ISSN 1512-1801

EXPLORING CALIBRATION OF SPORT PERFORMANCE: THEORETICAL AND EMPIRICAL EVIDENCE, CHALLENGES, AND FUTURE DIRECTIONS

Athanasios Kolovelonis
Department of Physical Education and Sport Science, University of Thessaly, 42 100 Karies, Trikala, Greece. E-mail: akolov@pe.uth.gr

Abstract
The aim of this paper was to explore the construct of calibration of sport performance in sport and physical education settings. First, the term calibration was defined, conceptualized, and distinguished from other related constructs while the various types of calibration and the levels that they work were presented. The peculiarities of calibration of sport performance were highlighted and theoretical approaches aiming to explain calibration and measurement issues were provided. Then, empirical evidence from calibration research in sport and physical education were critically reviewed. Finally, four areas for expanding calibration research in sport and physical education settings were suggested and specific research directions were discussed.

Key words: *calibration, accuracy, bias, metacognition, self-regulated learning, sport, physical education*

Introduction

Calibration is educationally important because it is inherent in various constructs and processes associated with human learning and development (Alexander, 2013) including strategic behavior, motivation (Schunk & Pajares, 2009), metacognitive control processes and self-regulated learning (Efklides & Misailidi, 2010). Actually, calibration is a core construct underlying in the self-regulated learning process reflecting students' metacognitive judgements of what they know or can do in specific tasks or skills and how accurate these judgements are compared to actual performance (Schraw, 2009). For example, students who overestimate their capabilities may attempt challenging tasks and fail decreasing their subsequent motivation, while students who underestimate their capabilities may feel less motivated avoiding challenging tasks limiting their potential for developing necessary skills (Schunk & Pajares, 2004). Moreover, calibration is associated with effort exertion as students who erroneously consider themselves as knowledgeable in a domain may be reluctant to try hard to further develop their skills (Efklides & Misailidi, 2010). Calibration is also associated with higher performance (e.g., Bol et al., 2005) and may be involved in strategic behaviour in the sense that miscalibration may provide a false sense of a strategy's effectiveness (Hacker, 1998). Considering these important implications for learning and performance, calibration has attracted researchers' interest in educational contexts. However, calibration research in sport and physical education settings only recently has been expanded and thus the understanding of calibration of sport performance is still limited.

The aim of this paper was to review theoretical, methodological, and empirical evidence regarding calibration focusing on calibration of motor and sport tasks and to provide practical implications and directions for expanding calibration research in sport and physical education. The paper consists of six main parts. First, definitions and conceptualizations of calibration including the various types and the levels that calibration works are presented. Next, theoretical approaches aiming to explain calibration and measurement issues are reviewed. Although these parts were mainly informed from evidence of calibration research in academic settings, special emphasis was also given in incorporating this evidence in calibration of sport performance considering the peculiarities of sport tasks. Then, empirical evidence regarding calibration research in sport and physical education were critically reviewed and practical implications were discussed. Finally,

ISSN 1512-1801

directions for overcoming challenges and expanding calibration research in sport and physical education settings were provided.

Defining and conceptualizing calibration

The term calibration has been widely used in various domains including the decision-making literature examining probability judgments and accuracy (e.g., Keren, 1991) and the area of metacognitive research focusing on students' awareness of their learning or performance (e.g., Hacker et al., 2008). This paper focuses on the accuracy of metacognitive judgments for learning or performance in educational settings with an emphasis in sport and physical education domains. The construct of calibration is defined and conceptualized next highlighting the differences between absolute and relative accuracy, the use of predictions or postictions and the levels of providing them, the multidimensional nature of calibration and the differences between calibration and other similar constructs (e.g., confidence and self-efficacy). Special emphasis in peculiarities of calibration of sport performance is also given.

Defining calibration

The construct of calibration entails a metacognitive judgment for learning or performance and an objective measure of this learning or performance. The discrepancy between the judgment and the actual learning or performance is the core notion in the construct of calibration (Hattie, 2013). Indeed, calibration has been defined as the degree of correspondence between judged and actual performance (Keren, 1991). Depending on the range of the discrepancy between judged and actual performance students can be considered as well-calibrated or miscalibrated (Schraw, 2009).

Absolute versus relative accuracy

The discrepancy between judged and actual performance can be approached either as absolute accuracy or relative accuracy (Dunlosky & Thiede, 2013; Schraw, 2009). The absolute accuracy focuses on the absolute match between actual and judged performance and a difference score is calculated while the relative accuracy focuses on the discrimination between different levels of performance across items or trials. Absolute and relative accuracy seems to be independent (Dunlosky & Thiede, 2013) reflecting different aspects of metacognitive monitoring (Maki et al., 2005). This distinction between absolute and relative accuracy is important for understanding calibration while it is closely associated with the approach of measuring and analyzing calibration data (Dunlosky & Thiede, 2013). More details on this issue are presented next, in the section measurement of calibration.

Predictions and postdictions

Considering the time students provide their judgments with respect to the task at hand, two types of judgments, predictions and postdictions, have been identified (Schraw, 2009). In predictions students provide their judgments before their involvement in a task or a test while in postdictions, students provide their judgments after completing a task or a test but before they are informed for their performance. Both predictions and postdictions are metacognitive judgments that have been widely used in calibration research (e.g., Chen & Rossi, 2013; Pieschl, 2009; Schraw, 2009) as they are considered useful measures of online monitoring of learning and performance (Griffin et al., 2013).

The levels of calibration

Calibration can be examined either at local or at global level depending on whether metacognitive judgments are provided for each specific item or trial included in a test or task (i.e., local level) or for the entire test or task (i.e., global level) (Schraw, 2009). Calibration at both levels is considered important because it provides unique information regarding learning and performance. Calibration at local level is closed to metacognitive self-monitoring providing online information regarding specific aspects of performance while calibration at global level is closed to more self-evaluative process focusing on more general aspects of performance and having implications with goal setting and self-evaluation (Chen & Rossi, 2013). Calibration can also work at a meta-level when students are asked to provide second-order judgments for their confidence regarding the accuracy of their judgments of learning and performance (Buratti & Allwood, 2015).

ISSN 1512-1801

The multidimensional nature of calibration

Although much of attention in calibration literature has been paid to calibration of learning and performance outcomes, it should be noted that calibration should be considered as a multidimensional construct focusing not only in students' learning and performance, but also in other related aspects such as learning strategies, problem-solving, and goal setting (Hattie, 2013). The implications of this multidimensional nature of calibration should be considered both from theoretical and empirical perspective. For example, well-calibrated students can accurately judge how effectively use a learning strategy and to adapt or change this strategy if necessary. The challenge however is to objectively measure these aspects of learning (e.g., effective use of a learning strategy) for examining how calibrated students are.

Distinguishing calibration from other related constructs

Calibration is distinguished from other related constructs such as confidence and self-efficacy. Some research has focused on confidence judgments of learning and performance suggesting that these judgments tend to define a single factor regardless tasks or domains (Kleitman et al., 2011; Kleitman et al., 2012). However, this line of research has focused only on confidence judgments regarding learning or performance and not on the discrepancy between confidence judgments and actual performance which is the core notion of calibration. There is a difference between how confident I feel for answering a question correctly and how accurate my confidence is compared to my actual performance. Both of these conditions have implications for learning and performance, but the most serious one is when the confidence is high and the accuracy low (Hattie, 2013).

Another similar but conceptually different to calibration construct is self-efficacy. Self-efficacy is students' beliefs regarding their learning or performance in a specific task reflecting future-oriented judgments of competence (Schunk & Pajares, 2005) whereas calibration highlights the discrepancy between judged and actual performance. That is, self-efficacy reflects what students believe can achieve while calibration capture how accurate these beliefs are. Empirical evidence has supported this distinction showing a low correlation between self-efficacy and accuracy (Chen & Zimmerman, 2007; Kolovelonis & Goudas, 2018). However, calibration accuracy and self-efficacy seems to interact as students who increased their calibration also increased, although modestly, their self-efficacy (Nietfeld et al., 2006).

Peculiarities of calibration of sport performance

Calibration in sport and physical education settings involves the accuracy of students' judgments about what they know or they can do with respect to motor and sport skills and tasks (Gasser & Tan, 2005). In the case of sport knowledge (e.g., key elements of sport skills, sport rules) the calibration paradigm used in academic settings can be used. However, in the case of sport of motor skills there are some peculiarities that should be considered in applying the calibration paradigm used in academic settings. In particular, feedback is often available after performance in sport skills either in the form of knowledge of results (e.g., successfulness of a trial) or as internal feedback produced during task execution (e.g., kinesthetic sense of performing the technique correctly) (Schmidt & Wrisberg, 2008). Thus, if outcomes of performance are used for judging and measuring sport performance, students' judgments can take only the form of predictions. Postdictions could be used in the case of calibrating qualitative aspects of performance of sport skills (e.g., postdicting the correct performance of the technique of a skill). Another issue is that feedback in the form of knowledge of the results for the previous trials or set of trials can affect subsequent predictions (Avugos et al., 2013). Thus, when a test with multiple trials is used (e.g., a basketball shooting accuracy test included 10 trials) the use of a single global judgment (i.e., how many successful shots out of 10 trials) rather than judgments specific to each trial (Pieschl, 2009) is considered more appropriate.

Theoretical approaches to calibration

Although the construct of calibration has been examined for more than three decades little has been done regarding the development of a comprehensive theory of calibration. Usually,

ISSN 1512-1801

calibration has been viewed through the lens of related theoretical frameworks. For example, calibration has been involved in theories of self-regulated learning and metacognition. Some other theoretical approaches focusing on explaining miscalibration have also been proposed. Some of these approaches are shortly presented next and discussed towards the development of a comprehensive model of calibration.

Calibration in self-regulated learning and metacognitive theories

Calibration is closely linked to self-regulating learning (Pieschl, 2009). By definition self-regulated learners are well-calibrated learners as they are aware of what they do and do not know or what strategies to use (Zimmerman, 1990). Actually, well-calibrated students can realize their actual level of learning or performance focusing on aspects of tasks that they have not mastered yet (Efklides, 2014). Zimmerman’s (2000) cyclical model of self-regulated learning and Efklides’ (2011) metacognitive and affective model of self-regulated learning (for a description of these models see Efklides, 2011; Kitsantas et al., 2018; Zimmerman, 2000) put special emphasis on students’ capacity to monitor their own learning and performance. For example, calibration should be considered un underlining process that play an essential role in all phases of Zimmerman’s (2000) model (forethought, performance, and self-reflection) in the sense that calibration accuracy informs effective circles of self-regulation (Chen & Rossi, 2013). In particular, through the lens of Zimmerman’s (2000) model, well-calibrated information derived from self-monitoring process at local level during learning and performance (performance phase) is vital for making the necessary adjustments (self-reflection phase) in using learning strategies or in focusing in aspects of performance needed more. At the global level, a well-calibrated student can set more challenging goals (forethought phase) to guide effective self-regulation. These processes of self-regulated learning at the local (e.g., involvement in the task at hand) and the global level (e.g., setting learning or performance goals for the next month) and the interactions between person and task related factors have also been described by Efklides’ (2011) model. This model also put special emphasis on the role of metacognitive experiences (e.g., metacognitive feelings and estimates) before, during, or after task performance that function interactively enhancing students’ awareness regarding learning and performance (Efklides, 2011). Considering that metacognitive feelings are associated with students’ performance (e.g., Goudas et al., 2017) the accuracy of the information derived from these metacognitive processes (i.e., calibration) matters for effective self-regulated learning.

Cues and heuristics

The cue utilization framework suggests that judgments are formed based on both theory-based cues (e.g., beliefs about ability) and online experience-based cues (e.g., shooting position) and explains miscalibration in terms of using cues that are not valid for judging the specific performance (Koriat, 1997). For example, students may base their judgments on the amount of the accessible information in their memory which may be independent of the correctness of this information leading in miscalibration. Thus, accessibility may not be a valid cue for judging learning because the quality (i.e., correctness) and not the quantity (i.e., amount) of information matters for calibration (Koriat, 1993). A similar approach, the heuristics and biases approach, suggests that miscalibration occurs because of errors people make when they use heuristics for estimating performance focusing on some but not on all relevant information. Heuristics are cognitive short-cuts that individuals use when estimating probabilities associated with various problems (Tversky & Kahneman, 1983). Examples of heuristics used in judging performance may be the response time, the memory for past performance, or the retrieval fluency (Bjork et al., 2013).

Random errors and uncertainty

Other views have suggested that confidence judgments include not only a true judgment component but also random errors (e.g., attentional or memory lapses) that may lead to miscalibration (e.g., Soll, 1996). These random errors may affect judgments in two main directions including a cognitive inconsistency in the process of forming a subjective feeling of confidence from the available cues or when translating this feeling in a judgment and a lack in relative experience regarding the problem or the task at hand (Soll, 1996). Miscalibration has been also associated with different modes of uncertainty underlying confidence judgments. These modes of uncertainty may vary across different task including noise in the nervous system (for perceptual tasks) and incomplete states of knowledge (for cognitive tasks) (Juslin & Olsson, 1997). However, the prediction of this approach for greater underconfidence in perceptual tasks because of sensory noise has not been supported (e.g., Pallier et al., 2002).

Individual differences versus environmental factors

Other views have explained miscalibration as peoples' tendency to provide consistent confidence levels when judge their performance in various tasks regardless of their accuracy level (Kleitman et al., 2011). Evidence has shown the existence of an independent metacognitive trait that mediates the accuracy of self-assessment (e.g., Kleitman & Stankov, 2001; Stankov & Crawford, 1996). However, this research line has focused on the confidence judgments and not in the accuracy of these judgments (aka calibration). On the other hand, the Probabilistic Mental Model (Gigerenzer et al., 1991) focuses on ecological factors external to individuals suggesting that people use their knowledge of the relative frequency of events in the natural environment to form confidence judgments. Through the interactions with their environments, individuals learn the validity and the effectiveness of each cue for judging learning and performance and use those cues they consider more valid and effective for each occasion. This approach suggests that miscalibration resulted from the disparity between cue and ecological validities. For example, the general knowledge tests used in calibration research lack from ecological validity and thus the cues used by students have low validity resulting in miscalibration.

Towards a comprehension model of calibration

All these approaches seek to explain miscalibration focusing in some but not all aspects of forming judgments. For example, some approaches (e.g., the heuristic and biases approach) focus more on judgmental biases within the individual considering individual differences in the process of forming confidence judgments while other (e.g., Probabilistic Mental Model) in external factors such as the procedures involved in the creation of knowledge tests. Thus, further theoretical work for synthesizing all these approaches into a comprehensive model of calibration is warranted. This model may include the factors associated with calibration (determinants and consequences), the mechanism of calibration accuracy and miscalibration, the relations of calibration with leaning and performance outcomes and links with self-regulation and metacognition. These theoretical tenets should be empirically tested and the respective evidence should be used for developing a comprehensive model for explaining calibration accuracy and miscalibration in both academic and sport and physical education settings. The use of cues in forming judgments is considered in most current approaches for explaining calibration and it may be the basis of developing such a model. Indeed, this model may be informed by evidence from research regarding the cues students' use for making judgments and the potential interactions between the use of cues and person (e.g., self-efficacy), task-related (e.g., task's characteristics) and environmental (e.g., feedback, instructions) factors. A comprehensive model of calibration should also enable links with models of self-regulation and metacognition. In particular, the construct of calibration should be highlighted as an integral component of self-regulated learning (Chen & Rossi, 2013) and should be integrated in the existing self-regulated learning theories (e.g., Efklides and Zimmerman's models).

Measurement of calibration

Various methods for measuring calibration have been used reflecting in most cases different approaches for conceptualizing calibration (Dunlosky &Thiede, 2013). Indeed, the way of approaching or defining calibration (e.g., absolute versus relative accuracy) may affect the way of measuring it. At the same time, for each of these types of calibration various measures has been used. For example, absolute accuracy can be measured through calibration curves or calculating the difference score between judged and actual performance either at the local (item by item) or at global level (average performance across items). Similarly, for measuring relative accuracy gamma correlation or discrimination measures can be used (Keren, 1991; Rutherford, 2017; Schraw, 2009). These different measures of calibration are usually high correlated. Indeed, Schraw et al. (2013) found that 10 different measures used to assess monitoring accuracy based on a 2 (performance) x 2

ISSN 1512-1801

(monitoring judgment) contingency table where both performance and judgments were provided in yes-no format, loaded on a single factor. This implies that the various measures may work complementary providing information regarding aspects of calibration accuracy of monitoring processes (e.g., absolute accuracy, bias, or discrimination between correct and incorrect items) (Schraw, 2009). Thus, it has been suggested that researchers should compute all the valid calibration measures and evaluate whether they yield the same qualitative conclusions (Dunlosky & Thiede, 2013) considering the purposes of the study and the underlying monitoring processes under examination (Schraw, 2009).

Calibration measures used in academic settings are not all appropriate for use in sport and physical education settings due to peculiarities of sport tasks described in a previous section. In particular, the available feedback following sport performance set challenges in measuring metacognitive judgments regarding this performance. Thus, most of the research in sport and physical education (see next section) approached calibration as the difference between judged and actual performance, with performance judgments to involve mainly predictions at the global level. However, the use of a difference score for measuring calibration has been criticized for having low reliability (Hattie, 2013) while the use of a single metacognitive judgment at the global level prevents from examining the internal consistency of these measure. Although single-item measures with clear and unambiguous purpose for the respondent and clear experiential focus can provide valid indicators of the state being investigated having the advantage of evaluating students' immediate experience with little interference to their on-going learning process (Ainley & Patrick, 2006), more elaborative calibration measures for sport tasks should be developed for expanding calibration in sport and physical education settings. An alternative paradigm resolved some of these issues (postdictions and multiple measures were used) involving students in dart-throwing over the top of a screen to hit a target lying in the floor behind it, and thus depriving them from having feedback regarding the successfulness of their throws (Gasser & Tan, 2005). However, the disadvantages of this approach are also strong, such as the lack of ecological validity of this kind of measure. Moreover, this alternative paradigm does not exclude other factors interfering in sport tasks including the internal feedback produced during sport performance (Schmidt & Wrisberg, 2008) and the environmental cues revealing the successfulness of sport tasks such as a basketball shot (e.g., sounds produced when the ball hits the backboard or passes through the net). These factors may interfere with students' metacognitive judgments of performance making the interpretation of the results difficult. All these issues highlight the need for developing new and more elaborative valid approaches for measuring calibration of sport performance considering the peculiarities of sport tasks. This issue is discussed in details in the section of future directions.

Calibration research in sport and physical education

Research regarding calibration of motor and sport performance in sport and physical education settings is reviewed in this section. Searches in databases (Scopus, Sportdiscus, ERIC, PsychInfo, and Google Scholar) using specific key words (i.e., calibration, accuracy, absolute accuracy, relative accuracy, bias, prediction, postdiction, estimation, judgment, metacognitive monitoring, motor skill, sport skill, sport, and physical education) and manual searches were conducted of articles in the English language literature. It should be noted, that this was not a systematic review of the literature. Studies that have been conducted in the fields of sport, school physical education and physical activity and involved a calibration measure in their design were included in this review. Research findings has been summarized in three parts including the description of students' or athletes' status of calibration, the factors associated with calibration, and the interventions conducted for improving calibration.

Calibration status among students and athletes

The overestimation of performance in motor or sport skills is a consistent finding in calibration research in sport and physical education settings. For example, recreational basketball players were overconfident regarding their shooting performance (McGraw et al., 2004) while golfers (Fogarty & Else, 2005) and tennis players (Fogarty & Ross, 2007) were overconfident

ISSN 1512-1801

mainly in the difficult form of the tasks. Moreover, undergraduate psychology students were poorly calibrated from three different dart-throwing positions (Gasser & Tan, 2005). Runners were generally well-calibrated, especially the older and most experienced ones, but a tendency to overestimation of performance was also found mainly among female athletes (Liverakos et al., 2018).

Similarly, in physical education, students overestimated their basketball dribbling (Kolovelonis et al., 2012; Kolovelonis et al., 2013), basketball chest-pass (Kolovelonis & Goudas, 2012, 2019), basketball shooting (Kolovelonis & Goudas, 2018, 2019), and soccer pass (Kolovelonis & Goudas, 2019) performance. The number of students who overestimated their performance in these studies ranged from 40% to 80%. Moreover, it was found that students were overconfident regarding their performance in both a basketball shooting test and sport-related knowledge tests regarding shooting skill technique and sport rules (Kolovelonis, 2019a).

These results are consistent with respective findings in academic settings showing that students are often inaccurate in judging their learning or performance (e.g., Chen, 2003; Destan & Roebers, 2015; Hacker & Bol, 2004) having a tendency to overestimate their performance (e.g., Bouffard et al., 2011; Hacker et al., 2008). A recent review verified children's tendency towards overestimation of performance (Xia et al., 2023).

Factors associated with calibration

Some research in physical education setting examined potential factors associated with students' calibration. In particular, Kolovelonis and Goudas (2018) examined students' calibration in relation to person-related factors and found that students' calibration accuracy was associated with their task orientation, self-efficacy, and perceived competence, but not with their global self-worth, sport competence, optimism, and pessimism. Kolovelonis and Goudas (2019) examined students' calibration in relation to task-related characteristic. They found that the magnitude of calibration error was similar across sport tasks and contexts while approximately half of students were consistent in the direction of calibration (most of them were overestimators). However, the type of the task may be associated with miscalibration. Indeed, Kolovelonis (2019a) found higher overconfidence in sport-related knowledge tests (i.e., shooting skill technique and sport rules) rather than in sport skill test (i.e., basketball shooting). Higher levels of calibration of sport tasks (e.g., golf, tennis) compared to cognitive tasks (e.g., numbers recall) were also reported by Hildenbrand and Sanchez (2022). Evidence regarding the hard-easy effect was also found while high performers were more accurate compared to low performers. Students' experiences from sport participation may be associated with their calibration. Indeed, Kolovelonis (2019b) found higher levels of calibration accuracy among students who participated in sports out of school compared to those who did not. Similarly, experienced runners estimated more accurately their performance compared to less experienced ones (Liverakos et al., 2018). Students' predictions regarding their peers' performance may also be involved in calibration process (Kolovelonis & Dimitriou, 2018).

These findings expanded calibration research providing new insights regarding the factors associated with students' calibration. For example, Kolovelonis and Goudas (2018, 2019) supported Dinsmore and Parkinson (2013) evidence that students based their confidence judgments on a combination of personal (e.g., prior knowledge) and task characteristics (e.g., item difficulty) factors. Moreover, Kolovelonis and Goudas (2018) supported and expanded previous evidence in academic (e.g., Chen & Zimmerman, 2007) and sport settings (Fogarty & Else, 2005; Fogarty & Ross, 2007) regarding the hard easy effect, highlighting the effects of task difficulty in the direction rather than in the magnitude of calibration error. A general finding in academic settings that high performers are usually better at judging their own performance compared to low performers (Bol et al., 2005; Hacker et al., 2008) was also supported in physical education (Kolovelonis & Goudas 2019). Regarding the relations between gender and calibration of performance, research in both academic and sport and physical education settings has shown generally mixed results. For example, previous evidence has shown no gender differences in calibration in academic (Chen, 2003), sport (Gasser & Tan, 2005) and physical education (Kolovelonis et al., 2012) settings. Recently, however, some variations in students' calibration accuracy and bias in relation to gender

ISSN 1512-1801

were found (Kolovelonis & Goudas, 2019). These inconsistent findings highlight the need for further research on this issue.

Interventions for improving calibration

Two studies have been conducted in physical education to examine the effects of interventions on students' calibration. Kolovelonis et al. (2013) examined the effectiveness of the four-level model of self-regulated learning (Goudas et al., 2013; Zimmerman, 2000) in teaching a sport skill in physical education including calibration measures (i.e., calibration accuracy and bias indexes). Results showed no differences in calibration accuracy between students who practiced basketball dribbling under different self-regulatory conditions (i.e., receiving feedback, setting goals). However, this intervention did not explicitly focus on improving students' calibration of sport performance. Thus, Kolovelonis et al. (2022) expanded this study to include self-regulatory processes (i.e., self-set goals, self-recording, self-reflection, attributions, self-talk) targeting in improving students' calibration. Results showed that experimental group students become more accurate in predicting their performance after the intervention compared to control group. These results suggested that an appropriately designed intervention can positively affect students' calibration of sport performance. Similarly, interventions in academic settings had positive effects on improving students' calibration through the use of feedback and practicing monitoring accuracy (Nietfeld et al., 2006), self-reflection (Zimmerman et al., 2011), informing for the consequences of making overconfident judgments (Roelle et al., 2017), providing guidelines and group settings (Bol et al., 2012), and strategy instruction combined with extrinsic incentives (Gutierrez de Blume, 2017; Gutierrez & Schraw, 2015). However, there is also evidence suggested that calibration remained unaffected even after practice or other interventions (e.g., Bol & Hacker, 2001; Bol et al., 2005; Nietfeld et al., 2005).

Expanding calibration research in sport and physical education

Calibration research in sport and physical education is rather limited. Although some recent studies have shed light in aspects of calibration in physical education, much more should be done. Previous research involved only a few sport skills and focused in some aspects and factors associated with calibration of sport performance while methodological issues should be resolved for expanding calibration research in sport and physical settings. In this section, four areas for expanding calibration research of sport performance are discussed including methodological and measurement issues, determinants and consequences of calibration (including justifications of metacognitive judgments), calibration in social learning environments, and the design of interventions for improving calibration (Bol & Hacker, 2012).

Methodological and measurement issues in calibration of sport performance

A critical issue for expanding calibration research is the development of valid methods of collecting and analyzing calibration data (Alexander, 2013). This issue is a real challenge for sport and physical education due to the fact that calibration measures used in academic settings are not all appropriate for examining calibration of sport performance. In particular, the availability of feedback immediately after performance in sport tasks restricts the use of postdictions or multiple trials of the same task because the knowledge of previous performance can affect the estimations for the future performance (Avugos et al., 2013). Indeed, using multiple sets of the same test (e.g., four sets of tests included 10 trials each) may permit the examination of the internal consistency of predictions, but the effects of feedback from the knowledge of the results of the firsts set of the tests on subsequent set of tests cannot be eliminated. However, the use of an alternative paradigm may reduce this effect. In particular, following this paradigm, students provide their judgments in multiple sets of a test of the same skill but each set to include a different number of trials (e.g., 5, 7, 10, 12 trials). Students provide their judgments in all these sets of the test before performance following a counterbalance approach in providing both predictions and performing the sets of the test. This approach should permit to check the internal consistency of both students' predictions and performance in this test providing more valid calibration data.

ISSN 1512-1801

Calibration research in sport and physical education should be expanded by focusing on qualitative aspects of sport performance (i.e., technique of a skill) measured through the use of both predictions and postdictions at the local (e.g., for each performance standard) or at global level (e.g., for the entire technique). A challenge in this process would be the development of objective measures of qualitative aspects of sport performance for comparing judged and actual performance. Probably, the first step should be the development of an objective measure of measuring qualitative aspects of performance and then to adapt this measure as a scale for providing metacognitive judgments for these aspects of performance. Moreover, future research should examine students' second-order judgments regarding their confidence for the accuracy of their predictions and postdictions of sport performance. Evidence from academic settings (Buratti & Allwood, 2015) can guide this totally unexplored area of calibration research in sport and physical education settings.

Statistical analysis of calibration data should consider the nature of calibration index used (e.g., calibration accuracy and calibration bias). The calibration accuracy (i.e., absolute values of the calibration bias) that represents an index of the magnitude of calibration error can be used for comparing groups or in correlational analysis. On the other hand, calibration bias (i.e., the signed difference between judged and actual performance) representing the direction of calibration is not a linear measure of the degree of accuracy and thus it is not considered appropriate for comparing group means (Griffin et al., 2013) or in correlational analyses (Stankov et al., 2012). This is because the bias index includes both positive and negative values that may be mutually exclusive when mean scores are calculated resulting in misleading results regarding the direction of miscalibration. For avoiding such issues, the bias index can be used for classifying students as accurates, underestimators, and overestimators, an approach used in previous research in physical education (e.g., Kolovelonis & Goudas, 2018, 2019) obtaining a gross picture of the direction of miscalibration. However, a more elaborative approach of the direction of miscalibration (e.g., low/medium/high overestimators or underestimators) may be warranted. Moreover, some research in academic (e.g., Gonida & Leondari, 2011) has classified students in bias groups (accurates, underestimators, overestimators) using the self-criterion residual strategy (for detail see Paulhus & John, 1998; Robins & John, 1997). Future research in sport and physical education may consider this approach for classifying students in groups of accurates, underestimators, or overestimators. Moreover, creating separate variables for underestimation (including only students with negative score in calibration bias) and overestimation (including only students with positive score in calibration bias) as well as reporting the absolute upper and lower bounds of calibration bias can provide a clear picture for the magnitude of underestimation and overestimation at group level. Recent research in academic (Gutierrez de Blume, 2017) and in physical education (Kolovelonis et al., 2022) has effectively used such measures to evaluate the effectiveness of interventions in improving sport performance calibration. Future research should further test the validity of all these measures and their strength in predicting students' learning and performance outcomes in sport and physical education settings.

Determinants and consequences of calibration

A fruitful area for future research involves the factors associated with calibration of sport performance including both determinants and consequences of calibration or miscalibration. Regarding the determinants of calibration future research should expand previous studies (e.g., Kolovelonis & Goudas, 2018, 2019) focusing on the interactions of person and task-related factors and their effects on calibration. For example, reciprocal interactions between calibration accuracy and self-efficacy should be examined and the role of personality characteristics may be considered. Environmental factors, such as providing feedback or instructions and the characteristics of the tasks used should also be considered. This research should adopt longitudinal designs involving various sport and motor tasks from both team and individual sport. Longitudinal designs should also examine developmental changes in calibration accuracy and bias while the role of gender should be further explored.

Regarding the consequences of calibration future research should focus on the associations between calibration accuracy and performance. Previous evidence in academic (e.g., Bol et al.,

ISSN 1512-1801

2005) and sport and physical education (e.g., Kolovelonis & Goudas, 2019) has shown that high performers are better calibrated. However, considering that most of the previous studies were correlational and cross sectional in nature, future research should involve longitudinal and experimental designs to establish a cause-and-effect relation between calibration accuracy and high performance. A recent study with sport students provided evidence regarding the longitudinal associations between calibration accuracy and academic achievement (Kolovelonis & Goudas, 2022). Such research should be expanded to involve students' performance in motor and sport tasks. Furthermore, interactions between calibration accuracy and performance should be examined to explore if well-calibrated students perform high or high performers are well-calibrated or a reciprocal interaction exists. Moreover, further research should explore associations between calibration and other learning and performance related outcomes such as motivation, effort exertion, persistence, and strategy selection (Efklides & Misailidi, 2010; Schunk & Pajares, 2009). Future research should also focus on potential negative effects of miscalibration exploring if the direction of miscalibration (i.e., overconfidence and underconfidence) is associated with different patterns of learning and performance or under what circumstances overconfidence or underconfidence may not deteriorate performance, metacognitive or executive functions (Destan & Roebers 2015).

The factors associated with calibration of sport performance represents a new, totally unexplored research area in sport and physical education. Some research in academic settings has examined this issue showing that students base their judgments in a variety of factors including prior knowledge and performance, characteristics of the tasks and person related factors (Bol et al., 2010; Dinsmore & Parkinson, 2013; Hacker et al., 2008; Kolovelonis, 2023). Only recently, a study explored university sport students' justification for making their metacognitive judgments of learning related to a sport psychology and a developmental psychology course (Kolovelonis, 2023). Future research should explore how students form their metacognitive judgments of sport performance and if these justifications are associated with different patterns of outcomes. In this line, the role of various types of cues (Koriat et al., 2008) that students may use in forming their metacognitive judgments should also be explored. The results of this qualitative research may inform the development of quantitative tools for measuring students' justifications and cues for forming metacognitive judgments for sport performance that can be used in large scale quantitative research.

Calibration in social learning environments

Learning and performance in sport and physical education usually take place in social environments. Thus, students may form metacognitive judgments of learning and performance in settings (e.g., school, sport) where the presence of other social agents (e.g., peers, teachers, or coaches) is salient and sometimes critical. A combination of group learning context and guiding questions has been found to promote metacognition and achievement (Kramarski and Dudai, 2009). From this perspective, social interactions may also be considered in calibration research in sport and physical education settings. Preliminary research in physical education has shown that students' scores in better or worse than average index could significantly predict their scores in calibration bias and accuracy (Kolovelonis & Dimitriou, 2018). However, this study was cross-sectional in nature and focused on students' beliefs regarding their peers' performance minimising the effects of social comparisons (students participated individually). Future research should involve experimental and longitudinal designs to explore the effects of social comparisons in students' metacognitive judgments and calibration. Most importantly, in the light of research focusing on co-regulated learning (Bransen et al., 2022; Hadwin et al., 2011), an intriguing line for future research may be to examine co-calibration processes in sport and physical education. For example, students or athletes may work together to co-calibrate their learning and performance. Reciprocal teaching style (Mosston & Ashworth, 2002) may guide this process. On the other hand, examining students' or athletes' calibration of sport performance in competitive sport environments may highlight potential association between calibration accuracy and aspects of competitive environments.

Calibration research should also focus on coaches and physical educators' capability to accurately judged their own performance as instructors and teachers (Gabriele et al., 2016). In

ISSN 1512-1801

particular, it would be of great interest to examine the accuracy of coaches and physical educators' beliefs regarding the effectiveness of their teaching and instruction and their evaluations of students' learning and performance. The associations between coaches and physical educators and students' calibration should also be examined. Coaches and physical educators' calibration may affect the quantity and the quality of the instructions and feedback they give to their athletes or students and thus it may affect their students' learning and performance (Südkamp et al., 2012; Thiede et al., 2018).

Interventions for improving calibration

The development, implementation, and evaluation of appropriate interventions for enhancing students or athletes' calibration accuracy is another fruitful area for future research in sport and physical education. Previous interventions implemented in academic setting suggested that feedback (Labuhn et al., 2010; Nederhand et al., 2019), practice opportunities for calibrating performance (Bol et al., 2012), self-reflection (Zimmerman et al., 2011), providing rubrics for writing tasks (Hawthorne et al., 2017) and strategy instruction combined with extrinsic incentives (Gutierrez de Blume, 2017, 2022; Gutierrez & Schraw, 2015) may enhance performance calibration. Such interventions focusing on improving calibration may also be tested in sport and physical education. Preliminary evidence in physical education have suggested that students' calibration can be improved after an appropriately designed intervention (Kolovelonis et al., 2013; Kolovelonis et al., 2022). These interventions were based on the four-level training model of self-regulated learning development (Goudas et al., 2013; Zimmerman, 2000). Further research should examine the effectiveness of this instructional model in enhancing students' calibration involving large scale interventions and larger samples and exploring the dynamics between calibration accuracy and sport performance (Digiacomo & Chen, 2016). Calibration interventions targeting specific populations in terms of the nature of their miscalibration may also be developed as overestimators may need different type of intervention compared to underestimators. For example, lower-achieving students gained more in improving calibration accuracy from a respective intervention (Hacker et al. 2008). Moreover, other interventions may focus on high performers who may underestimate their performance (Bol et al., 2005; Kruger & Dunning, 1999) and may put needles effort in skills that they have mastered. This research should also consider the effects of the hard-easy effect (Juslin et al., 2000; Wisniewski et al., 2019) on these associations.

Conclusion

This paper focused on calibration of sport performance to highlight the importance of this construct for learning and performance in sport and physical education. Calibration research in these fields only recently has been emerged and thus our understanding of calibration of sport performance is rather limited. The present paper contributed to this effort, defining and conceptualizing the construct of calibration, highlighting the peculiarities of calibration of sport performance, and presenting theoretical perspectives examining and explaining calibration. The research evidence presented suggested that calibration is associated with learning and performance in sport and physical education. However, research regarding calibration of sport performance is still in its infancy. Undoubtedly, examining calibration in sport and physical education is a fruitful research area and the suggested lines for future research presented here may expand our knowledge in this field.

ISSN 1512-1801

References

Ainley, M., & Patrick, L. (2006). Measuring self-regulated learning processes through tracking patterns of student interaction with achievement activities. *Educational Psychology Review, 18,* 267-286. doi:10.1007/s10648-006-9018-z

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1-3. doi:10.1016/j.learninstruc.2012.10.003

Avugos, S., Bar-Eli, M., Ritov, I., & Sher, E. (2013). The elusive reality of efficacy–performance cycles in basketball shooting: An analysis of players' performance under invariant conditions. *International Journal of Sport and Exercise Psychology*, *11*, 184-202. doi:10.1080/1612197X.2013.773661

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, *64*, 417-444. 10.1146/annurev-psych-113011-143823

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69,* 133-151. doi:10.1080/00220970109600653

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here?. *Frontiers in psychology*, *3,* 229. doi:10.3389/fpsyg.2012.00229

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73,* 269-290. doi:10.3200/JEXE.73.4.269-290.

Bol, L., Hacker, D., Walck, C., & Nunnery, J. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*, *37,* 280-287. doi:10.1016/j.cedpsych.2012.02.004

Bol, L., Riggs, R., Hacker, D., & Nunnery, J. (2010). The calibration accuracy of middle school students in math classes. *Journal of Research in Education, 21*, 81-96.

Bouffard, T., Vezeau, C., Roy, M., & Lengele, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*, 221-229. doi:10.1016/j.ijer.2011.08.003

Bransen, D., Govaerts, M. J., Panadero, E., Sluijsmans, D. M., & Driessen, E. W. (2022). Putting self -regulated learning. *Medical Education*, *56*(1), 29-36. doi:10.1111/medu.14566

Buratti, S., & Allwood, C. M. (2015). Regulating metacognitive processes—support for a meta-metacognitive ability. In A. Peña-Ayala (Ed.), *Metacognition: Fundaments, applications, and trends* (pp. 17-38). Springer. doi:10.1007/978-3-319-11062-2_2

Chen, P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh grade mathematics students. *Learning and Individual Differences*, *14*, 79-92. doi:10.1016/j.lindif.2003.08.003

Chen, P., & Rossi, P. (2013). Utilizing calibration accuracy information with adolescents to improve academic learning and performance. In H. Bembenutty, T. Cleary, & A. Kitsantas (Eds.), *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman* (pp.263-297). Greenwich, CT: Information Age.

Chen, P., & Zimmerman, B. (2007). A cross-national comparison study on the accuracy of self-efficacy beliefs of middle-school mathematics students. *The Journal of Experimental Education*, *75*, 221-244. doi:10.3200/JEXE.75.3.221-244

Destan, N., & Roebers, C. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition and Learning*, *10,* 347-374. doi:10.1007/s11409-014-9133-z

Digiacomo, G., & Chen, P. P. (2016). Enhancing self-regulatory skills through an intervention embedded in a middle school mathematics curriculum. *Psychology in the Schools*, *53*, 601-616. doi:10.1002/pits.21929

ISSN 1512-1801
Dinsmore, D., & Parkinson, M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24,* 4-14. doi:10.1016/j.learninstruc.2012.06.001

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24,* 58-61. doi:10.1016/j.learninstruc.2012.05.002

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46,* 6-25. doi:10.1080/00461520.2011.538645

Efklides, A. (2014). How does metacognition contribute to the regulation of learning? An integrative approach. *Psychological Topics*, *23*, 1-30.

Efklides, A., & Misailidi, P. (2010). Introduction: The present and the future in metacognition. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 1-18). New York: Springer.

Fogarty, G., & Else, D. (2005). Performance calibration in sport: Implications for self-confidence and metacognitive biases. *International Journal of Sport and Exercise Psychology, 3,* 41-57. doi:10.1080/1612197X.2005.9671757

Fogarty, G., & Ross, A. (2007). Calibration in tennis: The role of feedback and expertise. In K. Moore (Ed.), *Proceedings of the 2007 Conference of the Australian Psychological Society* (pp. 148-152). Brisbane, Australia.

Gabriele, A. J., Joram, E., & Park, K. H. (2016). Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes? *Learning and Instruction*, *45*, 49-60. doi:10.1016/j.learninstruc.2016.06.008

Gasser, M., & Tan, R. (2005). Performance estimates and confidence calibration for a perceptual-motor task. *North American Journal of Psychology, 3,* 457-468.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological review*, *98,* 506-528. doi:10.1037/0033-295X.98.4.506

Gonida, E., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research*, *50*, 209-220. doi:10.1016/j.ijer.2011.08.002

Goudas, M., Dermitzaki, I., & Kolovelonis, A. (2017). Self-regulated learning and students' metacognitive feelings in physical education. *International Journal of Sport and Exercise Psychology, 15,* 131-145. doi:10.1080/1612197X.2015.1079791

Goudas, M., Kolovelonis, A., & Dermitzaki, I. (2013). Implementation of self-regulation interventions in physical education and sports contexts. In Bembenutty, H., Cleary, T., & Kitsantas, A. (Eds.), *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman* (pp. 383-415). Greenwich, CT: Information Age.

Griffin, T. D., Wiley, J., & Salas, C. R. (2013). Supporting effective self-regulated learning: The critical role of monitoring. In R. Azevedo& V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 19-34). New York: Springer. doi:10.1007/978-1-4419-5546-3_2

Gutierrez de Blume, A. P. (2017). The effects of strategy training and an extrinsic incentive on fourth- and fifth-grade students' performance, confidence, and calibration accuracy. *Cogent Education, 4,*1-17. doi:10.1080/2331186X.2017.1314652

Gutierrez de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology*, *114*(4), 681. doi:10.1037/edu0000674

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration.*The Journal of Experimental Education*, *83,* 386-404. doi: 10.1080/00220973.2014.907230

ISSN 1512-1801
Hacker, D. J. (1998). Metacognition: Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1-23). Mahwah, NJ: Erlbaum.
Hacker, D. J., & Bol, L. (2004). Metacognitive theory: Considering the social–cognitive influences. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited: Vol. 4: Research on sociocultural influences on motivation and learning* (pp. 275-297). Greenwich, CT: Information Age.
Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*, 101-121. doi:10.1007/s11409-008-9021-5
Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65-84). New York: Routledge.
Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction, 24*, 62-66. doi:10.1016/j.learninstruc.2012.05.009
Hildenbrand, L., & Sanchez, C. A. (2022) Metacognitive accuracy across cognitive and physical task domains, *Psychonomic Bulletin & Review*, *29*(4), 1524-1530. doi:10.3758/s13423-022-02066-4
Hawthorne, K. A., Bol, L., & Pribesh, S. (2017). Can providing rubrics for writing tasks improve developing writers' calibration accuracy? *The Journal of Experimental Education*, *85*(4), 689-708. doi:10.1080/00220973.2017.1299081
Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344-366.
Juslin, P., Winman, A., & Olsson, H. (2000). Naïve empiricism and dogmatisim in confidence research: a critical examination of the hard-easy effect. *Psychological Review,* 107, 384-396. doi:10.1037/0033-295x.107.2.384
Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologia, 77,* 217-273. doi:10.1016/0001-6918(91)90036-Y
Kitsantas, A., Kolovelonis, A., Gorozidis, G., & Kosmidou, E. (2018). Connecting self-regulated learning and performance with high school instruction in health and physical education. In M. DiBenedetto (Ed.), *Connecting self-regulated learning and performance with instruction across high school content areas* (pp. 351-373). Dordrecht, The Netherlands: Springer International Publishing. doi:10.1007/978-3-319-90928-8_12
Kleitman, S., Mak, K., Young, S., Lau, P., & Livesey, D. (2011). Something about metacognition: Self-confidence factor(s) in school-aged children. In S. Boag & N. Tiliopoulos (Eds.), *Personality and individual differences: Theory, assessment, and application* (pp. 103-115). New York: Nova.
Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Journal of Applied Cognitive Psychology, 15*, 321-341. doi:10.1002/acp.705
Kleitman, S., Stankov, L., Allwood, C. M., Young, S., & Mak, K. K. L. (2012). Metacognitive self-confidence in school-aged children. In M. Mo Ching Mok (Ed.), *Self-directed learning oriented assessments in the Asia-Pacific* (pp. 139-153). Springer Netherlands. doi:10.1007/978-94-007-4507-0_8
Kolovelonis, A. (2023) Exploring university students' justifications for making metacognitive judgments of learning, *Trends Higher Education, 2,* 421-433. https://doi.org/10.3390/higheredu2030025
Kolovelonis, A. (2019a) Greek physical education students' calibration accuracy in sport and knowledge tasks – a comparison, *International Sports Studies, 41*, 16-28. doi:10.30819/iss.41-1.03

Kolovelonis, A. (2019b) Relating students' participation in sport out of school and performance calibration in physical education, *Issues in Educational Research, 29(3),* 774-789. http://www.iier.org.au/iier29/kolovelonis.pdf

Kolovelonis, A., & Dimitriou, E. (2018). Exploring performance calibration in relation to better or worse than average effect in physical education. *Europe's Journal of Psychology, 14,* 665-679, doi:10.5964/ejop.v14i3.1599

Kolovelonis, A., & Goudas, M. (2012). Students' recording accuracy in the reciprocal and the self-check styles in physical education. *Educational Research and Evaluation, 18,* 733-747. doi:10.1080/13803611.2012.724938

Kolovelonis, A., & Goudas, M. (2018). The relation of physical self-perceptions of competence, goal orientation, and optimism with students' performance calibration in physical education. *Learning and Individual Differences, 61,* 77-86. doi:10.1016/j.lindif.2017.11.013

Kolovelonis, A., & Goudas, M. (2019). Does performance calibration generalize across sport tasks? A multiexperiment study in physical education. *Journal of Sport and Exercise Psychology, 41*, 333-344. doi:10.1123/jsep.2018-0255

Kolovelonis, A., & Goudas, M. (2022). Physical education students' calibration accuracy and academic achievement: A longitudinal study. *Issues in Educational Research, 32(2),* 575-592. http://www.iier.org.au/iier32/kolovelonis.pdf

Kolovelonis, A., Goudas, M., & Dermitzaki, I. (2012). Students' performance calibration in a basketball dibbling task in elementary physical education. *International Electronic Journal of Elementary Education, 4,* 507-517.

Kolovelonis, A., Goudas, M., Dermitzaki, I., & Kitsantas, A. (2013). Self-regulated learning and performance calibration among elementary physical education students. *European Journal of Psychology of Education, 28,* 685-701. doi:10.1007/s10212-012-0135-4

Kolovelonis, A., Goudas, M., & Samara, E. (2022) The effects of a self-regulated learning teaching unit on students' performance calibration, goal attainment, and attributions in physical education, *The Journal of Experimental Education, 90,* 112-129. doi:10.1080/00220973.2020.1724852

Koriat, A. (1993). How do we know what we know: the accessibility model of the feeling of knowing. Psychological Review, 100, 609e639. doi:10.1037/0033-295X.100.4.609.

Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349-370. doi:10.1037/0096-3445.126.4.349

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information- based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117-135). New York: Psychology Press.

Kramarski, B., and Dudai, V. (2009). Group metacognitive support for on-line inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research,* 40, 377-404. doi:10.2190/EC.40.4.a

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134. doi:10.1037/0022-3514.77.6.1121

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning, 5,* 173-194. doi:10.1007/s11409-010-9056-2.

Liverakos, K., McIntosh, K., Moulin, C. J., & O'Connor, A. R. (2018) How accurate are runners' prospective predictions of their race times? *PloS ONE, 13(8).* doi:10.1371/journal.pone.0200744

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97,* 723-731. doi:10.1037/0022-0663.97.4.723

ISSN 1512-1801

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective cost of overconfidence. *Journal of Behavioral Decision Making, 17*, 281-295. doi:10.1002/bdm.472

Mosston, M., & Ashworth, S. (2002). *Teaching physical education* (5th ed.). San Francisco: Benjamin Cummings.

Nederhand, M. L., Tabbers, H. K., & Rikers, R. M. (2019). Learning to calibrate: Providing standards to improve calibration accuracy for different performance levels. *Applied Cognitive Psychology*, *33*(6), 1068-1079. doi:10.1002/acp.3548

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

Nietfeld, J., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education, 74,* 7-28.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1,* 159-179.doi:10.1007/s10409-006-9595-6

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, S., Stankov, L., et al. (2002). The role of individual differences in the realism of confidence judgements. *The Journal of General Psychology, 129,* 257-299.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025–1060.

Pieschl, S. (2009). Metacognitive calibration - an extended conceptualization and potential applications. *Metacognition and Learning, 4,* 3-31. doi:10.1007/s11409-008-9030-4

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on the accuracy of self-perceptions. In R. Hogan, J. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649–679). New York: Academic Press.

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of making overconfident judgments of learning. *Journal of Educational Psychology*, *109*, 99-117 doi:10.1037/edu0000132

Rutherford, T. (2017). The measurement of calibration in real contexts. *Learning and Instruction, 47*, 33-42. doi:10.1016/j.learninstruc.2016.10.006

Schmidt, R. A., & Wrisberg, C. A. (2008). *Motor learning and performance: A situation-based learning approach*. Champaign, IL: Human Kinetics.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4,* 33-45. doi:10.1007/s11409-008-9031-3

Schraw, G., Kuch, F., & Gutierrez, A. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24*, 48-57. doi:10.1016/j.learninstruc.2012.08.007

Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited, Vol. 4: Research on sociocultural influences on motivation and learning* (pp. 115-138). Greenwich, CT: Information Age.

Schunk, D., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. Elliot & C. Dweck (Eds.) *Handbook of competence and motivation* (pp. 85-104). New York: The Guilford Press.

Schunk, D., & Pajares, F. (2009). Self-efficacy theory. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35-53). New York: Routledge.

Soll, J. B. (1996). Determinants of overconfidence and mis-calibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*(2), 117-137. doi:10.1006/obhd.1996.0011

ISSN 1512-1801

Stankov, L., & Crawford, J. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, *21*, 971-986. doi:10.1016/S0191-8869(96)00130-4

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, *22*, 747-758. doi:10.1016/j.lindif.2012.05.013

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743. doi:10.1037/a0027627

Thiede, K. W., Brendefur, J. L., Carney, M. B., Champion, J., Turner, L., Stewart, R., & Osguthorpe, R. D. (2018). Improving the accuracy of teachers' judgments of student learning. *Teaching and Teacher Education*, *76*, 106-115. doi:10.1016/j.tate.2018.08.004

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315. doi:10.1037/0033-295X.90.4.293

Wisniewski, M. G., Church, B. A., Mercado, E., Radell, M. L., & Zakrzewski, A. C. (2019). Easy-to-hard effects in perceptual learning depend upon the degree to which initial trials are "easy". *Psychonomic Bulletin & Review*, *26*, 1889-1895. doi:10.3758/s13423-019-01627-4

Xia, M., Poorthuis A. M. G., & Thomaes, S. (2023). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development*, 00:1–22. doi:10.1111/cdev.14042

Zimmerman, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*, 3-17. doi:10.1207/s15326985ep2501_2

Zimmerman, B. (2000). Attaining self-regulation: A social-cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press. doi:10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling*, *53*, 141-160.

Article received 2024-02-02