



Article

Exploring University Students' Justifications for Making Metacognitive Judgments of Learning

Athanasios Kolovelonis 

Department of Physical Education and Sport Science, University of Thessaly, 42100 Trikala, Greece; akolov@uth.gr

Abstract: The accuracy of students' judgments has important implications for their learning and performance in educational settings. However, little is known about how students make these judgments. This study explored university students' justifications for making their judgments of learning in a developmental psychology course. Two independent samples were involved, including a total number of 115 senior sport students. Participants responded to a knowledge test and provided their judgments at the local (Sample 1) or at the global level (Sample 2) and then provided their justifications for making these judgments. Students' justifications for making their judgments were classified in ten categories, including the study of the learning materials, the confidence for answering (or not) correctly, the memory, the general knowledge, the knowledge of the answer and a general reference to common sense, experience, lectures, and judgment. Variations in the frequency of these justifications were found across the local and the global level, low and high accurate students, and low and high performers. These results are discussed regarding their theoretical and practical implications for undergraduate students' learning.

Keywords: calibration accuracy; local judgments; global judgments; undergraduate students; metacognition



Citation: Kolovelonis, A. Exploring University Students' Justifications for Making Metacognitive Judgments of Learning. *Trends High. Educ.* **2023**, *2*, 421–433. <https://doi.org/10.3390/higheredu2030025>

Academic Editor: Hani Morgan

Received: 31 May 2023

Revised: 28 June 2023

Accepted: 4 July 2023

Published: 6 July 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In educational settings, students are frequently involved in tasks, tests, or exams and face the challenge to judge their knowledge, learning, or performance. Although the accuracy of these judgments has attracted research interest [1,2], little is known about how students make these judgments. Thus, the present study, adopting the calibration research paradigm, focused on the justifications provided by undergraduate sport students regarding their judgments of learning in an academic course.

The term calibration has been widely used to describe the discrepancy between judged and actual performance [3] and can be measured either at the local or the global level [4]. In particular, students may judge their performance in every single item of a test (i.e., local level) or provide a cumulative judgment for all items of the test (i.e., global level). Then, they perform the test and compare their judgments with their actual performance [4]. If the judged performance is close to the actual performance, students are considered well-calibrated. In the case that judged performance is higher compared to the actual performance, students are considered overestimators, and if the judged performance is lower than the actual performance, students are underestimators.

Calibration has important implications for learning and performance in educational settings. Indeed, focusing on the discrepancy between judged and actual performance, calibration can inform students how close to reality their beliefs about learning and performance are [5]. Moreover, calibration accuracy seems to be associated with executive functions [6]. Most importantly, calibration can affect students' decisions during learning and performance, as it is associated with their motivation and self-regulated learning [7]. In particular, miscalibrated students who believe that their capabilities are lower than they actually are may avoid challenging tasks and set lower learning goals, thus limiting their

potential for mastering new skills [8]. In contrast, students who erroneously judged that they have reached a high level of performance may be reluctant to try hard to further develop their skills, or may set unrealistic and unachievable goals.

Considering the important implications of calibration, research has widely examined students' calibration accuracy in various settings involving various types of tasks. One prominent finding in calibration research is that students are usually inaccurate, with a tendency to overestimate their learning or performance. This tendency has been supported by research in academic settings [9,10], in sport and physical activity settings [1,11,12], and in physical education [13–15]. However, evidence regarding an underestimation of performance among university students has also been reported [16].

Some research has focused on investigating factors related to students' calibration to explain this tendency to miscalibration. This research has shown that miscalibration was associated with task difficulty and low performance. In the more difficult tasks, students tend to be less accurate [17] and overestimate their performance [18], while high performers are usually more accurate in predicting or postdicting their performance [2,19–21]. For example, Valdez [22] found that students' absolute accuracy was significantly correlated with exam performance in an undergraduate course in language acquisition. Research in physical education has shown that higher self-efficacy and task goal orientation was positively associated with calibration accuracy [23]. However, these associations between goal orientations and monitoring accuracy were not supported in a study with university students [24]. Moreover, task-related characteristics such as the shooting position [18], the sport participation out of school [25], and the predictions regarding their peers' performance [26] were associated with calibration accuracy.

Except for examining students' accuracy of metacognitive judgments of performance, an interesting research question is how students form these metacognitive judgments. Exploring the factors involved in the process of making metacognitive judgments will increase our understanding regarding the complex process of calibration accuracy. It has been theorized that the cues used for making metacognitive judgments can be classified into two broad categories, beliefs and experiences. The information-based cues refer to what students consciously believe about their own knowledge, competence, or memory capacities, while the experience-based cues involve students' experiences derived during task engagement, including how familiar an answer seems and the feelings of knowing [27]. However, only a few studies have focused on exploring this issue, examining students' origins of response confidence ratings. Dinsmore and Parkinson [28] examined students' justifications for making their metacognitive judgments using as a framework Bandura's [29] model of reciprocal determinism that emphasizes the reciprocal relations between personal, behavioral, and environmental influences in students' efforts to self-regulate their learning. Similarly, students may base their judgments not only on personal factors but also on the nature of the task. Indeed, they found that undergraduate students based their confidence judgments on prior knowledge, characteristics of the text or the item, guessing, and a combination of these categories. A recent study [30] supported these findings, showing that undergraduate students provided a variety of justifications for their self-evaluations' judgments regarding their written responses, composed based on multiple texts. These justifications included task (e.g., the number of texts used), context (e.g., perceived time limit), and person-related (e.g., prior knowledge) factors. Hacker et al. [2] used attributional style as a framework to examine college students' explanations of metacognitive judgments. They found that the most reported explanations focused on internal, student-centered constructs, such as how well they studied, test-taking ability, and prior performance. Similarly, Bol et al. [31] found that the most frequent categories of explanations provided by middle school students for their predictions were the time and effort spent in studying, a global perception of their abilities, and their past performance. Regarding postdictions, the most frequent explanations were the knowing of the answer, the effort exerted in studying, and their general self-confidence. Using the same approach, Bol et al. [20] found that explanatory style concerning student-centered factors related to studying and test-taking

and to a task-centered factor could predict their achievement, prediction accuracy, and postdiction accuracy. This research has provided preliminary evidence regarding students' justifications for making their judgments. However, further research is needed to explore how students make their judgments about their learning and performance, and how these justifications are related to calibration accuracy and performance.

The Present Study

The cues students use for making their judgments of learning and performance represent a fruitful area for further investigation in the field of calibration research. Thus, this study focused on undergraduate sport students' justifications for making judgments regarding their knowledge in a developmental psychology course. Although some research has examined students' explanations for forming their metacognitive judgments of performance [2,28,31], further research is warranted [32]. Most of these previous studies based their results on students' explanatory style measured through self-reported questionnaires that included a limited and predefined number of responses. The present study expanded the previous ones involving a qualitative approach (i.e., using an open-ended question for students to provide their justifications) [28]. This approach allowed students to provide a wider range of factors and, thus, it is considered appropriate for examining students' justifications for making judgments of learning [32].

Examining how students make judgments of learning and performance will further increase our understanding of the factors associated with calibration accuracy [32]. For example, the potential associations between students' justifications and their level of calibration accuracy (low versus high accuracy) or achievement can be explored. This evidence can inform interventions for enhancing calibration accuracy by increasing students' awareness regarding the factors associated with forming accurate judgments of learning, or the factors that should be avoided because they may generate misleading information regarding the current status of learning. Calibration accuracy is considered a key element for self-regulated learning and increased performance [33,34], and may help undergraduate students to manage their time and effort more effectively, avoiding either premature termination or prolonged duration of the study [21].

Considering that both local and global judgments are useful measures of online monitoring [32,35], the present study involved two independent samples of students who provided their judgments either at the local or the global level to examine if these two types of judgments are associated with different forms of justifications. The different nature of these two types of judgments suggests that they may be based on different cognitive processes, and thus, different cues may be involved in making judgments in each case [36]. Moreover, research on metacomprehension judgments has shown that there was little or no relation between absolute (i.e., judgments of overall performance) and relative accuracy (i.e., discrimination of performance across items) [37]. On the other hand, Grabe and Holfeld [38] found that both measures were significant and unique predictors of future performance in an introductory college course of online study environments, while Nietfeld et al. [39] found that local compared to global prediction accuracy was more strongly associated with test performance. All these mixed findings highlight the need for investigating further the nature of judgments provided at the local and global levels. Moreover, to the best of our knowledge, there is no evidence regarding the potential differences in the justifications provided by students for making their judgments at the local compared to the global level. Therefore, a closer look at how students make their judgments of learning and performance either at the local or at the global level may shed further light on the nature of these two types of judgments used widely in calibration research. Moreover, the present study following a recent trend of examining calibration in applied settings [2,39,40] was conducted in real-life learning environments involving learning materials and evaluation processes that are meaningful for students increasing the ecological validity of the results.

The aim of this study was to explore undergraduate students' justifications for making their judgments of learning provided either at the local or the global level. Moreover, it was examined if students' justifications varied according to their calibration accuracy (low versus high accurate students) or level of their performance (low versus high performers). The nature of this study was exploratory, and thus, no specific hypotheses were stated.

2. Materials and Methods

2.1. Settings and Procedures

Ethical approval for this study was granted by the University Ethics Review Committee. The study was conducted during a regular developmental psychology course delivered in the winter semester in the local department of physical education and sport science. The knowledge test used was part of the students' official evaluation process for this course. Students were informed about the study and those who agreed to voluntarily participate responded to the additional questions regarding calibration. No credits were provided to students for their participation. No student refused to participate. Students were told that the calibration questions would not be used in their evaluation.

The course of developmental psychology is an elective course delivered in the spring semester of the fourth year of studies. The course included 10 90-min lectures regarding introduction to developmental psychology, Piaget's theory of cognitive development, cognitive functions, metacognition, intelligence, emotional intelligence, self-esteem, play as a developmental process, and the role of family in children's development. The evaluation included three written tests (after the third, sixth, and tenth lecture) counting 30% each, and one written assignment counting 10% of the total degree. Attendance was compulsory and students could miss 30% of the sessions.

2.2. Participants

A total number of 115 Greek senior sport students of a physical education and sport science department participated in this study. These students comprised two independent samples that attended the developmental psychology course in the spring semester of two consecutive academic years. In particular, students in Sample 1 ($N = 53$, $M_{age} = 22.14$, $SD = 0.50$, 23 males) attended the course in the first academic year of the study and provided their judgments at the local level, while students in Sample 2 ($N = 62$, $M_{age} = 22.64$, $SD = 3.14$, 38 males) attended the course in the next academic year and provided their judgments at the global level.

2.3. Measures

2.3.1. Knowledge Test

The knowledge test consisted of 20 multiple-choice questions regarding the content of the first three lectures of the developmental psychology course. For each question, four potential answers were provided, with one of them being correct. Students had 30 min to complete the test. The number of the correct answers was each student's actual score on the test.

2.3.2. Judgments of Learning and Calibration Accuracy

Students provided their judgments, either at the local (Sample 1) or at the global level (Sample 2). In particular, students in Sample 1 provided their judgments at the local level (i.e., after each question of the test) by responding to the following question: "Did you answer this question correctly or erroneously?". Students were just asked to circle the appropriate word (correctly or erroneously) representing their judgment. The total number of the questions in which students' actual and judged performance matched indicated their score in the calibration accuracy at the local level (Schraw, 2009).

Students in Sample 2 provided their judgments at the global level (i.e., for their overall performance in the test) by responding to the following question: "I think I have answered

correctly ... out of 20 questions". The absolute difference between the global judgment and the actual score indicated students' calibration accuracy at the global level (Schraw, 2009).

2.3.3. Justifications of Confidence Judgments

In both samples, after answering all the questions and providing their judgments, students were asked to describe how they formed their judgments by responding to the following open-ended question: "Please describe what you considered when making your judgments regarding your responses in the above questions" (Dinsmore and Parkinson, 2013). Students' responses were coded by two independent coders in the 10 categories described in Table 1. Kappa analysis revealed a high (92%) inter-coder agreement [41]. Disagreements between coders were resolved after discussion and these responses were coded after consensus between coders was reached.

Table 1. Categories of students' justifications for their confidence judgments.

Justification Category	Definition	Examples
Study	Responses regarding how students have studied the learning materials or not	"I have studied the learning materials" "I did not study enough"
Confidence	Statements regarding how sure students felt about providing a correct or a wrong answer	"I was sure for the responses"
Common sense	Statements indicating the use of common sense for selecting the answer	"I select the logical answer with respect to the questions"
Memory	Statements regarding students' memory of the correct answer, of what they remembered from their study, or general reference to memory	"I was based on what I remembered from the lectures" "Lack of memory" (for incorrect answers)
General knowledge	Statements indicating a general reference in knowledge acquired through lectures attendance (excluding knowledge acquired by studying at home)	"Based on my knowledge from the lectures"
Lectures	General statements referring to lectures' notes, attendance, and comprehension of presented materials	"I have comprehended the materials presented in the lectures" "From my notes"
Judgment	Statements indicating that students used their judgment to estimate if they provided a correct or a wrong answer	"Based on my judgment"
Experience	Responses regarding students' personal experiences from their studies or life	"In my personal experiences"
Knowledge of the answer	Statements indicating that students knew the answer	"I knew the response"
Other	Statements not included in the above categories	"Luck" "How I am thinking"

3. Results

3.1. Preliminary Analyses

In Sample 1, students responded correctly to half of the questions ($M = 10.92$, $SD = 4.42$), with their performance varying significantly (range = 3–20). Regarding their accuracy at the local level, students judged accurately their performance in more than half of the questions ($M = 12.17$, $SD = 3.89$), but their accuracy level varied significantly (range = 6–20). Students' scores in the knowledge test and the accuracy index at the local level were highly correlated ($r = 0.84$, $p < 0.001$).

In Sample 2, students responded correctly to half of the questions ($M = 10.55$, $SD = 4.36$), with their performance varying significantly (range = 3–20). Regarding accuracy at the global level, students' calibration accuracy index (absolute values of judged performance minus actual performance) ranged between 0 and 9 ($M = 3.42$, $SD = 2.44$). Students' scores in the knowledge test and the accuracy index at the global level were highly correlated ($r = 0.59$, $p < 0.001$).

3.2. Justifications of Judgments

Students' responses to the open-ended question regarding the justifications of their judgments of learning were classified in the following 10 categories: experience, study, confidence, general knowledge, knowledge of the answer, memory, common sense, judgment, lectures, and other. The operational definitions and representative examples of these categories are presented in Table 1. It should be noted that the majority of students (37 out of 53 in Sample 1, and 37 out of 62 in Sample 2) provided more than one response (up to 4) that were classified in more than one category. Thus, the final number of classified responses was 222 (108 in Sample 1 and 114 in Sample 2). The frequencies and the within-sample percentages of students' responses in each category across two samples and in the total and separately for high and low-accurate students are presented in Table 2. Low- and high-accuracy groups of students were formed within each sample based on the median score in the respective accuracy index (Sample 1 median score = 11, Sample 2 median score = 3). Frequencies of students' responses in the ten categories across high and low performers are presented in Table 3. Groups of low and high performance were formed within each sample based on the median score in the respective knowledge test (median score in both samples = 9). This study was exploratory and descriptive in nature. Therefore, frequencies and percentages of students' justifications were presented and discussed for both samples and low and high accurate students and low and high performers. Statistical comparisons were avoided due to the relatively high number of categories and the small number of cases in some of the cells created.

Table 2. Frequencies and percentages of students' justifications for their confidence judgments separately for low and high accurate students and in total.

Justification Category	Sample 1			Sample 2			Total		
	Total	Low Accuracy	High Accuracy	Total	Low Accuracy	High Accuracy	Total	Low Accuracy	High Accuracy
Study	21 (19 + 2) 19.4%	12 (11 + 1) 11.1%	9 (8 + 1) 8.3%	32 (27 + 5) 28.1%	14 (10 + 4) 12.3%	18 (17 + 1) 15.8%	53 (46 + 7) 23.9%	26 (21 + 5) 11.7%	27 (25 + 2) 12.2%
Confidence	19 (11 + 8) 17.6%	8 (6 + 2) 7.4%	11 (5 + 6) 10.2%	12 (9 + 3) 10.5%	8 (7 + 1) 7%	4 (2 + 2) 3.5%	31 (22 + 9) 13.9%	16 (13 + 3) 7.2%	15 (7 + 8) 6.7%
Common sense	16 14.8%	10 9.3%	6 5.5%	15 13.1%	10 8.7%	5 4.4%	31 13.9%	20 9%	11 4.9%
Memory	12 (9 + 3) 11.1%	6 (5 + 1) 5.6%	6 (4 + 2) 5.5%	12 (10 + 2) 10.5%	7 (6 + 1) 6.1%	5 (4 + 1) 4.4%	24 (19 + 5) 10.8%	13 (11 + 2) 5.8%	11 (8 + 3) 5%
General Knowledge	17 15.7%	7 6.4%	10 9.3%	4 3.5%	1 0.9%	3 2.6%	21 9.5%	8 3.6%	13 5.9%
Lectures	6 5.6%	3 2.8%	3 2.8%	13 11.4%	9 7.9%	4 3.5%	19 8.6%	12 5.4%	7 3.2%
Judgment	5 4.6%	1 0.9%	4 3.7%	9 7.9%	3 2.6%	6 5.3%	14 6.3%	4 1.8%	10 4.5%
Experience	6 5.6%	5 4.6%	1 0.9%	6 5.3%	3 2.6%	3 2.6%	12 5.4%	8 3.6%	4 1.8%
Knowledge of the answer	3 2.8%	2 1.9%	1 0.9%	5 4.4%	1 0.9%	4 3.5%	8 3.6%	3 1.4%	5 2.2%
Other	3 2.8%	2 1.9%	1 0.9%	6 5.3%	4 3.5%	2 1.8%	9 4.1%	6 2.7%	3 1.4%

Note: In the parentheses, the first number refers to the positive and the second to negative aspect of the statements included in each category. All percentages have been calculated within each respective sample.

Table 3. Frequencies and percentages of students' justifications for their confidence judgments separately for low and high performers.

Justification Category	Sample 1		Sample 2		Total	
	Low Performers	High Performers	Low Performers	High Performers	Low Performers	High Performers
Study	10 (9 + 1) 9.2%	11 (10 + 1) 10.2%	18 (14 + 4) 15.8%	14 (11 + 1) 12.3%	28 12.6%	25 11.3%
Confidence	8 (5 + 3) 7.4%	11 (8 + 3) 10.2%	3 (2 + 1) 2.6%	9 (7 + 2) 7.9%	11 4.9%	20 9%
Common sense	11 10.2%	5 4.6%	9 7.9%	6 5.3%	20 9%	11 4.9%
Memory	6 (4 + 2) 5.6%	6 (5 + 1) 5.6%	6 (6 + 0) 5.3%	6 (3 + 3) 5.3%	12 5.4%	12 5.4%
General Knowledge	7 6.5%	10 9.2%	2 1.8%	2 1.8%	9 4.1%	12 5.4%
Lectures	4 3.7%	2 1.9%	6 5.3%	7 6.1%	10 4.5%	9 4.1%
Judgment	1 0.9%	4 3.7%	6 5.3%	3 2.6%	7 3.2%	7 3.2%
Experience	5 4.6%	1 0.9%	4 3.5%	2 1.8%	9 4.1%	3 1.3%
Knowledge of the answer	2 1.9%	1 0.9%	0 0%	5 4.4%	2 0.9%	6 2.7%
Other	2 1.9%	1 0.9%	3 2.6%	3 2.6%	5 2.2%	4 1.8%

Note: In the parentheses, the first number refers to the positive and the second to negative aspect of the statements included in each category. All percentages have been calculated within each respective sample.

Students provided a wide range of justifications for their judgments. The most frequent justification was the study of the learning materials, followed by the confidence for providing the correct (or not) answer, a general report of knowledge, and a general reference to common sense used for providing their judgments. The justifications with the lower frequency were the memory (that helped them to remember the correct answer), the general reference to lectures and the associated materials, the knowledge of the answer, the personal judgment, the general personal experiences, and the category other.

Some variations in the frequency of four of these justifications were found across the two samples of students who provided their judgments either at the local or the global level. In particular, the frequency of the justifications of confidence and general knowledge was higher at the local compared to the global level, while the frequency of the justifications of study of learning materials and lectures was higher at the global compared to the local level.

Regarding the frequency of justification across high and low-accurate students, the results showed that the less accurate students compared to more accurate students reported more frequently the justifications of experience, common sense, and lectures and less frequently the judgment. Moreover, the frequency of students who provided three or more justifications was more within low-accurate students compared to high-accurate students (18 versus 8). Regarding the frequency of justifications across high and low performers, the results showed that the high performers compared to low performers reported more frequently the justifications of confidence and knowledge and less frequently the justifications of experience and common sense. The frequency of students who provided three or more justifications was similar across low and high performers (i.e., 13). This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

4. Discussion

This study examined undergraduate sport students' justifications for making their judgments of learning in an academic course (i.e., developmental psychology). Two independent samples were involved providing their judgments, either at the local or the global level, and then explaining how they formed these judgments. Consistent with previous findings [2,28,30], students provided a wide range of explanations for making their metacognitive judgments for their learning and performance. Some variations in the frequency of these justifications were found across the two samples of students (providing judgments at the local or the global level), low and high accurate students, and low and high performers. Moreover, consistent with previous evidence [1,9,11,15], students were generally inaccurate in estimating their actual performance both at the local and the global level. All these results are discussed next regarding their theoretical and practical implications for undergraduate students' learning and performance.

The most frequent justification that students provided was the study of the learning materials with most of these justifications to be positive or neutral (i.e., students reported they have studied their materials or they just referred to the study of the materials) and a few of them negative (i.e., lack of study). This finding is consistent with previous evidence showing that students based their judgments on the fact that they knew the correct answer [31] or attributed their inaccurate calibration to insufficient study [2]. In this line, another two types of justification in the present study that were associated with knowledge were general knowledge and the knowledge of the answer. These justifications included statements of general reference in knowledge acquiring through participating in the lectures (excluding knowledge acquired by studying at home) or knowing or not the correct answer. Interestingly, the frequency of this kind of justification was higher among the more accurate students and the high-performers. The knowledge acquired through the study of the learning materials or participation in the lectures of a course may increase students' awareness about what they know and what they do not know [42]. Moreover, this knowledge may also enhance students' feelings of familiarity regarding the task or the test at hand and, thus, may help students to make their judgments of learning and performance [27].

The justifications of confidence and common sense were in second place among the most frequently reported justifications. Consistent with previous findings [31], students reported using a global sense of self-confidence for answering the question correctly. The use of confidence for making learning judgments was reported more frequently by high performers. High levels of performance are associated with higher levels of self-efficacy and confidence [43], which in turn are associated with high levels of calibration accuracy. Indeed, research evidence has shown that self-confidence and self-efficacy were associated with students' calibration accuracy [23,44], while students attributed miscalibration to a lack of confidence in their performance [2]. Thus, feelings of confidence may be a valid source for students to base on for making their judgments of learning. Moreover, almost one out of four students used common sense as a basis to judge his or her learning. The frequency of this justification was higher among the less accurate students and lower among the high-performing students. Probably students who did not know the correct answer selected the response that looked more logical or possibly to be correct. Previous findings have suggested that guessing is one of the factors students may use to form their judgments [28]. Moreover, students with high performance avoided using common sense for making their learning judgments, probably because they were better prepared for the test and may know most of the correct answers.

Another group of three justifications followed in the third position of frequency, including memory, lectures, and judgment. Regarding memory, previous evidence has suggested that students used memory for past tests for making metacognitive judgments of learning or performance in the subsequent tests [45]. It has also been suggested that predictions of performance are subjective experiences of memory [46] working either as a general sense of previous memory experience (e.g., providing judgments in other related

tests or circumstances) or a specific memory experience related to the test at hand (i.e., the memory of the materials resulting in providing the correct answer). The justification category of lectures included students' statements regarding attendance, taking notes, or comprehension of materials during lectures that help them to form their judgments of learning. The use of such justifications, that are associated with task-related issues, has also been reported in previous research [20]. Interestingly, the frequency of these justifications was higher among the less accurate students. Probably, less accurate students used these more general and vague justifications due to the lack of more concise and specific ones. Statements indicating that students used their judgment to estimate if they provided a correct or a wrong answer were also reported by some students, especially by those with lower levels of calibration accuracy. It seems to be a tautology for one to use as an explanation for forming his or her learning judgments the judgment process itself. Probably with this statement, students may want to describe their general ability to judge or to evaluate in conditions that required doing that. Future research may further explore this issue by involving more appropriate approaches such as in-depth interviews with students.

A small number of students used as justifications their general and personal experiences from their studies or life. Personal experiences that are related to educational materials may help students to form their judgments, as it has been suggested that experience-based cues can be used for generating judgments [27]. Moreover, a recent study in physical education has shown that previous experience in the field of sports has been associated with calibration accuracy in that field [25]. However, in the present study, the majority of the statements classified in this category were general and vague without associating with students' experiences with the educational materials of the course. This may also explain the fact that these justifications were reported more frequently by less accurate students and low performers.

The present findings suggest that students based their judgments more on general and personal-related factors and less on task-specific factors. Indeed, the majority of the justification categories identified in this study (i.e., experience, confidence, general knowledge, memory, common sense, judgment) could be considered as information-based cues [27] and were associated with personal-related factors in terms of Bandura's [29] model of reciprocal determinism. Three of the justifications were associated more with task-specific factors (i.e., the study of learning materials, knowledge of the answer, lectures) and may be considered as experience-based cues [27] in the sense that generated experiences during the knowledge test or the preparation for the test that students used to make their judgments of learning.

A main research question of this study was if students were based on different factors for making their judgments at the local or the global level. Involving two independent samples, this study showed some variations in the frequency of the four justifications that students reported for making their judgments at the local or the global level. The frequency of the justifications of confidence and general knowledge was higher at the local compared to the global level, while the frequency of the justifications of study of learning materials and lectures was higher at the global compared to the local level. To the best of our knowledge, this is the first study examining simultaneously the nature of the justifications for judgments provided at the local and the global level. The different nature of these two types of judgment [4] suggests that they are based on different cognitive processes and, thus, different cues may be involved in making judgments in each case [36]. For example, it has been suggested that students may base their local judgments more easily on task-specific or experience-based cues, and their global judgments more on general information-based cues (e.g., self-concept) [47]. However, these associations were not supported in the present study as students reported more frequently the use of some types of information-based cues (i.e., confidence and general knowledge) at the local level, and the use of some experience-based cues (i.e., the study of learning materials and lectures) at the global level. Regarding the use of experience-based cues at the global level, it

should be noted that this is also possible when global judgments are provided after the performance (i.e., postdictions) [47]. This was the case in the present study, with students postdicting their performance both at the local and the global level and, thus, they may base their judgments on both information- and experience-based cues [27]. Experimental studies involving comparisons between the justifications provided for predictions and postdictions both at the local and global level may further explore this issue. Another reason for this inconsistency may be the fact that in this study, students were directly asked to justify how they made their judgments of learning. In contrast, other studies [47] based their conclusions on the associations between students' accuracy of their judgments at the local or at the global level and factors representing information- or experience-based cues (e.g., self-concept). The present study involved this qualitative method (i.e., open-ended question) because it is considered the most appropriate approach for investigating students' justifications for making judgments of learning [32]. However, considering that only a few studies have used this qualitative approach, further similar research is needed for exploring the factors students use to make their metacognitive judgments. This research may also be combined with quantitative research methods for exploring further this issue. Moreover, it should be noted that in the present study, students provided their justifications at the end of the knowledge test, after they had provided their local judgments for each specific item. Future research may also involve students justifying their judgments provided for each specific item to further explore this issue. Considering that both local and global calibration accuracy can be useful measures of online monitoring [35], such research may shed further light on the nature of these two types of judgments and inform respective interventions.

Consistent with previous evidence [28], the majority of students (almost two out of three) provided more than one justification for making their judgments of learning. Moreover, among the students who provided three or more justifications, the majority were low-accurate students. One may expect that the students who made their judgments based on multiple factors may be more accurate compared to those based only on one factor [7]. It seems, however, that it is not the number of the justifications used that matters most, but how valid and relevant the justifications used by students are. However, it should be noted that the number of students who provided multiple justifications was small (26 out of 115). Moreover, although some students were able to incorporate multiple factors when making their judgments, it is still unclear if there is an underlining factor associated with students' tendency to provide one or more justifications for their judgments. These issues should be further explored in future research.

Focusing on an aspect of calibration that needs further investigation [32] and replicating and expanding previous research [2,28,31], this exploratory study provided further insights into how students make their judgments of learning and performance. Indeed, students provided a wide range of justifications for making their metacognitive judgements, including both experience-based and information-based cues [27]. Moreover, the present study provided preliminary findings showing that the frequency of the justifications varied to some degree across the local and the global level, high and low accurate students, and high and low performers. However, considering the exploratory nature of the present study, the question if the use of certain types of justifications are associated with higher calibration accuracy or higher performance either at the local or the global level remains open for further investigation.

Practical Implications, Limitations, and Future Research

This study was conducted in real-life learning environments involving learning materials from an academic course delivered to students as part of their program of study, thus increasing the ecological validity of the results. From an applied perspective, the results of this study can inform interventions focusing on improving students' calibration accuracy. In particular, the knowledge regarding the factors that students use to make their judgments and the associations of these factors with calibration is of great interest. In particular, teachers may act appropriately to guide their students to base their judgments

of learning on more valid sources, and thus increase the accuracy of these judgments. In this line, interventions for improving calibration accuracy [48–50] may be enriched with elements enabling students to reflect on the factors they use to make their judgments of learning. Most importantly, interventions should train students to adopt the most valid cues for making their metacognitive judgments, increasing, thus, the accuracy of these judgments. Increased calibration accuracy is considered a critical element in self-regulated learning [33] and thus, it may be beneficial for undergraduate students' academic achievement as their academic work is largely based on self-initiatives and self-regulated learning procedures [51].

Research regarding how students make their judgments of learning and performance is in its origins. Thus, exploring students' justifications for their metacognitive judgments of learning and performance and how these may be associated with their calibration accuracy is a fruitful area for further research. This study focused on exploring undergraduate students' justifications for their judgments of learning during a developmental psychology course, and thus the generalization of results out of this setting should be done with caution. Future studies should involve larger samples involving a variety of academic courses and expanding to other educational levels (e.g., primary and secondary schools) to compare students' justifications concerning their calibration accuracy and level of performance. Moreover, it should be examined if different justifications used for forming metacognitive judgments are associated with more accurate predictions or postdictions of performance. The types of justifications used concerning the learning settings and environments should also be examined. For example, it has been found that elementary and secondary students were more accurate in estimating their performance in a sport task compared to sport-related knowledge tests [52]. Thus, it should be examined if the use of justifications varies according to the type of tasks (e.g., academic versus motor or sport tasks) or the type of knowledge (conditional versus procedural knowledge). The role of other control variables (e.g., personality traits, and self-efficacy expectations) should also be explored. From a methodological perspective, both qualitative and quantitative research methods are needed for exploring students' justifications for their metacognitive judgments. Indeed, in-depth interviews may provide further insights into how students postdict or predict their performance, while the development of valid questionnaires may permit a large-scale investigation of students' justification of their judgments of learning.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Department Of Physical Education And Sport Science, University Of Thessaly.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data underlying the results presented in the study are available on reasonable request from the author (akolov@pe.uth.gr).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Fogarty, G.; Else, D. Performance calibration in sport: Implications for self-confidence and metacognitive biases. *Int. J. Sport Exerc. Psychol.* **2005**, *3*, 41–57. [\[CrossRef\]](#)
2. Hacker, D.; Bol, L.; Bahbahani, K. Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacogn. Learn.* **2008**, *3*, 101–121. [\[CrossRef\]](#)
3. Keren, G. Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychol.* **1991**, *77*, 217–273. [\[CrossRef\]](#)
4. Schraw, G. A conceptual analysis of five measures of metacognitive monitoring. *Metacogn. Learn.* **2009**, *4*, 33–45. [\[CrossRef\]](#)
5. Gasser, M.; Tan, R. Performance estimates and confidence calibration for a perceptual-motor task. *N. Am. J. Psychol.* **2005**, *3*, 457–468.
6. Samara, E.; Kolovelonis, A.; Goudas, M. Examining the associations between calibration Accuracy and executive functions in physical education. *Eur. J. Educ. Res.* **2023**, *12*, 359–369. [\[CrossRef\]](#)

7. Pieschl, S. Metacognitive calibration—An extended conceptualization and potential applications. *Metacogn. Learn.* **2009**, *4*, 3–31. [CrossRef]
8. Schunk, D.H.; Pajares, F. Self-Efficacy in Education Revisited: Empirical and Applied Evidence. In *Big Theories Revisited, Vol. 4: Research on Sociocultural Influences on Motivation and Learning*; McInerney, D.M., Van Etten, S., Eds.; Information Age: Greenwich, CT, USA, 2004; pp. 115–138.
9. Chen, P. Exploring the accuracy and predictability of the self-efficacy beliefs of seventh grade mathematics students. *Learn. Individ. Differ.* **2003**, *14*, 79–92. [CrossRef]
10. Cleary, T.; Konopasky, A.; La Rochelle, J.; Neubauer, B.; Durning, S.; Artino, A. First-year medical students' calibration bias and accuracy across clinical reasoning activities. *Adv. Health Sci. Educ.* **2019**, *24*, 767–781. [CrossRef]
11. Liverakos, K.; McIntosh, K.; Moulin, C.J.; O'Connor, A.R. How accurate are runners' prospective predictions of their race times? *PLoS ONE* **2018**, *13*, e0200744. [CrossRef]
12. McGraw, A.; Mellers, B.; Ritov, I. The affective cost of overconfidence. *J. Behav. Decis. Mak.* **2004**, *17*, 281–295. [CrossRef]
13. Kolovelonis, A.; Goudas, M. Students' recording accuracy in the reciprocal and the self-check styles in physical education. *Educ. Res. Eval.* **2012**, *18*, 733–747. [CrossRef]
14. Kolovelonis, A.; Goudas, M.; Dermitzaki, I. Students' performance calibration in a basketball dribbling task in elementary physical education. *Int. Electron. J. Elem. Educ.* **2012**, *4*, 507–517.
15. Kolovelonis, A.; Goudas, M.; Dermitzaki, I.; Kitsantas, A. Self-regulated learning and performance calibration among elementary physical education students. *Eur. J. Psychol. Educ.* **2013**, *28*, 685–701. [CrossRef]
16. Händel, M.; Fritzsche, E. Students' confidence in their performance judgments: A comparison of different response scales. *Educ. Psychol. Int. J. Exp. Educ. Psychol.* **2015**, *35*, 377–395. [CrossRef]
17. Chen, P.; Zimmerman, B. A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *J. Exp. Educ.* **2007**, *75*, 221–244. [CrossRef]
18. Kolovelonis, A.; Goudas, M. Does performance calibration generalize across sport tasks? A multiexperiment study in physical education. *J. Sport Exerc. Psychol.* **2019**, *41*, 333–344. [CrossRef] [PubMed]
19. Bol, L.; Hacker, D. A comparison of the effects of practice tests and traditional review on performance and calibration. *J. Exp. Educ.* **2001**, *69*, 133–151. [CrossRef]
20. Bol, L.; Hacker, D.; O'Shea, P.; Allen, D. The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *J. Exp. Educ.* **2005**, *73*, 269–290. [CrossRef]
21. Hacker, D.; Bol, L.; Horgan, D.D.; Rakow, E.A. Test prediction and performance in a classroom context. *J. Educ. Psychol.* **2000**, *92*, 160–170. [CrossRef]
22. Valdez, A. Student Metacognitive Monitoring: Predicting Test Achievement from Judgment Accuracy. *Int. J. High. Educ.* **2013**, *2*, 141–146. [CrossRef]
23. Kolovelonis, A.; Goudas, M. The relation of physical self-perceptions of competence, goal orientation, and optimism with students' performance calibration in physical education. *Learn. Individ. Differ.* **2018**, *61*, 77–86. [CrossRef]
24. Zhou, M. University student's goal profiles and metacomprehension accuracy. *Educ. Psychol. Int. J. Exp. Educ. Psychol.* **2013**, *33*, 1–13. [CrossRef]
25. Kolovelonis, A. Relating students' participation in sport out of school and performance calibration in physical education. *Issues Educ. Res.* **2019**, *29*, 774–789. Available online: <http://www.iier.org.au/iier29/kolovelonis.pdf> (accessed on 10 March 2023).
26. Kolovelonis, A.; Dimitriou, E. Exploring performance calibration in relation to better or worse than average effect in physical education. *Eur. J. Psychol.* **2018**, *14*, 665–679. [CrossRef] [PubMed]
27. Koriati, A.; Nussinson, R.; Bless, H.; Shaked, N. Information-Based and Experience-Based Metacognitive Judgments: Evidence from Subjective Confidence. In *A Handbook of Memory and Metamemory*; Dunlosky, J., Bjork, R., Eds.; Psychology Press: New York, NY, USA, 2008; pp. 117–136.
28. Dinsmore, D.; Parkinson, M. What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learn. Instr.* **2013**, *24*, 4–14. [CrossRef]
29. Bandura, A. *Social Foundations of Thought and Action: A Social Cognitive Theory*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1986.
30. Wang, Y.; List, A. Calibration in multiple text use. *Metacogn. Learn.* **2019**, *14*, 131–166. [CrossRef]
31. Bol, L.; Riggs, R.; Hacker, D.; Nunnery, J. The calibration accuracy of middle school students in math classes. *J. Res. Educ.* **2010**, *21*, 81–96.
32. Bol, L.; Hacker, D. Calibration research: Where do we go from here? *Front. Psychol.* **2012**, *3*, 229. [CrossRef]
33. Chen, P.; Rossi, P. Utilizing Calibration Accuracy Information with Adolescents to Improve Academic Learning and Performance. In *Applications of Self-Regulated Learning across Diverse Disciplines: A Tribute to Barry J. Zimmerman*; Bembenuddy, H., Cleary, T., Kitsantas, A., Eds.; Information Age: Greenwich, CT, USA, 2013; pp. 263–297.
34. Zimmerman, B. Attaining Self-Regulation: A Social-Cognitive Perspective. In *Handbook of Self-Regulation*; Boekaerts, M., Pintrich, P., Zeidner, M., Eds.; Academic Press: San Diego, CA, USA, 2000; pp. 13–39. [CrossRef]
35. Griffin, T.; Wiley, J.; Salas, C. Supporting Effective Self-Regulated Learning: The Critical Role of Monitoring. In *International Handbook of Metacognition and Learning Technologies*; Azevedo, R., Aleven, V., Eds.; Springer: New York, NY, USA, 2013; pp. 19–34. [CrossRef]

36. Gigerenzer, G.; Hoffrage, U.; Kleinbölting, H. Probabilistic mental models: A Brunswikian theory of confidence. *Psychol. Rev.* **1991**, *98*, 506–528. [[CrossRef](#)] [[PubMed](#)]
37. Maki, R.H.; Shields, M.; Wheeler, A.E.; Zacchilli, T.L. Individual differences in absolute and relative metacomprehension accuracy. *J. Educ. Psychol.* **2005**, *97*, 723–731. [[CrossRef](#)]
38. Grabe, M.; Holfeld, B. Estimating the degree of failed understanding: A possible role for online technology. *J. Comput. Assist. Learn.* **2014**, *30*, 173–186. [[CrossRef](#)]
39. Nietfeld, J.; Cao, L.; Osborne, J. Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *J. Exp. Educ.* **2005**, *74*, 7–28.
40. Hadwin, A.F.; Webster, E.A. Calibration in goal setting: Examining the nature of judgments of confidence. *Learn. Instr.* **2013**, *24*, 37–47. [[CrossRef](#)]
41. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
42. Efklides, A. How does metacognition contribute to the regulation of learning? An integrative approach. *Psychol. Top.* **2014**, *23*, 1–30.
43. Schunk, D.; Pajares, F. Self-Efficacy Theory. In *Handbook of Motivation at School*; Wentzel, K., Wigfield, A., Eds.; Routledge: New York, NY, USA, 2009; pp. 35–53.
44. Nietfeld, J.; Cao, L.; Osborne, J. The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacogn. Learn.* **2006**, *1*, 159–179. [[CrossRef](#)]
45. Finn, B.; Metcalfe, J. Judgments of learning are influenced by memory for past test. *J. Mem. Lang.* **2008**, *58*, 19–34. [[CrossRef](#)] [[PubMed](#)]
46. Hertzog, C.; Dixon, R.; Hultsch, D. Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychol. Aging* **1990**, *5*, 215–227. [[CrossRef](#)]
47. Händel, M.; de Bruin, A.B.; Dresel, M. Individual differences in local and global metacognitive judgments. *Metacogn. Learn.* **2020**, *15*, 51–75. [[CrossRef](#)]
48. Gutierrez, A.; Schraw, G. Effects of strategy training and incentives on students' performance, confidence, and calibration. *J. Exp. Educ.* **2015**, *83*, 386–404. [[CrossRef](#)]
49. Kolovelonis, A.; Goudas, M.; Samara, E. The effects of a self-regulated learning teaching unit on students' performance calibration, goal attainment, and attributions in physical education. *J. Exp. Educ.* **2022**, *90*, 112–129. [[CrossRef](#)]
50. Labuhn, A.S.; Zimmerman, B.J.; Hasselhorn, M. Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacogn. Learn.* **2010**, *5*, 173–194. [[CrossRef](#)]
51. Kitsantas, A.; Zimmerman, B. College students' homework and academic achievement: The mediating role of self-regulatory beliefs. *Metacogn. Learn.* **2009**, *4*, 97–110. [[CrossRef](#)]
52. Kolovelonis, A. Greek physical education students' calibration accuracy in sport and knowledge tasks—A comparison. *Int. Sport. Stud.* **2019**, *41*, 16–28. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.