









Systematic Review

# A Systematic Review of Meta-Analyses on the Impact of Formative Assessment on K-12 Students' Learning: Toward Sustainable Quality Education

Andrew Sortwell <sup>1,2,3,\*</sup> , Kevin Trimble <sup>4</sup> , Ricardo Ferraz <sup>3,5</sup> , David R. Geelan <sup>1</sup> , Gregory Hine <sup>6</sup> , Rodrigo Ramirez-Campillo <sup>7</sup> , Bastian Carter-Thuiller <sup>8,9</sup> , Evgenia Gkintoni <sup>10,11</sup>  and Qianying Xuan <sup>12</sup>

- <sup>1</sup> School of Education, The University of Notre Dame Australia, Sydney 2007, Australia; david.geelan@nd.edu.au
  - <sup>2</sup> School of Health Sciences and Physiotherapy, University of Notre Dame Australia, Fremantle 6160, Australia
  - <sup>3</sup> Research Center in Sports Sciences Health Sciences and Human Development (CIDESD), University of Beira Interior, 6201-001 Covilhã, Portugal; ricardompferraz@gmail.com
  - <sup>4</sup> Department of Research and Evaluation, National Catholic Education Commission, Sydney 2000, Australia; kevin.trimble@ncec.catholic.edu.au
  - <sup>5</sup> Department of Sports Sciences, University of Beira Interior, 6201-001 Covilhã, Portugal
  - <sup>6</sup> School of Education, The University of Notre Dame Australia, Fremantle 6160, Australia; gregory.hine@nd.edu.au
  - <sup>7</sup> Exercise and Rehabilitation Sciences Institute, Faculty of Rehabilitation Sciences, Universidad Andres Bello, Santiago 7991538, Chile; rodrigo.ramirez@unab.cl
  - <sup>8</sup> Department of Education, Universidad de Los Lagos, Osorno 5290000, Chile; bastiancarter@gmail.com
  - <sup>9</sup> Departamento de Didáctica y Práctica, Facultad de Educación, Universidad Católica de Temuco, Temuco 4780000, Chile
  - <sup>10</sup> Department of Educational Sciences & Social Work, University of Patras, 26504 Patras, Greece; evgintoni@upatras.gr
  - <sup>11</sup> University General Hospital of Patras, 26504 Rio, Greece
  - <sup>12</sup> School of English Education & Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou 510420, China; qyxuan@link.cuhk.edu.hk
- \* Correspondence: andrew.sortwell@nd.edu.au



**Citation:** Sortwell, A.; Trimble, K.; Ferraz, R.; Geelan, D.R.; Hine, G.; Ramirez-Campillo, R.; Carter-Thuiller, B.; Gkintoni, E.; Xuan, Q. A Systematic Review of Meta-Analyses on the Impact of Formative Assessment on K-12 Students' Learning: Toward Sustainable Quality Education. *Sustainability* **2024**, *16*, 7826. <https://doi.org/10.3390/su16177826>

Academic Editors: Shaljan Areepattamannil and David Manzano Sánchez

Received: 2 August 2024  
Revised: 2 September 2024  
Accepted: 5 September 2024  
Published: 8 September 2024



**Copyright:** © 2024 by the authors. 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/>).

**Abstract:** Formative assessment in K-12 education has been a notable teaching and learning focus area in schools over the last 20 years, as evidenced by numerous recent systematic reviews and meta-analyses investigating and summarizing the evidence for formative assessments' effectiveness and sustainability. This umbrella review systematically reviews meta-analyses investigating the effects of formative assessment on learning, summarizes the current findings, and assesses the quality and risk of bias in the published meta-analyses. Meta-analyses were identified using systematic literature searches in the following databases: Scopus, ERIC, Academic Research Complete, ProQuest, APA PsycArticles, SPORTDiscus, Web of Science, and Humanities International Complete. Thirteen meta-analyses, each of which examined the effects of formative assessment on learning in K-12 students, were included in this umbrella review. The review considered evidence for the potential effectiveness of using formative assessment in class with primary and secondary school students. Formative assessment was found to produce trivial to large positive effects on student learning, with no negative effects identified. The magnitude of effects varied according to the type of formative assessment. The 13 included meta-analysis studies showed moderate ( $n = 10$ ), high ( $n = 1$ ), and low ( $n = 2$ ) methodological quality (AMSTAR-2), although the robustness of the evidence (i.e., GRADE analysis) was very low ( $n = 9$ ), low ( $n = 3$ ), and moderate ( $n = 1$ ). These findings offer valuable insights for designing and implementing different types of formative assessment aimed at optimizing student learning and ensuring the sustainability of assessment practices. However, the low-to-very-low certainty of the available evidence precludes robust recommendations regarding optimal formative assessment strategies for learning in K-12 students.

**Keywords:** formative assessment; sustainable assessment; K-12; students; learning; meta-analysis; achievement; umbrella review; program sustainability

## 1. Introduction

Formative assessment is a continuous evaluation process focused on monitoring and adjusting teaching to maximize the potential for student learning [1,2]. Its implementation involves transcending traditional approaches that view assessment as an isolated event, detached from the teaching and learning process, primarily focused on control and verification [3]. Instead, formative assessment advances towards an approach that fosters and enables learning by continuously gathering information throughout the process, thus identifying areas for improvement for both students and teachers [4,5].

Indeed, formative assessment is an effective teaching strategy for creating a socio-ecological sustainable environment for learning that maximizes the implementation of strategies and processes of interventions whilst shaping student learning experiences in the classroom [6–8]. It involves a flexible and interactive assessment process that encourages the active involvement of educators and learners in systematically gathering, analyzing, and utilizing evidence to assess pedagogical effectiveness and students' learning progress. This process aims to facilitate ongoing student learning in the short term or midstream during an activity and fosters their self-regulated learning [9]. For example, teachers may use informal processes such as questioning during a set activity to probe students' level of understanding, identify learning gaps, and then decide to alter the sequencing of the teaching and learning activities to meet students' needs. Alternatively, formative assessment can be used to provide short-term insight into students' learning progress in relation to curriculum-based learning objectives [10], allowing timely modification of instruction and interventions as needed for subsequent teaching and learning activities within the current lesson or in the subsequent lesson [11]. Formative assessment can also aid in formulating contextual explanations of the content or learning tasks, supporting the success of classroom interventions, and improving student achievement of learning objectives through tailored approaches (e.g., advanced readings for students who master the material quickly) that can support specific student needs [12,13].

In the primary and secondary school education spheres, policymakers and teachers are interested in high-quality, robust research evidence that guides the effective implementation of various formative assessment approaches that reflect and support student learning [14–16]. Internationally, government education authorities view formative assessment as an essential element of the teacher's repertoire in monitoring and enhancing student understanding and skill and capability competencies [16–19]. Promotion and support for using formative assessment through school, government, and non-government educational authority policies (e.g., relevant professional learning) can also encourage teachers to implement formative assessment in classrooms [20]. To support this advocacy of formative assessment in policy and practice, policymakers and teachers need high-quality, robust research to be summarized, providing accessible, credible research findings, enabling them to identify evidence-based effective practices and promote their implementation in real-world settings.

Teacher education and training appear as fundamental factors in transforming assessment practices and determining effective implementation of formative assessment [19,21]. This is in line with the claim by Li et al. [22] that teachers need substantial time and professional support to become proficient in implementing formative assessment. Indeed, teacher training and professional support longitudinally are critical in changing teachers' attitudes and perceptions about formative assessment [15]. This change occurs by enhancing teachers' positive instrumental attitude [15,23] and self-efficacy [24,25], which in turn promotes the correct (i.e., to inform teaching and learning) and effective use of formative assessment. Furthermore, it can also improve teachers' knowledge, understanding of, and skills in using formative assessment as another strategy to support student learning [19,26,27].

Teachers play an important role in effectively using formative assessment strategies, using each lesson to assess students' learning progress, deciding how the assessment information is best used for individual student or cohort feedback, and altering teaching and learning activities to meet learning needs [16,20]. A growing body of recent published

intervention studies [28,29], systematic reviews [1,20], and meta-analyses [9,30] recognize the critical role played by teachers in effectively implementing formative assessment practices in the classroom to achieve the intended outcomes. Moreover, a number of studies have postulated a convergence between how teachers interpret the information or data collected via formative assessment and the resulting changes in teacher instruction and/or the structure of the teaching and learning activities for adaptation to the students' knowledge, understanding, and skills [20,31]. The interpretation of information and the changes that follow resulting from professional judgement are critical in the cycle of formative assessment.

Indeed, there is a plethora of educational research into the effect of formative assessment on the cognitive learning domain, such as academic achievement [1,9]. However, there is a dearth of empirical research on the effects on the affective learning domain and the associated underlying mechanisms. Considering that the core elements of formative assessment (i.e., feedback and adaptive teaching behavior) are associated with enhancing student competence, which in turn is a prerequisite for fostering student intrinsic motivation [32], the effect of formative assessment on affective learning should be explored alongside the cognitive learning domain within an umbrella review. This approach would provide a comprehensive overview of the impact of formative assessment on the student. Evidence from such a review can be used to identify potentially effective formative assessment approaches for those students who struggle or have learning gaps in the affective domain.

There are numerous systematic reviews and meta-analyses to date [33–35] reporting that different types of formative assessment tools (e.g., clicker questions, rubrics, scaffolds, students assessing themselves or peers) have the potential to improve learning (e.g., cognitive, psychomotor, or affective) and performance-related outcomes (e.g., students' abilities to take ownership of their learning) such as student agency and students' scholastic competence. Acquiring data and information through original research, comprising controlled, quasi-experimental, or randomized-controlled trials, represents the initial stages of inquiry to advancing our understanding in K-12 education. Researchers subsequently conduct systematic reviews to summarize the findings from the original research and then statistically aggregate the data in meta-analyses.

Although systematic reviews and meta-analysis in the field of education are considered the best evidence to obtain insights into the current state of the field concerning a research question, the findings are limited and conditional in as much as they tend to exhibit a propensity to concentrate narrowly on a specific outcome measure, a specific formative assessment tool type [30,36], within a broad population group (i.e., students from all three settings: primary, secondary, and higher education), which can be challenging when applying their findings to specific population groups (i.e., primary, secondary, types of single-sex school settings). Given these potential methodological limitations, formulating comprehensive recommendations and achieving pooled robust results describing the effectiveness of formative assessment for the overarching instructional area of K-12 teaching and learning can be more challenging. Moreover, extensive literature reviews indicate that there are conflicting outcomes stemming from meta-analyses on student learning [34,37], likely due to divergent methodological approaches (e.g., number of databases, database relevance, search syntax, poorly defined inclusion and exclusion criteria, assessment of the quality of primary papers) and applied methods such as different statistical methods. For example, many meta-analyses focus on the broad scope of students from kindergarten through higher education, do not report the effect of biological sex and the socio-economic demographics of the school as moderators, and neglect to assess the methodological quality of the primary papers using a critical appraisal tool.

One way to overcome the methodological limitations of meta-analyses is to conduct umbrella reviews [38]. Umbrella reviews are positioned at the top of the evidence pyramid, consolidating all available data pertaining to an area of education into a cohesive study, thereby serving as the key source of evidence-based knowledge in the chosen area of study [39]. Notably, umbrella reviews aim to summarize, synthesize, and critically

evaluate findings from peer-reviewed published meta-analyses to provide a comprehensive overview of the knowledge available on a given topic [39]. Thus, umbrella reviews help researchers understand and gain greater insight into the present strengths and limitations inherent in the entirety of the pertinent topical literature, in this case, formative assessment in K-12 school settings. To our knowledge, no published umbrella review has investigated the effects of formative assessment on measures of learning and achievement in K-12 students.

The overarching objective of this umbrella review is to systematically review, assess, and evaluate the available meta-analyses examining the effects of formative assessment on student learning and performance-related outcomes. Therefore, we aim to (i) systematically review and assess the available meta-analyses that have investigated the effects of formative assessment-based strategies or interventions on student achievement and learning within the primary and secondary school settings; (ii) address the methodological approaches used and the robustness and credibility of the meta-analytical evidence; and (iii) identify current limitations and suggest areas for further research. Our findings may be useful for educational scientists, policymakers, school leaders, and teachers in understanding the effects of formative assessment on K-12 school students across different age ranges (i.e., primary school, middle school, and secondary school students).

## 2. Materials and Methods

The present umbrella review was conducted in accordance with internationally recognized Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) and umbrella review guidelines [38,40]. The umbrella review protocol was registered in the OSF database: <https://osf.io/t7sx8>, associated with <https://osf.io/xzs6a> (19 December 2023).

### 2.1. Literature Search Strategy

A systematic literature review following the PRISMA review guidelines was conducted in the databases Scopus, ERIC, Academic Research Complete, ProQuest, APA PsycArticles, SPORTDiscus, Web of Science, and Humanities International Complete. Google Scholar was also searched for meta-analyses focused on formative assessment, and we scoured the reference lists of the included meta-analyses to identify potentially relevant studies for inclusion. The database searches commenced in December 2023 and were updated in January, February, March, and June 2024. Each database was searched from the inception of indexing. The following Boolean search syntax was used: ('formative assessment' OR 'Assessment for learning' OR 'Feedback' OR 'self-assessment' OR 'peer assessment' OR 'formative evaluation' OR 'assessment as learning') AND ('primary school' OR 'elementary school' OR 'high school' OR 'secondary school') AND ('meta-analysis'). Two authors (AS and KT) performed the literature search independently.

### 2.2. Selection Criteria

Studies were selected based on a *priori*-defined inclusion and exclusion criteria that were developed using the PICOS framework (PICOS = population, intervention, comparator, outcome, study design), as shown in Table 1. Two authors (AS and KT) independently reviewed and evaluated the eligibility of 71 papers, achieving a 94.37% inter-rater agreement. When AS and KT did not reach an agreement concerning the inclusion of an article, the disagreement was resolved by discussion with another author (BC). When needed, corresponding authors of papers were emailed seeking clarity around the inclusion/exclusion criteria information provided in their manuscripts. A total of seven authors were contacted, and all replied. The meta-analyses excluded during the reviewing and evaluation stage for eligibility are listed in the online Supplementary File, Table S1.

**Table 1.** Selection criteria applied in the umbrella review.

Category	Inclusion Criteria	Exclusion Criteria
Population	- Typical developing students in an academic K-12 setting or are primary or secondary school students.	- Atypical developing students. - Settings other than K-12, primary, or secondary schools (e.g., pre-school, higher education, or other forms of adult learning). - Intervention is unrelated to student learning in the school setting.
Intervention	- Includes formative assessment strategies designed to improve teaching and learning simultaneously.	- The study examines students' feedback on or evaluation of courses, teaching, or teachers' focus. - The study includes summative assessment. - Application of assessment approaches that are not intended or purposeful for actionable change in the lesson activities to enhance learning.
Comparator	- Control group (e.g., usual lessons and teaching and learning activities). - Measure of achievement, learning, academic performance, engagement, motivation or attitude, student understanding, and learning (affective, cognitive, and psychomotor).	- No control group.
Outcome	- Meta-analysis and included controlled or quasi-experimental trials.	- There is no measure of academic achievement/performance, learning engagement, motivation or attitude, or student understanding.
Study design	- Meta-analysis and included controlled or quasi-experimental trials.	- No meta-analysis. - Included observational studies.

### 2.3. Data Extraction

The following data were extracted from included meta-analyses: (i) first author and year of publication; (ii) aim of the study; (iii) school setting; (iv) the number of included primary studies; (v) study design; (vi) number of effect sizes; (vii) sample size; type of formative assessment and outcomes assessed; (viii) statistical model and effect metrics; (ix) effect sizes, confidence intervals (CI; e.g., 95% CI), significance level,  $p$  values, and  $I^2$  values (i.e., primary studies heterogeneity). Two authors (AS and KT) conducted the data extraction process. Data extracted were cross-checked for accuracy by RF and RC. When relevant data were unavailable in the respective papers, the corresponding authors were emailed with a request for the data. Furthermore, authors of the included papers were contacted as needed to verify data and other aspects of their included primary studies. A total of six authors were contacted, and while all replied, only five provided the necessary information.

### 2.4. Methodological Quality Assessment

The methodological quality of the included meta-analyses studies was examined using the Assessing the Methodological Quality of Systematic Reviews-2 (AMSTAR-2) tool [41,42], previously applied in the field of education [43,44]. The tool includes a checklist for 16 domains relating to the meta-analysis research question, review framework, search strategy, selection of primary studies, process for selection and data extraction, justification for included and excluded primary studies, description of eligible studies, the risk of bias, sources of funding, heterogeneity, publication bias, and conflicts of interest.

The AMSTAR-2 was independently applied by three reviewers (EG, DG, and AS), and disagreements between the reviewer's ratings were resolved through discussion with supporting evidence. Each item on the AMSTAR-2 tool checklist was answered with either

a *yes* (1 point), a *partial yes* (0.5 points), or *no* (0 points) as applicable. To allow comparison of the AMSTAR-2 scores, raw scores were transformed into percentages of the maximum possible scores. For each meta-analysis quality was rated as (i) high, with  $\geq 80\%$  of the 16 items satisfied; (ii) moderate, with 40–80% of the items satisfied; or (iii) low, with  $< 40\%$  of the items satisfied [45–47].

### 2.5. Robustness of the Results-Based Recommendations

The following aspects of the GRADE (Grading of Recommendations Assessment, Development and Evaluation) tool were assessed at the outcome and study level [48]: (1) risk of bias, determined by the quality of the primary studies, as assessed in the included meta-analyses; (2) inconsistency (heterogeneity), determined by the variation in effects across the included studies and the magnitude of statistical heterogeneity as measured by  $I^2$ ; (3) indirectness, determined by the generalizability of the findings to the study populations included in the primary studies of the selected meta-analyses; (4) imprecision, determined by the total sample size in the analysis and the number of included studies; and (5) publication bias, determined by comparing the effect size of the largest primary study with the pooled estimate from the meta-analysis and examining the asymmetry of the funnel plot. The GRADE assessment was conducted independently by three authors (AS, BC, and GH), with discussion and agreement to resolve any differences. A fourth author (RC) also critically reviewed GRADE assessment results for any anomalies. For each of the five aspects evaluated, each outcome was initially categorized as: not reported, neutral, serious, or very serious [48]. Meta-analyses were downgraded from initially four points by one point for each ‘Not reported’ or ‘Serious’ and by two points for each ‘Very Serious’ rating. Then, meta-analyses were rated as ‘High’ (4 points), ‘Moderate’ (3 points), ‘Low’ (2 points), or ‘Very Low’ ( $\leq 1$  point).

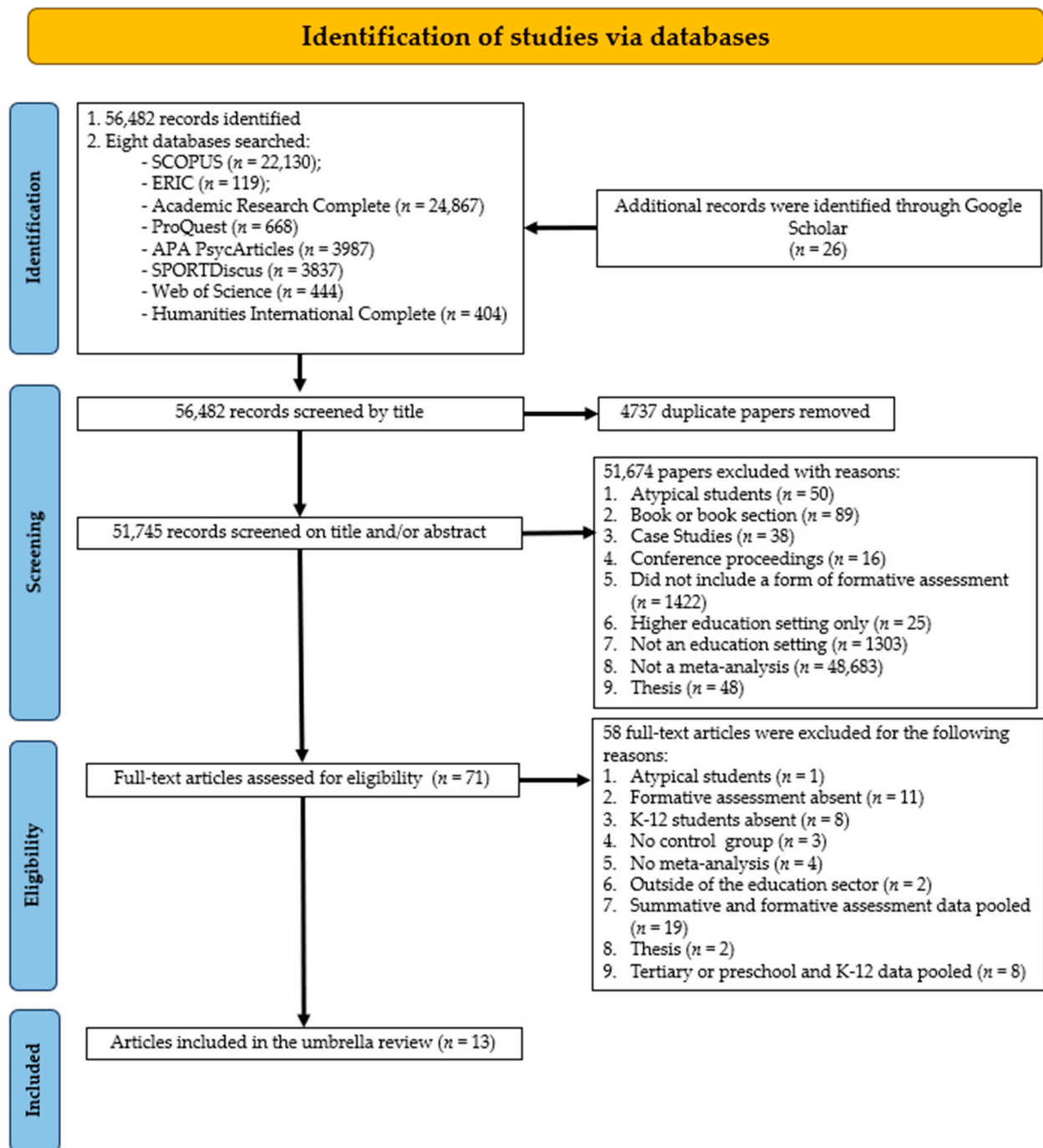
### 2.6. Data Interpretation

Comparisons were performed for the magnitude of effects across all the included meta-analyses using standardized mean differences (SMD). Appropriate conversions were considered for meta-analyses using different equations to compute SMDs (e.g., sample size adjustment with Hedges’  $g$ ) (see Supplementary File, Table S2, pp. 6–14). The SMDs were classified as trivial ( $< 0.20$ ), small (0.20–0.49), medium (0.50–0.79), or large ( $\geq 0.80$ ) [49]. If any of the studies provided confidence intervals (CI) for the effects sizes other than 95% CI, they were converted to a 95% CI by calculating the standard error and then the margin of error for the desired CI, followed by constructing the new CI around the mean weight effect size [50].

## 3. Results

### 3.1. Search Results

The systematic search yielded 56,482 documents (Figure 1). After removing duplicates and inspecting titles and abstracts, 71 full-text studies were read and assessed for eligibility. Thirteen meta-analyses were eligible for inclusion in this umbrella review. All included studies were published in peer-reviewed journals.



**Figure 1.** PRISMA flow diagram.

### 3.2. Characteristics of the Meta-Analyses

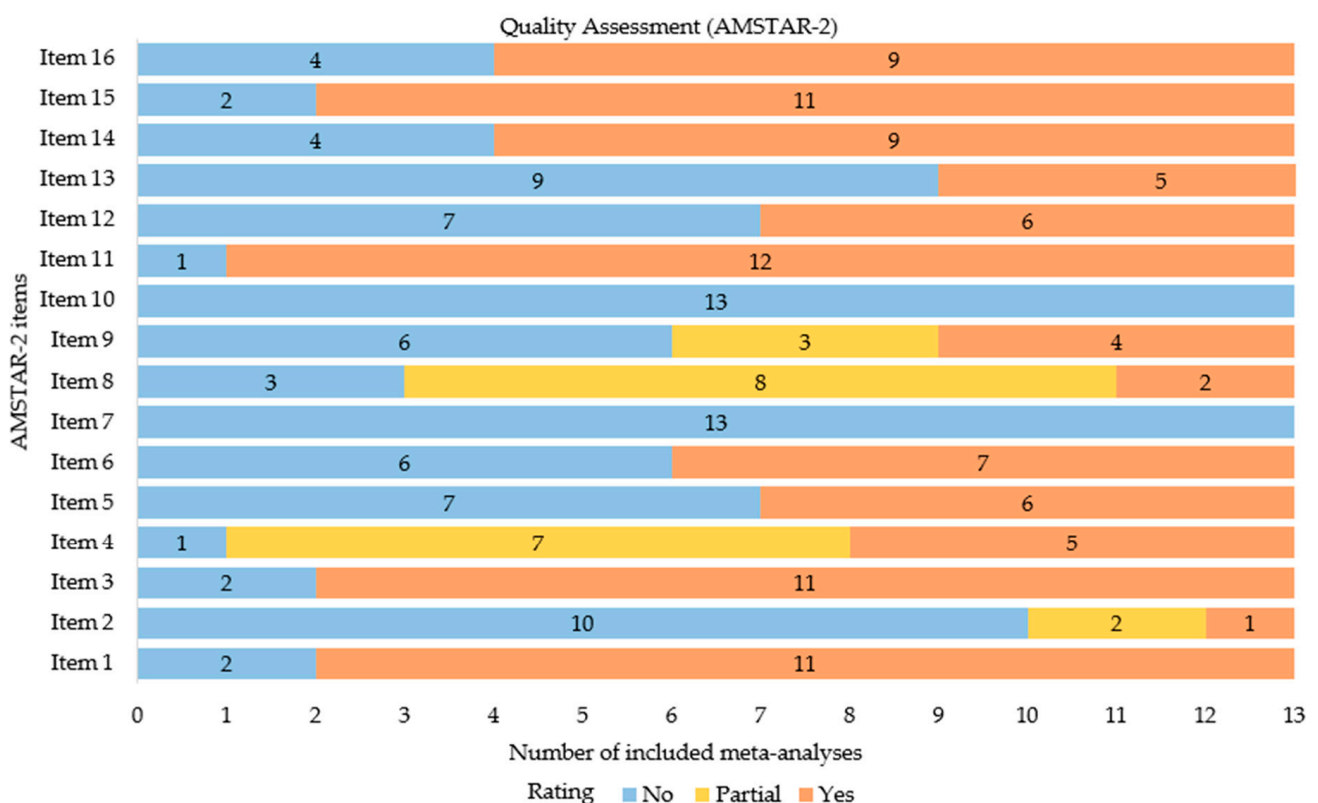
The 13 included meta-analyses were published from 2011 to 2023. The effect of formative assessment on K-12 students or primary and secondary school students was explicitly reported in the main analysis of four studies [1,9,51,52] and in the sub-analysis of nine studies [30,36,53–59] that examined the broad impact of formative assessment across all levels of education, including tertiary. The number of primary studies within the K-12 school setting ranged from four to 30 in the 13 included meta-analyses. Sample sizes ranged from 1047 to 116,051 students with a combined total of 256,659. One meta-analysis [58] did not provide sample size information. One meta-analysis [30] investigated the effect of biological sex as a moderating factor on learning resulting from formative assessment

interventions. Regarding the type of formative assessment, five meta-analyses [1,9,51,52,56] included more than one type of formative assessment, two meta-analyses included self-regulated learning scaffolds [58,59], and other meta-analyses included self-monitoring [54], real-time classroom interactive competition (i.e., gamified formative assessment) [36,55], computer-mediated feedback [57], rubrics [30], and student response systems [53].

Of the 13 meta-analyses included in the umbrella review (see Supplementary File, Table S2, pp. 6–14), 12 included calculations of summary effect estimates and corresponding 95% CIs, and one meta-analysis provided corresponding 90% CIs [58]. Twelve meta-analyses used a random effects model, with one [30] using a three-level model with robust variance correction. One meta-analysis [55] used mean difference only. The  $I^2$  heterogeneity estimates were reported in 46% ( $n = 6$ ) of included meta-analyses [9,30,36,51,55,57] and funnel plots in 54% ( $n = 7$ ) of the 12 included meta-analyses [30,36,51,53,54,56,59].

### 3.3. Methodological Quality Assessment

The 13 included meta-analyses were assessed using the AMSTAR-2 tool (see Figure 2); the completed AMSTAR-2 rating for the 13 included meta-analyses is available in the Supplementary File, pp. 15–16, Table S3. Most ( $n = 10$ ) [1,9,30,36,51,52,54,56–58] of the 13 included meta-analyses were rated as ‘moderate,’ two meta-analyses were rated as ‘low’ [53,59], and one meta-analysis was rated with a ‘high’ quality [55]. The 13 included meta-analyses received scores ranging from 34 to 81% of the maximum score (16 points).



**Figure 2.** AMSTAR-2 assessment for the 13 included meta-analyses.

Overall, the meta-analyses did not adequately report on Item 2 (prior written protocol) (85%,  $n = 11$ ); Item 4 (explanation of the choice of study design for inclusion) (62%,  $n = 8$ ), Item 5 (study selection in duplicate) (54%,  $n = 7$ ), Item 6 (data extraction in duplicate) (46%,  $n = 6$ ), Item 7 (provided a list of excluded studies with rationale for exclusion) (100%,  $n = 13$ ), Item 8 (description of the included studies in detail) (85%,  $n = 11$ ), Item 9 (assessment of the risk of bias) (69%,  $n = 9$ ), Item 10 (information about funding sources of included primary studies) (100%,  $n = 13$ ), Item 12 (assessed the impact of the risk of bias within the included

studies in the meta-analysis results) (54%,  $n = 7$ ), Item 13 (account for the impact of the risk of bias while interpreting/discussing the results) (69%,  $n = 9$ ).

### 3.4. Robustness of the Results-Based Recommendations

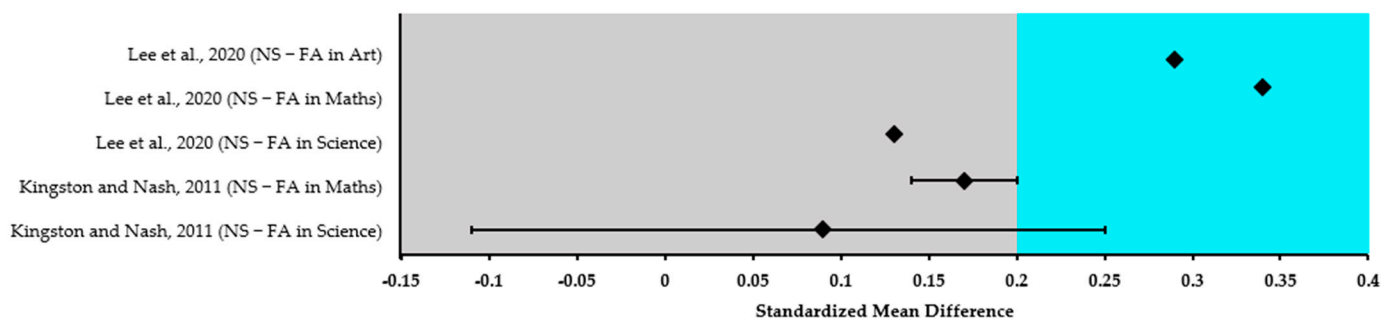
Based on the GRADE assessment (see Supplementary File, Tables S4–S8, pp. 17–21), the robustness of the results-based recommendation is considered low or very low.

### 3.5. Effect of Formative Assessment in Key Areas of Learning

Having considered the quality of the included meta-analyses using the AMSTAR 2 and GRADE instruments, the studies can be further explored in relation to the effect within key areas of learning or subject areas.

#### 3.5.1. Effects of Formative Assessment on Learning in STEAM-Related Subjects

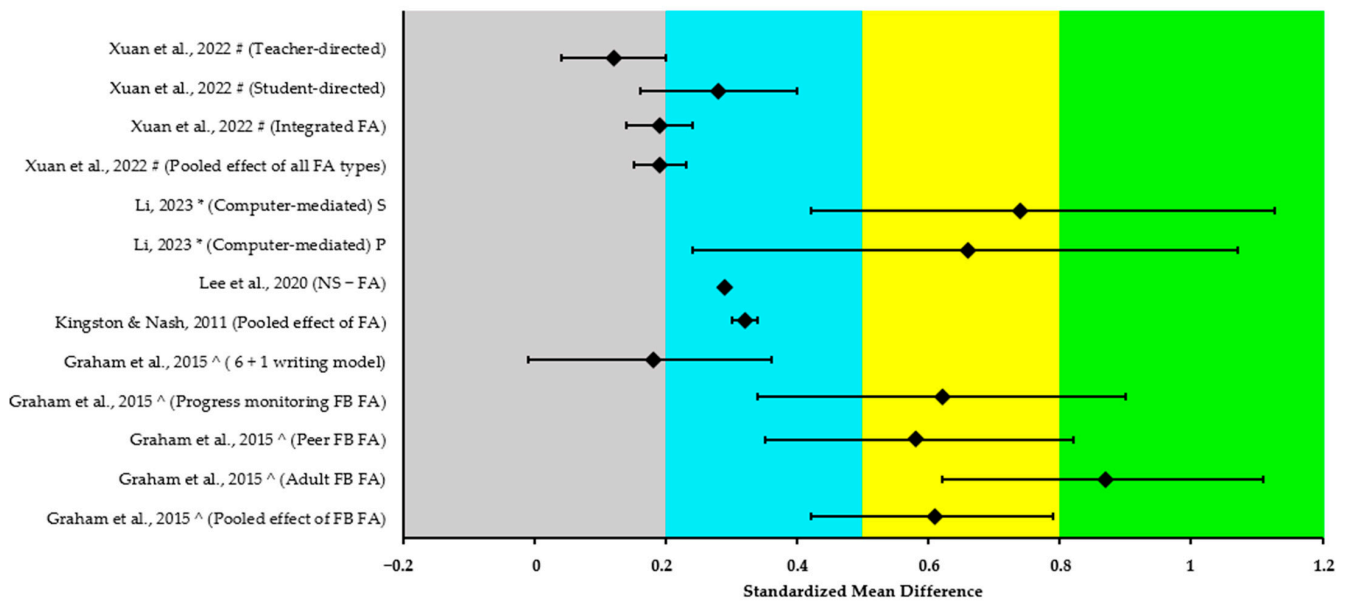
Three of the five effect sizes reported for formative assessment in STEAM-related subjects Mathematics, Science, and Art reported trivial effect sizes (SMD = 0.09–0.34) [1,52] (see Figure 3). One of the two meta-analyses examining the effect of formative assessment in Mathematics showed a small positive effect (SMD = 0.34;  $p < 0.001$ ) [1], whereas two meta-analyses showed trivial effect sizes (SMD = 0.09–0.13;  $p < 0.001$ ) in Science [1,52]. The study by Lee et al. [1] observed a small positive effect in Art (SMD = 0.29;  $p < 0.001$ ).



**Figure 3.** Standardized mean difference (black diamond) and 95% confidence intervals (black diamond horizontal lines) reported in meta-analyses comparing the baseline with post-formative assessment interventions on student performance in STEAM-related subjects: Art, Mathematics, and Science [1,52]. NS – FA = non-specific formative assessment. Note 1: Positive and negative standardized mean difference values denote favorable and detrimental effects of intervention, respectively. Note 2: Grey and blue areas = trivial and small magnitude, respectively. Note 3: Lee et al. [1] did not report the 95% confidence interval; therefore, only standardized mean difference is reported.

#### 3.5.2. Effect of Formative Assessment on Learning Literacy

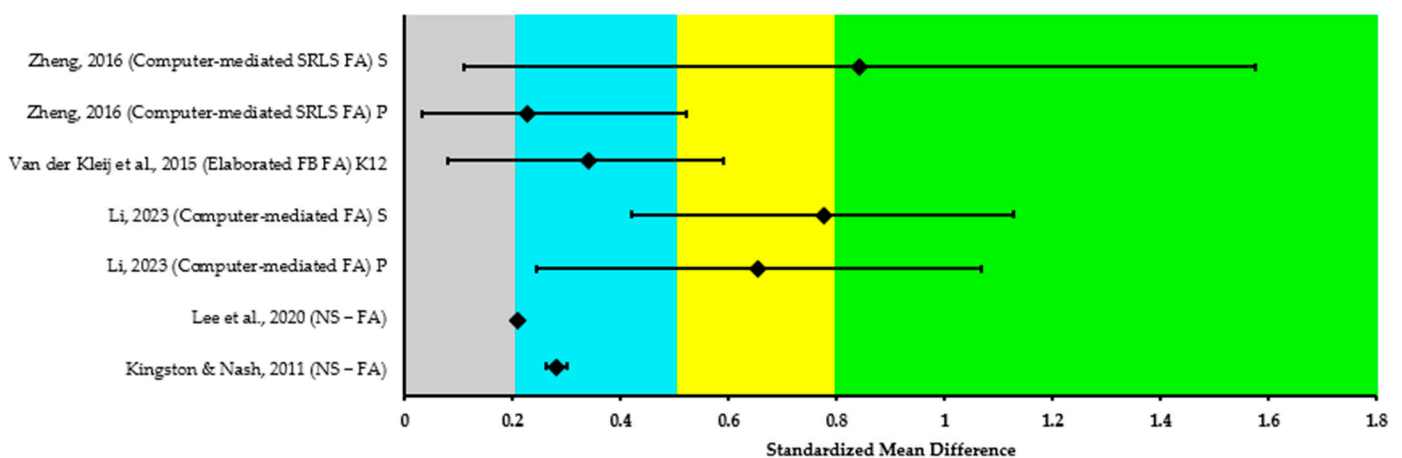
Five studies investigated the effects of formative assessment on student literacy (see Figure 4). One study reported a large positive effect for a formative assessment intervention that involved adult feedback on literacy (SMD = 0.87;  $p < 0.001$ ) [51]. Two studies [51,57] showed a moderate positive effect for formative assessment types on literacy learning, such as non-specific feedback (SMD = 0.61;  $p < 0.001$ ) [51], peer feedback (SMD = 0.58;  $p < 0.001$ ) [51], progress monitoring feedback (SMD = 0.18;  $p = 0.06$ ) [51] and computer-mediated formative assessment (SMD = 0.66–0.77;  $p < 0.01$ ) [57]. One study showed a small positive effect for student-directed formative assessment (SMD = 0.28;  $p < 0.001$ ) [9]. Trivial effects were observed for teacher-directed formative assessment (SMD = 0.12;  $p < 0.01$ ) [9], non-specific formative assessment (SMD = 0.19;  $p < 0.01$ ) [9], integrated (SMD = 0.19;  $p < 0.001$ ) [9] and also the 6 + 1 writing model (SMD = 0.05;  $p = 0.08$ ) [51] type of formative assessment. Only one study performed sub-analysis for school settings (primary and secondary) that indicated a trivial-to-small positive effect on literacy (SMD = 0.18–0.27;  $p < 0.001$ ) [9]. Summarized in Figure 4 are the formative assessment effects on student literacy in the five meta-analyses studies reporting SMD values.



**Figure 4.** Standardized mean difference (black diamond) and 95% confidence intervals (black diamond horizontal lines) reported in meta-analyses comparing the baseline with post-formative assessment interventions on changes in student literacy [1,29,51,52,57]. Lead author name is followed by the type of formative assessment. S: secondary school; P: primary school; ^: effect on writing; \*: effect on vocabulary; #: effect on reading; FA = formative assessment; FB = feedback. Note 1: Positive standardized mean difference values denote favorable effects. Note 2: Grey and blue areas = trivial and small magnitude, respectively, and yellow and green = moderate and large effects, respectively. Note 3: Lee et al. [1] did not report the 95% confidence interval; therefore, only standardized mean difference is reported.

### 3.6. Effect of Computer-Based Formative Assessment on Learning

Five studies investigated the effects of computer-based formative assessment interventions on student learning (see Figure 5). Two of the studies performed sub-analysis of school settings (secondary and primary), with one reporting a large effect in secondary students (SMD = 0.84) and a small effect in primary students (SMD = 0.23;  $p < 0.01$ ) [59]. Another reported moderate effect size in both secondary (SMD = 0.77;  $p < 0.01$ ) and primary students (SMD = 0.66;  $p < 0.01$ ) [57]. Three meta-analyses [1,52,58] indicated small effect sizes (SMD = 0.21–0.34;  $p < 0.05$ ) in computer-based formative assessment.



**Figure 5.** Standardized mean difference (black diamond) and 95% confidence intervals (black diamond horizontal lines) reported in meta-analyses comparing the baseline with post-computer-based

formative assessment interventions on student learning [1,52,57–59]. Lead author name is followed by the specific type of formative assessment. SRLS = self-regulated learning scaffold; P = primary school students; S = secondary school students; K12 = kindergarten to grade 12 students; NS – FA = non-specific formative assessment; FB = feedback; FA = formative assessment. Note: 1: Positive standardized mean difference values denote favorable effects. Note 2: Grey and blue areas = trivial and small magnitude, respectively, and yellow and green = moderate and large effects, respectively. Note 3: Lee et al. [1] did not report the 95% confidence interval; therefore, only standardized mean difference is reported.

### 3.7. Effect on Affective Learning

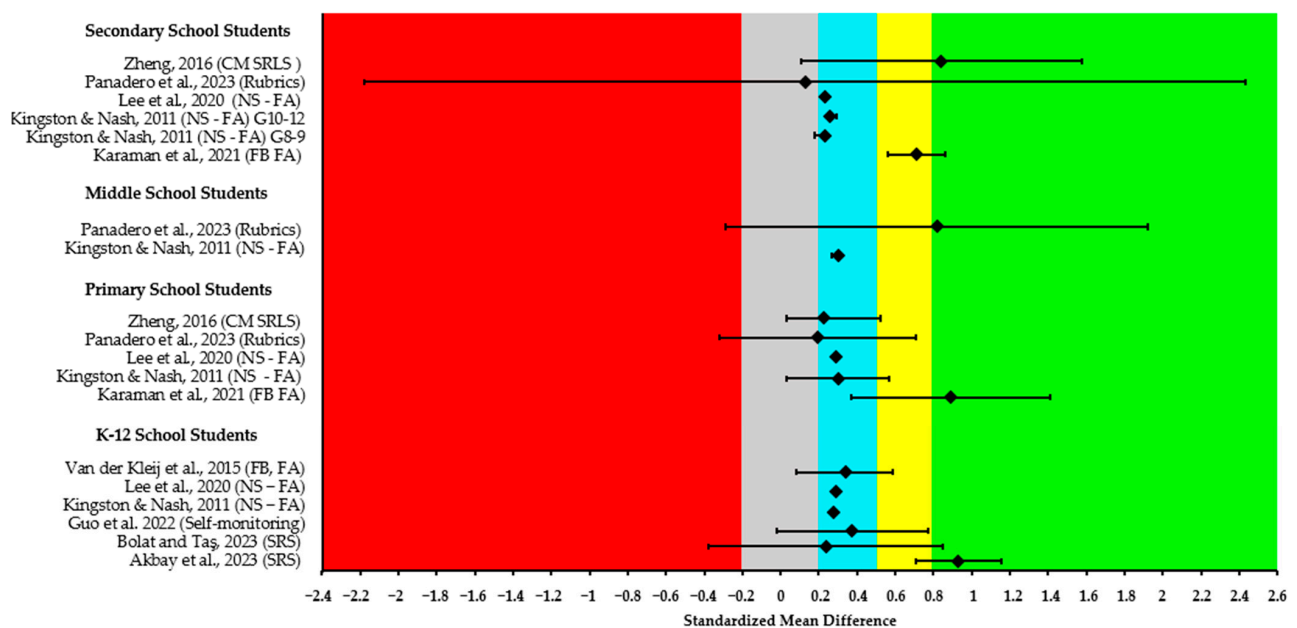
Two studies [30,54] included an examination of the affective learning domain. The impact of formative assessment rubrics on regulated learning varied across primary (SMD =  $-0.047$ ;  $p > 0.05$ ) and middle school students (SMD =  $0.246$ ;  $p < 0.001$ ), respectively [30]. The utilization of formative assessment self-monitoring resulted in a significant increase in strategy use among participants (SMD =  $0.57$ ;  $p < 0.05$ ) [54].

### 3.8. Effect of Formative Assessment on Learning within Settings

In terms of the effect of formative assessment on student learning or achievement, without reference to any subject area, there were 10 studies with 19 effect sizes ranging from trivial to large effect. The meta-analysis effect sizes are provided per school setting (i.e., K-12, primary and secondary).

#### 3.8.1. K-12 School Students

Six studies conducted meta-analyses across the K-12 school setting with six effect sizes for student learning (see Figure 6). In terms of effect, the included meta-analyses indicated large effects for the student response system (SMD =  $0.93$ ;  $p < 0.001$ ) [53], small effects for unspecified types of formative assessment (SMD =  $0.28$ – $0.29$ ;  $p < 0.01$ ) [1,52], and formative assessment in a computer-based learning environment (SMD =  $0.34$ ;  $p < 0.03$ ) [58]. However, Guo [54] and Bolat et al. [36] reported small and non-significant effects of self-monitoring on academic performance (SMD =  $0.37$ ) and gamified formative assessment tools (SMD =  $0.24$ ), respectively.



**Figure 6.** Standardized mean difference (black diamond) and 95% confidence intervals (black diamond horizontal lines) reported in meta-analyses comparing the baseline with post-formative assessment interventions on student academic achievement according to school setting [1,30,36,52–54,56,58,59].

Lead author name is followed by the type of formative assessment. NS – FA = non-specific formative assessment; CM = computer-mediated; SRLS = self-regulated learning scaffold; FB = feedback; FA = formative assessment; SRS = student response system. Note 1: Positive and negative standardized mean difference values denote the favorable and detrimental effects of intervention compared with the control condition. Note 2: Red area = detrimental, grey and blue areas = trivial and small magnitude, respectively, and yellow and green = moderate and large effects, respectively. Note 3: Lee et al. [1] and Kingston et al. [52] did not report the 95% confidence interval; therefore, only standardized mean difference is reported.

### 3.8.2. Primary School Students

Six studies conducted meta-analyses that included primary school students. One study reported a trivial positive effect (SMD = 0.20;  $p = 0.350$ ) [30] of using rubrics for formative assessment. Three studies indicated a small positive effect for the following types of formative assessment: computer-mediated self-regulated learning (SRL) (SMD = 0.23;  $p = 0.026$ ) [59], and unspecified type of formative assessment (SMD = 0.29;  $p < 0.001$ ) [1], (SMD = 0.30 [K-4];  $p < 0.001$ ) [52]. Two meta-analyses observed a large effect size, with feedback formative assessment (SMD = 0.89;  $p < 0.05$ ) [56] and real-time classroom competition (MD = 5.78;  $p < 0.001$ ) [55].

### 3.8.3. Middle School Students

Two studies included meta-analyses for middle school students. Panadero et al. [30] reported a large effect size (SMD = 0.82;  $p = 0.086$ ) for the use of rubrics, and Kingston et al. [52] reported a small effect size (SMD = 0.30;  $p < 0.001$ ) for non-specific formative assessment.

### 3.8.4. Secondary School Students

Five studies performed meta-analyses within the secondary school setting, with six effect sizes reported for learning. Computer-mediated SRL had a large effect (SMD = 0.84;  $p = 0.024$ ) [59]. A moderate effect (SMD = 0.71;  $p < 0.05$ ) was observed for feedback formative assessment [56]. The effectiveness of real-time classroom interactive competition on academic performance had a large effect (MD = 7.60;  $p < 0.001$ ) [55]. A small effect (SMD = 0.23–0.26;  $p < 0.001$ ) was indicated by [52].

## 3.9. Other Influencing Variables

### 3.9.1. Differentiation

Regarding the addition of differentiated instruction or instructional adjustment with formative assessment on learning, meta-regression analysis by Lee et al. [1] indicated that planned instructional adjustment had a small moderating effect (SMD = 0.35;  $p < 0.05$ ), while unplanned instructional adjustment had a trivial effect (SMD = 0.15;  $p > 0.05$ ), and a mixture of unplanned and planned adjustment resulted in a small significant effect (SMD = 0.29;  $p < 0.001$ ). No instructional adjustment indicated a trivial, non-significant effect (SMD = 0.16;  $p > 0.05$ ) on formative assessment in the study by Lee et al. [1]. For formative assessment with differentiated instruction on reading achievement, the subgroup analysis conducted by Xuan et al. [9] indicated that students had gained significantly more with a small effect (SMD = 0.241;  $p = 0.000$ ) compared with no differentiation (SMD = 0.05;  $p = 0.07$ ).

### 3.9.2. Professional Development

In terms of expertise level, two of the included meta-analyses [1,52] indicated professional development for teachers in the use of formative assessment had a small positive effect (SMD = 0.30;  $p < 0.001$ ). Lee et al. [1] conducted a meta-regression analysis regarding the level of teacher exposure to professional development and observed that one-time professional development (SMD = 0.27;  $p < 0.001$ ) was effective; however, ongoing professional development had the greatest effect (SMD = 0.30;  $p < 0.001$ ). Regarding teachers

who had engaged in no professional development, there was no significant effect ( $p > 0.05$ ) on learning.

### 3.9.3. Teacher and/or Student-Directed

Two meta-analyses [1,9] specifically examined the effects of formative assessment initiated or directed by the teacher and/or student. The subgroup analysis by Xuan et al. [9] observed a small positive effect (SMD = 0.28;  $p < 0.001$ ) of student-directed formative assessment on reading achievement, a trivial positive effect (SMD = 0.19;  $p < 0.001$ ) for integration of student and teacher-directed formative assessment, and teacher-directed also demonstrated a trivial positive effect (SMD = 0.12;  $p < 0.05$ ). Lee et al. [1] conducted a meta-regression showing that teacher-directed formative assessment had a trivial moderating effect (SMD = 0.18;  $p < 0.05$ ), student-directed feedback had a medium effect (SMD = 0.61;  $p < 0.001$ ), while interventions that focused on both teacher and student as the main initiators of the formative assessment had a trivial non-significant effect (SMD = 0.13;  $p > 0.05$ ). The meta-analyses by Graham et al. [51] included a sub-group analysis of writing interventions according to the delivery of formative assessment involving feedback. Graham et al. [51] reported teacher-directed formative assessment had the greatest effect (SMD = 0.87;  $p < 0.001$ ), compared with self-assessment (SMD = 0.62;  $p < 0.001$ ) and peer assessment (SMD = 0.58;  $p < 0.001$ ).

## 4. Discussion

This umbrella review aimed to systematically assess the available meta-analytical evidence to then examine the effects of formative assessment interventions on student learning. The present umbrella review analyzed 13 meta-analyses assessing the efficacy of formative assessment interventions on a range of outcomes, predominantly in the cognitive learning domain and to a lesser degree in the affective learning domains. Based on the meta-analytical evidence, the umbrella review's main finding is that formative assessment enhances students' learning.

Even though formative assessment interventions resulted in a positive effect for most outcomes, effect sizes were only one element of the evaluation of the included meta-analyses in this review. The search process and identification of eligible papers for the umbrella review highlighted a greater depth of formative assessment research in higher or tertiary education, where most evidence is available, than in the K-12 schooling sector. The most concerning findings from our study are the lack of methodological rigor, robustness, or quality of evidence GRADE in the available studies, the paucity of meta-analytical studies focused on primary and secondary schooling, and the lack of evidence available for examining biological sexes and economic and social demographics of schools as moderators.

### 4.1. Generalizability of the Results

Most of the meta-analyses examined primary studies that included participants from tertiary education and secondary and primary school settings [30,36,53–59]. A crude consideration of generalizability based on surface similarity may lead us to decide that while the formative assessment intervention may generally apply to students (tertiary, primary, and secondary), if we look more closely at the study populations and compare their characteristics, we might note that tertiary students are more educated (e.g., knowledge, understanding, skills) and have more years of learning experience. As a result, most moderator analyses could not be considered in this umbrella review due to the distortion of data resulting from the inclusion of tertiary education students. The exclusion of such moderator analyses is not only due to methodological considerations but also because different stages of lifespan development exhibit distinct variables that influence teaching and learning processes [60] and, consequently, affect the outcomes of formative assessment experiences.

Another consideration in analyzing the generalizability of the results is that most meta-analyses were conducted with general populations among a mixture of biological

sexes, thus limiting the generalizability of findings to single-sex-based schools and classrooms. Furthermore, most meta-analyses in this umbrella review only reported total effect sizes according to the level of education or education setting (primary, secondary, and K-12), with only one providing a sex-based subgroup effect [30]. Some primary studies examining the effect of formative assessment have found different effects for males and females [61], and also that formative assessment can be effective in improving results while also reducing the gender gap in literacy development and numeracy [62]. Because of the lack of information regarding the distribution of biological sexes, analysis of sex as a subgroup, and consideration for various school settings (i.e., single sex or coeducational, socio-economic demographics), all included meta-analyses were categorized as having ‘serious indirectness’ in the GRADE assessment [48]. Future meta-analysis studies should include an examination of both males and females and analyze their data separately to determine if there is a difference in responses to formative assessment intervention between the sexes. However, it is important to recognize that such comparisons would require larger sample sizes to achieve sufficient statistical power.

#### 4.2. Quality of the Included Meta-Analyses

Using the AMSTAR-2 assessment, the methodological quality of the included meta-analyses varied from low to high. However, most studies (77%) were of moderate quality, with two rated as low and only one rated as high. Although it is not mandatory to pre-register the meta-analysis protocols on a specific platform, it is a fundamental aspect recommended by a significant number of journals. Pre-registration of meta-analysis protocols is crucial for transparency, preventing duplication of efforts, ensuring that the research process is well-documented, and minimizing the risk of bias [63]. However, in this case, only one was registered on PROSPERO [55]. The paucity of meta-analysis papers in this review with pre-registered protocols can likely be attributed to the temporal context in which earlier studies were conducted prior to the common adoption of this practice and a time when there were few or non-existent general (i.e., no specific field of study) registration platforms (e.g., Open Source Framework—OSF) or the absence of databases for Supplementary Materials [64]. Furthermore, most eligible meta-analyses did not provide an explicit statement that the protocol used was established prior to the conduct of the review. However, the absence of this missing element in the eligible papers for this umbrella review has contributed to the low- to moderate methodological quality.

For AMSTAR-2 Item 7 (list of excluded studies with reasons for exclusion), the included meta-analyses did not provide a list of excluded studies to then justify their exclusion. This omission may be attributed to limitations such as table/figure/word restrictions or, as previously mentioned, the absence of databases for Supplementary Materials, affecting the authors’ ability to present all extracted information from the included primary studies. However, it is also plausible that the authors, whose research background is in education, may have been unaware of the significance of these methodological quality characteristics.

A low number of studies [30,36,51,52,55,57] conducted adequate evaluations with varying degrees of risk of bias, suggesting potential flaws in the evidence, thus reducing the overall certainty of the evidence. Moreover, meta-analysis studies in this umbrella that lacked an evaluation of the risk of bias in their primary studies examined (i.e., did not comply with AMSTAR-2 guidelines, Item 12), contributed to a downgrading of the robustness of the evidence (e.g., low-to-very-low GRADE). Therefore, the low-to-very-low certainty of the currently available evidence precludes robust recommendations regarding optimal formative assessment strategies for learning in K-12 students.

#### 4.3. Formative Assessment Practices Enhance Learning

Most studies indicated improvement in K-12 student learning due to various types of formative assessment interventions and no negative effects. In the study by Kingston et al. [52], a sub-group analysis identified curriculum-embedded assessment schemes that involved implementing open-ended formative assessments as having a negative trivial

effect. Considering the overall positive effects, applying the experimental protocols used for applying formative assessment may be an effective strategy to monitor, promote, and optimize student learning in general and within specific key learning areas (i.e., Mathematics, Science, English, and Art) [1,9,51,52,57]. Additionally, they provide teachers with a tool to evaluate the efficacy of their classroom practices, identify learning gaps, and make evidence-based decisions to implement necessary adjustments [65]. Consequently, formative assessment is recognized as an important component of effective pedagogy within daily lessons to increase student learning and enhance teacher quality, and it should be present in classrooms [66].

#### 4.4. Type of Intervention

The impact of formative assessment interventions on student learning was examined in the meta-analyses included in this review through both a broad curriculum lens (i.e., involving a range of key learning areas) and a subject-specific curriculum (i.e., one key learning area) lens. The studies identified three main sources of formative assessment practices: student-directed, teacher-directed, or integrated (teacher and student). In this umbrella review, six studies [1,9,30,51,54,59] examined the effect of student-directed types of formative assessment (i.e., peer- or self-assessment, rubrics, or scaffold) and six meta-analyses [9,36,51,53,55,58] examined teacher-directed types of formative assessment via additional analyses (i.e., sub-analyses). The remaining two studies examined integrated practices involving both teacher and student in the assessment process [1,9]. Only one study explored biological sex as a moderator and found no significant differences [30]. Differentiating between these types is crucial because each approach influences student learning in unique ways. For example, student-directed assessments, such as peer or self-assessment, empower learners to take an active role in their evaluation, promoting self-regulation and deeper understanding, while integrated formative assessments, which involve both teachers and students in the process, foster a collaborative approach, combining the strengths of both perspectives to enhance learning outcomes. By understanding these distinctions, educators can better implement formative assessment strategies aligned with their teaching goals and the needs of their students.

##### 4.4.1. Student Centered

Across a range of key learning areas, learning in general appears to be most positively affected by student-initiated formative assessment feedback and, to a lesser extent, by teachers' formative assessment feedback and mixed feedback from both students and teachers [1]. In the study by Lee et al. [1], student-initiated formative assessment had a moderate effect compared with teacher-directed or integrated assessment that had a trivial effect on learning. Specific types of student-directed forms of formative assessment, such as rubrics, have demonstrated a small to large positive effect on students' academic performance in primary (SMD = 0.196;  $p = 0.350$ ), middle (SMD = 0.819;  $p = 0.086$ ), and secondary school students (SMD = 0.127;  $p = 0.612$ ) [30]. However, in the study by Panadero et al. [30], the effect of rubrics on student self-regulated learning was small (SMD = 0.246;  $p < 0.001$ ) in middle school students and detrimental (SMD =  $-0.047$ ;  $p = 0.849$ ) in primary school students. The results from the meta-analyses by Panadero et al. [30] lend support to the conclusion that student-directed formative assessment strategies such as rubrics are more beneficial when used with middle and secondary school students.

Self-monitoring, another student-centered approach, showed small to moderate effects for academic performance (SMD = 0.37;  $p > 0.05$ ) and self-monitoring ability (SMD = 0.57;  $p < 0.05$ ), indicating significant enhancements in self-monitoring ability and a positive effect on academic performance [54]. This finding accords with the observations of Reid et al. [67], which found self-monitoring was effective in improving student behavior, productivity, and accuracy in the school classroom setting. The meta-analysis by Guo [54] revealed a moderate effect size in the association between self-monitoring and academic performance, but the findings did not reach statistical significance ( $p > 0.05$ ), indicating a need for

further investigation into the nuanced dynamics of this relationship. However, the results of the study by Guo [54] broadly support the work of Panadero et al. [30] in this area, which links self-regulated learning and metacognitive training. This is consistent with the meta-analysis by Zheng [59], which showed that the use of self-regulated learning scaffolds in computer-based learning environments produced a positive moderate-to-large effect on academic performance. These outcomes suggest that student-directed formative assessment strategies positively moderate academic performance more effectively than teacher-directed and mixed, while also improving affective learning. This combination of findings provides some support for the conceptual premise that learners need to be encouraged to have an active role in the formative assessment process to maximize impact on learning.

#### 4.4.2. Student-Directed Formative Assessment According to School Setting

Despite the consistent moderate effect of student-directed formative assessment on student academic performance, the findings cannot be extrapolated to all age groups. In the study by Panadero et al. [30], the effect of using rubrics increased with the increased age of students, benefiting secondary students more. Previous reviews concentrating on studies in higher education [68,69] have also indicated a link between a student's educational level and the effectiveness of rubrics. The meta-analysis by Zheng [59] indicated that student-directed formative assessment in the form of self-regulated scaffolds had a large effect in secondary students in contrast to a moderate effect in primary students. Several explanations exist for the greater effectiveness in secondary students, such as increased cognitive and metacognitive readiness to regulate their learning through monitoring and active adaptation [70]. This explanation is supported by a cognitive and constructivist view of learning, where students are at the formal operational stage of reasoning, suggesting that learners benefit from a more independent active role in formative assessment, which is considered essential for its success [3,33,61].

#### 4.4.3. Student-Directed Formative Assessment and Subject Area

In this umbrella review, two meta-analyses [9,51] included a sub-analysis to explore the effects of student-directed formative assessment on literacy learning, more specifically writing [51] and reading [9]. The results of these meta-analyses on formative assessment in reading indicate, similar to other findings [1,30,54,59], that student-directed formative assessment has a moderate positive effect. This moderate effect was larger than teacher-directed or integrated approaches in K-12 settings. Similarly, Graham et al. [51] observed that student-directed formative assessment such as peer feedback and self-assessment resulted in moderate effects. However, teacher-directed formative assessment was more favorable, demonstrating a large positive effect. The contrast between these results between reading and writing can partly be explained by the extreme complexity of writing skills [71–73]. Learning to write often relies on teacher-directed instructional writing practices [71,74,75]. Due to its complexities in content structuring, grammar, punctuation, spelling, and language mechanics, writing requires teachers to guide students through processes by providing formative assessment such as feedback, modeling effective writing techniques, and fostering a supportive environment. This process helps students develop their writing skills, whereas reading primarily focuses on comprehension.

#### 4.4.4. Teacher-Directed Approaches

The umbrella review findings revealed that teacher-directed approaches enhance students' academic achievement. Lee et al. [1] observed trivial positive effects for teacher-directed formative assessment in K-12; however, they were less significant compared with student-directed approaches that exhibited a moderate effect. It is important to consider the three recent meta-analyses [36,53,55] that examined a specific type of teacher-directed formative assessment and the effect on learning. Akbay et al. [53] examined the effects of student response systems (e.g., clicker, Kahoot) and found a large effect on both students'

cognitive and non-cognitive learning outcomes, with no significant difference between these domains. These findings are consistent with those of Jurado-Castro et al. [55], who found that real-time classroom interactive competition interventions also had a large positive and significant effect on student academic performance. Furthermore, a third meta-analysis that investigated the effect of gamified formative assessment tools also showed a small but positive effect on student achievement. These studies highlighted applications such as Kahoot and Socrative as highly effective due to their game-based features [36,53,55]. These tools serve the purpose of formative assessment while also engaging and motivating students more than traditional student response systems such as clickers [76,77]. The gaming elements may enhance student motivation and also performance [36].

The observed effect of teacher-directed formative assessment strategies such as student response systems could be attributed to increased means to address students' personal characteristics (i.e., preferred learning style, lack of self-esteem, hesitation, shyness, self-regulation) [78] or classroom contextual factors (i.e., time limit, class size, and dynamics) [79], which can pose pedagogical challenges. These challenges can reduce students' motivation, self-efficacy, participation, and interaction [80,81], which are crucial non-cognitive learning outcomes that impact cognitive processes (e.g., memorization, retention) and outcomes such as recall. The application of gamification can encourage students to be more motivated and ambitious to learn, resulting in enhanced academic achievement [36,53,55,82]. Using real-time interactive competitions in the learning process can help overcome the aforementioned challenges and enhance the quality of student learning, significantly impacting class dynamics, participation, motivation, and overall learning experiences [36,55,83]. In relation to using these types of formative assessment, teachers need to be aware of the potential novelty effect that could be a conditioning factor in short-term success. In addition, it is suggested teachers take into account the difficulties with memorizing each student's areas of strength and for improvement [80] to ensure all components of formative assessment are completed.

#### 4.5. Effect of Formative Assessment within Various Contexts

##### 4.5.1. Formative Assessment in STEAM-Related Subjects

This umbrella review indicates that formative assessment interventions have significant trivial positive to small effects on student learning in STEAM-related subjects [1,52]. For Mathematics and Science, one meta-analysis observed trivial positive effects for both Mathematics and Science [52] and another small positive effect for Mathematics and trivial positive effects for Science [1]. These results demonstrate the benefits of including formative assessment in the teaching and learning of these subjects. It is also important to highlight that some experimental aspects may have influenced the observed effects in the included studies, such as the type of study design and the students' pre-existing experience with formative assessment practices. In Mathematics, characterized by the need for correctness and determining a solution, formative assessment is inherently built into instruction through problem-solving activities, classroom discussions, peer collaboration, and teacher-directed quick checks for understanding. Student-directed quick checks (e.g., back-of-book answers) for understanding using prescribed sets of answers also allow for continual monitoring of student progress and adjustment of instruction accordingly. Therefore, formative assessment interventions delivered through research investigations do not present distinctly different approaches from what is already practiced.

##### 4.5.2. Formative Assessment Using Computers

One of the key findings of our umbrella review was the positive effect of computer-based formative assessment on student learning. Five meta-analyses examined the effect of computer-based assessment [1,52,57–59]. Van der Kleij et al. [58] conducted a meta-analysis to determine the effects of formative assessment feedback in computer-based learning, observing a small effect on learning. These positive results are similar to those in the meta-analysis by Li [57], which indicated a moderate effect of computer-mediated formative

assessment in both primary and secondary students. Li [57] also found that computer-mediated formative assessment is significantly more effective than traditional methods for learning vocabulary. However, the meta-analysis by Zheng [59] that examined the effects of self-regulated learning scaffolds on academic achievement in computer-based learning environments produced significant, large positive effects in secondary students and small positive effects in primary students. Other meta-analyses [1,52] examining the effect of computer-based formative assessment more broadly on learning observed a small positive effect ( $SMD = 0.27-0.28$ ) on student learning. The effect of computer-based formative assessment from the meta-analysis by Graham [71] is often cited in other papers; however, it was not included in this umbrella review, as the studies used did not include control groups. Therefore, computer-assisted formative assessment facilitates educators in focusing on instructional practices in response to the information collected more objectively while minimizing the time and effort required for data collection and analysis [84]. In addition, it also nurtures the development of student self-regulation skills and self-efficacy [58].

#### 4.5.3. Formative Assessment and Literacy

The findings of this umbrella review demonstrated that formative assessment has a significant trivial-to-large effect on literacy proficiency, with no negative effects, according to the included meta-analyses [1,9,52,57,71]. The writing domain of literacy was most positively affected by formative assessment, showing a large effect [71], followed by the vocabulary domain with a moderate effect [57], and lastly, reading with a trivial-to-small effect [9]. An important finding of this umbrella review is that formative assessment practices on literacy add a range of effective strategies for teachers to meet the wide variety of learning needs that students present with each day [85]. The professional judgement of teachers or intuition, along with the effect sizes of various interventions, can guide the effective use of formative assessment in the classroom context to teach a concept or skill [8]. In addition, the development of literacy skills is both complex and multifaceted and can vary significantly for each student. To help student literacy skills to constantly develop and improve, teachers need to frequently monitor their progress using various formative assessment strategies and address learning issues that arise through differentiation and other practices. Xuan et al. [9] indicated that complementing formative assessment with deliberate differentiation in teaching strategies results in a significant small positive effect, leading to better reading outcomes. Similarly, in the context of general student learning, Lee et al. [1] confirmed that formative assessment strategies with planned and unplanned instructional adjustment had a small effect size compared with those without instructional adjustment, resulting in a significant trivial effect. Responsive formative assessment strategies adapt and continually evolve to align with each student's unique stage of literacy growth and development. The iterative and ongoing nature of formative assessment, interconnected with responsive adjustments, is an important enabler for the teacher to foster continuous improvement and tailor support to meet individual student needs while also providing essential scaffolding for literacy learning.

#### 4.6. Overall Effect of Formative Assessment

Advocates of formative assessment contend that it is an effective classroom pedagogical technique for enhancing and supporting student learning [86]. While systematic reviews and meta-analyses have supported the benefits of formative assessment, this umbrella review systematically organizes and evaluates existing evidence from multiple meta-analyses on all learning outcomes associated with types of formative assessment educational contexts. Overall, the results suggest that formative assessment positively affects student learning in K-12 students, with an overall positive effect size ranging from trivial to large.

Recently, educational researchers have increasingly advocated for the important role of formative assessment in enhancing and promoting learning in educational practice [66]. Formative assessment is an important element of each lesson, providing new informa-

tion about the learning process to the teacher or student, thus facilitating progressive achievement and later performance [87]. The results of this umbrella review support the position held by researchers that formative assessment is an effective classroom technique for improving student learning [37]. Moreover, the most obvious finding to emerge from the results is that formative assessment is more effective than the absence of assessment intervention of this kind. Similar to the primary studies used in the meta-analyses, the findings suggest that deliberately structuring classroom activities using a planned formative assessment strategy may effectively promote learning. Furthermore, using student-directed formative assessment such as scaffolds, rubrics, and peer assessment where appropriate optimizes the use of teaching resources by freeing the teacher to focus on those students' experiencing difficulties or engaging in cognitively demanding complex tasks. The findings indicate that learners' active role is crucial for effective formative assessment implementation and that formative assessment can be effective across a wide range of subject areas. Pragmatically, as previously alluded to, this suggests that teachers can implement different forms of formative assessment in various ways, tailored to the specific characteristics and constraints of their classroom context.

#### 4.7. Strengths and Limitations

This umbrella review provides a rigorous, comprehensive, and current synthesis of evidence in the K-12 school context of formative assessment interventions for improving student learning. This review adhered to the prescribed methodology and guidelines [88,89] for conducting umbrella overviews of meta-analyses, including duplicate independent selection of studies, assessment of eligibility, and data extraction. The methods used in this review were registered prospectively with the Open Science Framework, and the search strategy was defined and deposited before conducting the initial search. Although the umbrella review has demonstrated the effectiveness of formative assessment, certain limitations need to be considered.

Firstly, only 13 meta-analyses met the inclusion criteria in this umbrella review. However, the total sample size was >256,000 students. Secondly, although most ( $n = 11$ ) of the included meta-analyses were of moderate methodological quality, the global assessment of the robustness of the evidence (i.e., GRADE) was very low ( $n = 9$ ), low ( $n = 3$ ), and moderate ( $n = 1$ ). Some AMSTAR-2 and GRADE criteria were either under-reported or under-represented. Moreover, it is worth noting that while the meta-analyses investigated similar research questions (i.e., the effect on learning or academic achievement), their methodological approaches varied, including search strategies, selection criteria, and analytical approaches. It is important to note that a minimal number (<5%) of primary studies were included in multiple meta-analyses. Also considering the differences in scope and published dates of the included meta-analyses, the authors consider the minimal overlap of primary studies not to be a critical flaw in this umbrella review.

Another major limitation that needs to be addressed is the need for more meta-analyses to be selective and to only include randomly controlled trials. In most of the meta-analyses in this umbrella review, the primary studies included quasi-experiments randomly allocated at the classroom level. This presents a challenge in isolating the intervention effect from classroom-level effects like teacher quality; consequently, statistical inferences may be biased [90]. The results indicate that randomized controlled trials produce more conservative estimates than do studies with quasi-experimental designs of the effect of formative assessment [9,30]. Addressing this limitation would enhance the overall quality of research and enable more accurate conclusions. However, despite these limitations, the current umbrella review shows that formative assessment positively impacts K-12 student learning. The umbrella review of evidence has highlighted new challenges, underscoring problematic areas in theory and application for researchers and teachers.

## 5. Conclusions

The aim of this umbrella review was to investigate the effects of formative assessment on K-12 student learning and to evaluate the methodological quality of eligible meta-analyses. This umbrella review has shown that the use of formative assessment as an instructional practice is beneficial for improving academic performance. These findings align with existing theories on formative assessment and instructional best practices, offering support for their ongoing implementation in K-12 classrooms. For school leaders and teachers, these findings are particularly useful for guiding decision-making and policy formulation. Practically speaking, the results encourage the use of formative assessment as an effective and sustainable strategy for monitoring and guiding teaching and learning processes, promoting continuous improvement for both students and teachers. In this regard, educators can better advocate for resources, training, and professional development that support its integration into daily teaching practices. Furthermore, this review highlights the need for cautious interpretation of existing data, prompting leaders to encourage and participate in additional research that examines the diverse factors influencing its effectiveness.

While most of the included meta-analysis studies achieved moderate quality (AMSTAR-2), the robustness of the evidence is mostly low or very low (GRADE); therefore, it is advised that the outcomes from this umbrella review should be considered cautiously. Further research should be conducted to understand the contextual and educational variables that moderate the effectiveness of formative assessment. More specifically, research is needed to distinguish effectiveness based on gender groups and school types, such as socio-economic demographics and coeducational and single-sex settings, to improve the generalizability of findings. Overall, the present findings are encouraging for those looking to utilize formative assessment to enhance teaching practices and student learning.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su16177826/s1>, Table S1: The list of meta-analyses excluded after full-text screening with reason; Table S2; Summary of the meta-analyses included in the umbrella review; Table S3: Results of the assessment of the methodological quality of the included meta-analyses using AMSTAR-2; Table S4: Global GRADE rating for all the meta-analyses papers included in the umbrella review; Table S5: GRADE rating for the meta-analyses papers examining the effect of formative assessment via computers; Table S6: GRADE rating of papers according to type of formative assessment examining the effect on literacy; Table S7: GRADE rating for papers examining the effect of formative assessment according to school setting; Table S8: GRADE rating for papers examining the effect on STEAM related subjects.

**Author Contributions:** Conceptualization, A.S., and K.T.; methodology, A.S., R.R.-C., and K.T.; validation, K.T., R.R.-C., R.F., and B.C.-T.; formal analysis, A.S., G.H., D.R.G., E.G., and B.C.-T.; investigation, A.S., R.F., K.T., R.R.-C., E.G., D.R.G., G.H., and B.C.-T.; writing—original draft preparation, A.S., R.F., K.T., R.R.-C., E.G., D.R.G., G.H., and B.C.-T.; writing—review and editing, A.S., R.F., K.T., R.R.-C., E.G., D.R.G., G.H., B.C.-T., and Q.X.; visualization, A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work is supported by national funding through the Portuguese Foundation for Science and Technology, I.P., under project UID04045/2020 and the National Research Agency of Chile for the support provided through the FONDECYT project (No. 1230609) 'When school regulations change, should teacher training also change? Impact of the new decree on assessment, grading, and school promotion in the preparation of future Physical Education teachers' (original name in Spanish: ¿Cuándo cambia la normativa de la escuela también debe hacerlo la formación docente?

Impacto del nuevo decreto sobre evaluación, calificación y promoción escolar en la preparación de los/as futuros/as profesores/as de Educación Física).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Lee, H.; Chung, H.; Zhang, Y.; Abedi, J.; Warschauer, M. The Effectiveness and Features of Formative Assessment in US K-12 Education: A Systematic Review. *Appl. Meas. Educ.* **2020**, *33*, 124–140. [\[CrossRef\]](#)
- McCallum, S.; Milner, M.M. The effectiveness of formative assessment: Student views and staff reflections. *Assess. Eval. High. Educ.* **2021**, *46*, 1–16. [\[CrossRef\]](#)
- Clark, I. Formative Assessment: Assessment Is for Self-regulated Learning. *Educ. Psychol. Rev.* **2012**, *24*, 205–249. [\[CrossRef\]](#)
- Dayal, H. How Teachers use Formative Assessment Strategies during Teaching: Evidence from the Classroom. *Aust. J. Teach. Educ.* **2021**, *46*, 1–21. [\[CrossRef\]](#)
- Näsström, G.; Andersson, C.; Granberg, C.; Palm, T.; Palmberg, B. Changes in Student Motivation and Teacher Decision Making When Implementing a Formative Assessment Practice. *Front. Educ.* **2021**, *6*, 616216. [\[CrossRef\]](#)
- Cowie, B.; Moreland, J. Leveraging disciplinary practices to support students' active participation in formative assessment. *Assess. Educ. Princ. Policy Pract.* **2015**, *22*, 247–264. [\[CrossRef\]](#)
- Cowie, B.; Khoo, E. An Ecological Approach to Understanding Assessment for Learning in Support of Student Writing Achievement. *Front. Educ.* **2018**, *3*, 11. [\[CrossRef\]](#)
- Black, P.; Wiliam, D. Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* **2009**, *21*, 5–31. [\[CrossRef\]](#)
- Xuan, Q.; Cheung, A.; Sun, D. The effectiveness of formative assessment for enhancing reading achievement in K-12 classrooms: A meta-analysis. *Front. Psychol.* **2022**, *13*, 990196. [\[CrossRef\]](#)
- Hornby, G.; Greaves, D. *Essential Evidence-Based Teaching Strategies: Ensuring Optimal Academic Achievement for Students*; Springer: Berlin/Heidelberg, Germany, 2022.
- Andersson, C.; Palm, T. Characteristics of improved formative assessment practice. *Educ. Inq.* **2017**, *8*, 104–122. [\[CrossRef\]](#)
- Granberg, C.; Palm, T.; Palmberg, B. A case study of a formative assessment practice and the effects on students' self-regulated learning. *Stud. Educ. Eval.* **2021**, *68*, 100955. [\[CrossRef\]](#)
- Ozan, C.; Kincal, R. The Effects of Formative Assessment on Academic Achievement, Attitudes toward the Lesson, and Self-Regulation Skills. *Educ. Sci. Theory Pract.* **2018**, *18*, 85–118. [\[CrossRef\]](#)
- Johnson, C.C.; Sondergeld, T.A.; Walton, J.B. A Study of the Implementation of Formative Assessment in Three Large Urban Districts. *Am. Educ. Res. J.* **2019**, *56*, 2408–2438. [\[CrossRef\]](#)
- Darling-Hammond, L.; Flook, L.; Cook-Harvey, C.; Barron, B.; Osher, D. Implications for educational practice of the science of learning and development. *Appl. Dev. Sci.* **2020**, *24*, 97–140. [\[CrossRef\]](#)
- Birenbaum, M.; DeLuca, C.; Earl, L.; Heritage, M.; Klenowski, V.; Looney, A.; Smith, K.; Timperley, H.; Volante, L.; Wyatt-Smith, C. International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures Educ.* **2015**, *13*, 117–140. [\[CrossRef\]](#)
- Clark, I.D. Formative Assessment: Policy, Perspectives and Practice. *Fla. J. Educ. Adm. Policy* **2011**, *4*, 158–180.
- Gordon, E.; McGill, M.; Sands, D.; Kalinich, K.; Pellegrino, J.; Chatterji, M. Bringing formative classroom assessment to schools and making it count. *Qual. Assur. Educ.* **2014**, *22*, 339–352. [\[CrossRef\]](#)
- Oo, C.Z.; Alonzo, D.; Asih, R.; Pelobillo, G.; Lim, R.; San, N.M.H.; O'Neill, S. Implementing school-based assessment reforms to enhance student learning: A systematic review. *Educ. Assess. Eval. Account.* **2023**, *36*, 7–30. [\[CrossRef\]](#)
- Yan, Z.; Li, Z.; Panadero, E.; Yang, M.; Yang, L.; Lao, H. A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assess. Educ. Princ. Policy Pract.* **2021**, *28*, 228–260. [\[CrossRef\]](#)
- Lutovac, S.; Flores, M.A. Conceptions of assessment in pre-service teachers' narratives of students' failure. *Camb. J. Educ.* **2022**, *52*, 55–71. [\[CrossRef\]](#)
- Li, Z.; Yan, Z.; Chan, K.K.Y.; Zhan, Y.; Guo, W.Y. The role of a professional development program in improving primary teachers' formative assessment literacy. *Teach. Dev.* **2023**, *27*, 447–467. [\[CrossRef\]](#)
- Ahmedi, V. Teachers' Attitudes and Practices Towards Formative Assessment in Primary Schools. *J. Soc. Stud. Educ. Res.* **2019**, *10*, 161–175.
- DeLuca, C.; Chapman-Chin, A.; Klinger, D.A. Toward a Teacher Professional Learning Continuum in Assessment for Learning. *Educ. Assess.* **2019**, *24*, 267–285. [\[CrossRef\]](#)
- Yan, Z.; King, R.B. Assessment is contagious: The social contagion of formative assessment practices and self-efficacy among teachers. *Assess. Educ. Princ. Policy Pract.* **2023**, *30*, 130–150. [\[CrossRef\]](#)
- Schütze, B.; Rakoczy, K.; Klieme, E.; Besser, M.; Leiss, D. Training effects on teachers' feedback practice: The mediating function of feedback knowledge and the moderating role of self-efficacy. *ZDM* **2017**, *49*, 475–489. [\[CrossRef\]](#)
- Lobos, K.; Bustos, C.; Saez-Delgado, F.; Cobo-Rendon, R.; Bruna, C. Promoting ASC in the primary education classroom: The role of teacher training. *Int. J. Sch. Educ. Psychol.* **2023**, *11*, 233–244. [\[CrossRef\]](#)

28. Hebbecker, K.; Souvignier, E. Formative Assessment im Leseunterricht der Grundschule—Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Z. Für Erzieh.* **2018**, *21*, 735–765. [[CrossRef](#)]
29. Aust, L.; Schütze, B.; Hochweber, J.; Souvignier, E. Effects of formative assessment on intrinsic motivation in primary school mathematics instruction. *Eur. J. Psychol. Educ.* **2023**, *39*, 1–24. [[CrossRef](#)]
30. Panadero, E.; Jonsson, A.; Pinedo, L.; Fernández-Castilla, B. Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: A Meta-analytic Review. *Educ. Psychol. Rev.* **2023**, *35*, 113. [[CrossRef](#)]
31. Hebbecker, K.; Förster, N.; Forthmann, B.; Souvignier, E. Data-based decision-making in schools: Examining the process and effects of teacher support. *J. Educ. Psychol.* **2022**, *114*, 1695–1721. [[CrossRef](#)]
32. Luo, W.; Lim, S.Q.W. Perceived formative assessment and student motivational beliefs and self-regulation strategies: A multilevel analysis. *Educ. Psychol.* **2024**, *44*, 1–19. [[CrossRef](#)]
33. Andrade, H.L. A Critical Review of Research on Student Self-Assessment. *Front. Educ.* **2019**, *4*, 87. [[CrossRef](#)]
34. Bennett, R.E. Formative assessment: A critical review. *Assess. Educ. Princ. Policy Pract.* **2011**, *18*, 5–25. [[CrossRef](#)]
35. Double, K.S.; McGrane, J.A.; Hopfenbeck, T.N. The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educ. Psychol. Rev.* **2020**, *32*, 481–509. [[CrossRef](#)]
36. Bolat, Y.I.; Taş, N. A meta-analysis on the effect of gamified-assessment tools' on academic achievement in formal educational settings. *Educ. Inf. Technol.* **2023**, *28*, 5011–5039. [[CrossRef](#)]
37. McMillan, J.H.; Venable, J.C.; Varier, D. Studies of the effect of formative assessment on student achievement: So much more is needed. *Pract. Assess. Res. Eval.* **2013**, *18*, 1–15.
38. Aromataris, E.; Fernandez, R.; Godfrey, C.M.; Holly, C.; Khalil, H.; Tungpunkom, P. Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *JBI Evid. Implement.* **2015**, *13*, 132–140. [[CrossRef](#)]
39. Fusar-Poli, P.; Radua, J. Ten simple rules for conducting umbrella reviews. *BMJ Ment. Health* **2018**, *21*, 95–100. [[CrossRef](#)]
40. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *74*, 790–799. [[CrossRef](#)]
41. Shea, B.J.; Reeves, B.C.; Wells, G.; Thuku, M.; Hamel, C.; Moran, J.; Moher, D.; Tugwell, P.; Welch, V.; Kristjansson, E.; et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* **2017**, *358*, j4008. [[CrossRef](#)]
42. Hariton, E.; Locascio, J.J. Randomised controlled trials—the gold standard for effectiveness research. *BJOG Int. J. Obstet. Gynaecol.* **2018**, *125*, 1716. [[CrossRef](#)] [[PubMed](#)]
43. Barbosa, A.; Whiting, S.; Simmonds, P.; Scotini Moreno, R.; Mendes, R.; Breda, J. Physical Activity and Academic Achievement: An Umbrella Review. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5972. [[CrossRef](#)] [[PubMed](#)]
44. O'Brien, K.M.; Barnes, C.; Yoong, S.; Campbell, E.; Wyse, R.; Delaney, T.; Brown, A.; Stacey, F.; Davies, L.; Lorien, S.; et al. School-Based Nutrition Interventions in Children Aged 6 to 18 Years: An Umbrella Review of Systematic Reviews. *Nutrients* **2021**, *13*, 4113. [[CrossRef](#)]
45. Johnson, B.T.; MacDonald, H.V.; Bruneau, M.L., Jr.; Goldsby, T.U.; Brown, J.C.; Huedo-Medina, T.B.; Pescatello, L.S. Methodological quality of meta-analyses on the blood pressure response to exercise: A review. *J. Hypertens.* **2014**, *32*, 706–723. [[CrossRef](#)]
46. Monasta, L.; Batty, G.D.; Cattaneo, A.; Lutje, V.; Ronfani, L.; Van Lenthe, F.J.; Brug, J. Early-life determinants of overweight and obesity: A review of systematic reviews. *Obes. Rev.* **2010**, *11*, 695–708. [[CrossRef](#)] [[PubMed](#)]
47. Grgic, J.; Grgic, I.; Pickering, C.; Schoenfeld, B.J.; Bishop, D.J.; Pedisic, Z. Wake up and smell the coffee: Caffeine supplementation and exercise performance—an umbrella review of 21 published meta-analyses. *Br. J. Sports Med.* **2020**, *54*, 681–688. [[CrossRef](#)]
48. Guyatt, G.; Oxman, A.D.; Akl, E.A.; Kunz, R.; Vist, G.; Brozek, J.; Norris, S.; Falck-Ytter, Y.; Glasziou, P.; deBeer, H.; et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J. Clin. Epidemiol.* **2011**, *64*, 383–394. [[CrossRef](#)]
49. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Routledge: London, UK, 2013.
50. Hazra, A. Using the confidence interval confidently. *J. Thorac. Dis.* **2017**, *9*, 4125–4130. [[CrossRef](#)]
51. Graham, S.; Hebert, M.; Harris, K.R. A Meta-Analysis Formative Assessment and Writing. *Elem. Sch. J.* **2015**, *115*, 523–547. [[CrossRef](#)]
52. Kingston, N.; Nash, B. Formative Assessment: A Meta-Analysis and a Call for Research. *Educ. Meas. Issues Pract.* **2011**, *30*, 28–37. [[CrossRef](#)]
53. Akbay, T.; Sevim-Cirak, N.; Erol, O. Re-Examining the Effect of Audience Response Systems on Learning Outcomes: Evidence from the Last Decade. *Int. J. Hum. Comput. Interact.* **2023**, *40*, 1–15. [[CrossRef](#)]
54. Guo, L. The effects of self-monitoring on strategy use and academic performance: A meta-analysis. *Int. J. Educ. Res.* **2022**, *112*, 101939. [[CrossRef](#)]
55. Jurado-Castro, J.M.; Vargas-Molina, S.; Gómez-Urquiza, J.L.; Benítez-Porres, J. Effectiveness of real-time classroom interactive competition on academic performance: A systematic review and meta-analysis. *PeerJ Comput. Sci.* **2023**, *9*, e1310. [[CrossRef](#)] [[PubMed](#)]
56. Karaman, P. The Effect of Formative Assessment Practices on Student Learning: A Meta- Analysis Study. *Int. J. Assess. Tools Educ.* **2021**, *8*, 801–817. [[CrossRef](#)]

57. Li, R. Investigating effects of computer-mediated feedback on L2 vocabulary learning. *Comput. Educ.* **2023**, *198*, 104763. [[CrossRef](#)]
58. Van der Kleij, F.M.; Feskens, R.C.W.; Eggen, T.J.H.M. Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Rev. Educ. Res.* **2015**, *85*, 475–511. [[CrossRef](#)]
59. Zheng, L. The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pac. Educ. Rev.* **2016**, *17*, 187–202. [[CrossRef](#)]
60. Lövdén, M.; Fratiglioni, L.; Glymour, M.M.; Lindenberger, U.; Tucker-Drob, E.M. Education and Cognitive Functioning Across the Life Span. *Psychol. Sci. Public Interest* **2020**, *21*, 6–41. [[CrossRef](#)]
61. Andrade, H.; Wang, X.; Du, Y.; Akawi, R. Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *J. Educ. Res.* **2009**, *102*, 287–302. [[CrossRef](#)]
62. Genlott, A.A.; Grönlund, Å. Closing the gaps—Improving literacy and mathematics by ict-enhanced collaboration. *Comput. Educ.* **2016**, *99*, 68–80. [[CrossRef](#)]
63. Quintana, D.S. From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. *Front. Psychol.* **2015**, *6*, 1549. [[CrossRef](#)]
64. Kolaski, K.; Logan, L.R.; Ioannidis, J.P.A. Guidance to best tools and practices for systematic reviews. *Syst. Rev.* **2023**, *12*, 96. [[CrossRef](#)] [[PubMed](#)]
65. Lamberg, T.; Gillette-Koyen, L.; Moss, D. Supporting Teachers to Use Formative Assessment for Adaptive Decision Making. *Math. Teach. Educ.* **2020**, *8*, 37–58. [[CrossRef](#)]
66. Schildkamp, K.; van der Kleij, F.M.; Heitink, M.C.; Kippers, W.B.; Veldkamp, B.P. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *Int. J. Educ. Res.* **2020**, *103*, 101602. [[CrossRef](#)]
67. Reid, R.; Trout, A.L.; Schartz, M. Self-regulation interventions for children with attention deficit/hyperactivity disorder. *Except. Child.* **2005**, *71*, 361–377.
68. Brookhart, S. Appropriate Criteria: Key to Effective Rubrics. *Front. Educ.* **2018**, *3*, 1–12. [[CrossRef](#)]
69. Reddy, Y.M.; Andrade, H. A review of rubric use in higher education. *Assess. Eval. High. Educ.* **2010**, *35*, 435–448. [[CrossRef](#)]
70. Dignath, C.; van Ewijk, R.; Perels, F.; Fabriz, S. Let Learners Monitor the Learning Content and Their Learning Behavior! A Meta-analysis on the Effectiveness of Tools to Foster Monitoring. *Educ. Psychol. Rev.* **2023**, *35*, 62. [[CrossRef](#)]
71. Graham, S. Changing How Writing Is Taught. *Rev. Res. Educ.* **2019**, *43*, 277–303. [[CrossRef](#)]
72. Saravanan, A.; Palanisamy, L.; Aziz, A.A. Systematic review: Challenges in teaching writing skills for upper secondary in ESL classrooms and suggestions to overcome them. *Malays. J. Soc. Sci. Humanit.* **2021**, *6*, 262–275.
73. Troia, G.A.; Olinghouse, N.G.; Mo, Y.; Hawkins, L.; Kopke, R.A.; Chen, A.; Wilson, J.; Stewart, K.A. Academic standards for writing: To what degree do standards signpost evidence-based instructional practices and interventions? *Elem. Sch. J.* **2015**, *116*, 291–321. [[CrossRef](#)]
74. Coker, D.L.; Farley-Ripple, E.; Jackson, A.F.; Wen, H.; MacArthur, C.A.; Jennings, A.S. Writing instruction in first grade: An observational study. *Read. Writ.* **2016**, *29*, 793–832. [[CrossRef](#)]
75. Graham, S.; Alves, R.A. Research and teaching writing. *Read. Writ.* **2021**, *34*, 1613–1621. [[CrossRef](#)]
76. Zainuddin, Z.; Chu, S.K.W.; Shujahat, M.; Perera, C.J. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educ. Res. Rev.* **2020**, *30*, 100326. [[CrossRef](#)]
77. Wang, A.I. The wear out effect of a game-based student response system. *Comput. Educ.* **2015**, *82*, 217–227. [[CrossRef](#)]
78. Radosevich, D.; Salomon, R.; Radosevich, D.; Kahn, P. Using Student Response Systems to Increase Motivation, Learning, and Knowledge Retention. *Innov. J. Online Educ.* **2008**, *5*, 7.
79. Licorish, S.A.; Owen, H.E.; Daniel, B.; George, J.L. Students' perception of Kahoot!'s influence on teaching and learning. *Res. Pract. Technol. Enhanc. Learn.* **2018**, *13*, 9. [[CrossRef](#)]
80. Cowan, N. The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Curr. Dir. Psychol. Sci.* **2010**, *19*, 51–57. [[CrossRef](#)]
81. Chamizo-Nieto, M.T.; Arrivillaga, C.; Rey, L.; Extremera, N. The Role of Emotional Intelligence, the Teacher-Student Relationship, and Flourishing on Academic Performance in Adolescents: A Moderated Mediation Study. *Front. Psychol.* **2021**, *12*, 695067. [[CrossRef](#)]
82. Smiderle, R.; Rigo, S.J.; Marques, L.B.; Peçanha de Miranda Coelho, J.A.; Jaques, P.A. The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learn. Environ.* **2020**, *7*, 3. [[CrossRef](#)]
83. Bicen, H.; Kocakoyun, Ş. Perceptions of Students for Gamification Approach: Kahoot as a Case Study. *Int. J. Emerg. Technol. Learn.* **2018**, *13*, 72. [[CrossRef](#)]
84. Tomasik, M.J.; Berger, S.; Moser, U. On the Development of a Computer-Based Tool for Formative Student Assessment: Epistemological, Methodological, and Practical Issues. *Front. Psychol.* **2018**, *9*, 2245. [[CrossRef](#)]
85. van der Steen, J.; van Schilt-Mol, T.; van der Vleuten, C.; Joosten-ten Brinke, D. Designing Formative Assessment That Improves Teaching and Learning: What Can Be Learned from the Design Stories of Experienced Teachers? *J. Form. Des. Learn.* **2023**, *7*, 182–194. [[CrossRef](#)]
86. Topping, K.J. Peer Assessment. *Theory Into Pract.* **2009**, *48*, 20–27. [[CrossRef](#)]
87. Ismail, S.M.; Rahul, D.R.; Patra, I.; Rezvani, E. Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Lang. Test. Asia* **2022**, *12*, 40. [[CrossRef](#)]

88. Choi, G.J.; Kang, H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. *J. Lipid Atheroscler.* **2023**, *12*, 3–11. [[CrossRef](#)]
89. Lazaros, B.; Vanesa, B.; John, P.A.I. Conducting umbrella reviews. *BMJ Med.* **2022**, *1*, e000071. [[CrossRef](#)]
90. Peugh, J.L. A practical guide to multilevel modeling. *J. Sch. Psychol.* **2010**, *48*, 85–112. [[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.