitoring measure represents each test. “+” represents a prediction while “-” represents a postdiction.

N = 65

Table 4. Zero-Order Correlation Matrix for the Difference between Expected Grade and Actual Grade by Test

| Variable  | 1 | 2     | 3   | 4     | 5    | 6     | 7      | 8     |
|-----------|---|-------|-----|-------|------|-------|--------|-------|
| 1. Test1+ | - | .80** | .01 | -.08  | .16  | .10   | .19    | -.04  |
| 2. Test1- |   | -     | .02 | .01   | .03  | .02   | -.01   | -.15  |
| 3. Test2+ |   |       | -   | .84** | .24* | .21*  | -.14   | -.19  |
| 4. Test2- |   |       |     | -     | .08  | .11   | -.30** | -.19  |
| 5. Test3+ |   |       |     |       | -    | .79** | .09    | -.09  |
| 6. Test3- |   |       |     |       |      | -     | -.09   | -.08  |
| 7. Test4+ |   |       |     |       |      |       | -      | .38** |
| 8. Test4- |   |       |     |       |      |       |        | -     |

\*\*  $p < .01$ \*  $p < .05$ 

Note. “+” represents a prediction while “-” represents a postdiction.

N = 65

### Main Analyses

#### RQ1: Relation between Grade Difference Scores and Monitoring Accuracy and Bias

Regarding the first research question, correlation patterns show that monitoring accuracy and bias were inversely related and that correlations were stronger within predictions and postdictions than between predictions and postdictions. Also, accuracy was positively related to predictions and postdictions, a pattern that was consistent for bias as well. Difference scores between

expected grade and actual grade were also positively related between predictions and postdictions. Interestingly, correlational patterns between monitoring accuracy and the difference score between expected grade and actual grade were negative across both predictions and postdictions, indicating that lower difference scores between expected grade and actual grade coincided with greater monitoring accuracy. Likewise, the positive association between monitoring bias and difference scores suggests that higher difference scores between expected grade and actual grade coincided with greater bias or error.

Table 5. Zero-Order Correlation Matrix between Composite Monitoring Accuracy, Bias, and Difference between Expected Grade and Actual Grade for Predictions and Postdictions

| Variable                | 1 | 2     | 3      | 4      | 5      | 6      |
|-------------------------|---|-------|--------|--------|--------|--------|
| 1. Prediction Accuracy  | - | .68** | -.73** | -.56** | -.64** | -.37** |
| 2. Postdiction Accuracy |   | -     | -.52** | -.82** | -.50** | -.63** |
| 3. Prediction Bias      |   |       | -      | .71**  | .79**  | .58**  |
| 4. Postdiction Bias     |   |       |        | -      | .68**  | .82**  |
| 5. Prediction EG-AG     |   |       |        |        | -      | .65**  |
| 6. Postdiction EG-AG    |   |       |        |        |        | -      |

\*\*  $p < .01$ 

Note. EG = expected grade; AG = actual grade.

N = 65

## RQ2: Predictive Effect of Grade Difference Scores on Monitoring Accuracy and Bias

The first regression model with composite monitoring accuracy as the criterion was statistically significant,  $F(2,61) = 20.50, p < .001, R^2 = .52$ . Both composite prediction ( $b = -.33 [CI_{95\%} = -.53, -.14], \beta = -.49$ ) and postdiction ( $b = -.22 [CI_{95\%} = -.45, -.05], \beta = -.29$ ) difference scores between expected grade and actual grade negatively predicted composite monitoring accuracy, albeit composite prediction difference scores between expected grade and actual grade was the best predictor. The second regression model with monitoring bias as the criterion was also significant,  $F(2,61) = 73.72, p < .001, R^2 = .69$ . As with the model with monitoring accuracy, both composite prediction ( $b = .24 [CI_{95\%} = .06, .42], \beta = .24$ ) and postdiction ( $b = .75 [CI_{95\%} = .55, .97], \beta = .66$ ) difference scores between expected grade and actual grade positively predicted composite monitoring bias; however, unlike the model with monitoring accuracy, composite postdiction difference scores between expected grade and actual grade was the best predictor.

## Qualitative

Regarding the answer to the third research question, participants' qualitative data was coded to develop themes in an iterative process of inductive reasoning between the two groups of extreme cases (those with very high and very low monitoring accuracy based on the quantitative data). The in-depth analysis led to the identification of four themes that permeated the processes selected by individuals with very high and very low monitoring accuracy: 1) effort/preparation; 2) strategies; 3) planning; and 4) evaluation. These themes align with the two theoretical frameworks employed in the present study, Ekflides' (2011) MASRL and Winne and Hadwin's (2008) MPM, and they also align with previous qualitative research on this topic (Gutierrez de Blume et al., 2017).

### Effort/Preparation

Individuals use cognitive skills and strategies as well as metacognitive knowledge and regulation to successfully prepare for assessments of their learning. Typi-

cally, students are viewed as effective self-regulators of their learning when they can accurately determine what they know and do not know about a given topic or content area. Being able to determine what they know and do not know about a given topic allows learners to focus attention and other cognitive resources on material they have not yet mastered and spend less time reviewing material they already know, thereby effectively demonstrating self-regulated learning behavior.

When describing their effort and preparation for the exams, individuals with very high monitoring accuracy stated:

“Estudiar y luego revisar y comparar con lo que preguntan. Aplicar el método de estudiar haciéndome preguntas yo misma.” (Participant 55). (English: “Study and then review and compare with what the item asks. Apply the method of studying by asking myself questions.”)

When asked how well they will perform on some future assessment of their knowledge, students with very high accuracy came closer to accurately predicting their actual performance because they have superior comprehension monitoring regarding the knowledge of their past and present performance.

Individuals with very low monitoring accuracy, conversely, struggled to express their level of effort and preparation:

“El nivel de confianza que tengo frente a los exámenes a presentar.” (Participant 60) (English: “The level of confidence that I have regarding the exams while presenting.”)

“El considerar si mis respuestas son buenas o no, y qué tan segura estoy de ellas.” (Participant 35) (English: “Considering if my answers are good or not, and how sure I am in them.”)

Learners who are less metacognitively aware do not always accurately understand what they know or do not know about a topic, and thus, often demonstrate they are less capable of regulating their learning (i.e., they may be lacking in planning, evaluation, information management, or comprehension monitoring skills) and are prone to too much confidence or insufficient

confidence when it comes to developing metacognitive monitoring.

### *Strategy Selection/Implementation*

Learning strategy use refers to students' ability to invoke and apply strategies that are conducive to enhanced learning outcomes. The literature on this topic has distinguished between shallow strategies (e.g., surface-level strategies such as rote learning and rehearsal) and deep or meaningful strategies that are more closely aligned to accurate metacognitive monitoring (Dinsmore & Alexander, 2012) such as reflecting, planning, and evaluation.

Students with very high monitoring accuracy responded as follows:

“Estudiar juiciosamente. Tener en cuenta los aspectos importantes que podrían entrar en el examen. Tener en cuenta lo que la profesora dice que va a evaluar cuando esta explicando una tematica.” (Participant 55). (English: “Study judiciously. Take into account the important aspects that could go into the test. Take into account what the teacher says that she will evaluate when she is explaining a topic.”)

As described in previous research (Dinsmore & Alexander, 2012), deep cognitive strategies such as reflecting, planning, and evaluating, are connected to more accurate metacognitive monitoring. Participants shared their intentionality in these strategies: reflecting (“Elaborar mapas ...”/Elaborate maps ...), planning (“... ademas antes de revisar las respuestas lo realice con tiempo ...”/besides, before reviewing my answers I took my time), and evaluating (“Realizar un proceso de comparación entre lo que estudie ...”/ Undergo a process of comparison between what I studied).

Individuals with high accuracy placed a high priority on consistently invoking reflective practices into their learning process, including the importance of self-awareness.

Students with very low monitoring accuracy did not evince any adaptive learning strategies. Instead, they blamed external forces for their lack of understanding or lack of preparation:

“... si bien es diferente cuando es un taller pero con los exámenes por mi parte hay cierta insertidumbre incluso si me sienten preparado.” (Participant 60) (English: “... although it is different when it is a workshop, but with the tests, on my part, there is a certain uncertainty even if I feel prepared.”)

In addition, students with low accuracy struggled with identifying strategies based on the demands of the exams (i.e., conditional knowledge):

“Leer, subrayar ... pero no se cuando debo cambiar de estrategia.” (Participant 49) (English: “Read, underline ... but I don't know when I should change strategy.”)

Evidently, students with very high and very low monitoring accuracy employed substantively different learning strategies, with accurate monitors not only recognizing the importance of incorporating different educational strategies, but also implementing deep strategies, while inaccurate monitors struggled with using ineffective, maladaptive learning strategies.

### *Planning*

Students with very high monitoring accuracy plan on many different levels, including attending class regularly, knowing personal strengths and weaknesses, and developing a deep, personal sense of ownership of their learning process:

“Siempre tomo notas y estoy presente.” (Participant 55) (English: “I always take notes and I am present.”)

Accurate monitors consistently expressed specificity in their planning process:

“Tiempo y horario específico dedicado a cada temática, algunas veces me siento a hablar con algunas compañeras para así dar claridad a lo que cada una entendió.” (Participant 13). (English: “Specific time and schedule dedicated to each topic, sometimes I sit down to talk with some classmates to clarify what each one of us understood.”)

Individuals with very low monitoring accuracy, on the other hand, seemed inflexible and immutable in their planning process:

“Intento no cambiar mucho la forma en que estudio. Se mantiene prácticamente igual ... eso me ayuda a mantenerme constante en mis notas.” (Participant 35) (English: “I try not to change the way I study too much. It stays pretty much the same ... that helps me stay consistent on my grades.”)

This exemplifies how inaccurate monitors did not seem to understand the importance of the complexities and nuances of learning that necessitate flexibility in planning. Planning differently, according to task demands, allows learners to adapt strategies for success. When inaccurate monitors are unable to adapt, they are unable to be as successful as their accurate counterparts.

### ***Evaluation***

Evaluation is described as the metacognitive act of reflecting after a learning episode and making appropriate adjustments for more effective future learning (Schraw & Dennison, 1994).

Individuals with very low monitoring accuracy, for example, tended to have an overabundance of confidence in their performance:

“... considero que la manera en que lo hice siempre es la adecuada.” (Participant 4) (English: “... I think the way I did it is always the right way.”)

“... no use ninguna estrategia diferente.” (Participant 35) (English: “... I don't use any different strategy.”)

Accurate monitors, in contrast, evaluate their understanding through reflection and make changes when necessary:

“Dependiendo de la situación, uso mapas, resúmenes, hablar con compañeros, y hacerme una evaluación antes de cualquier tarea.” (Participant 55) (English: “Depending on the situation, I use maps, summaries, talking with classmates, and do a self-evaluation before any task.”)

## **Integration of Quantitative and Qualitative Findings**

The first two questions of the present study, both of which were quantitative, sought to explore the relation between students' perceptions about their grade and their actual grade and metacognitive monitoring accuracy and error (bias) as well as examining the predictive effect of difference scores between expected grade and actual grade on monitoring accuracy and bias. Results revealed that difference scores between expected grade and actual grade and monitoring accuracy and bias were correlated, such that smaller differences between expected grade and actual grade corresponded with greater accuracy and lower bias. Further, composite difference scores between expected grade and actual grade significantly predicted composite monitoring accuracy and bias. However, prediction composite difference scores between expected grade and actual grade were a better predictor of composite monitoring accuracy whereas postdiction composite difference scores between expected grade and actual grade were a better predictor of composite monitoring bias. Regarding monitoring accuracy, a lower difference score between expected grade and actual grade significantly predicted higher accuracy whereas greater difference between expected grade and actual grade significantly predicted monitoring bias. The qualitative findings of the 10 participants selected via the extreme case approach (five students with the highest monitoring accuracy and five students with the lowest accuracy) supported these quantitative findings and help explain why more precise expected grade compared to actual grade differences coincide with greater monitoring accuracy and decreased bias. Students with very high monitoring accuracy not only manifested superior effort and preparation as they learn, but they also employ deeper, more adaptive learning strategies, plan their learning more successfully, and evaluate future learning based on previous learning episodes more effectively.

### **Discussion**

The objectives of the present study were to: 1) examine the relation between difference scores between

expected grade and actual grade and metacognitive monitoring accuracy and bias; 2) investigate the predictive effects of the difference between students' expected grade and their actual grade on monitoring accuracy and bias; and 3) explore the process students undertake to develop and refine their metacognitive judgments and the types of strategies they invoke during this process. Regarding the first objective, results in Tables 1 and 2 showed that students were not only generally consistent in their expectations about their grade and monitoring accuracy and bias across tests, but also between predictions and postdictions. Further, results for the second objective indicated that composite differences in expected grade and actual grade significantly predicted composite monitoring accuracy and bias in the theoretically expected direction (i.e., lower differences coincided with increased accuracy and higher differences coincided with increased bias). However, differences in expected and actual grade at prediction was the best predictor of monitoring accuracy whereas differences in expected and actual grade at postdiction was the best predictor of bias. These results are interesting from two perspectives. The first is that the work in the classroom with metacognitive judgments before (predictions) and after a task (postdictions) contribute to generating greater accuracy in the monitoring process invoked by students regarding their grade expectations on the assessments. The second is that it shows that students can be accurate in their monitoring prior to engaging in a task. Both findings are consistent with studies that have described improvements in students' adjustments about their expected grade derived from the opportunities for self-reflection and self-generated feedback practice that are afforded by each of the different tests throughout the semester, which contributes to improving monitoring accuracy (Cogliano et al., in press).

These results also align with the increase in accuracy that students exhibited in the fourth test (in both predictions and postdictions) compared to previous tests, a result that coincides with studies that highlight the importance of self-generated feedback (Moore & Chang, 2009). In the present study, students received substantive, individualized feedback throughout the semester regarding performance on previous assessments, and hence, were able to benefit from this performance feedback loop. Likewise, it coincides with research that

has described an improvement in the monitoring accuracy as students complete each assessment because, in the last one, students may be more familiar with the structure of the test or with the type of task (Dunlosky et al., 2013).

Regarding confidence judgments regarding their grade, manifested as the presence of bias in metacognitive monitoring, students showed a lack of confidence (*illusion of not knowing*) not only within each test, but also between the different tests, including between predictions and postdictions. These findings are congruent with studies that conclude that the estimation of confidence is one of the most stable traits within people, such as personality, self-concept, and cognitive style, among others (Ozturk, 2020). The excess in confidence (*illusion of knowing*) in predictions of the fourth test is consistent with the little adjustment that students can make to their initial confidence level, more specifically, before starting the assessment and having an opportunity to process the demands of the task. This outcome occurs because students rely mainly on their domain-specific self-concept for the type of task (test) and not necessarily on their performance on the test itself (Händel et al., 2020).

Regarding the differences between students' expected grade and their actual grade, Table 2 shows that students' expectations about their grade were more accurate in the prediction and in the postdiction of the third test when compared to the fourth test. This result could be explained by the hypothesis of the effect of "lack of confidence with practice" proposed by Koriat et al. (2002), known by its acronym, "underconfidence-with-practice" (UWP). The UWP effect has been associated with the repeated presentation of a stimulus, in which the effects of practice on learning judgments and working memory were compared. Findings revealed that judgments showed successive decreases in confidence, such that recall predictions became markedly lower than recall performance because study and test practice affected monitoring accuracy, thereby reducing the difference between general judgments of recall and actual recall (Koriat et al., 2002).

Finally, regarding the third research objective, qualitative evidence suggests that there was relative consistency in how students with the lowest accuracy (and thus, highest monitoring bias) and the highest accuracy

expressed the process they undergo to develop and refine metacognitive judgments, including the types of strategies they employ throughout this process. Nevertheless, there were stark differences in the scope and depth of the quality of responses between the two groups of extreme cases. Those with the highest monitoring accuracy not only better understood their metacognition, but the strategies, cues, and evaluative criteria they employed during learning were much more sophisticated and based on previous feedback and experiences. Those with the lowest accuracy, on the other hand, provided superficial responses and strategies, and they also did not truly engage in refining their metacognitive judgments, tending to ignore previous experiences in feedback in developing future metacognitive judgments.

Research conducted in support of SRL theory indicates that effort plays a key role in learners successfully meeting their self-choice goals. For example, learners who are more capable of managing and controlling their effort on learning tasks are more apt to show improved performance, confidence in their performance, and a more accurate monitoring (Efklides, 2001; Winne & Hadwin, 2008). More specifically, these learners exert additional effort on learning more complex information and less effort on simpler information, and they know when information has been learned so as not to expend unnecessary effort on already-learned information. Thus, the qualitative findings of the present investigation support the notion that individuals with greater monitoring accuracy are also more proficient in their self-regulated learning, a conclusion that coincides with Winne and Hadwin's (2008) MPM and Efklide's (2011) MASRL.

### **Blending Quantitative and Qualitative Findings**

Quantitative findings from the present study suggest that students are generally consistent in their predictions and postdictions across different tests. However, consistency is greatest within tests than across tests, and that predictions are more highly related to predictions in other tests than postdictions (the pattern was the same for postdictions). Equally as important, quantitative data demonstrated that monitoring accuracy and bias differed between tests, in which students showed varying degrees

of accuracy and bias as a function of test. Nevertheless, knowing that differences in metacognitive monitoring accuracy and bias differs, while interesting, does not help researchers understand *why* and *how* this occurs.

The qualitative results help explain why and how differences are evident between students who exhibit accurate monitoring and those who exhibit poor monitoring, and that these differences are due to four aspects of learning: 1) effort/preparation; 2) strategy type and selection; 3) planning; and 4) evaluation. More accurate monitors more adequately prepare for learning, and thus, can better gauge how much effort they need to expend to succeed in learning whereas poor monitors engage in little, if any, preparation for learning, and hence, apply too little effort or too much effort. Those with very high monitoring accuracy were also more skilled at strategy selection and sequencing, choosing deeper learning strategies more suitable to the task, especially for more complex material. Those with very low monitoring accuracy, on the other hand, selected superficial or shallow learning strategies and were unable to apply those strategies successfully during learning. Finally, students with high monitoring accuracy were able to employ two key elements of the regulation component of metacognition, planning and evaluation, more effectively. By understanding the demands of the task more completely, these students can more effectively know needed resources and anticipate potential pitfalls in task completion and, once the learning episode concluded, were able to evaluate learning and adjust accordingly for enhanced future learning more successfully. Thus, the qualitative findings helped elucidate some of the reasons behind the quantitative findings, and they support the only other qualitative study on these topics to date (Gutierrez de Blume et al., 2017).

### **Implications and Avenues for Future Research**

Optimal monitoring accuracy is arguably necessary for sustained effort while learning. Learners are likely to persevere while tackling a difficult problem if previous experience has demonstrated that they will ultimately succeed in solving it. Nevertheless, if students believe that efforts to master a subject or solve a pro-

blem are fruitless, the likelihood that they will persist in such efforts decreases. However, if students feel that they have mastered a topic, they are less likely to expend additional time studying it. Therefore, poor monitoring accuracy can be expected to result in students misallocating effort in wasteful endeavors. On the other hand, improved monitoring accuracy should permit learners to become more aware of their own cognitive strengths and weaknesses, thereby improving their ability to determine where to best expend effort (Gutierrez & Schraw, 2015).

In a series of studies, Gutierrez and colleagues (Schraw et al., 2013; Schraw et al., 2014; Gutierrez et al., 2016; Gutierrez de Blume et al., 2021) called attention to the need to better understand the processes underlying metacognitive monitoring, which forms the foundation for students' ability to develop accurate monitoring. These studies demonstrated that learners experience and engage in related yet distinct metacognitive processes when making accurate and erroneous judgments. Findings suggest that, by more deeply understanding metacognitive monitoring processes, researchers and practitioners could develop more specific, effective, and targeted educational interventions tailored to specific metacognitive profiles. According to the quantitative and qualitative findings of the present study, accurate and inaccurate monitors experience metacognition in fundamentally different ways. Thus, finding ways to better support, model, and scaffold more effective metacognitive monitoring across the lifespan, especially for those with poor monitoring accuracy, is essential. For teachers this means providing better, more individualized instruction to either specifically target the reduction of erroneous monitoring, increase accuracy, or both, depending on the needs of the individual learner. Presumably, these strategies could assist learners in appropriately adjusting confidence in what they know and do not know to coincide with what they *actually* know and do not know.

Future research on metacognitive monitoring should more closely examine the role of metacognitive monitoring accuracy and error in the development of metacognitive judgments and how these influence learning outcomes. As previous studies have demonstrated, monitoring accuracy and error not only develop in distinct ways, but errors in judgment are also unique insofar

as overconfidence or *illusions of knowing* appear to develop differently than underconfidence or *illusions of not knowing* (Gutierrez de Blume, 2020; Gutierrez et al., 2016; Gutierrez de Blume, 2021). These studies can subsequently inform educational interventions more focused on the metacognitive profile of the student, and thus, honor individual differences. Likewise, future studies should strive to control what metacognition researchers have called the “study granularity” problem (Pieschl, 2009; Rovers et al., 2019), as this can influence metacognitive monitoring. Study granularity refers to the control of extremely fine aspects such as the different ways of measuring the construct, the difficulty of the test items, the difficulty of the texts, the cognitive skills underlying metacognitive performance, and personality factors, as well as the categorization of the samples between subgroups that discriminate students with high and low difference scores between expected grade and actual grade. This should help researchers to see the relation between metacognitive skills and other variables at a finer grain than in previous work.

### Methodological Reflections and Limitations

Every study has limitations. A significant limitation of the present study is the lack of observational data. Communication is more than a verbal interaction; no interaction with the students was recorded, and therefore, these significant data could not be analyzed. Much can be gained by examining the personality of a participant in an interview. However, because the qualitative data was collected via online open-ended semi-structured questions, observations could not be captured. Moreover, the quantitative portion of the study only included 65 participants, a relatively small sample size. Nevertheless, despite these limitations, the study has strengths worth mentioning.

First, the study employed a truly mixed method research design. No study in metacognitive monitoring to date has employed such a design. In addition, the study employed objective measures for its quantitative component rather than self-report surveys. Finally, the study occurred in an ecologically valid setting, and thus, the inferences and conclusions drawn from the data are



more contextually valid. Therefore, the present study contributes substantively to research on metacognitive monitoring not only empirically, but methodologically as well.

Considering the preponderance of quantitative studies on metacognitive monitoring and the broad findings about the topic, the evidence reported in this mixed method study allows researchers to delve into a finer and more detailed explanation of the phenomenon. Mixed method studies such as this one better elucidate the way in which students from the two extremes of the metacognitive monitoring continuum (i.e., very high monitoring accuracy and very low accuracy) experience their decision-making process during learning. Further, findings show the process of online monitoring of students' expectations of their grades and the adjustment of the regulation of effort in relation to the establishment of goals, the value of the task, and the degree of motivation of these two groups of learners. Finally, the present study demonstrates some of the possibilities that students have at their disposal to be progressively more aware of what they know and of what they do not know about a topic, which can lead them to make better directed and more efficient decisions during learning.

## References

- Azevedo, R. (2020). Reflections on the field of metacognition: issues, challenges, and opportunities. *Metacognition and Learning, 15*(2), 91–98. <https://doi.org/10.1007/s11409-020-09231-x>
- Avhustiuk, M. M., Pasichnyk, I. D., & Kalamazh, R. V. (2018). The Illusion of knowing in metacognitive monitoring: Effect of the type on information and of personal, cognitive, metacognitive, and individual psychological characteristics. *Europe's Journal of Psychology, 14*(2), 317–341. <https://doi.org/10.5964/ejop.v14i2.1418>
- Cogliano, M., Bernacki, M., & Kardash, C. (in press). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. *Journal of Educational Psychology. Advance Online Publication*. <https://doi.org/https://doi.org/10.1037/edu0000624>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Creswell, J., & Plano Clark, V. (2017). *Designing and conducting Mixed methods research*. SAGE Publications, Inc. <https://us.sagepub.com/en-us/nam/designing-and-conducting-mixed-methods-research/book241842>
- Dinsmore, D., & Alexander, P. (2012). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review, 24*(4), 499–567. <https://doi.org/10.1007/s10648-012-9198-7>
- Dinsmore, D., & Parkinson, M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what means for calibration. *Learning and Instruction, 24*, 4–14. <http://dx.doi.org/10.1016/j.learninstruc.2012.06.001>
- Dunlosky, J., & Rawson, K. A. (2019). *The Cambridge handbook of cognition and education*. Cambridge University Press. <https://doi.org/10.1017/9781108235631>
- Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, Supplement, 14*(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. In A. Efklides, J. Kuhl, & R. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 297–324). Kluwer Academic. <https://psycnet.apa.org/record/2006-20747-016>
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*(1), 6–25. <https://doi.org/10.1080/00461520.2011.538645>
- Efklides, A., Schwartz, B., & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role? In D. Schunk & J. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp.

- 64–82). Taylor & Francis Group. <https://psycnet.apa.org/record/2017-45259-005>
- Follmer, D. J., & Clariana, R. (2020). Predictors of adults' metacognitive monitoring ability: The roles of task and item characteristics. *Journal of Experimental Education*, 1–23. <https://doi.org/10.1080/00220973.2020.1783193>
- Gutierrez, A. P., & Price, A. F. (2017). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *Journal of Experimental Education*, 85(3), 486–500. <https://doi.org/10.1080/00220973.2016.1180278>
- Gutierrez, A., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *Journal of Experimental Education*, 83(3), 386–404. <https://doi.org/10.1080/00220973.2014.907230>
- Gutierrez, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction*, 44, 1–10. <https://doi.org/10.1016/j.learninstruc.2016.02.006>
- Gutierrez de Blume, A. P. (2020). Efecto de la instrucción de estrategias cognitivas en la precisión del monitoreo de los alumnos universitarios estadounidenses. *Revista Tesis Psicológica*, 15(2), 1–29. <https://doi.org/https://doi.org/10.37511/tesis.v15n2a9>
- Gutierrez de Blume, A., & Montoya, D. (2020). Relación entre factores de personalidad y metacognición en una muestra de estudiantes del último semestre de formación de programas de licenciatura en Educación en Colombia. *Educación y Humanismo*, 22(39), 1–20. <https://doi.org/10.17081/eduhum.22.39.4048>
- Gutierrez de Blume, A. P.; Montoya, D., Hederich, C. (in press). An exploratory study of the relation between cognitive style and metacognitive monitoring in a sample of Colombian university students. *Psicología desde el Caribe*.
- Gutierrez de Blume, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2021). General accuracy a general error factors in metacognitive monitoring and the role of time on task un predicting metacognitive judgments. *Revista CES Psicología*, 14(2), 179–208 <https://doi.org/10.21615/cesp.5494>
- Gutierrez de Blume, A. P., Wells, P., Davis, C., & Parker, J. (2017). “You can sort of feel it”: Exploring metacognition and the feeling of knowing among undergraduate students. *The Qualitative Report*, 22(7), 2017–2032. <https://doi.org/10.46743/2160-3715/2017.2802>
- Händel, M., de Bruin, A., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, 15(1), 51–75. <https://doi.org/10.1007/s11409-020-09220-0>
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131(2), 147–162. <https://doi.org/10.1037/0096-3445.131.2.147>
- Mathabathe, K. (2019). Factors underlying metacognitive judgements in foundation chemistry. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(5). <https://doi.org/10.29333/ejmste/105868>
- Mathabathe, K., & Potgieter, M. (2014). Metacognitive monitoring and learning gain in foundation chemistry. *Chemistry Education Research and Practice*, 15(1), 94–104. <https://doi.org/10.1039/c3rp00119a>
- Ministry of HealthRovers (1993). Resolución Número 8430 de Octubre 4 de 1993. Bogotá, Colombia: Republica de Colombia. Retrieved from <https://www.minsalud.gov.co/sities/rid/Lists/Biblioteca-Digital/RIDE/DE/DIJ/RESOLUCION-8430>
- Moores, T. T., & Chang, J. C. J. (2009). Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. *Information and Management*, 46(2), 69–76. <https://doi.org/10.1016/j.im.2008.11.006>
- Nelson, T. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item: Comments on Schraw (1995). *Applied Cognitive Psychology*, 10(3), 257–260. [https://doi.org/10.1002/\(SICI\)1099-0720\(199606\)10:3<257::AID-ACP400>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1099-0720(199606)10:3<257::AID-ACP400>3.0.CO;2-9)

- Ozturk, N. (2020). An analysis of teachers' metacognition and personality. *Psychology and Education*, 57(1), 40–44. <https://doi.org/10.17762/pae.v57i1.6>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 1–28. <https://doi.org/10.3389/fpsyg.2017.00422>
- Pieschl, S. (2009). Metacognitive calibration-an extended conceptualization and potential applications. *Metacognition and Learning*, 4(1), 3–31. <https://doi.org/10.1007/s11409-008-9030-4>
- Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, 14(1). <https://doi.org/10.1007/s11409-019-09188-6>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2<sup>nd</sup> ed.). SAGE Publications, Inc. <https://us.sagepub.com/en-us/nam/the-coding-manual-for-qualitative-researchers/book243616>
- Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460–475. <https://doi.org/https://doi.org/10.1006/ceps.1994.1033>
- Schraw, G. (2009a). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>
- Schraw, G. (2009b). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 415–429). Routledge.
- Schraw, G., Kuch, F., & Gutierrez, A. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, 24, 48–57. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2012.08.007>
- Schraw, G., Kuch, F., Gutierrez, A., & Richmond, A. (2014). Exploring a three-level model of calibration accuracy. *Journal of Educational Psychology*, 106(4), 1192–1202. <https://doi.org/10.1037/a0036653>
- Serra, M. J., & Metcalfe, J. (2009). Effective Implementation of Metacognition. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 278–298). Erlbaum. <https://doi.org/10.4324/9780203876428>
- Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6<sup>th</sup> ed.). Pearson. <https://www.pearson.com/us/higher-education/program/Tabachnick-Using-Multivariate-Statistics-6th-Edition/PGM332849.html>
- Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Taylor & Francis. <https://www.routledge.com/Motivation-and-Self-Regulated-Learning-Theory-Research-and-Applications/Schunk-Zimmerman/p/book/9780805858983>