children

MDPI

Systematic Review

Cross-Informant Agreement and Discrepancies in ADHD Assessment During Adolescence: A Systematic Review and Meta-Analysis

Georgios Giannakopoulos [1,*] and Afroditi Prassou [2]

[1] Department of Child and Adolescent Psychiatry, School of Medicine, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, 115 27 Athens, Greece
[2] 4th Directorate of Secondary Education of Athens, 171 21 Athens, Greece; aprassou@sch.gr
* Correspondence: ggiannak@med.uoa.gr

Highlights

What are the main findings?

- Cross-informant agreement for adolescent attention-deficit/hyperactivity disorder symptoms was positive but modest and highly heterogeneous across caregiver, adolescent and teacher reports.
- Caregiver–adolescent pooled correlations averaged $r = 0.33$ for inattention and $r = 0.32$ for hyperactivity/impulsivity, with prediction intervals crossing zero; caregiver–teacher estimates were similar in magnitude, highly heterogeneous, and imprecise.

What are the implications of the main findings?

- Discrepant ratings should be treated as source- and context-specific clinical information rather than resolved by selecting one informant as universally more accurate.
- Assessment and research should document informant identity, setting, independence, timing and both dimensional and categorical agreement using age-stratified results.

Abstract

Background/Objectives: Assessment of attention-deficit/hyperactivity disorder (ADHD) requires information across settings, but reports from adolescents, caregivers, teachers, and clinicians are often only partly concordant. This review synthesized cross-informant agreement and discrepancies in ADHD assessment among adolescents aged 10–19 years. Methods: MEDLINE, CINAHL, ERIC and Academic Search Complete were searched through EBSCO, together with PubMed and Scopus, on 19 July 2026. Eligible primary quantitative studies compared at least two independent informants assessing the same adolescent ADHD domain. Methodological quality was appraised with an adapted 11-item QAREL framework. Pearson correlations were pooled after Fisher z transformation using restricted maximum-likelihood random-effects models and Hartung–Knapp–Sidik–Jonkman confidence intervals. Results: The searches identified 1447 records; 30 reports (282 outcomes) were included. Caregiver–adolescent pooled r was 0.33 for inattention ($k = 7$, 95% CI 0.19–0.46; $I^2 = 88.1\%$; 95% prediction interval −0.07 to 0.64) and 0.32 for hyperactivity/impulsivity ($k = 7$, 95% CI 0.12–0.50; $I^2 = 94.8\%$; prediction interval −0.28 to 0.74). Despite the positive pooled point estimates, both prediction intervals were wide and crossed zero. Caregiver–teacher estimates were $r = 0.29$ for inattention ($k = 4$, 95% CI −0.02 to 0.55; $I^2 = 94.1\%$) and $r = 0.27$ for hyperactivity/impulsivity ($k = 4$, 95% CI −0.03 to 0.52; $I^2 = 83.0\%$); both confidence intervals crossed zero. Diagnostic agreement was generally low and sensitive to impairment and consensus rules. Conclusions: Informant

Check for updates

Academic Editor: Heather Gordish-Dressman

Received: 5 August 2026
Revised: 22 August 2026
Accepted: 30 August 2026
Published: 1 September 2026

Copyright: © 2026 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.

reports were positively related on average, but the magnitude of agreement was highly heterogeneous and imprecisely estimated. Wide intervals, including caregiver–adolescent prediction intervals crossing zero and caregiver–teacher confidence intervals crossing zero, indicate that little or no agreement remains plausible in some comparable settings. Reports should therefore not be treated as clinically interchangeable, and pooled estimates should not be interpreted as universal levels of agreement.

Keywords: ADHD; adolescence; informant agreement; informant discrepancies; parent report; teacher report; self-report; meta-analysis

1. Introduction

Assessment of attention-deficit/hyperactivity disorder (ADHD) is intrinsically multicontextual. Symptoms must be evaluated alongside impairment and their expression across settings, which makes reports from adolescents, caregivers, teachers and clinicians central to diagnostic formulation and treatment planning [1–4]. These sources do not observe the same behavior under the same demands. Caregivers mainly observe routines and relationships at home, teachers observe performance under structured academic and peer demands, and adolescents have direct access to effort, distraction and internal experiences that may not be visible to adults [1,4–6].

Cross-informant differences are common throughout child and adolescent mental health assessment. Classic and more recent work indicates that agreement is generally lower when informants observe young people in different settings and that discrepancies may reflect context-specific behavior, different thresholds for concern, distinct relationships with the young person, and characteristics of the assessment method [7,8]. Disagreement should therefore not automatically be interpreted as evidence that one informant is inaccurate. The value of a multi-informant assessment lies not in expecting different sources to reproduce one another, but in using their convergence and divergence to clarify where, when, and to whom difficulties are apparent [8,9]. Privileging a single report may overlook clinically important difficulties, whereas combining reports without examining their differences may obscure meaningful variation across settings [8–10].

Adolescence makes this issue particularly consequential. Parental observation becomes less complete, school-based information may be distributed across several teachers, and adolescents assume a more prominent role in describing their own functioning [1,5,6]. Adolescents may be better positioned to report the effort required to sustain attention or regulate behavior, while caregivers and teachers may be better placed to evaluate observable behavior, consistency over time, and interference with everyday functioning [11–15]. Informants may also apply different thresholds when deciding whether a behavior is problematic, depending on their expectations, opportunities for observation, and relationship with the adolescent [16–18]. Cross-informant discrepancies during adolescence may therefore contain information about the contextual distribution and functional significance of ADHD symptoms rather than merely representing measurement noise [8,13,14,19–21].

Agreement is also not a unitary construct. Correlations between rating scales indicate whether two informants rank adolescents in a similar order, but they do not establish agreement in absolute symptom levels or diagnostic classification [10,22,23]. Conversely, paired mean differences can identify a systematic tendency for one informant to report more symptoms than another without showing whether the two reports correspond at the individual level. Intraclass correlations, kappa coefficients, agreement proportions, and latent-variable models address related but distinct aspects of concordance [22,24–27]. These

outcomes should not be treated as interchangeable, particularly because clinical decisions depend on symptom thresholds, impairment, and cross-setting evidence rather than solely on relative rank ordering [8,28,29].

Cross-informant disagreement, clinical non-interchangeability, and psychometric non-equivalence are distinct concepts. In this review, cross-informant disagreement refers to observed differences in symptom scores, classifications, or other agreement outcomes across reporters. Clinical non-interchangeability means that reports from different informants should not be assumed to provide the same information or to be substitutable in clinical formulation. Psychometric non-equivalence refers more narrowly to evidence that items or latent constructs function differently across informants, usually evaluated through measurement-invariance or related analyses [30]. Modest correlations or discrepant scores alone do not demonstrate psychometric non-equivalence.

Adjacent reviews have examined the validity of teacher rating scales against diagnostic interviews or classroom observations [31] and the measurement invariance of ADHD symptom criteria across parent and teacher reports [30]. These questions concern criterion validity and psychometric equivalence, respectively, and are related to but distinct from direct cross-informant agreement. Neither review directly synthesized cross-informant agreement across the range of informant pairs relevant to adolescent assessment. The pertinent primary literature includes parent–adolescent, parent–teacher, teacher–adolescent, teacher–teacher, and clinician or consensus comparisons and employs markedly different definitions and indices of agreement [8,9,30,31]. An adolescence-specific synthesis is therefore needed to determine whether observed discrepancies represent a broadly consistent pattern or are mainly attributable to particular informant pairs, instruments, outcome definitions, or diagnostic rules.

This review aimed to synthesize the magnitude, direction, and methodological determinants of cross-informant agreement and discrepancies in ADHD assessment among adolescents aged 10–19 years. The primary questions concerned agreement between caregivers and adolescents, and between caregivers and teachers, for the core domains of inattention and hyperactivity/impulsivity. Secondary questions concerned diagnostic and threshold agreement, teacher–adolescent and teacher–teacher agreement, comparisons with clinicians or consensus diagnoses, systematic mean differences, assessment format effects, and proposed moderators of disagreement.

2. Materials and Methods

2.1. Protocol and Reporting Framework

This systematic review and exploratory meta-analysis was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [32]. The completed PRISMA 2020 checklist is provided in Table S1. The review was registered in the Open Science Framework (OSF) Registries on 22 July 2026 (https://doi.org/10.17605/OSF.IO/49H5R (accessed on 20 August 2026)). Database searching and initial study retrieval had already been completed before registration. The review question, age range, broad eligibility framework, information sources, search strings, and search date had therefore already been established and are not presented as prospectively registered decisions.

Registration occurred before formal data extraction, methodological quality appraisal, effect grouping, and quantitative synthesis. The registered synthesis framework specified the separate treatment of conceptually different agreement and discrepancy metrics, domain-specific analyses, a minimum of three independent studies for quantitative pooling, a random-effects synthesis, and planned robustness checks. Thus, the minimum threshold of three independent studies for pooling was specified in the registered plan, whereas the

designation of the four specific domain-by-informant Pearson correlation models as the primary quantitative syntheses was made after registration.

After registration, but before interpretation of the pooled results, several operational decisions were documented in the review workbook. These included preference for DSM or core-domain scales over broader content scales, selection of parallel rating-scale effects where multiple formats were available, implementation of the dependence rule for non-overlapping age strata, specification of restricted maximum likelihood estimation and Hartung–Knapp–Sidik–Jonkman confidence intervals, and designation of four domain-specific Pearson-correlation models as the primary quantitative syntheses.

2.2. Eligibility Criteria

Eligibility was structured around the population, phenomenon of interest, comparison and outcome. Studies were eligible when they met all of the following criteria:

- Participants were aged 10–19 years, or the report presented a result that was separable for this age range.
- At least two independent informants assessed the same adolescent, including adolescents, parents or caregivers, teachers, clinicians, consensus teams, or two raters of the same eligible type. For synthesis-level reporting, "caregiver" is used throughout as an umbrella term encompassing parents, mothers, fathers, primary caregivers, caretakers, and legal guardians. Source-specific informant labels are retained when individual studies are described.
- The compared construct was an ADHD diagnosis, ADHD symptom total, inattention, hyperactivity/impulsivity, or a clearly ADHD-specific scale or classification.
- The report presented a quantitative cross-informant agreement or discrepancy analysis, such as Pearson correlation, ICC, kappa, agreement proportion, diagnostic concordance, paired classification, paired mean difference, or a compatible model-based estimate.
- The study was primary quantitative research. Baseline and secondary analyses of trials or cohorts were eligible when the relevant agreement data were available.

Studies were excluded when the eligible adolescent component of a mixed-age sample could not be separated; only one informant was used; the comparison concerned two instruments completed by the same person, chart coders, administrative records or treatment-response ratings; the outcome was not ADHD-specific; no direct agreement or discrepancy analysis was reported; or the record was a review, protocol, commentary, case report or other non-primary publication. No publication-year or language restriction was applied.

2.3. Information Sources and Search Strategy

Searches were completed on 19 July 2026 in MEDLINE, CINAHL with Full Text, ERIC and Academic Search Complete through the EBSCO interface; PubMed through the PubMed interface; and Scopus through the Scopus interface. APA PsycInfo was not available through the accessible institutional interface and could therefore not be searched directly. To broaden coverage beyond biomedical indexing, the search included Scopus and Academic Search Complete in addition to MEDLINE, PubMed, CINAHL and ERIC, using broad terms for informants, agreement, discrepancy and reliability. No separate formal backward- or forward-citation-chasing stage was conducted. The strategy combined free-text terms for ADHD, adolescence, informants or reporters, and agreement, concordance, discrepancy or inter-rater reliability; PubMed Medical Subject Headings (MeSH) terms were added where applicable. Database syntax and field tags were adapted to each interface; the complete database-specific search strategies and yields are presented in Table S2.

The database yields were MEDLINE 243, CINAHL 75, ERIC 29, Academic Search Complete 165, PubMed 531 and Scopus 404, for a gross total of 1447 records. Bibliographic records were exported in RIS or NBIB format. Full-text retrieval used publisher platforms, DOI links, institutional access, PubMed Central and university or author repositories; authors were approached when an eligible report could not otherwise be obtained.

2.4. Record Management and Study Selection

Records from all sources were merged into a master corpus. Deduplication used persistent identifiers where available and normalized title matching, followed by manual review of potential collisions. The resulting 796 unique records were initially screened at title and abstract level by one reviewer. Uncertain records were retained for full-text assessment rather than excluded.

During peer-review revision, the second reviewer first completed a calibration exercise on 50 title-and-abstract records to ensure consistent application of the eligibility criteria. She then independently rescreened all 796 unique records without access to the original screening decisions. The second reviewer also independently reassessed all 64 retrieved full-text reports without access to the original final decisions or exclusion reasons. Disagreements were resolved through discussion using the registered eligibility criteria and exclusion hierarchy.

Four records that had initially been excluded at title-and-abstract level were advanced and subsequently met the eligibility criteria. The final workflow therefore comprised 68 reports sought for retrieval, 64 reports assessed in full text, four reports not retrieved, 34 full-text exclusions, and 30 included reports.

2.5. Data Extraction

A standardized extraction form was piloted on five reports and then applied to the remaining included reports. Extraction captured bibliographic information; country, setting and design; recruitment and sample characteristics, including the eligible age range or stratum and mean age where reported; informant pairs; instruments, administration procedures and scoring; the exact outcome definition; metric, estimate, uncertainty and analysis sample; time point; effect direction; and source location in the report. Results were retained at the most granular level needed to distinguish symptom domain, informant pair, age stratum, assessment method and case definition. One reviewer performed the initial extraction and a same-reviewer second pass against the full reports. Four additional eligible reports identified during revision were extracted using the same coding rules. During revision, the second reviewer independently verified all 22 study-domain effect entries contributing to the four final primary meta-analyses against the source reports. These entries originated from nine unique reports. Verification covered the analysis-specific sample size, informant pair, ADHD domain, reported Pearson correlation, source page or table, primary-effect selection, and any within-study combination or transformation. Disagreements were resolved by returning to the source report and applying the prespecified extraction and effect-selection rules. The remaining result-level rows used in the structured narrative synthesis were not independently extracted in duplicate. Missing values were left as missing and were not imputed. The final dataset comprised 30 included reports and 282 result-level rows.

2.6. Methodological Quality Appraisal

Methodological quality was assessed at report level using an adapted 11-item Quality Appraisal of Reliability Studies (QAREL) framework [33]. QAREL was selected because it was developed specifically for studies of diagnostic reliability and agreement and therefore addresses methodological features that are directly relevant to cross-informant assessment,

including the suitability of the participant and rater spectra, independence of ratings, potential information or order effects, temporal proximity of assessments, and the appropriate application and interpretation of agreement statistics.

The framework was adapted operationally to the cross-informant ADHD literature rather than used as a generic risk-of-bias scale. Q1 and Q2 addressed the suitability and representativeness of the adolescent sample and the informants being compared. Q3–Q8 were interpreted in relation to independence of informant completion, exposure to another informant's ratings or other clinical cues, and possible order effects. Q9 addressed whether the compared assessments were sufficiently close in time for meaningful agreement to be expected. Q10 evaluated whether the reported agreement statistic was correctly applied and interpreted, including whether a correlation was described more broadly than its rank-order meaning justified. Q11 evaluated the appropriateness of the statistical analysis for the stated agreement question. Items that were structurally irrelevant to a particular assessment design were coded Not applicable rather than treated as methodological failures.

Each item was judged Yes, No, Unclear, or Not applicable. Overall judgments of Low concern, Some concerns, or High concern were qualitative and reflected the likely methodological impact of the identified issues rather than a summed item score. Reports were judged High concern when one or more limitations could materially affect the interpretation of the agreement estimate, whereas Some concerns generally reflected incomplete reporting or limitations of less certain impact. No numerical QAREL score or cutoff was calculated.

The initial appraisal was completed by one reviewer. During peer-review revision, the second reviewer independently repeated the item-level QAREL appraisal for all 30 included reports using the same operational interpretation guide and without access to the original overall judgments. Disagreements were resolved through discussion and re-examination of the source reports before the final report-level judgments were assigned.

2.7. Outcome Grouping and Effect Selection

After data extraction and effect grouping, but before interpretation of the pooled results, the quantitative synthesis was operationalized as four primary Pearson-correlation models: caregiver–adolescent agreement for inattention and hyperactivity/impulsivity, and caregiver–teacher agreement for the same two domains. The registered synthesis framework had specified domain-specific analyses, separate treatment of conceptually different agreement metrics, and a minimum of three independent studies for quantitative pooling; designation of these four models as the primary quantitative syntheses was a post-registration operational decision. One domain-aligned effect was selected per independent study. DSM or core symptom-domain scales were preferred to broader content scales. When a study reported multiple eligible assessment formats, the parallel rating-scale result was primary and the alternative format was retained for sensitivity analysis. Non-overlapping age strata from the same report were combined on the Fisher-z scale before entering the model.

Kappa coefficients, ICCs, agreement proportions, mean discrepancies and latent-variable estimates were not converted to Pearson correlations because they quantify different properties and often lacked the information required for defensible variance estimation. These outcomes were synthesized in structured narrative form, organized by informant pair, domain, assessment method and diagnostic rule.

2.8. Statistical Analysis

Pearson correlations were transformed to Fisher z values, with sampling variance calculated as $1/(n-3)$. Inverse-variance random-effects models were fitted using restricted

maximum likelihood to estimate between-study variance. Because each model contained few studies, uncertainty around pooled effects was calculated with Hartung–Knapp–Sidik–Jonkman t intervals using $k-1$ degrees of freedom. Summary estimates and confidence limits were back-transformed to Pearson r. Statistical heterogeneity was described using Cochran's Q, between-study variance (τ^2) on the Fisher-z scale, and the conventional Q-based I^2 statistic, calculated as max $[0, (Q - df)/Q] \times 100$. Ninety-five percent prediction intervals were calculated when at least five studies were available.

Formal subgroup analyses and meta-regression were not undertaken because the four primary models contained only four or seven studies and no candidate moderator was reported in a sufficiently consistent form across contributing studies. Age was generally available as broad, overlapping ranges, whereas mean age, sex distribution, parental ADHD, medication status, family characteristics, school context, and recruitment setting were not available in a form that could be harmonized across the studies in each meta-analysis. Potential sources of heterogeneity were therefore examined descriptively through study characteristic comparisons and through QAREL-restricted, alternative-measure, and leave-one-out sensitivity analyses.

Robustness checks comprised fixed-effect models; exclusion of reports judged High concern on QAREL when at least three studies remained; substitution of alternative eligible instruments or formats; and leave-one-out analyses. Funnel plots and regression-based tests of asymmetry were not performed because every meta-analysis included fewer than 10 studies. Meta-analyses were conducted in IBM SPSS Statistics, version 31.0 (IBM Corp., Armonk, NY, USA), using the META ES CONTINUOUS procedure with a random-effects model, restricted maximum-likelihood estimation, and the untruncated Knapp–Hartung standard-error adjustment (ADJUSTSE = KNAPP_HARTUNG). In SPSS this option is distinct from TRUNCATED_KNAPP_HARTUNG and therefore does not impose a lower bound of 1 on the variance adjustment. To verify the implementation, the four primary models were independently recalculated on the Fisher-z scale using REML weights, the standard untruncated Hartung–Knapp variance estimator, and t critical values with $k-1$ degrees of freedom. The independently calculated pooled estimates and confidence limits reproduced the SPSS results to the reported precision. The analysis syntax, input data and output files were retained for reproducibility.

3. Results

3.1. Study Selection

The searches yielded 1447 records. After removal of 651 duplicates, 796 unique records were screened. Following independent rescreening during revision, 68 reports were sought for full-text assessment; 64 were retrieved and four were not retrieved. Thirty reports met the review criteria and 34 were excluded after full-text assessment. The most common reason for full-text exclusion was the absence of a separable adolescent result in a mixed-age sample (n = 29), followed by wrong age (n = 2), wrong outcome (n = 1), no eligible ADHD-specific agreement outcome (n = 1), and no informant agreement analysis (n = 1). The final evidence set comprised 30 reports, all of which were extracted (Figure 1).

3.2. Study Characteristics

Included reports were published from 1996 to 2025 and represented clinical, community, school, epidemiologic and follow-up samples across 13 countries or national settings. Report-level sample sizes ranged from 35 to 10,476 participants, with several reports contributing multiple cohorts, age strata or repeated assessments. Informant comparisons included adolescent–caregiver, caregiver–teacher, adolescent–teacher, teacher–teacher, and comparisons with clinicians or consensus diagnoses. Instruments included

DSM-based symptom scales, Conners forms, Vanderbilt and SNAP-IV ratings, structured or semi-structured diagnostic interviews, broadband behavior measures, and multi-trait multi-method or latent-class models. The included reports are summarized in Table S3. No language restriction was applied. Of the 30 included reports, 29 were published in English and one was published in Spanish. No report was excluded solely on the basis of publication language.

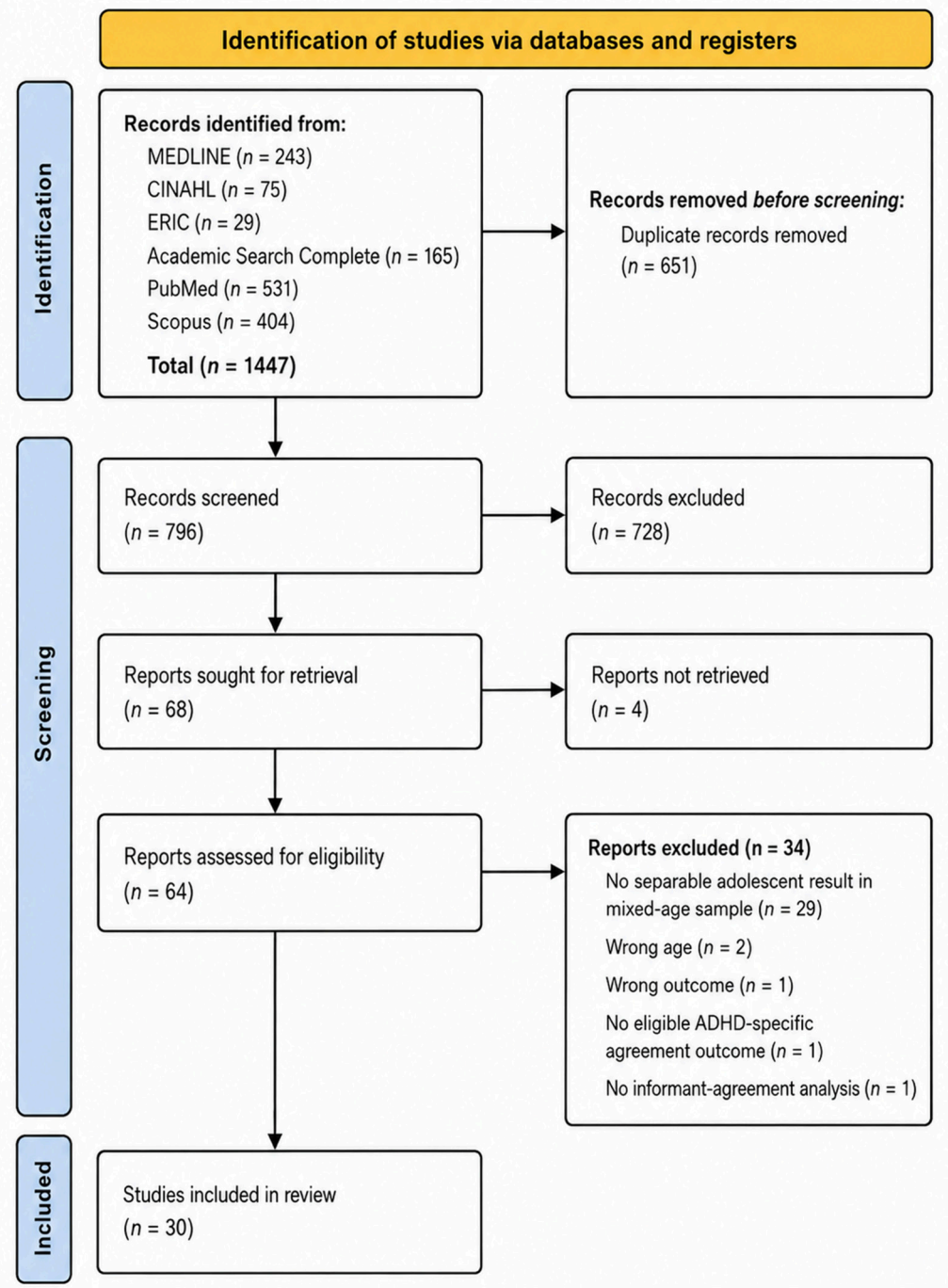


Figure 1. PRISMA 2020 flow diagram for study selection.

Age coverage was heterogeneous and is reported in detail in Table S3. The eligible evidence was concentrated between ages 10 and 18; no included report provided an age-19-specific agreement estimate. Among the seven caregiver–adolescent studies contributing to the primary correlation models [34–40], age ranges overlapped substantially, spanning

11–18 years overall, and mean ages ranged from 12.78 to 16.24 years. Among the four studies contributing to each caregiver–teacher model [3,37,39,41], mean age was reported for three studies (12.78, 15.00, and 16.24 years), while Montiel-Nava and Peña [41] reported separate 12–14 and 15–17-year strata. These data make the developmental coverage of the evidence visible but do not support a reliable age-gradient analysis.

The 282 extracted outcomes comprised Pearson correlations, ICCs, kappa coefficients, exact or positive-case agreement proportions, paired classifications, mean discrepancies, regression-based discrepancy effects, latent-variable estimates, and model-based class proportions. This variation supported quantitative pooling only for the four Pearson-correlation groups.

3.3. Methodological Quality

All 30 included reports were assessed with QAREL: three were judged "Low concern", 15 "Some concerns" and 12 "High concern". Recurrent concerns were insufficient reporting of masking or independent completion, fixed interview order or the same interviewer having access to both reports, nonparallel instruments, age-stratum uncertainty, and use of correlation alone to describe "agreement." Several studies had otherwise appropriate same-period measurements and suitable participant and rater spectra, but failed to describe procedures that would prevent influence between informants. Report-level methodological quality judgments are presented in Table S4. Because 27 of the 30 included reports had at least some methodological concern, the primary estimates were interpreted alongside the QAREL-restricted sensitivity analyses rather than as equally robust across all contributing studies. The influence of methodological quality on the quantitative findings is reported in Section 3.7.

3.4. Caregiver–Adolescent Agreement

Seven studies contributed to each caregiver–adolescent meta-analysis [34–40]. For inattention, the random-effects pooled correlation was $r = 0.33$ (95% CI 0.19–0.46, $p = 0.001$). Heterogeneity remained high, $Q(6) = 50.33$, $p < 0.001$, $I^2 = 88.1\%$, $\tau^2 = 0.022$, and the 95% prediction interval extended from −0.07 to 0.64. For hyperactivity/impulsivity, the pooled correlation was $r = 0.32$ (95% CI 0.12–0.50, $p = 0.009$), with $Q(6) = 115.25$, $p < 0.001$, $I^2 = 94.8\%$, $\tau^2 = 0.050$ and a prediction interval from −0.28 to 0.74. Thus, the pooled estimates represent average associations across highly heterogeneous studies rather than stable levels of agreement. Both prediction intervals crossed zero and extended to substantially positive correlations, indicating that a future comparable study could plausibly observe little, no, or much stronger correspondence. Summary estimates are presented in Table 1 and study-level estimates in Figures 2 and 3. Study-level inputs for all primary meta-analyses are provided in Table S5.

Table 1. Primary random-effects meta-analyses of cross-informant Pearson correlations.

Informant Pair	Domain	*k*	Pooled *r*	95% CI	*p*	I^2	τ^2	95% Prediction Interval
Caregiver–adolescent	Inattention	7	0.33	0.19–0.46	0.001	88.1%	0.022	−0.07–0.64
Caregiver–adolescent	Hyperactivity/impulsivity	7	0.32	0.12–0.50	0.009	94.8%	0.050	−0.28–0.74
Caregiver–teacher	Inattention	4	0.29	−0.02–0.55	0.060	94.1%	0.036	Not calculated ($k < 5$)
Caregiver–teacher	Hyperactivity/impulsivity	4	0.27	−0.03–0.52	0.064	83.0%	0.030	Not calculated ($k < 5$)

Note: Correlations were pooled on the Fisher *z* scale using restricted maximum likelihood and back-transformed to Pearson *r*. Pooled confidence intervals use the Hartung–Knapp–Sidik–Jonkman method. Prediction intervals were calculated only when $k \geq 5$. Because heterogeneity was high, the pooled correlations should be interpreted as average effects across studies rather than expected values for a specific population or setting.

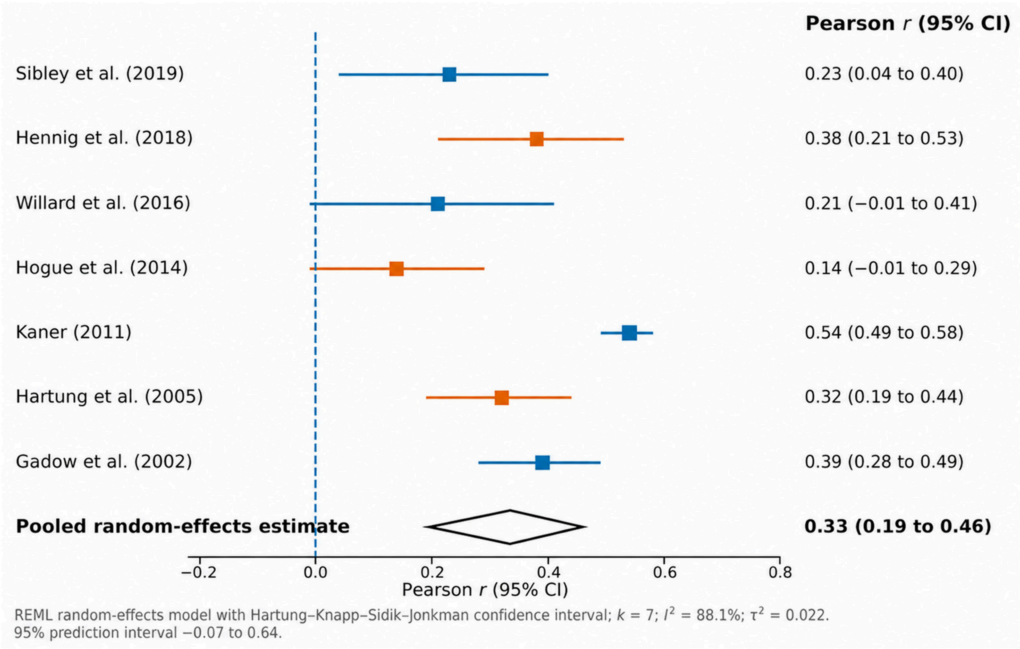


Figure 2. Forest plot of caregiver–adolescent agreement for inattention [34–40]. Note: Individual confidence intervals were calculated by normal approximation on the Fisher *z* scale; the pooled confidence interval used the Hartung–Knapp–Sidik–Jonkman method. Effect sizes and confidence intervals were back-transformed to Pearson *r* for presentation.

Diagnostic and threshold outcomes did not support a common meta-analysis. Kappa estimates for caregiver–adolescent ADHD classifications ranged from approximately zero in a clinically referred inner-city sample to 0.51 in a community epidemiologic cohort, and they changed when impairment or subtype rules changed [36,42–45]. For example, Jensen et al. [44] reported κ = 0.16 without diagnosis-specific impairment and κ = 0.08 when impairment was required, whereas Rohde et al. [45] reported κ = 0.45 and Cantwell et al. [42] reported κ = 0.51. Hope et al. [46] likewise found limited mother–adolescent diagnostic concordance (39.4%; κ = 0.30), with mothers reporting more ADHD symptoms than adolescents (M = 9.42 vs. 8.34, $p < 0.05$). Using independent mother and adolescent K-SADS-E interviews, Gau et al. [47] found small-to-moderate agreement at adolescent follow-up (ICC = 0.32–0.56), compared with higher agreement at the corresponding childhood baseline assessment (ICC = 0.44–0.71). These values should not be treated as direct equivalents because prevalence, reference rules, interview procedures and case definitions differed.

3.5. Caregiver–Teacher Agreement

Four studies contributed to each caregiver–teacher model [3,37,39,41]. The pooled point estimate for inattention was r = 0.29, but its HKSJ interval included zero (95% CI −0.02–0.55, p = 0.060). Heterogeneity was high, $Q(3)$ = 51.16, $p < 0.001$, I^2 = 94.1%, τ^2 = 0.036. The hyperactivity/impulsivity estimate was r = 0.27 (95% CI −0.03–0.52, p = 0.064), with $Q(3)$ = 17.65, $p < 0.001$, I^2 = 83.0% and τ^2 = 0.030. Prediction intervals were not calculated because only four studies contributed to each model. With only four studies per model and I^2 values of 94.1% and 83.0%, these estimates should be interpreted as highly uncertain average effects rather than stable expected levels of caregiver–teacher agreement. This caution is reinforced by the methodological quality analysis: after exclusion of High-concern reports, only two studies remained in each model, which was insufficient for a meaningful quality-restricted meta-analysis. The caregiver–teacher pooled estimates should therefore

be regarded as more tentative than the caregiver–adolescent pooled estimates. Study-level estimates are shown in Figures 4 and 5.

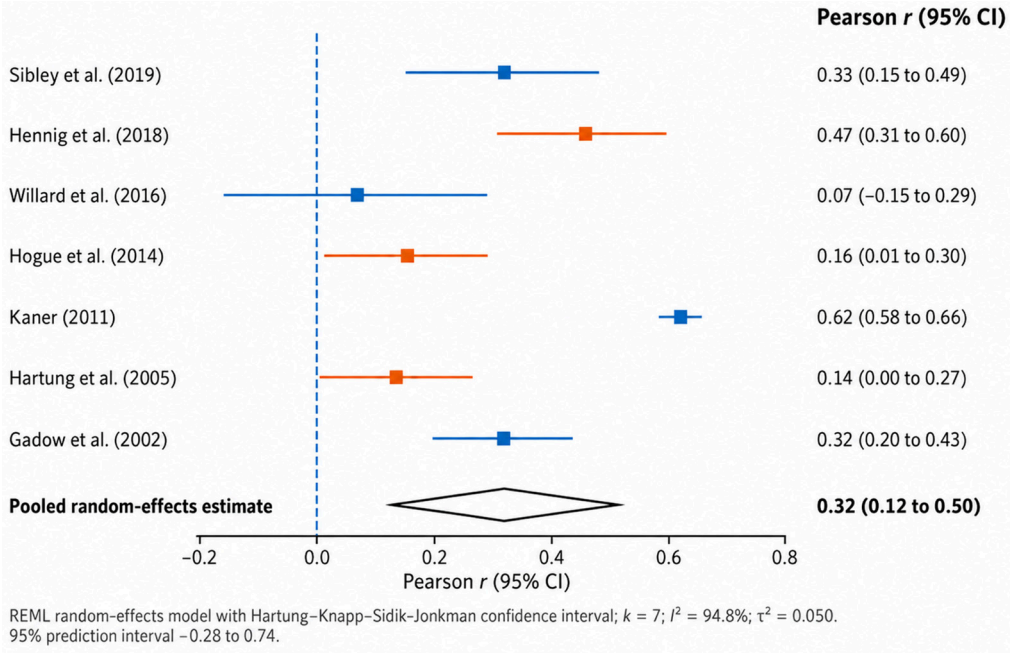


Figure 3. Forest plot of caregiver–adolescent agreement for hyperactivity/impulsivity [34–40]. Note: Individual confidence intervals were calculated by normal approximation on the Fisher *z* scale; the pooled confidence interval used the Hartung–Knapp–Sidik–Jonkman method. Effect sizes and confidence intervals were back-transformed to Pearson *r* for presentation.

Additional parent–teacher evidence used metrics that were not compatible with the Pearson-correlation models. Juneja et al. [48] reported paired parent and teacher Conners classifications in 500 adolescents aged 10–15 years; the reported margins and overlap yielded exact agreement of 0.94 and a derived Cohen's κ of 0.66. Wan Salwina et al. [49] reported weak parent–teacher correspondence in 12-year-old adolescents, including coefficients of 0.107 for inattention, 0.241 for hyperactivity, and 0.093 for combined symptoms. These coefficients were retained narratively because the report described the statistic inconsistently as an intraclass correlation and a Spearman correlation. Frick et al. [50] used parent and teacher DSM symptom ratings in a latent-class model at age 10 and identified both pervasive and setting-specific ADHD patterns, including Pervasive Combined (10.5%), School Combined (9.2%), School Inattentive (15.3%), and Home Combined (18.2%) classes.

Other parent–teacher reports used kappa, ICC, latent correlations or systematic discrepancy scores. The direction and magnitude of agreement varied by developmental stratum, instrument, reporter and definition. Correlation-based results sometimes indicated low-to-moderate rank-order consistency while categorical agreement remained low, reinforcing the need to keep dimensional and threshold results separate [51,52].

3.6. Teacher–Adolescent, Teacher–Teacher and Clinician Comparisons

Teacher–adolescent evidence was too heterogeneous for pooling. Gadow et al. [40] reported that all youth–teacher ADHD symptom correlations were below 0.18, whereas Wan Salwina et al. [49] reported weak or near-zero adolescent–teacher coefficients across hyperactive, inattentive, and combined symptom domains. Studies used Pearson correlations, ICCs, model-based shared-trait variance, signed-rank tests and broader attention-problem measures, often with different age restrictions. In an adolescent twin subgroup, Martin et al. [53] found higher self-reported hyperactivity scores than both parent ratings

($Z = -3.102$, $p = 0.002$) and teacher ratings ($Z = -7.244$, $p < 0.001$). The overall pattern suggested lower and less consistent convergence across school and self-report contexts than within more similar settings, but the available estimates could not be reduced to a single valid coefficient [35,54–56].

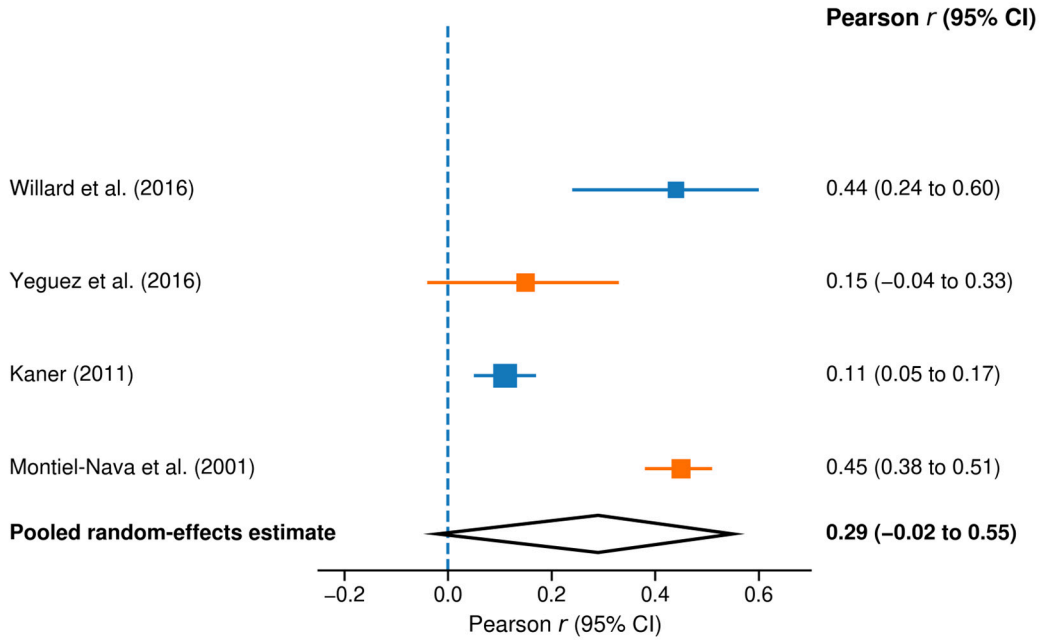


Figure 4. Forest plot of caregiver–teacher agreement for inattention [3,37,39,41]. Note: Individual confidence intervals were calculated by normal approximation on the Fisher *z* scale; the pooled confidence interval used the Hartung–Knapp–Sidik–Jonkman method. Effect sizes and confidence intervals were back-transformed to Pearson *r* for presentation.

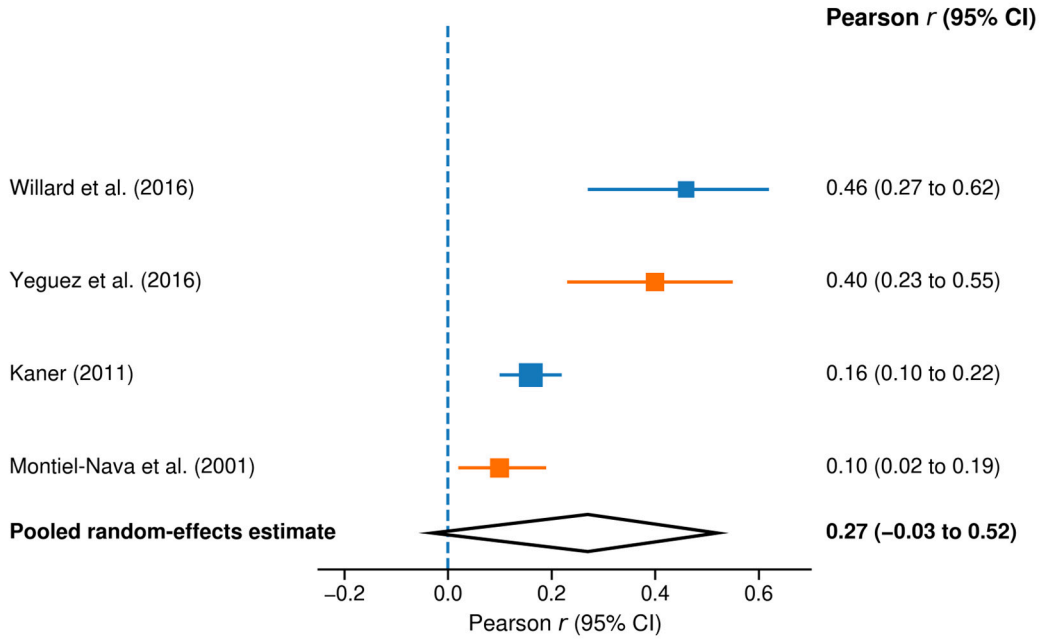


Figure 5. Forest plot of caregiver–teacher agreement for hyperactivity/impulsivity [3,37,39,41]. Note: Individual confidence intervals were calculated by normal approximation on the Fisher *z* scale; the pooled confidence interval used the Hartung–Knapp–Sidik–Jonkman method. Effect sizes and confidence intervals were back-transformed to Pearson *r* for presentation.

Teacher–teacher agreement varied according to the teacher pair, instrument and whether consistency or absolute agreement was estimated. Morris et al. [57] reported average-measures absolute-agreement ICCs of 0.57–0.71 across Math and English teachers in two adolescent cohorts. Molina et al. [58] reported dimensional ICCs around 0.35–0.51 and categorical kappas of 0.17–0.40 across instruments. Danforth and DuPaul [59] reported stronger correlations in an older subgroup, but the exact subgroup size and age cutoff were unavailable.

Comparisons with clinicians or consensus diagnoses also required cautious interpretation. A clinician rating was not an independent reference standard when it incorporated the same adolescent or parent report under evaluation. Joint interviews, fixed informant order and consensus procedures could either increase apparent agreement or redistribute discrepancies [42,45,60].

3.7. Sensitivity Analyses

For caregiver–adolescent inattention, exclusion of High-concern reports reduced the pooled estimate to r = 0.24 (k = 4, 95% CI 0.06–0.40) and I^2 to 33.1%. For hyperactivity/impulsivity, the corresponding estimate was r = 0.27 (k = 4, 95% CI −0.03–0.52), with residual heterogeneity of I^2 = 74.8%. Alternative-measure analyses remained positive but heterogeneous: r = 0.37 (95% CI 0.20–0.53; I^2 = 90.7%) for inattention and r = 0.30 (95% CI 0.16–0.43; I^2 = 84.0%) for hyperactivity/impulsivity. In leave-one-out analyses, omission of Kaner [37] reduced I^2 from 88.1% to 46.9% for inattention and from 94.8% to 68.7% for hyperactivity/impulsivity. Omission of the other reports generally left heterogeneity high or very high. Because Kaner [37] differed simultaneously in sample size, setting, instrument, and methodological quality, these influence results do not identify a single causal moderator.

The caregiver–teacher quality-restricted models could not be estimated because only two studies remained. Alternative-measure estimates were r = 0.32 (95% CI 0.08–0.53; I^2 = 87.4%) for inattention and r = 0.24 (95% CI −0.10–0.54; I^2 = 87.3%) for hyperactivity/impulsivity. Across leave-one-out analyses, all intervals crossed zero and I^2 remained between 78.5% and 96.0%, indicating that no single report accounted for the heterogeneity. Fixed-effect estimates were more precise but were retained only as sensitivity analyses because they do not represent between-study variation. The random-effects HKSJ results remained primary. Complete results are reported in Table S6.

3.8. Integrated Synthesis

Across all evidence streams, informant reports were positively related on average but were not clinically interchangeable. Agreement depended on the pair of observers, setting, assessment format, symptom domain and case definition. Dimensional correlations did not guarantee diagnostic agreement, and diagnostic agreement could change after adding impairment or consensus rules. Formal subgroup analyses and meta-regression were not performed. Each primary model contained only four or seven studies, and no candidate moderator was available in a sufficiently consistent form across the contributing reports. Age ranges were broad and overlapping. Table S3 reports the eligible age range or stratum and mean age where available. Across the seven caregiver–adolescent primary studies, mean ages ranged from 12.78 to 16.24 years, but the contributing age ranges overlapped substantially (11–18 years overall). Among the four caregiver–teacher studies, mean age was available for three and ranged from 12.78 to 16.24 years; the remaining report provided separate 12–14 and 15–17-year strata. These data permit descriptive inspection of developmental coverage but do not support a defensible test of an age gradient. Sex distribution was not available for all studies in a form that could be harmonized, and

parental ADHD, family characteristics, medication status, school context, and recruitment setting were either unavailable or defined differently. Study-specific moderator analyses also used different outcomes and interaction specifications. Any post hoc subgrouping would therefore have produced very small, non-comparable groups and unstable estimates. The available moderator findings were retained narratively. Rothen et al. [61] found no clear evidence that diagnostic agreement differed between adolescents aged 13–17 years and children aged 7–12 years (OR = 1.46, 95% CI 0.76–2.83). In the adolescent interview subgroup examined by Faraone et al. [62], mothers consistently reported more ADHD symptoms than youths, but the magnitude of this discrepancy did not differ according to maternal or paternal ADHD status (interaction $F = 0.50$, $p = 0.61$). The most defensible overall conclusion is therefore not that one informant is generally more accurate, but that each source contributes a partly overlapping and partly context-bound account. The integrated evidence streams are summarized in Table 2.

Table 2. Structured narrative synthesis of evidence streams not represented fully by the primary meta-analyses.

Evidence Stream	Outcome Scope	Synthesis Finding	Confidence
Caregiver–adolescent dimensional	Core inattention and hyperactivity/impulsivity	Average pooled $r = 0.32$–0.33; very high heterogeneity ($I^2 = 88.1$–94.8%); prediction intervals ranged from −0.28 to 0.74 across the two core-domain models and crossed zero.	Very cautious
Caregiver–teacher dimensional	Core inattention and hyperactivity/impulsivity	Point estimates $r = 0.27$–0.29; HKSJ confidence intervals included zero; high heterogeneity ($I^2 = 83.0$–94.1%); four studies per model; quality-restricted models could not be estimated because only two studies remained after exclusion of High-concern reports.	Very cautious
Caregiver–adolescent diagnostic	Attention-deficit/hyperactivity disorder diagnoses, subtype and impairment thresholds	Generally limited chance-corrected agreement; estimates changed with impairment and consensus rules.	Very cautious
Teacher–adolescent	Core symptoms and broader attention scales	Direction generally positive, but metrics and constructs were incompatible for pooling.	Very cautious
Teacher–teacher	Dimensional and categorical school ratings	Agreement depended on teacher pair, instrument, subgroup and reliability definition.	Very cautious
Clinician or consensus	Symptoms and diagnostic classifications	Interpretation was limited when the reference decision incorporated the informant under comparison.	Very cautious
Moderators	Age, sex, parental attention-deficit/hyperactivity disorder, family and school characteristics	Study-specific findings were too inconsistently defined and analyzed for pooled inference.	Exploratory

Note: Confidence labels are descriptive judgments about interpretive caution and are not GRADE certainty ratings.

4. Discussion

4.1. Principal Findings

This systematic review provides an adolescence-specific synthesis of cross-informant agreement in ADHD assessment across dimensional symptom ratings, diagnostic classifications and several informant pairings. Caregiver–adolescent agreement was positive but modest for both core domains, with pooled correlations of 0.33 for inattention and 0.32 for hyperactivity/impulsivity. Caregiver–teacher estimates were of a similar magnitude, although only four studies contributed to each model and the Hartung–Knapp–Sidik–Jonkman confidence intervals included zero. These findings are consistent with the broader multi-informant literature, which indicates that reports from different observers usually share meaningful variance but rarely provide clinically interchangeable descriptions of child or adolescent psychopathology [7–9,24]. These pooled correlations should be interpreted as average effects across heterogeneous studies rather than as stable expected levels of agreement.

The similarity of the pooled estimates for inattention and hyperactivity/impulsivity is noteworthy. Hyperactive or impulsive behaviors might be expected to be more directly observable than difficulties involving sustained attention, organization or mental effort. However, greater behavioral visibility does not necessarily produce greater agreement when the behavior is expressed differently across settings or when informants apply different thresholds for judging it as developmentally inappropriate. The present results do not support the conclusion that either core ADHD domain is consistently less vulnerable to informant disagreement. Both domains showed very high heterogeneity, and the prediction intervals ranged from negative to strong positive values. Thus, although the average association was positive, the evidence does not support a single expected magnitude of caregiver–adolescent agreement across clinical, educational, or cultural settings.

The caregiver–teacher estimates require particularly cautious interpretation. Heterogeneity was also high in these models (I^2 = 94.1% and 83.0%), and prediction intervals could not be estimated because only four studies contributed to each analysis. Their confidence intervals included zero under the prespecified random-effects model, but this should not be interpreted as evidence that reports from home and school are unrelated. The point estimates were comparable with those observed for caregiver–adolescent agreement and with low-to-moderate associations reported in previous developmental studies [2,31,51,52]. The imprecision mainly reflects the small number of contributing studies and the marked between-study variation. Methodological quality further weakens confidence in these estimates: exclusion of High-concern reports left only two studies per domain, preventing estimation of a quality-restricted pooled effect. The caregiver–teacher estimates therefore have a weaker evidential basis than the caregiver–adolescent pooled estimates and should not be interpreted as having comparable precision or robustness. Fixed-effect analyses produced more precise estimates because large studies received dominant weight, but such models do not adequately represent variation across populations, instruments and school contexts. The random-effects findings therefore provide a more conservative account of the available evidence.

Diagnostic and threshold-based findings strengthened the conclusion that informant reports cannot be treated as substitutes. Chance-corrected agreement was frequently low, and estimates changed when impairment requirements, subtype rules or consensus procedures were modified [42–45]. This apparent difference between dimensional and categorical findings is not contradictory. A correlation describes whether informants rank adolescents in a similar order, whereas diagnostic agreement depends on whether both reports cross a specified threshold. Even a relatively small difference near that threshold may change the resulting classification. Diagnostic concordance is consequently influenced

not only by the similarity of the underlying ratings but also by symptom prevalence, the placement of the threshold, the requirement for impairment and the rule used to integrate information [23–29].

4.2. Relation to Previous Literature

The present findings extend rather than contradict earlier work on multi-informant assessment. Broad reviews of child and adolescent psychopathology have repeatedly shown that informant discrepancies are common and cannot be explained solely as measurement error [7–9]. The multi-informant approach is valuable precisely because observers encounter the young person under different conditions and may provide information about different portions of the same clinical problem [8,17]. Our results demonstrate that this principle remains highly relevant when the analysis is restricted to ADHD during adolescence, a developmental period in which observation opportunities and the adolescent's role in assessment change substantially.

Previous reviews have addressed related but distinct questions. Staff et al. [31] examined the validity of teacher rating scales against interviews and classroom observations, whereas Garcia-Rosales et al. [30] reviewed whether ADHD symptom criteria function equivalently across parent and teacher ratings. The current review synthesized observed agreement and discrepancy between informants; it did not test measurement invariance. These evidence streams are complementary but should not be conflated. A teacher scale may show acceptable validity against an external criterion while agreeing only modestly with a parent rating. Conversely, evidence of measurement non-invariance would indicate that items or latent dimensions function differently across reporters, but such non-invariance cannot be inferred from a modest correlation or a score difference alone. The present findings therefore support the clinical non-interchangeability of informant reports, while psychometric equivalence must be evaluated through dedicated invariance or related analyses.

Adolescence also distinguishes the present evidence from much of the earlier child-focused literature. As adolescents spend more time outside direct parental supervision, caregivers may have less access to school behavior, peer interactions and unstructured activities. At the same time, secondary-school teachers often observe an adolescent for only one subject or a limited part of the week. Adolescents have access to effort, distractibility and subjective impairment that may not be evident to adults, but they may have less opportunity for comparison with same-age peers or may interpret symptoms differently from parents and teachers [11–14]. Studies of adolescent self-report have accordingly shown that it is neither wholly unreliable nor closely concordant with adult reports [34–40,60]. The modest pooled correlations observed here are compatible with the view that adolescent self-report contains clinically relevant information while remaining incomplete when considered in isolation.

4.3. Why Agreement Varies

Contextual specificity is one likely contributor to the high heterogeneity. ADHD symptoms are not expressed independently of environmental demands. A structured classroom, an examination, an unstructured peer setting and a familiar home routine place different demands on sustained attention, inhibition, organization and emotional regulation. Symptoms may also vary within the same adolescent across school and non-school days and across different parts of the day [21]. Divergent ratings may therefore reflect real variation in the conditions under which difficulties emerge rather than an error that must be attributed to one informant [8,9,17].

Observation opportunities also differ substantially. Parents usually observe behavior over longer periods and across family routines, but may have little direct knowledge of classroom functioning. Teachers can compare an adolescent with many same-age peers, although the selected teacher, subject and classroom structure can materially affect the resulting assessment [3,31,51]. Evidence from studies involving more than one teacher confirms that agreement varies according to the teacher pair, instrument and statistical definition employed [57–59]. In secondary education, a single teacher report should not automatically be assumed to represent functioning throughout the entire school environment.

Informants also differ in their frames of reference. A caregiver may judge behavior against the adolescent's earlier functioning or the behavior of siblings. A teacher is more likely to compare the adolescent with classmates and to emphasize conduct that interferes with classroom management or academic completion. Adolescents may focus on internal effort, distress or situations that adults do not observe. Expectations about age-appropriate behavior, family relationships, school climate and sociocultural context may further influence the threshold at which a behavior is regarded as problematic [8,16–19]. The resulting discrepancy may therefore reveal whose functioning is affected and under which conditions, even when it does not identify a single objectively correct score.

Parent–adolescent relationship processes may also influence caregiver–adolescent agreement. Open communication is particularly relevant because caregivers can only report some aspects of an adolescent's experience to the extent that these experiences are observable or disclosed. Conversely, adolescents who experience communication with a parent as less open may be less likely to discuss difficulties involving attention, effort, organization or impairment that occur outside direct parental observation. Recent cross-sectional evidence indicates that adolescents with ADHD who were not receiving pharmacological treatment reported lower open communication with both mothers and fathers than typically developing adolescents after adjustment for age and problematic internet use, whereas group differences were not consistently observed across broader positive and negative relationship dimensions [63]. Adolescents receiving pharmacological treatment did not differ significantly from controls in open communication in that sample [63]. This study did not examine cross-informant ADHD agreement directly and therefore cannot establish that communication explains informant discrepancies. Rather, it identifies parent–adolescent communication as a plausible relational factor that future agreement studies should test explicitly.

Methodological variation is another important explanation. Parallel rating scales improve comparability, but informants may still receive differently worded items, use different reference periods or complete forms with different normative frameworks. Measurement invariance across parent and teacher ADHD ratings cannot be assumed and must be tested directly; the agreement estimates synthesized here do not themselves establish non-invariance [30]. Interviews allow clarification and explicit assessment of impairment but can introduce interviewer, order and information-contamination effects. In several diagnostic studies, the clinician or consensus classification incorporated information from one or both informants being evaluated, meaning that it could not serve as an entirely independent reference standard [42,44,45,60].

The sensitivity analyses did not isolate a single explanation for the observed heterogeneity. Omission of Kaner [37] reduced I^2 to 46.9% for caregiver–adolescent inattention and to 68.7% for hyperactivity/impulsivity, but this report differed simultaneously in sample size, national setting, instrument, and methodological quality. In the caregiver–teacher analyses, heterogeneity remained high after every single-study omission. These patterns suggest that several overlapping population and design differences may contribute to

the variation, but the small number of studies prevents their independent effects from being separated.

The choice of agreement statistic also shaped the evidence. Pearson correlations measure rank-order consistency but cannot detect a systematic tendency for one informant to assign higher scores. Intraclass correlations depend on whether consistency or absolute agreement is specified, while kappa coefficients depend on the distribution of categorical decisions. Mean differences can demonstrate directional bias without showing correspondence between individuals. Latent multi-trait multi-method models can separate shared trait variance from informant-specific variance, but their parameters do not translate directly into clinical classification agreement [34,54]. Studies reporting more than one of these indices provided a more informative description than studies that labeled a single correlation as "agreement."

4.4. Clinical Implications

The findings support retaining source-specific information throughout adolescent ADHD assessment. Diagnostic guidelines require evidence concerning symptoms and impairment across settings, but they do not require parents, adolescents and teachers to provide identical ratings [1,10]. Clinicians should initially record each report separately, including the setting observed, duration of observation, medication status, functional demands and any opportunity for informants to influence one another. Integration should occur after these contextual features have been examined, rather than by immediately averaging scores or privileging one source as the default truth.

When ratings diverge, the next clinical task is to determine the structure of the discrepancy. Item-level review may show that informants agree about some behaviors but differ about others. Follow-up questions should examine whether symptoms occur only under particular academic, social or family demands; whether one informant has had limited opportunity to observe the behavior; whether the disagreement concerns symptom presence or functional impact; and whether scale items were interpreted in the same way. This approach is consistent with models that treat informant discrepancies as potentially meaningful data about the young person's context rather than merely as an obstacle to diagnosis [8,17,18].

Adolescent self-report should be included but interpreted within this broader framework. Modest agreement with caregivers does not justify dismissing the adolescent's account. Adolescents can describe the effort needed to maintain concentration, subjective restlessness, compensatory strategies, experiences outside adult observation and the personal consequences of symptoms [11–14]. Conversely, self-report alone may not capture longitudinal change, comparative deviation from same-age peers or impairment visible to others. The most defensible interpretation therefore combines the adolescent's perspective with external observations while preserving areas of disagreement [36,38,39,60].

Teacher information remains particularly important because ADHD diagnostic formulation requires evidence beyond the home environment [1,10]. Nevertheless, the findings argue against treating any single teacher as a complete representation of school functioning. When feasible, information should be obtained from teachers who know the adolescent well and across more than one subject or classroom context. When substantial discrepancies remain after differences in setting, observation opportunities, timing, medication status, and item interpretation have been considered, additional information should be sought from sources that are as independent as possible of the discrepant ratings. Relevant evidence may include attendance and assignment-completion records, changes in academic functioning, disciplinary information, reports from another teacher or school professional, and, when feasible, direct observation. These sources should not be used to "vote" between

informants or to create a mechanical average of their reports. Instead, they should be mapped to the specific symptoms, settings, and functional consequences described by each informant to determine where independent corroboration exists. If clinically important discrepancies remain unresolved, that uncertainty should be retained explicitly in the formulation, and diagnostic or treatment decisions should be based on the overall pattern of symptoms, impairment, developmental history, and cross-setting evidence rather than on forced agreement between reporters [1,10,31,51,57,58].

Integration rules should also be matched to the clinical purpose. Requiring agreement between all informants may miss difficulties that are severe but setting-specific, whereas classifying an adolescent whenever any informant endorses a symptom may increase the number of positive classifications. Averaging informant scores can predict broader impairment in some circumstances, but may conceal clinically meaningful contrasts between home, school and self-report [2,24]. No single mathematical rule removes the need for clinical evaluation of symptom duration, developmental appropriateness, cross-setting expression and impairment [1,10].

4.5. Implications for Research

Future studies should report adolescent results separately from those of younger children. Twenty-nine otherwise relevant full-text reports were excluded because findings for participants aged 10–19 years could not be separated. This loss of evidence could be avoided by presenting age-stratified coefficients or by depositing sufficiently detailed analytic data. Age should not be treated only as a covariate in broad child samples, because the transition into adolescence changes school organization, autonomy, parental observation and the role of self-report. In the present evidence base, age was also confounded with study-level differences in instrument, setting, sampling and informant composition; the descriptive age information in Table S3 should therefore not be interpreted as evidence for or against a developmental effect on agreement.

A minimum reporting set for cross-informant studies should include the analysis sample for every estimate, the exact symptom domain and instrument, the informant pair, confidence intervals, administration timing, medication condition, independence of completion and the statistical definition of agreement. For categorical outcomes, complete cross-tabulations and the classification rule should be provided. For parallel continuous scales, correlations should be accompanied by information about systematic mean differences and, where appropriate, absolute-agreement ICCs. These procedures would address several of the recurrent methodological concerns identified through the QAREL appraisal [33].

Studies should also distinguish between consistency, absolute agreement, diagnostic concordance and agreement about impairment. These are separate questions and may have different clinical implications. Before interpreting score differences, investigators should examine whether the instrument functions comparably across informants [30]. When clinician or consensus decisions are used as comparators, reports should specify exactly which informant data were available to the decision makers so that incorporation of the index report into the reference classification can be recognized.

The evidence base did not permit reliable conclusions about whether disagreement is moderated by age, sex, parental ADHD, family characteristics, school context or cultural background. A further candidate moderator, highlighted by recent relational research but not directly tested in the agreement studies synthesized here, is the quality of parent–adolescent communication [63]. Future caregiver–adolescent agreement studies should therefore assess open communication and broader relationship quality directly, preferably separately for mothers and fathers, and test whether these factors account for variation in

symptom agreement. Medication status and problematic internet use may also warrant consideration as related contextual variables, while avoiding causal interpretation unless supported by longitudinal or experimental designs [63]. Future moderator analyses should be prespecified and use comparable definitions. Large longitudinal studies would be particularly informative because they could distinguish developmental change in agreement from changes caused by the transition to multiple secondary-school teachers or reduced parental observation. Repeated parallel assessments could also determine whether particular discrepancy patterns predict later impairment, treatment engagement or diagnostic stability rather than merely describing cross-sectional disagreement [2,3,20,51].

4.6. Strengths and Limitations

This review has several strengths. It addressed a clearly defined adolescence-specific question, searched medical, nursing, educational and multidisciplinary databases, and included comparisons involving adolescents, caregivers, teachers and clinicians. Extraction was conducted at result level, preserving distinctions between symptom domain, informant pair, instrument, age stratum and diagnostic rule. Correlations were not combined with kappas, ICCs or agreement proportions, because these statistics quantify different properties. The use of restricted-maximum-likelihood random-effects models and Hartung–Knapp–Sidik–Jonkman intervals avoided giving unwarranted precision to meta-analyses containing only a small number of studies. The review also documented four reports that could not be retrieved in full rather than silently omitting them from the study selection record.

Several limitations should be considered. The original title-and-abstract screening, data extraction, and methodological quality appraisal were performed by one reviewer. During peer-review revision, a second reviewer independently rescreened all 796 unique records, reassessed all 64 retrieved full-text reports, verified every study-domain effect entering the four final primary meta-analyses, and independently repeated the adapted QAREL appraisal of all included reports. These procedures substantially strengthen confidence in the final evidence set and primary quantitative analyses, but they do not fully replace a prospectively implemented duplicate workflow. In particular, the complete set of 282 result-level rows used across the narrative and quantitative evidence streams was not independently extracted in duplicate.

Registration was not fully prospective. Database searching and initial retrieval preceded registration, so the search process and eligibility framework already used to identify records were not constrained by the registered protocol. Registration did, however, precede formal extraction, methodological quality appraisal, effect grouping, and quantitative synthesis, and the revision now distinguishes registered plans from subsequent operational specifications. A chronology of methodological decisions relative to registration is provided in Table S7.

APA PsycInfo was not available through the accessible institutional interface. Although the search included Scopus and Academic Search Complete to broaden coverage beyond biomedical databases, no separate formal citation-chasing stage was undertaken. Psychology-focused studies indexed preferentially in PsycInfo, as well as grey literature reports, may therefore have been missed. This limits confidence that the search captured the complete relevant evidence base. Although no language restriction was imposed and one Spanish-language report was included, the final evidence base was overwhelmingly English-language (29 of 30 reports). Thus, the geographical spread of the included studies across 13 countries or national settings should not be interpreted as equivalent linguistic or cultural coverage.

Four potentially eligible reports could not be retrieved in full, and each meta-analysis contained only four or seven studies. Formal subgroup analysis or meta-regression would therefore have produced very small groups or unstable coefficients, particularly because candidate moderators were incompletely and inconsistently reported. The absence of formal moderator analysis should not be interpreted as evidence that age, sex, parental ADHD, family characteristics, medication status, or school context are unimportant. Rather, the very high and largely unexplained heterogeneity means that the summary correlations are average effects across diverse studies, not universal constants for adolescent ADHD assessment. The wide confidence and prediction intervals should be given greater weight than the numerical similarity of the pooled point estimates.

The methodological quality of the primary evidence also limits the precision of the conclusions. Only three of the 30 included reports were judged Low concern, whereas 15 had Some concerns and 12 were judged High concern. Recurrent problems included incomplete reporting of independent completion and masking, fixed assessment order or access to another informant's information, nonparallel instruments, age-stratum uncertainty, and use of correlation alone to characterize agreement. These issues may either inflate apparent correspondence through information contamination or make estimates difficult to interpret because the informants, instruments, or statistical definitions are not fully comparable.

The quality-restricted sensitivity analyses illustrate the practical importance of these concerns. Excluding High-concern reports reduced the caregiver–adolescent pooled estimate to $r = 0.24$ for inattention, with $I^2 = 33.1\%$, and to $r = 0.27$ for hyperactivity/impulsivity, with $I^2 = 74.8\%$. For the caregiver–teacher models, exclusion of High-concern reports left only two studies, precluding a meaningful quality-restricted meta-analysis. The numerical pooled correlations should therefore be interpreted cautiously, particularly for caregiver–teacher agreement. Greater confidence can be placed in the broader conclusion that reports from different informants provide partly overlapping but non-interchangeable information than in any single pooled estimate as a precise expected level of agreement.

Finally, the review did not apply GRADE certainty ratings. QAREL was useful for identifying methodological problems in individual reliability studies, but its judgments do not provide a complete assessment of certainty across a heterogeneous evidence stream [33]. The conclusions have therefore been framed cautiously: the evidence indicates that informant reports are related but not clinically interchangeable, whereas the exact magnitude of agreement and the effects of particular moderators remain uncertain. Psychometric equivalence across informant versions was not directly assessed in this review.

5. Conclusions

Cross-informant agreement in adolescent ADHD assessment was positive on average but highly heterogeneous and imprecisely estimated. The caregiver–adolescent prediction intervals crossed zero, and the caregiver–teacher confidence intervals also included zero; the pooled correlations should therefore be interpreted as average effects across diverse studies rather than stable levels expected in a particular setting. Caregiver, adolescent, teacher, and clinician reports should be treated as related but not clinically interchangeable sources of evidence; this conclusion does not establish psychometric non-equivalence of the instruments. The clinically useful response to disagreement is not to select a universally superior informant, but to investigate what each source observed and how the discrepancy relates to context and impairment. Better age-stratified reporting and harmonized moderator data are needed before narrower conclusions can be made about sources of heterogeneity or specific assessment procedures.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/children13091176/s1, Table S1. PRISMA 2020 checklist [32]. Table S2. Database search strategies and yields. Table S3. Characteristics of included reports. Table S4. Methodological quality appraisal of included reports using the adapted QAREL framework. Table S5. Study-level inputs for the primary meta-analyses. Table S6. Sensitivity and robustness analyses. Table S7. Timing and status of methodological decisions relative to OSF registration.

Author Contributions: Conceptualization, G.G.; methodology, G.G. and A.P.; software, G.G.; validation, G.G. and A.P.; formal analysis, G.G.; investigation, G.G. and A.P.; data curation, G.G. and A.P.; writing—original draft preparation, G.G.; writing—review and editing, G.G. and A.P.; visualization, G.G.; supervision, G.G.; project administration, G.G. 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. This systematic review analyzed data reported in published or publicly available studies.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new primary data were generated. The OSF registration and the study selection/full-text screening workbook are available in the associated Open Science Framework project (https://doi.org/10.17605/OSF.IO/49H5R (accessed on 20 August 2026)). Study characteristics, methodological quality judgments, primary meta-analytic inputs, and sensitivity analyses are reported in the Supplementary Materials (Tables S3–S6). Copyrighted full-text reports will not be redistributed. Additional review materials are available from the corresponding author upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

ADHD	Attention-deficit/hyperactivity disorder
CI	Confidence interval
DSM	Diagnostic and Statistical Manual of Mental Disorders
H/I	Hyperactivity/impulsivity
HKSJ	Hartung–Knapp–Sidik–Jonkman
ICC	Intraclass correlation coefficient
OSF	Open Science Framework
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QAREL	Quality Appraisal of Reliability Studies
REML	Restricted maximum likelihood

References

1. Peterson, B.; Trampush, J.; Brown, M.; Maglione, M.; Bolshakova, M.; Rozelle, M.; Miles, J.; Pakdaman, S.; Yagyu, S.; Motala, A.; et al. Tools for the Diagnosis of ADHD in Children and Adolescents: A Systematic Review. *Pediatrics* **2024**, *153*, e2024065854. [CrossRef] [PubMed]
2. Martel, M.; Eng, A.; Bansal, P.; Smith, T.; Elkins, A.; Goh, P. Multiple informant average integration of ADHD symptom ratings predictive of concurrent and longitudinal impairment. *Psychol. Assess.* **2021**, *33*, 443–451. [CrossRef] [PubMed]
3. Yeguez, C.E.; Sibley, M.H. Predictors of Informant Discrepancies Between Mother and Middle School Teacher ADHD Ratings. *Sch. Ment. Health* **2016**, *8*, 452–460. [CrossRef] [PubMed]
4. Kennerley, S.; Jaquiery, B.; Hatch, B.; Healey, M.; Wheeler, B.; Healey, D. Informant Discrepancies in the Assessment of Attention-Deficit/Hyperactivity Disorder. *J. Psychoeduc. Assess.* **2018**, *36*, 136–147. [CrossRef]
5. Izzo, V.; Donati, M.; Primi, C. Assessing ADHD Through the Multi-Informant Approach: The Contribution of the Conners' 3 Scales. *J. Atten. Disord.* **2019**, *23*, 641–650. [CrossRef] [PubMed]

6. Esposito, D.; Gigliotti, F.; Colotti, B.; Di Brina, C.; Pisani, F.; Romani, M. ADHD Reporting in Developmental Age: The Role of the Informants. *Children* **2025**, *12*, 914. [CrossRef] [PubMed]
7. Achenbach, T. Commentary: Definitely More Than Measurement Error: But How Should We Understand and Deal With Informant Discrepancies? *J. Clin. Child. Adolesc. Psychol.* **2011**, *40*, 80–86. [CrossRef] [PubMed]
8. De Los Reyes, A.; Augenstein, T.M.; Wang, M.; Thomas, S.A.; Drabick, D.A.G.; Burgers, D.E.; Rabinowitz, J. The Validity of the Multi-Informant Approach to Assessing Child and Adolescent Mental Health. *Psychol. Bull.* **2015**, *141*, 858–900. [CrossRef] [PubMed]
9. De Los Reyes, A.; Kazdin, A.E. Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study. *Psychol. Bull.* **2005**, *131*, 483–509. [CrossRef] [PubMed]
10. Wolraich, M.L.; Hagan, J.F., Jr.; Allan, C.; Chan, E.; Davison, D.; Earls, M.; Evans, S.W.; Flinn, S.K.; Froehlich, T.; Frost, J.; et al. Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics* **2019**, *144*, e20192528. [CrossRef] [PubMed]
11. Hareendran, A.; Setyawan, J.; Pokrzywinski, R.; Steenrod, A.; Madhoo, M.; Erder, M. Evaluating functional outcomes in adolescents with attention-deficit/hyperactivity disorder: Development and initial testing of a self-report instrument. *Health Qual. Life Outcomes* **2015**, *13*, 133. [CrossRef] [PubMed]
12. Meyer, J.; Alaie, I.; Ramklint, M.; Isaksson, J. Associated predictors of functional impairment among adolescents with ADHD—A cross-sectional study. *Child. Adolesc. Psychiatry Ment. Health* **2022**, *16*, 29. [CrossRef] [PubMed]
13. Andersen, A.C.; Sund, A.; Thomsen, P.; Lydersen, S.; Haugan, A.-L.J.; Nøvik, T. Executive function measured by BRIEF in adolescents diagnosed and treated for ADHD: Problem profiles and agreement between informants. *Child. Neuropsychol.* **2024**, *30*, 45–59. [CrossRef] [PubMed]
14. Andersen, A.C.; Thomsen, P.H.; Lydersen, S.; Sund, A.; Haugan, A.-L.J.; Nøvik, T. Behaviour ratings of executive functions in adolescents with ADHD: Correlation with core symptoms and functional impairment measured by multiple informants. *Nord. J. Psychiatry* **2025**, *79*, 347–353. [CrossRef] [PubMed]
15. Janicaud, N.; Chen, Y.; McAuliffe, T.; So, J.; Wilkes-Gillan, S.; Littlefair, D.; Graham, P.; Cordier, R. The construct components of self-regulation from a systematic review of self-regulation measures in children and adolescents. *Curr. Psychol.* **2026**, *45*, 497. [CrossRef]
16. Adhikari, S.; Ma, J.; Shakya, S.; Brøndbo, P.; Handegård, B.; Javo, A.C. Cross-informant ratings on emotional and behavioral problems in Nepali adolescents: A comparison of adolescents' self-reports with parents' and teachers' reports. *PLoS ONE* **2024**, *19*, e0303673. [CrossRef] [PubMed]
17. De Los Reyes, A.; Wang, M.; Lerner, M.; Makol, B.; Fitzpatrick, O.; Weisz, J. The Operations Triad Model and Youth Mental Health Assessments: Catalyzing a Paradigm Shift in Measurement Validation. *J. Clin. Child. Adolesc. Psychol.* **2023**, *52*, 19–54. [CrossRef] [PubMed]
18. Genachowski, K.; Starin, N.; Cummings, C.; Alvord, M.; Rich, B. Interpretation of Informant Discrepancy in School-Based Psychological Assessment of Internalizing and Externalizing Symptoms. *J. Emot. Behav. Disord.* **2022**, *31*, 248–259. [CrossRef]
19. Wexler, D.; Salgado, R.; Gornik, A.; Peterson, R.; Pritchard, A. What's race got to do with it? Informant rating discrepancies in neuropsychological evaluations for children with ADHD. *Clin. Neuropsychol.* **2022**, *36*, 264–286. [CrossRef] [PubMed]
20. Ketvel, L.; Vuoksimaa, E.; Pulkkinen, L.; Rose, R.; Vedenkannas, U.; Rapeli, P.; Raevuori, A.; Latvala, A. Dimensional ADHD symptoms and executive functioning in adolescence: A multi-informant, population-based twin study. *Neuropsychology* **2025**, *39*, 16–27. [CrossRef] [PubMed]
21. Pedersen, S.; Kennedy, T.; Joseph, H.; Riston, S.; Kipp, H.; Molina, B. Real-world changes in adolescents' ADHD symptoms throughout the day and across school and non-school days. *J. Abnorm. Child. Psychol.* **2020**, *48*, 1543–1553. [CrossRef] [PubMed]
22. Rohail, I.; Nisar, F.; Naz, M.D.; Fatima, H.; Zahra, R. Functional Divergence and Indicators of Adolescent Psychopathology: Evidence of Situational Specificity in a Pakistani Multi-Informant Cohort. *Nat.-Nurtur. J. Psychol.* **2026**, *6*, 44–55. [CrossRef]
23. Jónsdóttir, H.; Agnarsdóttir, H.; Jóhannesdóttir, H.; Smárason, O.; Harðardóttir, H.H.; Højgaard, D.; Skarphedinsson, G. Parent–youth agreement on psychiatric diagnoses and symptoms: Results from an adolescent outpatient clinical sample. *Nord. J. Psychiatry* **2022**, *76*, 466–473. [CrossRef] [PubMed]
24. Martel, M.; Markon, K.; Smith, G. Research Review: Multi-informant integration in child and adolescent psychopathology diagnosis. *J. Child. Psychol. Psychiatry* **2017**, *58*, 116–128. [CrossRef] [PubMed]
25. Booth, C.; Moreno-Agostino, D.; Fitzsimons, E. Parent-adolescent informant discrepancy on the Strengths and Difficulties Questionnaire in the UK Millennium Cohort Study. *Child. Adolesc. Psychiatry Ment. Health* **2023**, *17*, 57. [CrossRef] [PubMed]
26. Thöne, A.-K.; Görtz-Dorten, A.; Altenberger, P.; Dose, C.; Geldermann, N.; Hautmann, C.; Jendreizik, L.; Treier, A.-K.; Von Wirth, E.; Banaschewski, T.; et al. Toward a Dimensional Assessment of Externalizing Disorders in Children: Reliability and Validity of a Semi-Structured Parent Interview. *Front. Psychol.* **2020**, *11*, 1840. [CrossRef] [PubMed]
27. McDonald, E.; Whitney, S.; Horricks, L.; Lipman, E.; Ferro, M. Parent-Child Agreement on the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *J. Can. Acad. Child Adolesc. Psychiatry* **2021**, *30*, 264–272. [PubMed]

28. Kotelnikova, Y.; Clark, L.A.; Ruggero, C.; Ringwald, W.; Narrow, W.; Reed, G.; Skodol, A.; Tyrer, P.; Zimmerman, M.; Kotov, R. Strategies for Establishing Clinical-Decision Thresholds in Psychiatry: A Review. *JAMA Psychiatry* **2026**, *83*, 870–879. [CrossRef] [PubMed]
29. Clark, L.; Cuthbert, B.; Lewis-Fernández, R.; Narrow, W.; Reed, G. Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychol. Sci. Public Interest.* **2017**, *18*, 145–172. [CrossRef] [PubMed]
30. Garcia-Rosales, A.; Cortese, S.; Vitoratou, S. Measurement Invariance of Attention-Deficit/Hyperactivity Disorder Symptom Criteria as Rated by Parents and Teachers in Children and Adolescents: A Systematic Review. *PLoS ONE* **2024**, *19*, e0293677. [CrossRef] [PubMed]
31. Staff, A.I.; Oosterlaan, J.; van der Oord, S.; Hoekstra, P.J.; Vertessen, K.; de Vries, R.; van den Hoofdakker, B.J.; Luman, M. The Validity of Teacher Rating Scales for the Assessment of ADHD Symptoms in the Classroom: A Systematic Review and Meta-Analysis. *J. Atten. Disord.* **2021**, *25*, 1578–1593. [CrossRef] [PubMed]
32. 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**, *372*, n71. [CrossRef] [PubMed]
33. Lucas, N.P.; Macaskill, P.; Irwig, L.; Bogduk, N. The Development of a Quality Appraisal Tool for Studies of Diagnostic Reliability (QAREL). *J. Clin. Epidemiol.* **2010**, *63*, 854–861. [CrossRef] [PubMed]
34. Hartung, C.; McCarthy, D.; Milich, R.; Martin, C. Parent-Adolescent Agreement on Disruptive Behavior Symptoms: A Multitrait-Multimethod Model. *J. Psychopathol. Behav. Assess.* **2005**, *27*, 159–168. [CrossRef]
35. Hennig, T.; Schramm, S.A.; Linderkamp, F. Cross-Informant Disagreement on Behavioral Symptoms in Adolescent Attention-Deficit/Hyperactivity Disorder and Its Impact on Treatment Effects. *Eur. J. Psychol. Assess.* **2018**, *34*, 79–86. [CrossRef]
36. Hogue, A.; Dauber, S.; Lichvar, E.; Spiewak, G. Adolescent and caregiver reports of ADHD symptoms among inner-city youth: Agreement, perceived need for treatment, and behavioral correlates. *J. Atten. Disord.* **2014**, *18*, 212–225. [CrossRef] [PubMed]
37. Kaner, S. ADHD symptoms in national samples of Turkish adolescents: Self, parent, and teacher reports. *Procedia Soc. Behav. Sci.* **2011**, *15*, 3342–3348. [CrossRef]
38. Sibley, M.H.; Campez, M.; Raiker, J.S. Reexamining ADHD-Related Self-Reporting Problems Using Polynomial Regression. *Assessment* **2019**, *26*, 305–314. [CrossRef] [PubMed]
39. Willard, V.W.; Conklin, H.M.; Huang, L.; Zhang, H.; Kahalley, L.S. Concordance of parent-, teacher- and self-report ratings on the Conners 3 in adolescent survivors of cancer. *Psychol. Assess.* **2016**, *28*, 1110–1118. [CrossRef] [PubMed]
40. Gadow, K.D.; Sprafkin, J.; Carlson, G.A.; Schneider, J.; Nolan, E.E.; Mattison, R.E.; Rundberg-Rivera, V. A DSM-IV-referenced, adolescent self-report rating scale. *J. Am. Acad. Child Adolesc. Psychiatry* **2002**, *41*, 671–679. [CrossRef] [PubMed]
41. Montiel-Nava, C.; Peña, J.A. Discrepancy Between Parents and Teachers in the Evaluation of Behavior and Educational Problems in Children and Adolescents. *Rev. Neurol.* **2001**, *32*, 506–511. [CrossRef]
42. Cantwell, D.P.; Lewinsohn, P.M.; Rohde, P.; Seeley, J.R. Correspondence between adolescent report and parent report of psychiatric diagnostic data. *J. Am. Acad. Child Adolesc. Psychiatry* **1997**, *36*, 610–619. [CrossRef] [PubMed]
43. Colins, O.; Vermeiren, R.; Schuyten, G.; Broekaert, E.; Soyez, V. Informant agreement in the assessment of disruptive behavior disorders in detained minors in Belgium: A diagnosis-level and symptom-level examination. *J. Clin. Psychiatry* **2008**, *69*, 141–148. [CrossRef] [PubMed]
44. Jensen, P.S.; Rubio-Stipec, M.; Canino, G.; Bird, H.R.; Dulcan, M.K.; Schwab-Stone, M.E.; Lahey, B.B. Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary? *J. Am. Acad. Child Adolesc. Psychiatry* **1999**, *38*, 1569–1579. [CrossRef] [PubMed]
45. Rohde, L.A.; Biederman, J.; Knijnik, M.P.; Ketzer, C.; Chachamovich, E.; Vieira, G.M.; Pinzon, V. Exploring different information sources for DSM-IV ADHD diagnoses in Brazilian adolescents. *J. Atten. Disord.* **1999**, *3*, 91–96. [CrossRef]
46. Hope, T.L.; Adams, C.; Reynolds, L.; Powers, D.; Perez, R.A.; Kelley, M.L. Parent vs. self-report: Contributions toward diagnosis of adolescent psychopathology. *J. Psychopathol. Behav. Assess.* **1999**, *21*, 349–363. [CrossRef]
47. Gau, S.S.-F.; Lin, Y.-J.; Cheng, A.T.-A.; Chiu, Y.-N.; Tsai, W.-C.; Soong, W.-T. Psychopathology and symptom remission at adolescence among children with attention-deficit-hyperactivity disorder. *Aust. N. Z. J. Psychiatry* **2010**, *44*, 323–332. [CrossRef] [PubMed]
48. Juneja, M.; Sairam, S.; Jain, R. Attention deficit hyperactivity disorder in adolescent school children. *Indian Pediatr.* **2014**, *51*, 151–152. [CrossRef] [PubMed]
49. Wan Salwina, W.I.; Baharudin, A.; Nik Ruzyanei, N.J.; Midin, M.; Rahman, F.N. Attention deficit hyperactivity disorder symptoms reporting in Malaysian adolescents: Do adolescents, parents and teachers agree with each other? *Asian J. Psychiatry* **2013**, *6*, 483–487. [CrossRef] [PubMed]

50. Frick, M.A.; Hesser, H.; Sonuga-Barke, E. Pervasive versus situational childhood ADHD: Latent classes and their clinical characteristics, based on parent and teacher ratings in a large longitudinal population sample. *Eur. Child. Adolesc. Psychiatry* **2024**, *33*, 2253–2266. [CrossRef] [PubMed]
51. Narad, M.E.; Garner, A.A.; Peugh, J.L.; Tamm, L.; Antonini, T.N.; Kingery, K.M.; Simon, J.O.; Epstein, J.N. Parent-teacher agreement on ADHD symptoms across development. *Psychol. Assess.* **2015**, *27*, 239–248. [CrossRef] [PubMed]
52. Serra-Pinheiro, M.A.; Mattos, P.; Regalla, M.A. Inattention, hyperactivity, and oppositional-defiant symptoms in Brazilian adolescents: Gender prevalence and agreement between teachers and parents in a non-English speaking population. *J. Atten. Disord.* **2008**, *12*, 135–140. [CrossRef] [PubMed]
53. Martin, N.; Scourfield, J.; McGuffin, P. Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *Br. J. Psychiatry* **2002**, *180*, 260–265. [CrossRef] [PubMed]
54. Gomez, R.; Gomez, A. Agreement of adolescent ratings with mother ratings and teacher ratings of ADHD symptom groups: A correlated trait-correlated method minus one analysis. *Personal. Individ. Differ.* **2015**, *82*, 131–135. [CrossRef]
55. Pierrehumbert, B.; Bader, M.; Thévoz, S.; Kinal, A.; Halfon, O. Hyperactivity and attention problems in a Swiss sample of school-aged children: Effects of school achievement, child gender, and informants. *J. Atten. Disord.* **2006**, *10*, 65–76. [CrossRef] [PubMed]
56. Steinhausen, H.-C.; Drechsler, R.; Foldenyi, M.; Imhof, K.; Brandeis, D. Clinical course of attention-deficit/hyperactivity disorder from childhood toward early adolescence. *J. Am. Acad. Child. Adolesc. Psychiatry* **2003**, *42*, 1085–1092. [CrossRef] [PubMed]
57. Morris, S.; Ling, M.; Sheen, J.; Sciberras, E. The interteacher reliability of assessments of adolescents. *Psychol. Assess.* **2021**, *33*, 904–910. [CrossRef] [PubMed]
58. Molina, B.S.; Pelham, W.E.; Blumenthal, J.; Galiszewski, E. Agreement among teachers' behavior ratings of adolescents with a childhood history of attention deficit hyperactivity disorder. *J. Clin. Child. Psychol.* **1998**, *27*, 330–339. [CrossRef] [PubMed]
59. Danforth, J.S.; DuPaul, G.J. Interrater reliability of teacher rating scales for children with attention-deficit hyperactivity disorder. *J. Psychopathol. Behav. Assess.* **1996**, *18*, 227–237. [CrossRef]
60. Frick, M.A.; Lindman, L.; Meyer, J.; Isaksson, J. Are adolescent males or females more proficient self-raters of symptoms of Attention-Deficit/Hyperactivity Disorder? *J. Psychiatr. Res.* **2025**, *187*, 248–253. [CrossRef] [PubMed]
61. Rothen, S.; Vandeleur, C.L.; Lustenberger, Y.; Jeanprêtre, N.; Ayer, E.; Gamma, F.; Halfon, O.; Fornerod, D.; Ferrero, F.; Preisig, M. Parent-child agreement and prevalence estimates of diagnoses in childhood: Direct interview versus family history method. *Int. J. Methods Psychiatr. Res.* **2009**, *18*, 96–109. [CrossRef] [PubMed]
62. Faraone, S.V.; Monuteaux, M.C.; Biederman, J.; Cohan, S.L.; Mick, E. Does parental ADHD bias maternal reports of ADHD symptoms in children? *J. Consult. Clin. Psychol.* **2003**, *71*, 168–175. [CrossRef] [PubMed]
63. Özaslan, A.; Songür, İ.A.; Demirci, S.C.; Gürbüz, Ş.; Yıldırım, M.; Kaya, A.; Uras, M.C.; La Rosa, V.L.; Vehbi, A.; Güney, E.; et al. Examining the impact of ADHD, pharmacological treatment, and internet addiction on the parent-adolescent relationships quality. *Front. Psychol.* **2026**, *17*, 1761478. [CrossRef] [PubMed]

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.