

Assessment of child self-regulation: A review of ecological and structured approaches

Nachon, Juan Ignacio^{*a} ; Segretin, María Soledad ^{a,b} ; Fernández Liporace, Mercedes ^c ; Tisocco, Franco ^c ; Lipina, Sebastián Javier ^{a,b} 

Artículo de Revisión 

Abstract

The extent to which developmental science has mapped associations between cognitive, academic and socio-emotional functioning; caregiver reports and non-participant observation of young children was examined. Initially, 88 studies comprising over 300 effects were identified. The relationships between outcome measures and research design was described. Findings show that recurrent associations between task performance, behavioral report and naturalistic observation appear in this literature. The most frequently occurring comparisons in school settings were those between children's disruptive-aggressive behavior, on-task/off-task behavior, prosocial behavior, externalizing problems, prosocial skills and early math and language achievement. In home settings, those between children's positive/negative affect, engagement, compliance, temperament, cognitive and motor development were most common. Despite conceptual overlap, limited access to correlational data and between-study heterogeneity hinders cross-sample comparisons.

Keywords:

naturalistic observation, self-regulation, ecological validity, child development.

Resumen

Evaluación de la autorregulación infantil: Una revisión de enfoques ecológicos y estructurados. Se examinaron asociaciones entre el funcionamiento cognitivo, académico y socio-emocional, el reporte de cuidadores e indicadores observacionales, con el objetivo de establecer el estado del arte de los métodos ecológicos en las ciencias del desarrollo. Se identificaron 88 estudios que reúnen información sobre más de 300 efectos. Se describieron las relaciones entre los instrumentos de evaluación y el diseño de los estudios. Las comparaciones predominantes en escuelas fueron aquellas entre conductas disruptivas, pro-sociales y de involucramiento; el reporte docente de habilidades sociales y problemas externalizantes; y el desempeño en lengua y matemática. En contextos domésticos, primaron las comparaciones entre la expresividad emocional y el compromiso en actividades diádicas; los reportes del temperamento; y medidas del desarrollo cognitivo y motriz. A pesar de la convergencia conceptual, el acceso limitado a datos correlacionales y la heterogeneidad entre estudios dificulta la combinación de muestras.

Palabras clave:

observación no-participante, autorregulación, validez ecológica, desarrollo infantil.

Tabla de Contenido

Introduction	1
Materials and Methods	4
Results	5
Discussion	7
Acknowledgements	10
References	10
Supplementary Material	18

Recibido el 1 de junio de 2024; Aceptado el 11 de diciembre de 2024
 Editaron este artículo: Ana García Coni, Débora Mola, Belén Vera y Julieta Moltrasio

Developmental research relies on concepts and methods that are partially based on cultural representations of everyday life and typical behavior. For instance, most scales and tests that are administered in educational and clinical settings have the purpose of gathering data on

individual characteristics that are considered adaptive during children's progressive socialization (Bornstein & Arterberry, 2022; Lerner, 2006). As such, having access to evidence-based and ecologically valid instruments remains a concerning issue for scientists and practitioners

^a Centro de Educación Médica e Investigaciones Clínicas Norberto Quirno (CEMIC-CONICET), Unidad de Neurobiología Aplicada, Buenos Aires, Argentina
^b Instituto Universitario CEMIC (IUC), Buenos Aires, Argentina
^c CONICET e Instituto de Investigaciones en Psicología, Universidad de Buenos Aires, Buenos Aires, Argentina
 *Enviar correspondencia a: Nachon, Juan Ignacio. E-mail: juaninachon@gmail.com

Citar este artículo como: Nachon, J. I., Segretin, M. S., Fernández Liporace, M., Tisocco, F., & Lipina, S. J. (2025). Assessment of child self-regulation: A review of ecological and structured approaches. *Revista Argentina de Ciencias del Comportamiento*, 17(3), 1-45



who need to discern relevant and useful ways to measure change, growth, adaptation and dysfunction in human behavior across a variety of life stages and settings.

Ecological validity is often cited as a desirable attribute for psychological measures, one that ensures that an assessment is able to capture processes that reflect real-world demands and behaviors, or that what is being measured in the controlled environment of an experiment correlates to or predicts behavior in real-life situations (McCoy, 2019; Price et al., 2003). It is not always clear, however, if the claim that a measure is ecologically valid should be backed by evidence or if it is subjective, as face validity. From a psychometric perspective, ecological validity could also be appraised as an extension of content, criterion and external validities: Whether a test effectively measures what researchers intend to measure; whether the score that is obtained from a testing procedure correlates with a criterion measure of interest; and whether an outcome can reliably predict or be generalized to behavior in other settings, respectively (Jackson, 1996; McDermott, 2011).

An ecologically valid assessment for a certain psychological skill or trait in a population may be best thought of as a procedure in which the measurement of that construct is operationalized as a score, one that is found to be consistently associated with other previously agreed upon metrics (e.g., the rates of occurrence or the frequency counts for certain types of behavior), which are drawn from the naturally unfolding activity of this group of individuals in their daily lives. For instance, a laboratory based measure of pre-academic skills (e.g., aspects of early numeracy and literacy) could be considered ecologically relevant if the performance in this measure can accurately predict if preschool-aged children will later excel in their first grade math and language schoolwork (Thompson & Brassard, 1984). Similarly, a direct assessment of sustained attention that is carried out with a computerized continuous performance task may be considered ecologically valid for a school-aged sample if individual scores produced by this task significantly correlate with ratings from a questionnaire in which teachers are asked to report on each child's recurrent level of sustained attention during day-to-day classroom activities (Price et al., 2003).

Traditionally, research across academic disciplines has largely relied on various sources of empirical data when studying early childhood development, each with its distinctive strengths and caveats (Ahmed et al., 2021; McCoy, 2019): Direct cognitive and socioemotional assessments that are administered one-on-one to individual children in a controlled setting; experimental group-based tasks that are conducted with the collaboration of peers and caregivers; naturalistic or non-participant observations that take place in household and educational settings; and indirect reports of typical behavior that are provided by teacher and parent informants. Curiously, the convergent validity among all of these different approaches rarely has been explored systematically (Duckworth & Kern, 2011; Lindström et al., 2021; Silver, 2014). That is to say, quantitative evidence is not particularly abundant on the association between task performance, report scales and direct observational tools designed to measure children's naturally occurring behavior in their everyday life scenarios. Since most direct observational tools are relatively new and researchers can find complex circumstances that constrain their widespread roll-out with representative samples, less attention has been received by these instruments than other forms of assessments with easier scalability (McCoy, 2019; Nachon et al., 2020). Consequently, their psychometric properties have yet to be studied extensively across age and sociocultural groups.

This issue is important because developmental theories are committed to describe, explain and optimize how children's characteristics interact with their experiences to shape their cognition and behavior over time (Lerner, 2006). As most of children's experiences, thoughts and actions take place outside controlled research contexts, claims about the nature and distribution of what children do, what they experience and how they develop, to a large extent, refer to events that occur outside the laboratory (Dahl, 2017). Developmental science could turn to systematic naturalistic observation for a complementary point of view on the contextualized expression of processes that are typically being dissected in controlled experimental settings. Furthermore, direct observation procedures could address specific limitations present in most indirect report measures (e.g., social desirability bias, selective

reporting, memory recall inaccuracy). Then, if a positive correlation is consistently found across sustained attention scores from a computerized task and a highly reliable, but indirect report measure of sustained attention during school activities, one could expect to come to a similar or more reliable result if teacher reports were to be substituted for a representative amount of direct observation time inside the classroom. Unfortunately, this sort of hypothesis is rarely tested explicitly. As technical breakthroughs continue to permeate across behavioral research, however, testing such hypotheses offers an exciting prospect for teams working on novel psychometric solutions.

In this direction, the present study aims to help bridge a knowledge gap with regard to research on children's self-regulation (SR). SR refers to the ability to activate, monitor, adjust and inhibit behavior, attention, emotion and other cognitive and physiological processes in a flexible, adaptive and goal-oriented manner (Blair & Raver, 2012). During early childhood, SR is considered a critically developing asset for children's school readiness and effective social functioning. Several measures of self-regulatory processes that occur bidirectionally across levels of organization (e.g., hormonal markers related to the hypothalamic-pituitary-adrenal axis, cardiovagal autonomic function, temperament and emotional reactivity, executive functioning, patterns of dyadic oscillatory entrainment) have been found to predict academic and socioemotional outcomes throughout childhood, adolescence, and sometimes well into adulthood, in culturally diverse samples (McClelland et al., 2010; Wass et al., 2020). As an umbrella term that encompasses domain general and domain specific skills, SR is intertwined with constructs stemming from distinct multidisciplinary research programs, such as Executive Function (EF), socio-emotional competence, temperament and school readiness (Santos et al. 2024). EF are a set of skills involved in the regulation of thoughts, emotions, and behaviors (Diamond, 2013). Socio-emotional competences are a wide set of communicational abilities that draw on self-efficacy, motivation and knowledge about people, society and emotion to enable successful navigation of social challenges (Rose-Krasnor & Denham, 2009). Temperament stands for biologically-based individual variability in emotional, motor, and attentional reactivity and

regulation (Rothbart & Bates, 2006). School readiness can be thought of as the skills that children have developed at the time of school entry that — alongside family, peer, school, and community factors — facilitate learning (Blair, 2002).

The developmental window between birth and 96 months of age, canonically identified with the early childhood period by multilateral organizations (World Health Organization, 2012), stands as one of particular interest, as previous research has suggested that qualitative shifts in SR occur as children begin to traverse formal schooling, progressing from reactive, co-regulated forms of behavior to more proactive, complex and independent behaviors that are underpinned by change, growth and integration of fundamental language, socioemotional and cognitive skills (Montroy et al., 2016).

Given the multidimensional, context-dependent and relational aspects of SR, questions should be raised concerning the epistemological limits of non-participant observational methods in the study of partially covert processes, and the usefulness of said methods as complementary forms of measurement, ones that offset biases that are inherent to other sources of behavioral data, as could be explored through a multi-trait-multi-method matrix (Brown, 2015; Kline, 2023, Podsakoff et al., 2012).

The first objective of this review is to gain insight into the different ways in which non-participant observation in ecological settings has been studied alongside achievement tasks and indirect reports of young children's behavior (i.e., What domains of behavior have been assessed? How has behavior been operationalized? What types of activities have been observed?).

A second goal is to check for convergent validity and potential sources of variability in the associations between naturalistic observation of behavior and other types of assessment (i.e., Are there positive correlations between observations, reports and tasks used to measure similar theoretical constructs? Are correlation values the same when behavior is globally rated as opposed to when rates of occurrence or counts are estimated across time intervals? Do sociodemographic or developmental characteristics of children moderate the associations between these assessments?).

Materials and Methods

A group of peer-refereed articles was reviewed to determine if robust empirical relationships could be established between children's performance in controlled individual or group-based tasks that measure components of executive function, socio-emotional behavior, temperament or school readiness, caregiver reports of these same or other conceptually relevant constructs, and naturalistic observational data from children's cognitive and emotional SR, in the context of their homes, daycares, schools, community spaces or any other real-world behavioral setting. Additionally, there were intentions to clarify if such relationships might be modulated by socio-environmental conditions and individual characteristics (e.g., maternal education, child's age or biological sex).

This review was modeled on guidelines stated in the Preferred Reporting Items for Systematic reviews and Meta-Analyses statement (PRISMA; Moher, 2009) and a protocol was pre-registered in the PROSPERO International prospective register of systematic reviews database under the identifier CRD42020208409.

The PubMed.gov and Scielo.org databases were searched in October 2020 for peer-reviewed empirical articles, based on predetermined inclusion criteria (for the complete list of search terms used in each database, see supplementary file "0 - Search terms.docx"): (a) Studies with quantitative correlational or descriptive designs (either cross-sectional or longitudinal); (b) studies that recruited typically developing samples of children under 96 months of age (i.e., children that had not been diagnosed with a developmental disorder or had a presumed developmental difficulty; studies that did not exclusively sample children characterized by patterns of atypical behavior or stunted development, premature or low-weight births, chronic diseases, disabilities or that had been exposed to domestic violence, drug abuse, mental illness or forced displacements); (c) studies that informed the use of naturalistic observational measures in which the presence, frequency, duration, intensity or quality of children's individual-level behavior was defined as the unit for coding, in tandem with controlled direct assessments (i.e., tests/tasks), caregiver behavioral reports, or physiological measurements. Experimental and quasi-

experimental studies that met sample and outcome criteria were also considered for review, if they had included a passive control group or if they had made baseline (i.e., pre-intervention) assessments that could serve as comparisons for effect size estimation.

Each result was screened, eligibility checks were performed, duplicate records were removed, and researchers were contacted for unpublished results. Gray literature was sought through the Cognitive Development Society mailing list (<https://cogdevsoc.org/listserv>), and a snowball search strategy was deployed, based on the references collected during screening, personal readings and communications.

Data for each selected article was extracted into a frame with the following variables: (1) Article reference/doi; (2) year of publishing; (3) mean age; (4) predominant socioeconomic status and (5) size of the sample at baseline; (6) region/country of origin; (7) study design (cross-sectional/longitudinal); (8) description of observational proceedings and variables; (9) type of observed activities (e.g., free play, clean-up); (10) operationalization type of observational variables (e.g., global ratings, frequencies, rates of occurrence, hereafter referred to as primary outcomes); (11) approximate duration of observation sessions; (12) description of secondary outcomes (e.g., controlled direct assessments, caregiver behavioral reports); (13) individual level covariates; (14) contextual level covariates; and (15) presence of zero-order correlation coefficients reported by the authors.

For the risk of bias (i.e., quality) assessment, the first author modified the Cochrane Collaboration's Risk of bias tool (Higgins & Altman, 2008) and judged each individual study (at the study level, not outcome level) on the following items: (16) participant sampling bias (i.e., using random/stratified representative sampling vs. convenience sampling methods); (17) assessor bias (i.e., examiners being blind to study hypotheses and participant characteristics vs. not being blind); (18) inter-rater reliability bias (i.e., ICCs or Kappas reported above 0.8 for most observational variables vs. multiple coefficients reported below 0.8 values); and (19) attrition bias (i.e., appropriate methods for handling missing data and non-significant differences between participants that completed follow-up data collection rounds and those who dropped out vs.

Results

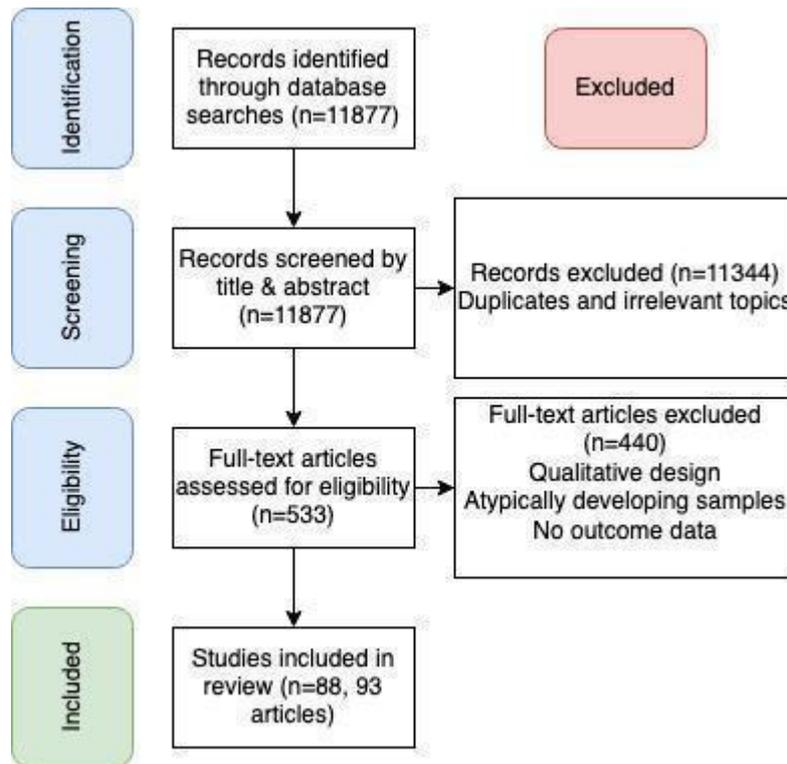
any of the opposite scenarios). A “Risk of bias” table is presented in the results section (scores are presented as “low risk of bias”, “high risk of bias”, and “unclear risk of bias”), and details on the available information that led to each individual judgment can be requested from the corresponding author.

Descriptive statistics were calculated for each variable in the data frame. Outcomes were then categorized according to their conceptual and technical definitions. To assess the amount of evidence available for each recurrent association between primary and secondary outcomes, the number of studies that reported on an association and the proportion of them which provided correlational data in the form of matrices was computed.

Database queries resulted in the identification of 11.877 records. The contents of their titles and abstracts were screened to filter duplicates and irrelevant studies. Five hundred and thirty-three records were considered of potential interest and were assessed again for eligibility on a full-text basis. Ninety-three articles reporting on 88 individual studies met all the pre-established inclusion criteria (Figure 1).

In this sample of 88 studies (see supplementary Table 1), the median sample size was 100 children (range = 10-3587). Nearly all took place in North America and Western Europe (see supplementary Figure 1). The most commonly studied age groups were toddlers and preschool-aged children (*Mdn* = 47.45 months,

Figure 1.
Flow diagram representing each step in the selection process



M = 47.84, *SD* = 31.96). There was a statistically significant difference in age composition in studies conducted in particular settings (Kruskal-Wallis $\chi^2(3) = 51.108, p < .001$), with a mean age of 25 months for home studies and 73 months for school studies (*Dunn Z* = -7.011, *p_{adj}* < .001) (see supplementary Figure 2).

Most studies recruited participants from diverse socioeconomic backgrounds (*n* = 34), although some samples did consist of distinct strata (Low SES: *n* = 16, Mid/high SES: *n* = 18, Unclear: *n* = 20). The number of studies which adopted longitudinal designs (*n* = 52) surpassed cross-sectional studies (*n* = 36). The most

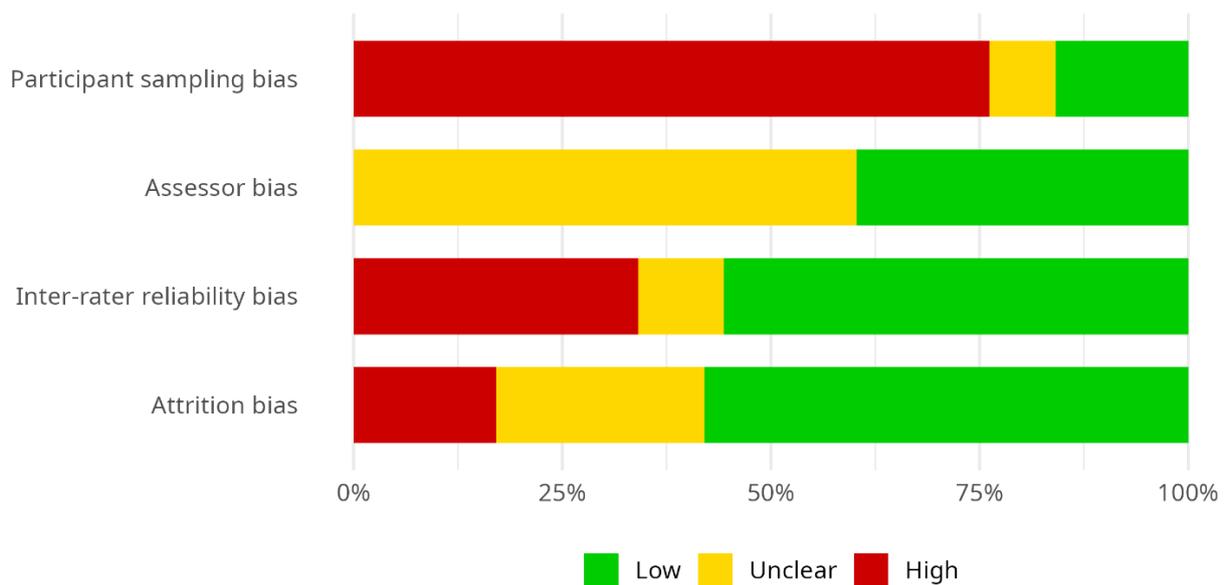
commonly observed scenarios were teacher-directed and child-directed activities in school settings, and dyadic play and household routines in home settings. The operationalization of behavioral variables was predominantly that of time sampling/frequentist approaches (i.e., frequency rates of an indicator's occurrence across time intervals), followed by global rating scales (i.e., ratings based upon the rater's integration of many attributes into single unified ratings) (see [supplementary Figure 3](#)). Observation sessions had a median duration of approximately 20 minutes ($M = 42.77$, $SD = 69.36$, $Min. = 2$, $Max. = 480$) (see [supplementary Figure 4](#)). Correlational data between primary and secondary outcomes of interest was not reported by most studies (Yes: $n = 24$, No: $n = 64$). The main

individual and contextual level covariates reported by these studies, which were often included in multiple regression models, were *age, sex, ethnicity, birth order, bilingualism, caretaker's level of formal education, household composition and levels of income and occupation* (see [supplementary Table 2](#)).

Regarding the risk of bias assessment ([Figure 2](#), also see [supplementary Table 3](#)), two-thirds of the studies did not recruit participants through representative or randomized sampling procedures. In this sense, it could be argued that the generalizability of the findings from most of these studies is limited and does not necessarily represent the full spectrum of typically developing children populations that live in those communities.

Figure 2.

Risk-of-bias judgment distribution for each domain



Then, almost 60% did not explicitly inform that people in charge of observations had been blind to hypotheses or other participant outcome data. This could be interpreted as a considerable risk of bias in their data collection. On the other hand, 55% reported very high levels of inter-rater reliability (e.g., kappa values over 0.8) for all their observational variables. This suggests that observations in some of these studies took place with a high degree of agreement between raters. The level of inter-rater agreement was not significantly related to the choice of approach to measurement (Pearson's $\chi^2(2) = 5.298$, $p = 0.07$)

(see [supplementary Figure 5](#)). Finally, 58% reported low levels of attrition and the use of appropriate methods when handling missing data (Pigott, 2001). This again restricts the generalizability of their findings.

Onto the types of interactions that were explored in this group of studies (see [supplementary Table 4](#)), it was found that associations between observational measures and other instruments mainly focused on domains stemming from psychopathology, personality and learning theory. In this sample, the most frequently reported interactions by studies conducted in

school settings were those between observed *disruptive-aggressive behavior*, observed *engagement behavior* (i.e., *on-task/off-task*), observed *prosocial behavior*, teacher reports of children's *externalizing problems* and *prosocial skills*, and children's early *math* and *language* achievement. In home settings, the most recurring associations reported were those between observed *positive/negative affect*, observed *engagement and compliance*, parental reports of children's *temperament*, and direct measures of children's *cognitive* and *motor development*.

When considering the methods involved in the measurement of these associations, it was found that these studies rarely relied on common instruments and procedures to assess similar constructs (i.e., diverse scales and tests are used to probe similarly defined aspects of behavior and development; see [supplementary Table 5](#)). For example, studies that explored the most frequently reported association, the one between disruptive/aggressive behavior (as observed in school settings) and teacher reports of externalizing problems, show the recurrent use of various well established report scales to measure externalizing problems (see [supplementary Table 6](#)): The Child Behavior Checklist (CBCL; [Achenbach, 1999](#)); The Conners Parent-Teacher Rating Scale (CPTRS; [Conners et al., 1998](#)) and the Strengths and Difficulties Questionnaire (SDQ; [Vostanis, 2006](#)). However, they also display differences in how they define and measure disruptive/aggressive behavior. For example, some might calculate the proportion of 30-second intervals in which a target child is observed in heated verbal disputes with peers, across a 10-minute period of free play, while others count the total number of times a target child is observed hurling objects at classmates, across a 5-minute period of seated teacher-directed instruction (see [supplementary Figure 6](#)).

As far as the authors are aware, these different observational procedures have not been previously tested for convergence, so they probably should not be treated as comparable measures of the same effect in the context of a quantitative synthesis. Although they might tap into the same behavioral construct from a conceptual point of view, they certainly do not share the same unit or scale of measurement.

Thus, they are not necessarily interchangeable.

Discussion

The first objective of this review was to gain insight into the different ways in which non-participant observation in ecological settings has been studied alongside achievement tasks and indirect reports of young children's behavior (i.e., What domains of behavior have been assessed? How has behavior been operationalized? What types of activities have been observed?). A second goal was to check for convergent validity and potential sources of variability in the associations between naturalistic observation of behavior and other types of assessment (i.e., Are there positive correlations between observations, reports and tasks used to measure similar theoretical constructs? Are correlation values the same when behavior is globally rated as opposed to when rates of occurrence or counts are estimated across time intervals? Do sociodemographic or developmental characteristics of children moderate the associations between these assessments?).

First, the most frequently reported interactions in school settings were noted to be those between the observation of engagement, disruptive and prosocial behavior during a variety of teacher-directed and children-directed activities, teacher report measures of externalizing problems and prosocial skills and direct assessment of children's early math and language achievement. In home settings, the most recurrently studied associations were those between the observation of children's emotional expression, compliance and engagement during dyadic play and household routines, parent report of children's temperamental traits and direct measures of children's cognitive and motor development.

A great diversity was observed in the ways in which behavioral constructs were defined across the studies that were reviewed: In the types of indirect and direct measures that were used, and in how observations in naturalistic settings were conducted. The relative high heterogeneity in how cognitive, behavioral and socio-emotional outcomes are conceptually defined and operationalized is a well-known issue that hinders evidence integration and synthesis in developmental science ([Fredricks & McColsky, 2012](#); [Nocita, 2021](#)). Across cultural

contexts and disciplinary fields, the measurement and interpretation of complex behaviors are typically bound to normative frameworks that can vary between behavioral settings and populations (e.g., certain behavior might be considered maladaptive for a certain age group in a determined culture, but expected or even valued in another setting).

With regard to observational units of analysis, it was found that studies predominantly coded behavior by adopting and combining frequentist and qualitative approaches (Hintze et al., 2002; Volpe et al., 2005). However, important differences persisted among these procedures, especially in relation to the total amount of time children were observed, in the length of observation intervals, the range of Likert scale values selected, and in other methods of computation (e.g., using rates of occurrence, partial interval sampling, unweighted averages of qualitative ratings, etc.). It is still unclear if some procedures are more relevant than others when it comes to measuring certain behavioral constructs in specific situations. For example, it has been ascertained that global ratings constitute a practical way to combine observations from different activities across multiple observation sessions (Downer et al., 2010), but frequency counts might still offer a more objective and time sensitive approach, one that is less prone to inter-rater variability (another usual critique that is posed on indirect report measures).

Second, the methodological dissimilarities within the sample halted progress towards the goal of estimating pooled effect sizes and checking for sources of variability. Noteworthy, most studies did not report correlation coefficients and their authors tended not respond to requests for unpublished data. In consequence, studies reporting correlation values for the same association pair were almost impossible to find, and the statistical requirements of meta-analytical models were not met. Combining outcome metrics that are too diverse could have eventually resulted in artificially high levels of heterogeneity and results that are difficult to interpret (Higgins et al., 2019). Different measures for similar developmental constructs have previously failed to meet the commensurability assumption that underlies meta-analysis (Pace & Brannick,

2010) and meta-regression models have revealed that combining measure types (e.g., considering that observation tools and indirect report scales measure the same attribute without previous evidence of their convergent validity) can introduce bias in pooled correlation estimates (Lindström et al., 2021). For these reasons, the decision to forgo a narrative synthesis of the directionality and variability in said effects was reached as well.

Limitations

The exploratory nature of this review should be clearly pointed out. The results and their interpretation are circumscribed to a relatively small sample of studies that do not necessarily represent the whole scientific corpus that exists on the topic of naturalistic observation. Despite the fact that systematic procedures were followed in order to guarantee a minimum degree of methodological rigor and reproducibility, the inclusion of every single article that exists on the subject cannot be claimed. Initially, the search protocol included more databases and journals that were later dropped because of time and resource constraints. Also, protocol guided searches are considered insufficient when it comes to obtaining representative literature samples in a given area of review (Greenhalgh & Peacock, 2005). Snowball searches hint to the possibility that hundreds or even thousands of articles compatible with the predefined inclusion criteria might exist elsewhere, in addition to the ones that were successfully identified. This could be attributable to the search strategy being imperfect, or to articles not being indexed in the consulted bibliographic resources. In this sense, the use of observational methods in the developmental literature is particularly difficult to identify solely from the content of articles' titles and abstracts, and tools for performing queries on a full-text basis are not commonly available in most search engines.

The other important caveat that must be addressed here is that this review was not able to access sufficient correlational evidence to conduct a thorough synthesis of all the empirical associations present in this literature. Without such data, it is not possible to confidently and systematically summarize the direction, magnitude and potential sources of variance in these associations. Here, content can only be

found in the thought that this experience—one that had outlined a meta-analysis but could not pursue it— might guide the design of future studies.

Conclusion

Measures derived from non-participant observation of children in ecological settings are still in an incipient stage of research and development. They hold the prospect of complementing the information that is typically obtained by more ubiquitous psychometric instruments (i.e., controlled laboratory tasks and indirect report scales). They will probably allow developmental scientists to gain a better understanding of how cognitive and relational processes express themselves in children's everyday lives, processes that traditionally have been studied indirectly or *in vitro*.

It is possible that the next decade of research will see great leaps in the fields of computer vision and pattern recognition technology, ensuing the automation of observation protocols and an exponential growth of massive behavioral data repositories (Ramakrishnan et al., 2019; Woolverton & Pollastri, 2021). This scenario will pose new ethical and legal challenges for behavioral researchers conducting studies in societies that choose to embrace these technologies (e.g., the potential use that observational procedures might have for unconsented surveillance, or how to best mask and protect biometric data).

As the ecological validity of instruments that are used to assess self-regulatory processes in human relational systems is constantly shifting with the tides of socio-cultural change, the ways in which adaptive behavior is defined and studied must transcend the pitfalls of both cultural bias and mechanistic conceptions, by constantly reexamining its empirical foundations and integrating subaltern perspectives that coexist with hegemonic values in a society. Adaptive behavior should not be defined unilaterally by academics, it should include the views held in the community that partake in the study of their children's development (DeJoseph et al., 2021).

At this point in time, the landscape of systematic naturalistic observation looks boggy. Raising the terrain should be easier by having multiple orchestrated teams working together to test hypotheses across different socio-cultural

settings. Some findings of this review might have practical implications for the design of future studies, in particular, those pointing to significant gaps in the literature. In the articles that were reviewed here, it was noted that some demographic groups and behavioral settings remain underrepresented. For example, comparatively few studies have taken place outside North America and Western Europe; children are less likely to be directly observed in household settings as they age; and—perhaps more notably— little research of this nature has been conducted in community settings other than homes and schools. As typical behavior is expected to change with children's age and with the context in which they develop, it would be useful to consider building transcultural and longitudinal behavioral databases, so as to better understand which behaviors are more common across age groups and social settings frequented by children (Valsiner, 2007). Behaviors that better characterize certain populations and that significantly predict other aspects of their development could be identified in this way, similarly to how other psychometric instruments are adapted to different cultural contexts and rescaled for their attuned interpretation. Multi-site, transcultural and longitudinal studies with representative samples of participants would be an ideal step in this direction.

It would also be interesting to see future observational studies contrast operationalization methods and activity settings. For instance, it is still unclear if there is any benefit in defining observational units for engagement as the proportion of 20-second intervals in which a child is partially observed to be disengaged from a certain task, as opposed to rating his global level of engagement in a 1 to 5 scale. Does engagement in book-sharing activities with peers predict reading proficiency better than engagement during periods of teacher-led literacy instruction? How long should observation sessions or intervals last for in order to capture a significant portion of children's behavior? Could methods from time series analysis be brought into play to better understand how a child's behavior unfolds during an activity, with regard to itself and to the behavior of other individuals?

Finally, since the lack of openly accessible

databases constitutes a major hurdle for quantitative synthesis efforts, it would certainly not be a bad idea for journals in this field to further promote data sharing culture in their editorial practices.

Pre-register

A protocol was pre-registered in the PROSPERO International prospective register of systematic reviews database under the identifier CRD42020208409..

Data Availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Acknowledgements

Este trabajo ha sido realizado con el aporte de subsidios de CONICET [PIP 2015 N° 0794, PIP-2014 N°0500, Beca interna doctoral: "Desarrollo autorregulatorio: Exploración y comparación de abordajes de laboratorio y ecológicos". Responsable: Nachon, Juan Ignacio y FONCYT- [PICT 2017: "Análisis del impacto de las experiencias infantiles asociadas a diferentes condiciones socioeconómicas en el desempeño cognitivo, emocional y académico". Investigadora Responsable: Segretin, María Soledad].

Conflict of interest

Los autores declaran la ausencia de cualquier relación comercial o financiera que pudiera interpretarse como un posible conflicto de intereses.

References

Articles included in the review are marked with an asterisk).

- *Aboud, F. E. (2007). Evaluation of an early childhood parenting programme in rural Bangladesh. *Journal of health, population, and nutrition*, 25(1) 3-13. <https://pubmed.ncbi.nlm.nih.gov/17615899/>
- Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed., pp. 429–466). Lawrence Erlbaum Associates Publishers.
- Ahmed, S. F., Grammer, J., & Morrison, F. (2021).

Cognition in context: Validating group-based executive function assessments in young children. *Journal of Experimental Child Psychology*, 208, 105131. <https://doi.org/10.1016/j.jecp.2021.105131>

- *Akdemir, D., Chadaroğlu-Cetin, F., Ozusta, S., & Karadağ, F. (2010). Relationship disorders and cognitive functioning in young children. *The Turkish Journal of Pediatrics*, 52(5), 512-519. <https://pubmed.ncbi.nlm.nih.gov/21434537/>
- *Apter-Levi, Y., Pratt, M., Vakart, A., Feldman, M., Zagoory-Sharon, O., & Feldman, R. (2016). Maternal depression across the first years of life compromises child psychosocial adjustment; relations to child HPA-axis functioning *Psychoneuroendocrinology*, 64, 47-56. <https://doi.org/10.1016/j.psyneuen.2015.11.006>
- *Arevalo, A., Kolobe, T. H., Arnold, S., & DeGrace, B. (2014). Early childrearing practices and their relationship to academic performance in Mexican American children. *Pediatric Physical Therapy*, 26(2), 214-222. <https://doi.org/10.1097/PEP.0000000000000033>
- *Assel, M. A., Landry, S. H., Swank, P. R., Steelman, L., Miller-Loncar, C., & Smith, K. E. (2002). How do mothers' childrearing histories, stress and parenting affect children's behavioural outcomes? *Child: Care, health and development*, 28(5), 359-368. <https://doi.org/10.1046/j.1365-2214.2002.00285.x>
- *Baiao, R., Baptista, J., Carneiro, A., Pinto, R., Toscano, C., Fearon, P., Soares, I., & Mesquita, A. R. (2018). Assessing preschoolers interactive behaviour: A validation study of the "Coding System for Mother-Child Interaction". *Child: Care, Health and Development*, 44(4), 644-650. <https://doi.org/10.1111/cch.12572>
- *Belsky, J., Woodworth, S., & Crnic, K. (1996). Trouble in the second year: three questions about family interaction. *Child development*, 67(2), 556-578. <https://doi.org/10.1111/j.1467-8624.1996.tb01751.x>
- *Berger, R. H., Valiente, C., Eisenberg, N., Hernandez, M. M., Thompson, M., Spinrad, T., VanSchyndel, S., Silva, K., & Southworth, J. (2017). Effortful control and school adjustment: The moderating role of classroom chaos. *Journal of applied developmental psychology*, 53, 108-119. <https://doi.org/10.1016/j.appdev.2017.10.001>
- *Bernier, A., Carlson, S. M., Deschênes, M., & Matte-Gagné, C. (2012). Social factors in the development of early executive functioning: a closer look at the caregiving environment. *Developmental science*, 15(1), 12-24. <https://doi.org/10.1111/j.1467-7687.2011.01093.x>
- Blair, C. (2002). School readiness: Integrating

- cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *The American psychologist*, 57(2), 111-127. <https://doi.org/10.1037//0003-066x.57.2.111>
- Blair, C., & Raver, C. C. (2012). Individual development and evolution: Experiential canalization of self-regulation. *Developmental Psychology*, 48(3), 647-657. <https://doi.org/10.1037/a0026472>
- *Bocknek, E. L., Brophy-Herb, H. E., Fitzgerald, H., Burns-Jager, K., & Carolan, M. T. (2012). Maternal psychological absence and toddlers' social-emotional development: interpretations from the perspective of boundary ambiguity theory. *Family process*, 51(4), 527-541. <https://doi.org/10.1111/j.1545-5300.2012.01411.x>
- Bornstein, M. H., & Arterberry, M. E. (2022). *Infancy: The Basics*. (1st ed.). Routledge. <https://doi.org/10.4324/9781003172802>
- *Brandes-Aitken, A., Braren, S., Swingler, M., Voegtline, K., & Blair, C. (2019). Sustained attention in infancy: A foundation for the development of multiple aspects of self-regulation for children in poverty. *Journal of experimental child psychology*, 184, 192-209. <https://doi.org/10.1016/j.jecp.2019.04.006>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.
- *Buckley, M. E., Klein, D. N., Durbin, C. E., Hayden, E. P., & Moerk, K. C. (2002). Development and validation of a Q-sort procedure to assess temperament and behavior in preschool-age children. *Journal of Clinical Child and Adolescent Psychology*, 31(4), 525-539. https://doi.org/10.1207/s15374424jccp3104_11
- *Cerniglia, L., Cimino, S., & Ballarotto, G. (2014). Mother-child and father-child interaction with their 24-month-old children during feeding, considering paternal involvement and the child's temperament in a community sample. *Infant Mental Health Journal*, 35(5), 473-481. <https://doi.org/10.1002/imhj.21466>
- *Chuang, C. C., Reinke, W. M., & Herman, K. C. (2020). Effects of a universal classroom management teacher training program on elementary children with aggressive behaviors. *School psychology*, 35(2), 128-136. <https://doi.org/10.1037/spq0000351>
- Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998). The Revised Conners' Parent Rating Scale (CPRS-R): Factor Structure, Reliability, and Criterion Validity. *Journal of Abnormal Child Psychology*, 26(4), 257-268. <https://doi.org/10.1023/A:1022602400621>
- *Coplan, R. J., Prakash, K., O'Neil, K., & Armer, M. (2004). Do you "want" to play? Distinguishing between conflicted shyness and social disinterest in early childhood. *Developmental psychology*, 40(2), 244-258. <https://doi.org/10.1037/0012-1649.40.2.244>
- *Crittenden, P. M. (1992). Children's strategies for coping with adverse home environments: an interpretation using attachment theory. *Child Abuse & Neglect*, 16(3), 329-343. [https://doi.org/10.1016/0145-2134\(92\)90043-q](https://doi.org/10.1016/0145-2134(92)90043-q)
- *Crittenden, P. M., & DiLalla, D. L. (1988). Compulsive compliance: the development of an inhibitory coping strategy in infancy. *Journal of abnormal child psychology*, 16(5), 585-599. <https://doi.org/10.1007/BF00914268>
- Dahl, A. (2017). Ecological Commitments: Why Developmental Science Needs Naturalistic Methods. *Child Development Perspectives*, 11(2), 79-84. <https://doi.org/10.1111/cdep.12217>
- *Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade classroom environment. *Journal of school psychology*, 53(5), 409-428. <https://doi.org/10.1016/j.jsp.2015.07.004>
- DeJoseph, M. L., Sifre, R. D., Raver, C. C., Blair, C. B., & Berry, D. (2021). Capturing Environmental Dimensions of Adversity and Resources in the Context of Poverty Across Infancy Through Early Adolescence: A Moderated Nonlinear Factor Model. *Child Development*, 92(4), e457-e475. <https://doi.org/10.1111/cdev.13504>
- *DeLuzio, J., & Girolametto, L. (2011). Peer interactions of preschool children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, 54(4), 1197-1210. [https://doi.org/10.1044/1092-4388\(2010/10-0099\)](https://doi.org/10.1044/1092-4388(2010/10-0099))
- Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64(1), 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- *Dollberg, D., Feldman, R., Keren, M., & Guedeney, A. (2006). Sustained withdrawal behavior in clinic-referred and nonreferred infants. *Infant mental health journal*, 27(3), 292-309. <https://doi.org/10.1002/imhj.20093>
- Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*, 25(1), 1-16. <https://doi.org/10.1016/j.ecresq.2009.08.004>
- Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45(3), 259-268.

- <https://doi.org/10.1016/j.jrp.2011.02.004>
- *Duggan, A. K., McFarlane, E. C., Windham, A. M., Rohde, C. A., Salkever, D. S., Fuddy, L., Rosenberg, L. A., Buchbinder, S. B., & Sia, C. C. (1999). Evaluation of Hawaii's Healthy Start Program. *The Future of Children*, 9(1), 66-90. <https://doi.org/10.2307/1602722>
- *Dunn, J., & Kendrick, C. (1980). Studying temperament and parent-child interaction: comparison of interview and direct observation. *Developmental Medicine & Child Neurology*, 22(4), 484-496. <https://doi.org/10.1111/j.1469-8749.1980.tb04353.x>
- *DuPaul GJ, McGoey KE, Eckert TL, VanBrakle J. (2001). Preschool children with attention-deficit/hyperactivity disorder: impairments in behavioral, social, and school functioning. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(5), 508-515. <https://doi.org/10.1097/00004583-200105000-00009>.
- *DuPaul, G. J., & Rapport, M. D. (1993). Does methylphenidate normalize the classroom performance of children with attention deficit disorder?. *Journal of the American Academy of Child & Adolescent Psychiatry*, 32(1), 190-198. <https://doi.org/10.1097/00004583-199301000-00027>
- *DuPaul, G. J., Ervin, R. A., Hook, C. L., & McGoey, K. E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: effects on classroom behavior and academic performance. *Journal of applied behavior analysis*, 31(4), 579-592. <https://doi.org/10.1901/jaba.1998.31-579>
- *Eisenberg, N., Shepard, S. A., Fabes, R. A., Murphy, B. C., & Guthrie, I. K. (1998). Shyness and children's emotionality, regulation, and coping: Contemporaneous, longitudinal, and across-context relations. *Child development*, 69(3), 767-790. <https://doi.org/10.1111/j.1467-8624.1998.tb06242.x>
- Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 763–782). Springer. https://doi.org/10.1007/978-1-4614-2018-7_37
- *Gibson, J., Hussain, J., Holsgrove, S., Adams, C., & Green, J. (2011). Quantifying peer interactions for research and clinical use: The Manchester Inventory for Playground Observation. *Research in Developmental Disabilities*, 32(6), 2458-2466. <https://doi.org/10.1016/j.ridd.2011.07.014>
- *Gouze, K. R. (1987). Attention and social problem solving as correlates of aggression in preschool males. *Journal of abnormal child psychology*, 15(2), 181-197. <https://doi.org/10.1007/BF00916348>.
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ: British Medical Journal*, 331(7524), 1064-1065. <https://doi.org/10.1136/bmj.38636.593461.68>
- *Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K. N., Beland, K., Frey, K., & Rivara, F. P. (1997). Effectiveness of a Violence Prevention Curriculum Among Children in Elementary School: A Randomized Controlled Trial. *JAMA*, 277(20), 1605–1611. <https://doi.org/10.1001/jama.1997.03540440039030>.
- *Gueron-Sela, N., Camerota, M., Willoughby, M. T., Vernon-Feagans, L., Cox, M. J., & The Family Life Project Key Investigators. (2018). Maternal depressive symptoms, mother-child interactions, and children's executive function. *Developmental Psychology*, 54(1), 71–82. <https://doi.org/10.1037/dev0000389>
- *Hane, A. A., Fox, N. A., Polak-Toste, C., Ghera, M. M., & Guner, B. M. (2006). Contextual basis of maternal perceptions of infant temperament. *Developmental Psychology*, 42(6), 1077–1088. <https://doi.org/10.1037/0012-1649.42.6.1077>
- *Harel-Gadassi, A., Friedlander, E., Yaari, M., Bar-Oz, B., Eventov-Friedman, S., Mankuta, D., & Yirmiya, N. (2020). Do developmental and temperamental characteristics mediate the association between preterm birth and the quality of mother-child interaction?. *Infant Behavior and Development*, 58, Article 101421. <https://doi.org/10.1016/j.infbeh.2020.101421>.
- *Hatch, B., Galland, B. C., Gray, A. R., Taylor, R. W., Sayers, R., Lawrence, J., & Taylor, B. (2019). Consistent use of bedtime parenting strategies mediates the effects of sleep education on child sleep: secondary findings from an early-life randomized controlled trial. *Sleep Health*, 5(5), 433-443. <https://doi.org/10.1016/j.sleh.2019.03.002>.
- *Herman, K. C., & Reinke, W. M. (2017). Improving teacher perceptions of parent involvement patterns: Findings from a group randomized trial. *School Psychology Quarterly*, 32(1), 89-104. <https://doi.org/10.1037/spq0000169>.
- *Hernández, M. M., Eisenberg, N., Valiente, C., Vanschyndel, S. K., Spinrad, T. L., & Silva, K. M. (2016). Emotional expression in school context, social relationships. *Emotion*, 16(4), 553–566. <https://doi.org/10.1037/emo0000147>
- *Hernández, M. M., Valiente, C., Eisenberg, N., Berger, R. H., Spinrad, T. L., VanSchyndel, S. K., Silva, K. M., Southworth, J., & Thompson, M. S. (2017). Elementary students' effortful control

- and academic achievement: The mediating role of teacher–student relationship quality. *Early Childhood Research Quarterly*, 40(3rd Quarter 2017), 98–109.
<https://doi.org/10.1016/j.ecresq.2016.10.004>
- Higgins, J. P., & Altman, D. G. (2008). Assessing Risk of Bias in Included Studies. In J. P. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 187–241). John Wiley & Sons.
<https://doi.org/10.1002/9780470712184.ch8>
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.
- *Hinde, R. A., Easton, D. F., & Meller, R. E. (1984). Teacher questionnaire compared with observational data on effects of sex and sibling status on preschool behaviour. *Child Psychology & Psychiatry & Allied Disciplines*, 25(2), 285–303.
<https://doi.org/10.1111/j.1469-7610.1984.tb00150.x>
- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best Practices in the Systematic Direct Observation of Student Behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 993–1006). National Association of School Psychologists.
- *Hoerger, M. L., & Mace, F. C. (2006). A computerized test of self-control predicts classroom behavior. *Journal of applied behavior analysis*, 39(2), 147–159.
<https://doi.org/10.1901/jaba.2006.171-04>
- *Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of school psychology*, 51(5), 571–585.
<https://doi.org/10.1016/j.jsp.2013.08.001>
- *Ialongo, N. S., Domitrovich, C., Embry, D., Greenberg, M., Lawson, A., Becker, K. D., & Bradshaw, C. (2019). A randomized controlled trial of the combination of two school-based universal preventive interventions. *Developmental psychology*, 55(6), 1313–1325.
<https://doi.org/10.1037/dev0000715>
- *Imeraj, L., Antrop, I., Sonuga-Barke, E., Deboutte, D., Deschepper, E., Bal, S., & Roeyers, H. (2013). The impact of instructional context on classroom on-task behavior: A matched comparison of children with ADHD and non-ADHD classmates. *Journal of school psychology*, 51(4), 487–498.
<https://doi.org/10.1016/j.jsp.2013.05.004>
- Jackson, J. S. (1996). *A critique of measurement theory in psychology: The case of the Sensation Seeking Scale* [unpublished PhD Thesis]. Simon Fraser University.
- *Katz, L. F., & Gottman, J. M. (1997). Buffering children from marital conflict and dissolution. *Journal of clinical child psychology*, 26(2), 157–171.
https://doi.org/10.1207/s15374424jccp2602_4
- *Klein, P. S., & Feldman, R. (2007). Mothers' and caregivers' interactive and teaching behavior with toddlers. *Early Child Development and Care*, 177(4), 383–402.
<https://doi.org/10.1080/03004430600551682>
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.
- *Kochanska, G., Aksan, N., Penney, S. J., & Doobay, A. F. (2007). Early positive emotionality as a heterogenous trait: Implications for children's self-regulation. *Journal of personality and social psychology*, 93(6), 1054–1066.
<https://doi.org/10.1037/0022-3514.93.6.1054>
- *Koniak-Griffin, D., & Verzemnieks, I. (1994). Relationship between patterns of infant temperament, child behavior ratings, and interactions during toddlerhood. *Journal of Child and Adolescent Psychiatric Nursing*, 7(4), 26–37.
<https://doi.org/10.1111/j.1744-6171.1994.tb00240.x>
- *Laible, D., Panfile, T., & Makariev, D. (2008). The quality and frequency of mother–toddler conflict: Links with attachment and temperament. *Child Development*, 79(2), 426–443.
<https://doi.org/10.1111/j.1467-8624.2007.01134.x>
- *Leckman-Westin, E., Cohen, P. R., & Stueve, A. (2009). Maternal depression and mother–child interaction patterns: Association with toddler problems and continuity of effects to late childhood. *Journal of Child Psychology and Psychiatry*, 50(9), 1176–1184.
<https://doi.org/10.1111/j.1469-7610.2009.02083.x>
- *Lee, C. L., & Bates, J. E. (1985). Mother-child interaction at age two years and perceived difficult temperament. *Child development*, 56(5), 1314–1325.
<https://doi.org/10.1111/j.1467-8624.1985.tb00199.x>
- *Leflot, G., Van Lier, P. A., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of abnormal child psychology*, 38(6), 869–882.
<https://doi.org/10.1007/s10802-010-9411-4>
- Lerner, R. M. (2006). Developmental Science, Developmental Systems, and Contemporary Theories of Human Development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child*

- psychology: *Theoretical models of human development* (6th ed., pp. 1–17). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9780470147658.chpsy0101>
- *Lewis, S.N., West, A.F., Stein, A., Malmberg, L.-E., Bethell, K., Barnes, J., Sylva, K., Leach, P. and The Families, Children and Child Care (FCCC) project team (2009). A comparison of father–infant interaction between primary and non-primary care giving fathers. *Child: Care, Health and Development*, 35(2), 199–207.
<https://doi.org/10.1111/j.1365-2214.2008.00913.x>
- Lindström, E. R., Chow, J. C., Zimmerman, K. N., Zhao, H., Settanni, E., & Ellison, A. (2021). A Systematic Review and Meta-Analysis of the Relation Between Engagement and Achievement in Early Childhood Research. *Topics in Early Childhood Special Education*, 41(3), 221–235.
<https://doi.org/10.1177/02711214211032720>
- *Lohaus, A., Keller, H., Völker, S., Cappenberg, M., & Chasiotis, A. (1997). Intuitive parenting and infant behavior: Concepts, implications, and empirical validation. *The Journal of genetic psychology*, 158(3), 271–286.
<https://doi.org/10.1080/00221329709596667>
- *Low, S., Cook, C. R., Smolkowski, K., & Buntain-Ricklefs, J. (2015). Promoting social–emotional competence: An evaluation of the elementary version of Second Step®. *Journal of school psychology*, 53(6), 463–477.
<https://doi.org/10.1016/j.jsp.2015.09.002>. Epub 2015 Oct 29.
- *Lunkenheimer, E. S., Olson, S. L., Hollenstein, T., Sameroff, A. J., & Winter, C. (2011). Dyadic flexibility and positive affect in parent–child coregulation and the development of child behavior problems. *Development and psychopathology*, 23(2), 577–591.
<https://doi.org/10.1017/S095457941100006X>
- *Lytton, H., & Youngman, M. B. (1980). Natural Groups And Centroid Clusters Of Mother-Child Dyads Observed In The Home. *Multivariate Behavioral Research*, 15(4), 459–474.
https://doi.org/10.1207/s15327906mbr1504_5
- *Mammen, M. A., Moore, G. A., Scaramella, L. V., Reiss, D., Shaw, D. S., Leve, L. D., & Neiderhiser, J. M. (2016). Infant patterns of reactivity to tactile stimulation during parent-child interaction. *Infant Behavior and Development*, 44, 121–132.
<https://doi.org/10.1016/j.infbeh.2016.06.004>
- *Mäntymaa, M., Puura, K., Luoma, I., Salmelin, R. K., & Tamminen, T. (2006). Mother's early perception of her infant's difficult temperament, parenting stress and early mother–infant interaction. *Nordic journal of psychiatry*, 60(5), 379–386.
<https://doi.org/10.1080/08039480600937280>
- McClelland, M. M., Ponitz, C. C., Messersmith, E. E., & Tominey, S. (2010). Self-regulation: Integration of cognition and emotion. In W. F. Overton & R. M. Lerner (Eds.), *The handbook of life-span development, Vol. 1. Cognition, biology, and methods* (pp. 509–553). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9780470880166.hlsd001015>
- McCoy, D. C. (2019). Measuring Young Children's Executive Function and Self-Regulation in Classrooms and Other Real-World Settings. *Clinical Child and Family Psychology Review*, 22(1), 63–74. <https://doi.org/10.1007/s10567-019-00285-1>
- McDermott, R. (2011). Internal and External Validity. In J. N. Druckman, D. P. Greene, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 27–40). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511921452.003>
- *Meisinger, E. B., Schwanenflugel, P. J., Bradley, B. A., & Stahl, S. A. (2004). Interaction quality during partner reading. *Journal of Literacy Research*, 36(2), 111–140.
https://doi.org/10.1207/s15548430jlr3602_1
- *Melnick, S. M., & Hinshaw, S. P. (2000). Emotion regulation and parenting in AD/HD and comparison boys: Linkages with social behaviors and peer preference. *Journal of abnormal child psychology*, 28(1), 73–86.
<https://doi.org/10.1023/a:1005174102794>
- *Mize, J., & Cox, R. A. (1990). Social knowledge and social competence: Number and quality of strategies as predictors of peer behavior. *The Journal of Genetic Psychology*, 151(1), 117–127.
<https://doi.org/10.1080/00221325.1990.9914648>
- *Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M., & Morrison, F. J. (2016). The development of self-regulation across early childhood. *Developmental psychology*, 52(11), 1744–1762. <https://doi.org/10.1037/dev0000159>
- Nachon, J. I., Segretin, M. S., & Lipina, S. J. (2020). Conceptual and methodological approaches to the study of Self-Regulation: An inquiry within Developmental Science. *Revista Argentina de Ciencias Del Comportamiento*, 12(3), 13–31.
<https://doi.org/10.32348/1852.4206.v12.n3.25739>

- *NICHD Early Child Care Research Network. (1999). Child care and mother–child interaction in the first three years of life. *Developmental Psychology*, 35(6), 1399–1413. <https://doi.org/10.1037/0012-1649.35.6.1399>
- Nocita, G. A. (2021). *Quality of Early Childhood Education and Care: Examining Different Conceptualizations and Levels of Quality* [Thesis]. <https://tspace.library.utoronto.ca/handle/1807/108870>
- *O'Campo, P., Caughy, M. O., & Nettles, S. M. (2010). Partner abuse or violence, parenting and neighborhood influences on children's behavioral problems. *Social Science & Medicine*, 70(9), 1404–1415. <https://doi.org/10.1016/j.socscimed.2009.11.036>.
- *Olson, S. L., Bates, J. E., Sandy, J. M., & Schilling, E. M. (2002). Early developmental precursors of impulsive and inattentive behavior: From infancy to middle childhood. *Journal of Child Psychology and Psychiatry*, 43(4), 435–447. <https://doi.org/10.1111/1469-7610.00035>.
- *Olson, S. L., Lopez-Duran, N., Lunkenheimer, E. S., Chang, H., & Sameroff, A. J. (2011). Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and psychopathology*, 23(1), 253–266. <https://doi.org/10.1017/S0954579410000775>.
- Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the “same” construct? A meta-analytic investigation. *Personality and Individual Differences*, 49(7), 669–676. <https://doi.org/10.1016/j.paid.2010.06.014>
- *Parade, S. H., Wong, K., Belair, R., Dickstein, S., & Seifer, R. (2019). Infant sleep moderates the effect of infant temperament on maternal depressive symptoms, maternal sensitivity, and family functioning. *Infant Behavior and Development*, 57, Article 101343. <https://doi.org/10.1016/j.infbeh.2019.101343>.
- *Pauli-Pott, U., Mertesacker, B., & Beckmann, D. (2005). Comparing assessment methods of infant emotionality. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 33(2), 123–135. <https://doi.org/10.1024/1422-4917.33.2.123>.
- *Pellegrini, A. D., Galda, L., & Flor, D. (1997). Relationships, individual differences, and children's use of literate language. *British Journal of Educational Psychology*, 67(2), 139–152. <https://doi.org/10.1111/j.2044-8279.1997.tb01233.x>.
- Pigott, T. D. (2001). A Review of Methods for Missing Data. *Educational Research and Evaluation*, 7(4), 353–383. <https://doi.org/10.1076/edre.7.4.353.8937>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, 63, 539–569. <https://doi.org/10.1037/0021-9010.88.5.879>
- Price, K. J., Joschko, M., & Kerns, K. (2003). The Ecological Validity of Pediatric Neuropsychological Tests of Attention. *The Clinical Neuropsychologist*, 17(2), 170–181. <https://doi.org/10.1076/clin.17.2.170.16506>
- *Puura, K., Leppänen, J., Salmelin, R., Mäntymaa, M., Luoma, I., Latva, R., Peltola, M., Lehtimäki, T., & Tamminen, T. (2019). Maternal and infant characteristics connected to shared pleasure in dyadic interaction. *Infant Mental Health Journal*, 40(4), 459–478. <https://doi.org/10.1002/imhj.21786>.
- *Qi, C. H., & Kaiser, A. P. (2004). Problem behaviors of low-income children with language delays. *Journal of Speech, Language, and Hearing Research*, 47(3), 595–609. [https://doi.org/10.1044/1092-4388\(2004\)046](https://doi.org/10.1044/1092-4388(2004)046).
- Ramakrishnan, A., Ottmar, E., LoCasale-Crouch, J., & Whitehill, J. (14-18 May 2019). *Toward Automated Classroom Observation: Predicting Positive and Negative Climate* [Conference]. 14th IEEE International Conference on Automatic Face Gesture Recognition, Lille, France. <https://doi.org/10.1109/FG.2019.8756529>
- *Ramos-Martínez, K., Alfaro-Zolá, L. M., Tirado-Amador, L. R., & González-Martínez, F. (2015). Social skills and behavior in children during the dental appointment in a school teaching and dental assistance, Cartagena (Colombia). *Revista Facultad de Odontología Universidad de Antioquia*, 27(1), 86–107. <https://doi.org/10.17533/udea.rfo.v27n1a5>
- *Rickel, A. U., & Burgio, J. C. (1982). Assessing social competencies in lower income preschool children. *American Journal of Community Psychology*, 10(6), 635–645. <https://doi.org/10.1007/BF01312596>
- *Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental psychology*, 45(4), 958. <https://doi.org/10.1037/a0015861>.
- Rose-Krasnor, L., & Denham, S. (2009). Social-emotional competence in early childhood. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 162–179). The Guilford Press.
- Rothbart, M. K., & Bates, J. E. (2006). Temperament.

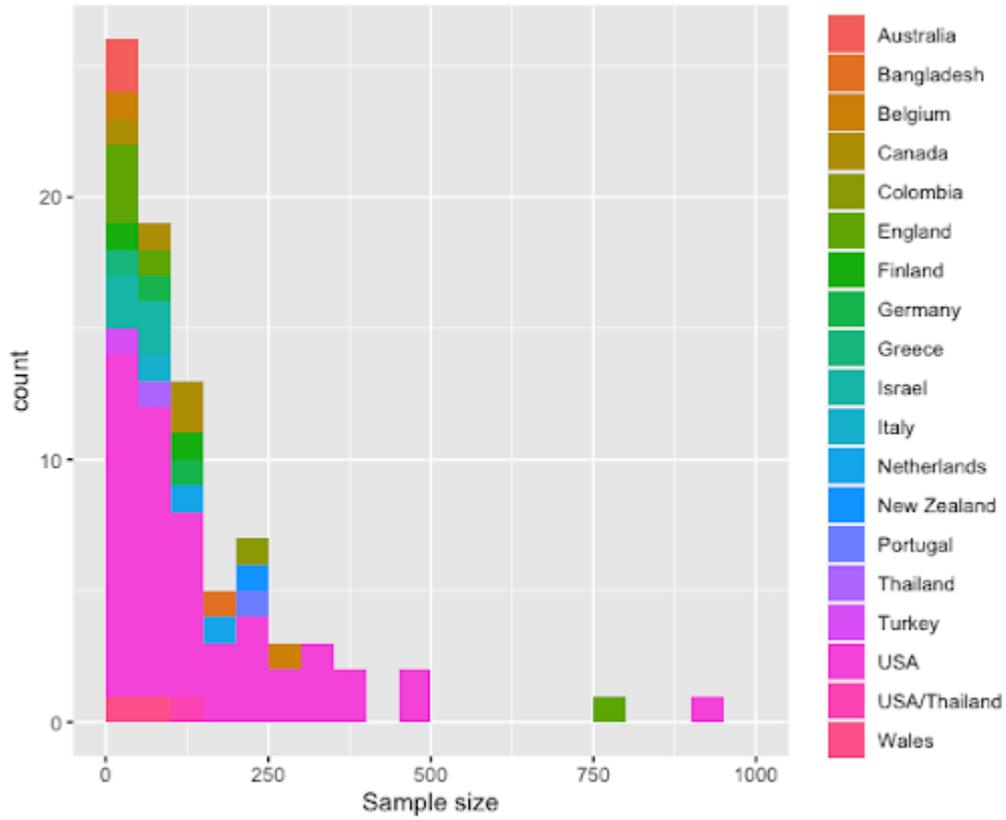
- In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (6th ed., pp. 99–166). John Wiley & Sons.
- *Sabourin Ward, C., & Haskett, M. E. (2008). Exploration and validation of clusters of physically abused children. *Child Abuse & Neglect*, 32(5), 577-588. <https://doi.org/10.1016/j.chiabu.2007.07.012>.
- *Sanders, M. R., Turner, K. M., Wall, C. R., Waugh, L. M., & Tully, L. A. (1997). Mealtime behavior and parent-child interaction: A comparison of children with feeding problems, and nonclinic controls. *Journal of pediatric Psychology*, 22(6), 881-900. <https://doi.org/10.1093/jpepsy/22.6.881>.
- Santos, J. A., Giovannetti, F., Smulski, M. C., Hermida, M. J., Petetta, D. R., Segretin, M. S., & Lipina, S. (2024). A systematic review of the concept of self-regulation in infants between 0 and 36 months in Latin America. *Infant Behavior and Development*, 75, Article 101954. <https://doi.org/10.1016/j.infbeh.2024.101954>
- *Scope, A., Empson, J., & McHale, S. (2010). Executive function in children with high and low attentional skills: Correspondences between behavioural and cognitive profiles. *British Journal of Developmental Psychology*, 28(2), 293-305. <https://doi.org/10.1348/026151009x410371>.
- Silver, C. H. (2014). Sources of data about children's executive functioning: Review and commentary. *Child Neuropsychology*, 20(1), 1–13. <https://doi.org/10.1080/09297049.2012.727793>
- *Slagt, M., Dubas, J. S., Ellis, B. J., van Aken, M. A., & Deković, M. (2019). Linking emotional reactivity “for better and for worse” to differential susceptibility to parenting among kindergartners. *Development and psychopathology*, 31(2), 741-758. <https://doi.org/10.1017/S0954579418000445>.
- *Smeekens, S., Marianne Riksen-Walraven, J., & Van Bakel, H. J. (2007). Cortisol reactions in five-year-olds to parent–child interaction: the moderating role of ego-resiliency. *Journal of Child Psychology and Psychiatry*, 48(7), 649-656. <https://doi.org/10.1111/j.1469-7610.2007.01753.x>.
- *Snyder, J., Cramer, A., A Frank, J., & Patterson, G. R. (2005). The contributions of ineffective discipline and parental hostile attributions of child misbehavior to the development of conduct problems at home and school. *Developmental psychology*, 41(1), 30. <https://doi.org/10.1037/0012-1649.41.1.30>.
- *Spence, S. H., Donovan, C., & Brechman-Toussaint, M. (1999). Social skills, social outcomes, and cognitive features of childhood social phobia. *Journal of abnormal psychology*, 108(2), 211-221. <https://doi.org/10.1037//0021-843x.108.2.211>.
- *Stifter, C. A., Willoughby, M. T., & Towe-Goodman, N. (2008). Agree or disagree to disagree? Assessing the convergence between parents and observers on infant temperament. *Infant and Child Development: An International Journal of Research and Practice*, 17(4), 407-426. <https://doi.org/10.1002/icd.584>
- *Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: ready to learn but further to go. *Developmental psychology*, 33(4), 711-723. <https://doi.org/10.1037//0012-1649.33.4.711>.
- *Susman, E. J., Huston-Stein, A., & Friedrich-Cofer, L. (1980). Relation of conceptual tempo to social behaviors of head start children. *The Journal of Genetic Psychology*, 137(1), 17-20. <https://doi.org/10.1080/00221325.1980.10532798>.
- *Szabo-Reed, A. N., Willis, E. A., Lee, J., Hillman, C. H., Washburn, R. A., & Donnelly, J. E. (2019). The influence of classroom physical activity participation and time on task on academic achievement. *Translational Journal of the American College of Sports Medicine*, 4(12), 84-95. <https://doi.org/10.1249/TJX.000000000000087>.
- Thompson, P. L., & Brassard, M. R. (1984). Validity of the Woodcock-Johnson tests of cognitive ability: A comparison with the WISC-R in LD and normal elementary students. *Journal of School Psychology*, 22(2), 201–208. [https://doi.org/10.1016/0022-4405\(84\)90040-2](https://doi.org/10.1016/0022-4405(84)90040-2)
- *Tulananda, O., & Roopnarine, J. L. (2001). Mothers' and fathers' interactions with preschoolers in the home in northern Thailand: Relationships to teachers' assessments of children's social skills. *Journal of Family Psychology*, 15(4), 676–687. <https://doi.org/10.1037/0893-3200.15.4.676>
- Valsiner, J. (2007). Becoming Integrative in Science: Re-building Contemporary Psychology through Interdisciplinary and International Collaboration. *Integrative Psychological and Behavioral Science*, 41(1), 1–5. <https://doi.org/10.1007/s12124-007-9002-2>
- *Vihman, M. M., & Greenlee, M. (1987). Individual differences in phonological development: Ages one and three years. *Journal of Speech, Language, and Hearing Research*, 30(4), 503-521. <https://doi.org/10.1044/jshr.3004.503>.
- Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing Students in Classroom Settings: A Review of Seven Coding Schemes. *School Psychology Review*, 34(4), 454–474. <https://doi.org/10.1080/02796015.2005.12088009>

- *Vorraia, P., Papaligoura, Z., Dunn, J., Van IJzendoorn, M. H., Steele, H., Kontopoulou, A., & Sarafidou, Y. (2003). Early experiences and attachment relationships of Greek infants raised in residential group care. *Journal of child psychology and psychiatry*, *44*(8), 1208-1220. <https://doi.org/10.1111/1469-7610.00202>.
- Vostanis, P. (2006). Strengths and Difficulties Questionnaire: Research and clinical applications. *Current Opinion in Psychiatry*, *19*(4), 367–372. <https://doi.org/10.1097/01.yco.0000228755.72366.05>
- Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in cognitive sciences*, *24*(4), 329-342. <https://doi.org/10.1016/j.tics.2020.01.006>
- *Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental psychology*, *35*(2), 386-391. <https://doi.org/10.1037//0012-1649.35.2.386>.
- *Webster-Stratton, C., Jamila Reid, M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: evaluation of the incredible years teacher and child training programs in high-risk schools. *Journal of child psychology and psychiatry*, *49*(5), 471-488. <https://doi.org/10.1111/j.1469-7610.2007.01861.x>.
- *Weisz, J. R., Chaiyasit, W., Weiss, B., Eastman, K. L., & Jackson, E. W. (1995). A multimethod study of problem behavior among Thai and American children in school: Teacher reports versus direct observations. *Child Development*, *66*(2), 402-415. <https://doi.org/10.1111/j.1467-8624.1995.tb00879.x>.
- *Williford, A. P., Vick Whittaker, J. E., Vitiello, V. E., & Downer, J. T. (2013). Children's engagement within the preschool classroom and their development of self-regulation. *Early Education & Development*, *24*(2), 162-187. <https://doi.org/10.1080/10409289.2011.628270>.
- *Winsler, A., De Leon, J. R., Wallace, B. A., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behaviour. *Journal of Child Language*, *30*(3), 583-608. <https://doi.org/10.1017/S0305000903005671>
- Woolverton, G. A., & Pollastri, A. R. (2021). An Exploration and Critical Examination of How “Intelligent Classroom Technologies” Can Improve Specific Uses of Direct Student Behavior Observation Methods. *Educational Measurement: Issues and Practice*, *40*(3), 7–17. <https://doi.org/10.1111/emip.12421>
- World Health Organization. (2012). Early childhood development and disability: A discussion paper. <https://iris.who.int/bitstream/handle/10665/75355/?sequence=1>
- *Zentall, S. S. (1980). Behavioral comparisons of hyperactive and normally active children in natural settings. *Journal of Abnormal Child Psychology*, *8*(1), 93-109. <https://doi.org/10.1007/BF00918164>.

Supplementary Material

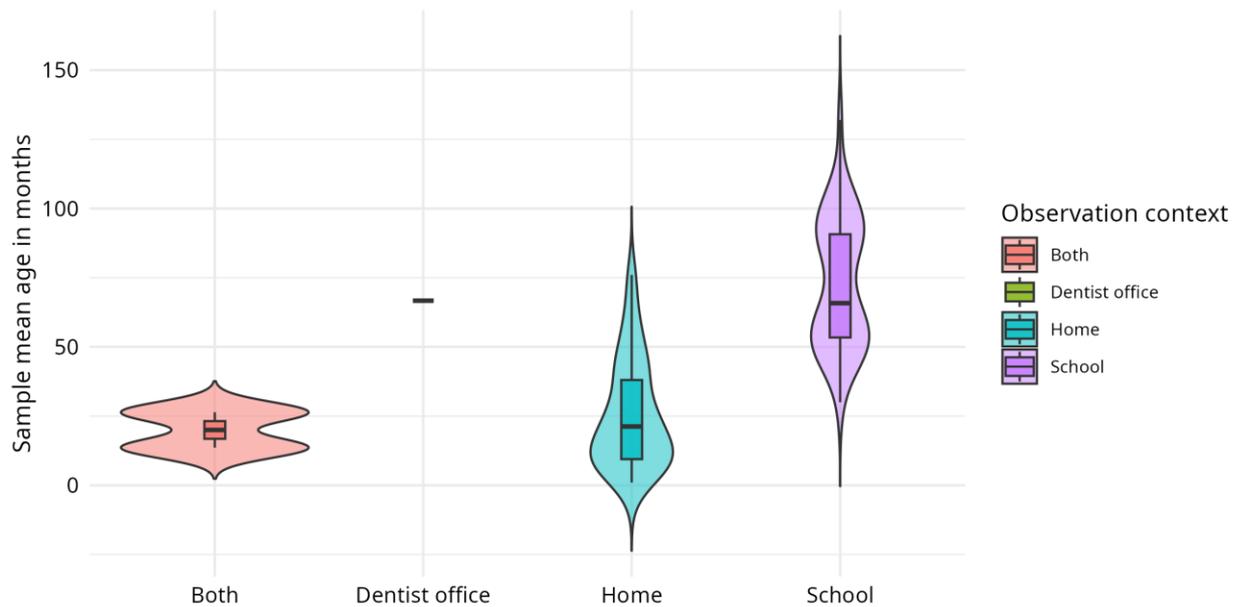
Supplementary Figure 1.

Sample sizes and their geographical origin

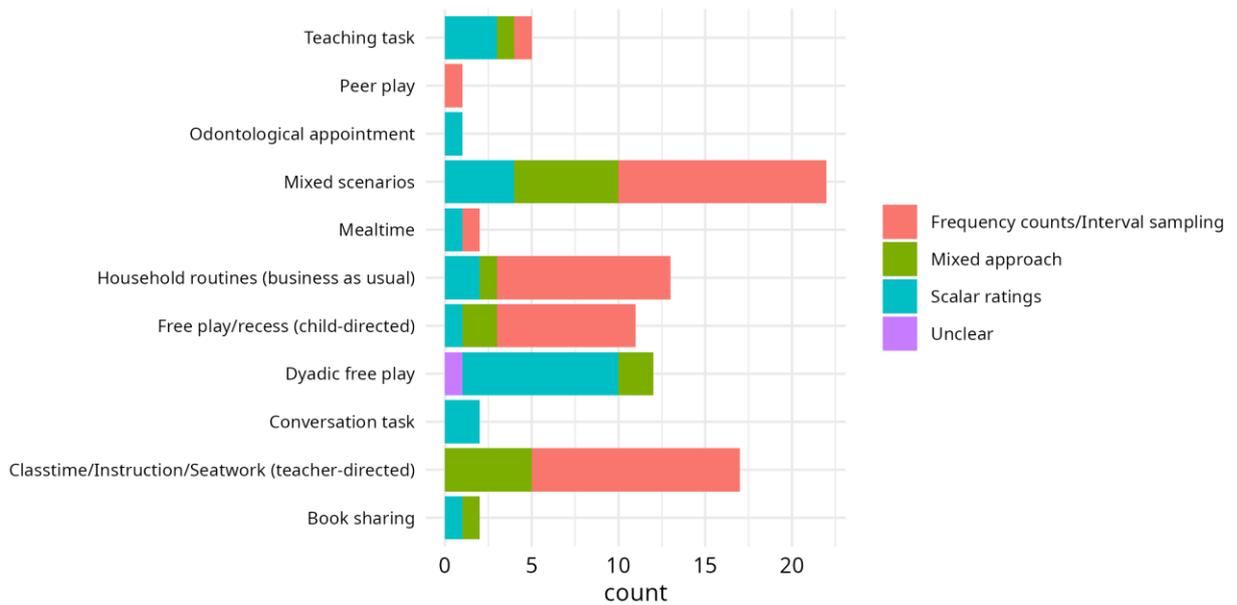


Supplementary Figure 2.

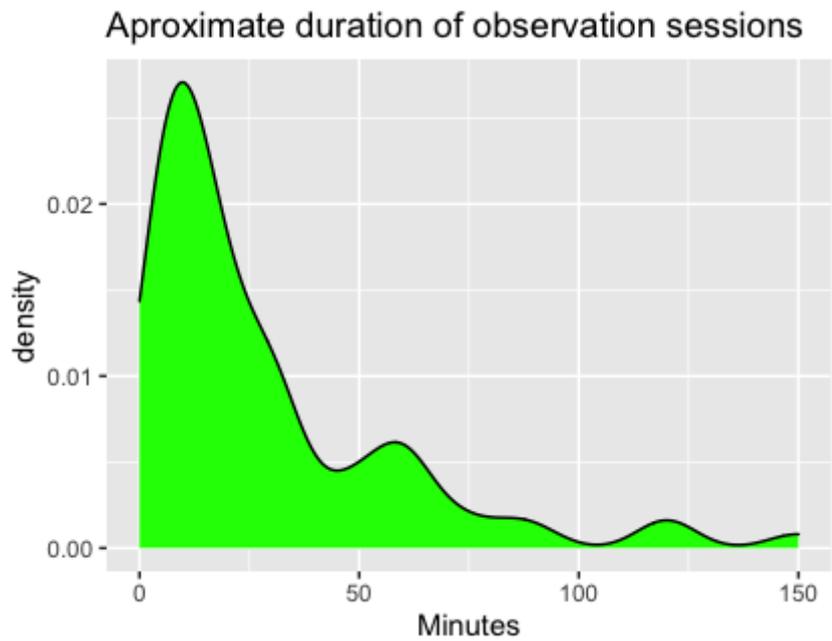
Sample mean age by observation context



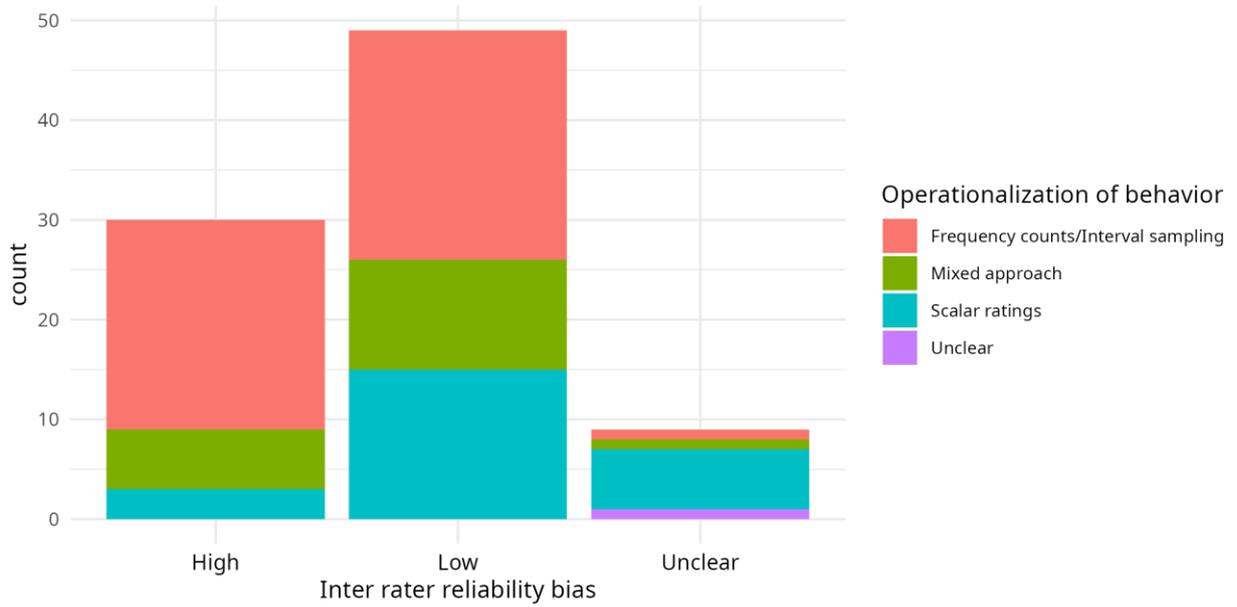
Supplementary Figure 3.
Observed activities by operationalization of behavior



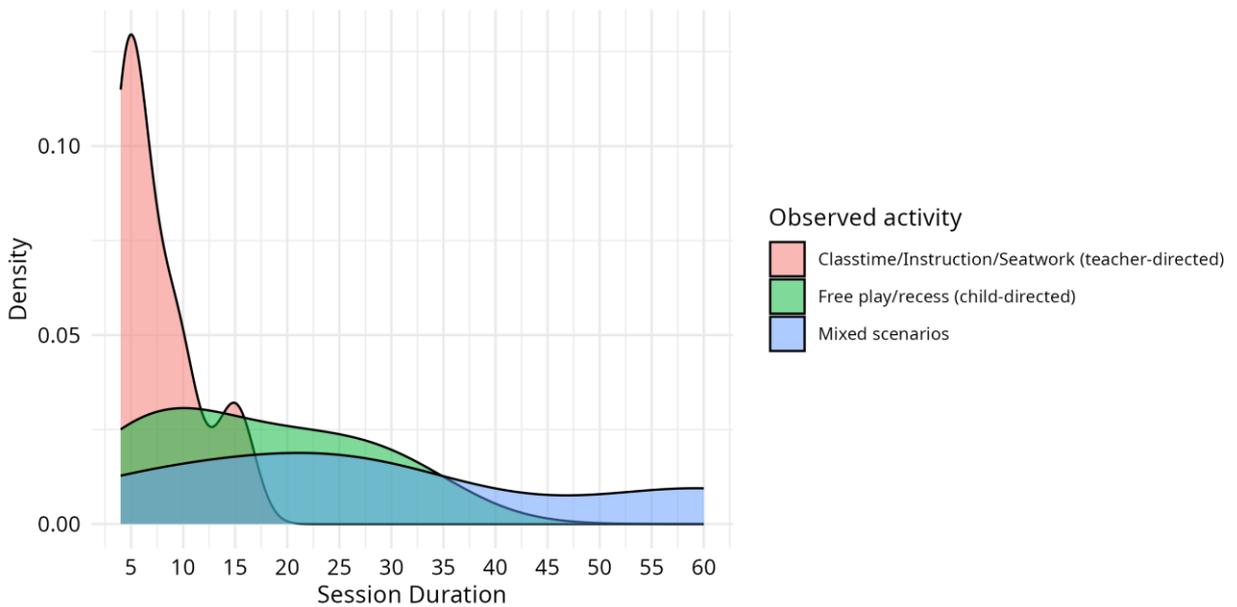
Supplementary Figure 4.
Probability distribution of observational session duration



Supplementary Figure 5.
Inter-rater bias distribution grouped by operationalization type



Supplementary Figure 6.
Probability distribution of observational session duration for studies that measured disruptive-aggressive behavior in schools



Supplementary Table 1.
Main characteristics of reviewed studies

Study ID	Observation context	Sample size	Mean age at baseline	Country	SES	Design	Observed activity	Operationalization of behavior	Approximate duration of observational	Correlations reported
Aboud FE 2007	Home	159	39.1	Bangladesh	Low SES	Cross-sectional	Conversation task	Scalar ratings	5	No
Akdemir D 2010	Home	16	46	Turkey	Unclear	Longitudinal	Dyadic free play	Unclear	30	No
Apter-Levi Y 2016	Home	91	76	Israel	Mid/High SES	Longitudinal	Dyadic free play	Scalar ratings	10	Yes
Arevalo A 2014	Home	52	11.3	USA	Diverse sample	Longitudinal	Teaching task	Scalar ratings		No
Assel MA 2002	Home	112	48	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	70	No
Baiao R 2018	Home	220	58.2	Portugal	Diverse sample	Cross-sectional	Mixed scenarios	Scalar ratings	10	Yes
Belsky J 1996	Home	69	15	USA	Unclear	Longitudinal	Household routines (business as usual)	Mixed approach	120	No
Berger RH 2017; Hernández MM 2016; Hernández	School	301	65.8	USA	Diverse sample	Longitudinal	Mixed scenarios	Mixed approach	30	Yes
Bernier A 2012	Home	62	12.7	Canada	Mid/High SES	Longitudinal	Dyadic free play	Scalar ratings	10	Yes
Bocknek EL 2012	Home	1474	37.5	USA	Low SES	Longitudinal	Dyadic free play	Scalar ratings	10	Yes
Brandes-Aitken A 2019; Gueron-Sela N 2018;	Home	1292	7	USA	Low SES	Longitudinal	Book sharing	Mixed approach	10	Yes
Buckley ME 2002	Home	100	43.2	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Scalar ratings	152	No
Cerniglia L 2014	Home	77	24	Italy	Mid/High SES	Cross-sectional	Mealtime	Scalar ratings	20	No

Chuang CC 2020	School	1817	78	USA	Diverse sample	Cross-sectional	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5	No
Coplan RJ 2004	School	119	48.6	Canada	Diverse sample	Cross-sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	20	Yes
Crittenden 1988	PM Home	25	18.5	USA	Low SES	Cross-sectional	Dyadic free play	Scalar ratings	3	No
Crittenden 1992	PM Home	30	24	USA	Low SES	Cross-sectional	Dyadic free play	Scalar ratings	3	No
Day SL 2015	School	500	87.6	USA	Diverse sample	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	120	No
DeLuzio J 2011	School	12	48.7	Canada	Unclear	Cross-sectional	Free play/recess (child-directed)	Mixed approach	20	No
Dollberg D 2006	Home	43	24	Israel	Unclear	Cross-sectional	Mixed scenarios	Scalar ratings	20	No
Duggan AK 1999	Home	270	12	USA	Low SES	Longitudinal	Teaching task	Scalar ratings		No
Dunn J 1980	Home	41	25	England	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	60	No
DuPaul GJ 1993	School	25	102.7	USA	Low SES	Cross-sectional	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	15	No
DuPaul GJ 1998	School	10	92.4	USA	Low SES	Cross-sectional	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	12	No
DuPaul GJ 2001	School	36	49.7	USA	Mid/High SES	Cross-sectional	Mixed scenarios	Frequency counts/Interval sampling	60	No
Eisenberg N 1998	School	82	55.3	USA	Diverse sample	Longitudinal	Free play/recess (child-directed)	Frequency counts/Interval sampling	10	No
Gibson J 2011	School	44	105.6	England	Unclear	Cross-sectional	Free play/recess (child-directed)	Scalar ratings	10	No
Gouze KR 1987	School	43	53.9	USA	Diverse sample	Cross-sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	50	No

Grossman 1997	DC	School	372	99.6	USA	Diverse sample	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	5	No
Hane AA 2006		Home	59	9.5	USA	Unclear	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	34	Yes
Harel-Gadassi 2020	A	Home	39	18	Israel	Mid/High SES	Longitudinal	Dyadic free play	Scalar ratings	5	No
Hatch B 2019		Home	209	12	New Zealand	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	480	No
Herman KC 2017		School	917	84	USA	Diverse sample	Longitudinal	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5	No
Hinde RA 1984		School	762	30	England	Diverse sample	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	85	No
Hoerger ML 2006		School	15	89	Wales	Unclear	Cross-sectional	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	8	No
Hutchings J 2013		School	54	56.8	Wales	Diverse sample	Longitudinal	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	15	No
Ialongo NS 2019		School	2055	90	USA	Low SES	Longitudinal	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5	No
Imeraj L 2013		School	31	96	Belgium	Unclear	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	274	No
Katz LF 1997		Home	56	60	USA	Unclear	Longitudinal	Peer play	Frequency counts/Interval sampling	30	Yes
Klein PS 2007		Both	90	26.4	Israel	Mid/High SES	Cross-sectional	Dyadic free play	Scalar ratings	60	Yes
Kochanska 2007 (a)	G	Home	112	9	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	60	No
Kochanska 2007 (b)	G	Home	102	7	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	60	No
Koniak-Griffin 1994	D	Home	81	24	USA	Mid/High SES	Longitudinal	Teaching task	Scalar ratings		Yes

Laible D 2008	Home	64	30	USA	Unclear	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	90	Yes
Leckman-Westin E 2009	Home	215	24	USA	Unclear	Longitudinal	Book sharing	Scalar ratings		No
Lee CL 1985	Home	111	24	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Frequency counts/Interval sampling	150	Yes
Leflot G 2010	School	285	89	Belgium	Unclear	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	2	No
Lewis SN 2009	Home	100	11.9	England	Diverse sample	Longitudinal	Mixed scenarios	Mixed approach	25	No
Lohaus A 1997	Home	62	3	Germany	Mid/High SES	Cross-sectional	Dyadic free play	Mixed approach	15	No
Low S 2015	School	3587	46.5	USA	Diverse sample	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	4	No
Lunkenheimer ES 2011	Home	167	37.7	USA	Mid/High SES	Longitudinal	Teaching task	Mixed approach	6	Yes
Lytton H 1980	Home	136	32.4	Canada	Diverse sample	Cross-sectional	Household routines (business as usual)	Frequency counts/Interval sampling	180	No
Mammen 2016	MA Home	497	9.3	USA	Mid/High SES	Longitudinal	Teaching task	Frequency counts/Interval sampling		Yes
Mäntymaa 2006	M Home	33	1	Finland	Diverse sample	Longitudinal	Dyadic free play	Scalar ratings	5	No
Meisinger 2004	EB School	86	96	USA	Low SES	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Mixed approach	13	Yes
Melnick SM 2000	School	34	109	USA	Diverse sample	Cross-sectional	Mixed scenarios	Frequency counts/Interval sampling		No
Mize J 1990	School	32	56.5	USA	Mid/High SES	Cross-sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	2	Yes
NICHD 1999; Huston AC 2005	Home	1260	6	USA	Diverse sample	Longitudinal	Dyadic free play	Scalar ratings	15	No

O'Campo P 2010	Home	383	72	USA	Diverse sample	Cross-sectional	Conversation task	Scalar ratings		Yes
Olson SL 2002	Home	168	6	USA	Diverse sample	Longitudinal	Household routines (business as usual)	Scalar ratings	180	No
Olson SL 2011	School	240	41.4	USA	Diverse sample	Longitudinal	Free play/recess (child-directed)	Frequency counts/Interval sampling	30	No
Parade SH 2019	Home	131	8	USA	Mid/High SES	Longitudinal	Mixed scenarios	Scalar ratings	45	Yes
Pauli-Pott U 2005	Home	101	12	Germany	Unclear	Longitudinal	Mixed scenarios	Mixed approach	50	Yes
Pellegrini AD 1997	School	56	65	USA	Unclear	Cross-sectional	Mixed scenarios	Frequency counts/Interval sampling	20	No
Puura K 2019	Home	113	3.7	Finland	Unclear	Longitudinal	Dyadic free play	Mixed approach	15	No
Qi CH 2004	School	28	53.4	USA	Low SES	Cross-sectional	Mixed scenarios	Frequency counts/Interval sampling	20	No
Ramos-Martínez 2015	Dentist office	250	66.7	Colombia	Diverse sample	Cross-sectional	Odontological appointment	Scalar ratings		No
Rickel AU 1982	School	110	51	USA	Low SES	Cross-sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	20	No
Rimm-Kaufman SE 2009	School	172	64.9	USA	Low SES	Cross-sectional	Classroom/Instruction/Seatwork (teacher-directed)	Mixed approach	15	Yes
Sabourin Ward C 2008	School	77	88	USA	Diverse sample	Cross-sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	30	No
Sanders MR 1997	Home	20	43.8	Australia	Unclear	Cross-sectional	Mealtime	Frequency counts/Interval sampling		No
Scope A 2010	School	24	102	England	Unclear	Cross-sectional	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	8	No
Slagt M 2019	Home	185	56.4	Netherlands	Mid/High SES	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	10	No

Smeekens S 2007; van Bakel HJ 2002	Home	124	15	Netherlands	Unclear	Longitudinal	Mixed scenarios	Scalar ratings	11	No
Snyder J 2005	School	266	66	USA	Low SES	Longitudinal	Free play/recess (child-directed)	Mixed approach	5	No
Spence SH 1999	School	27	132	Australia	Mid/High SES	Cross- sectional	Mixed scenarios	Frequency counts/Interval sampling	75	No
Stipek DJ 1997	School	223	58	USA	Diverse sample	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Mixed approach	32	No
Susman EJ 1980	School	121	50	USA	Low SES	Cross- sectional	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	50	No
Szabo-Reed AN 2019	School	314	90.7	USA	Diverse sample	Longitudinal	Classtime/Instruction/Seatwork (teacher-directed)	Mixed approach	20	No
Tulananda O 2001	Home	53	46	Thailand	Mid/High SES	Cross- sectional	Household routines (business as usual)	Frequency counts/Interval sampling	30	No
Vihman MM 1987	Home	10	9	USA	Mid/High SES	Longitudinal	Mixed scenarios	Mixed approach	30	No
Vorria P 2003	Both	41	13.6	Greece	Unclear	Cross- sectional	Household routines (business as usual)	Frequency counts/Interval sampling	40	No
Watson AC 1999	School	26	63	USA	Mid/High SES	Cross- sectional	Free play/recess (child-directed)	Frequency counts/Interval sampling	5	Yes
Webster-Stratton C 2008	School	216	54	USA	Diverse sample	Longitudinal	Mixed scenarios	Mixed approach	30	No
Weisz JR 1995	School	144	97.6	USA/Thailand	Unclear	Cross- sectional	Classtime/Instruction/Seatwork (teacher-directed)	Mixed approach	10	No
Williford AP 2013	School	341	46.9	USA	Low SES	Longitudinal	Mixed scenarios	Mixed approach	60	Yes
Winsler A 2003	School	32	49	USA	Diverse sample	Longitudinal	Mixed scenarios	Frequency counts/Interval sampling	10	Yes
Zentall SS 1980	School	31	88	USA	Mid/High SES	Cross- sectional	Mixed scenarios	Frequency counts/Interval sampling	10	No

Supplementary Table 2.*Number of studies (k) that reported on specific individual and contextual covariates*

Covariate	k
Child's age	76
Child's sex	75
Child's ethnicity	51
Mother's highest level of completed education	48
Mother's age	31
Father's highest level of completed education	29
Family's income	24
Mother's employment status	23
Household composition (e.g., mono/biparental)	19
Father's employment status	18
Father's age	16
Parental marital status	15
Child's birth order	14
Child qualifies for free or reduced meals at school	11
Household's Hollingshead SES index	11
Number of siblings	11
Number of children per classroom	9
Teacher's years of experience	9
Teacher's ethnicity	8
Teacher's highest level of completed education	8
Teacher's sex	7
Household's primary language	6
Child's weight	4
Child's weight at birth	4
Head Start, state or privately funded institution	4
Low income household	4
Teacher's age	4
Child's height	3
Child's nationality	3
Child's gestational age	3
Bilingual or monolingual classroom instruction	3
Child assisted preschool enrollment	2
Child's ethnic minority status	2
Classroom's ethnic composition	2
Paternal income distribution	2
Household's SES (other)	2
Mother's age at delivery	2
Mother's Beck Depression Index (BDI; Beck et al, 1961)	2
Center for Epidemiological Studies Depressive Symptoms (CES-D) short form	2
Welfare recipient household	2
Home Observation for Measurement of the Environment Inventory (HOME)	2
Maternal Symptom Checklist	2
Someone in family has problems with substance abuse	2
Someone in family has problems with police	2
Maternal depression symptoms (other)	2

Supplementary Table 3.
Study-level risk of bias judgements

Study	Participant sampling bias	Assessor bias	Inter-rater reliability bias	Attrition bias
Aboud FE 2007	High	Unclear	Unclear	Low
Akdemir D 2010	High	Low	Unclear	Low
Apter-Levi Y 2016	High	Unclear	Low	Low
Arevalo A 2014	High	Unclear	Unclear	High
Assel MA 2002	Unclear	Low	Low	Low
Baiao R 2018	High	Low	Low	Low
Belsky J 1996	High	Low	High	Low
Berger RH 2017; Hernández MM 2016; Hernández MM 2017	High	Low	Low	High
Bernier A 2012	Low	Unclear	High	Low
Bocknek EL 2012	Unclear	Low	Low	Unclear
Brandes-Aitken A 2019; Gueron-Sela N 2018; Stifter CA 2008	Low	Unclear	Unclear	Low
Buckley ME 2002	Low	Low	Low	Low
Cerniglia L 2014	High	Unclear	Unclear	Low
Chuang CC 2020	High	Low	Low	Low
Coplan RJ 2004	High	Unclear	Low	Unclear
Crittenden PM 1988	High	Low	Low	Low
Crittenden PM 1992	High	Low	Low	Low
Day SL 2015	High	Low	Low	High
DeLuzio J 2011	High	Low	Low	Low
Dollberg D 2006	High	Low	Low	Low
Duggan AK 1999	Unclear	Unclear	Unclear	High
Dunn J 1980	High	Unclear	High	Low
DuPaul GJ 1993	High	Low	Low	Unclear
DuPaul GJ 1998	High	Unclear	Low	Unclear
DuPaul GJ 2001	High	Low	Low	Low
Eisenberg N 1998	High	Unclear	High	High
Gibson J 2011	High	Unclear	High	High
Gouze KR 1987	High	Low	High	Low
Grossman DC 1997	Low	Low	High	Unclear
Hane AA 2006	High	Low	Low	High
Harel-Gadassi A 2020	High	Unclear	Low	Low
Hatch B 2019	High	Low	Low	Low
Herman KC 2017	High	Unclear	Low	Low
Hinde RA 1984	High	Unclear	High	Unclear
Hoerger ML 2006	High	Unclear	Low	Unclear
Hutchings J 2013	High	Low	High	Low
Ialongo NS 2019	Low	Unclear	Low	Low
Imeraj L 2013	High	Low	Low	Unclear
Katz LF 1997	High	Unclear	High	Unclear
Klein PS 2007	Low	Unclear	Low	Low
Kochanska G 2007 (a)	High	Unclear	High	High
Kochanska G 2007 (b)	High	Unclear	High	Low
Koniak-Griffin D 1994	High	Low	Low	Low
Laible D 2008	High	Unclear	High	Unclear
Leckman-Westin E 2009	Low	Unclear	Unclear	High
Lee CL 1985	Unclear	Low	High	Unclear

Leflot G 2010	High	Unclear	Low	High
Lewis SN 2009	High	Low	Low	Unclear
Lohaus A 1997	High	Unclear	High	Low
Low S 2015	Low	Unclear	High	Low
Lunkenheimer ES 2011	High	Unclear	Low	Unclear
Lytton H 1980	High	Unclear	High	Low
Mammen MA 2016	Unclear	Unclear	Low	High
Mäntymaa M 2006	High	Low	Low	Low
Meisinger EB 2004	High	Unclear	Low	Unclear
Melnick SM 2000	High	Low	High	Unclear
Mize J 1990	High	Unclear	Low	Low
NICHD 1999; Huston AC 2005	Low	Low	High	High
O'Campo P 2010	Low	Unclear	Low	Unclear
Olson SL 2002	High	Unclear	Low	Low
Olson SL 2011	High	Unclear	Low	Low
Parade SH 2019	High	Low	Low	Low
Pauli-Pott U 2005	High	Unclear	Low	Low
Pellegrini AD 1997	High	Low	High	Low
Puura K 2019	Low	Unclear	Low	Low
Qi CH 2004	High	Unclear	Low	Low
Ramos-Martínez 2015	High	Unclear	Unclear	Low
Rickel AU 1982	High	Unclear	Low	Unclear
Rimm-Kaufman SE 2009	High	Low	Low	Unclear
Sabourin Ward C 2008	High	Low	Low	High
Sanders MR 1997	High	Unclear	High	Low
Scope A 2010	High	Unclear	Unclear	Low
Slagt M 2019	High	Low	High	Low
Smeekens S 2007; van Bakel HJ 2002	Low	Unclear	Low	Unclear
Snyder J 2005	Unclear	Unclear	High	Low
Spence SH 1999	High	Unclear	Low	Low
Stipek DJ 1997	High	Unclear	High	Unclear
Susman EJ 1980	High	Unclear	High	Low
Szabo-Reed AN 2019	Low	Unclear	Low	Unclear
Tulananda O 2001	High	Unclear	High	Low
Vihman MM 1987	High	Unclear	Low	Low
Vorria P 2003	High	Unclear	High	Low
Watson AC 1999	High	Unclear	Low	High
Webster-Stratton C 2008	Unclear	Low	Low	Low
Weisz JR 1995	High	Unclear	High	Low
Williford AP 2013	High	Unclear	High	High
Winsler A 2003	High	Low	Low	Unclear
Zentall SS 1980	High	Low	High	Low

Supplementary Table 4.

Number of studies (k) that reported associations between primary and secondary outcomes at the conceptual level

Interactions between primary and secondary outcomes	k
School observed disruptive/aggressive behavior and teacher report of externalizing problems	15
School observed disruptive/aggressive behavior and tests of language development/literacy	12
School observed disruptive/aggressive behavior and teacher report of prosocial skills	11
Home observed negative affect behavior and parental report of negative affect	9
Home observed engagement behavior and parental report of negative affect	9
School observed off-task behavior and tests of language development/literacy	9
School observed off-task behavior and teacher report of externalizing problems	8
School observed off-task behavior and tests of math skills	8
School observed disruptive/aggressive behavior and teacher report of peer aggression	8
School observed disruptive/aggressive behavior and tests of math skills	8
Home observed negative affect behavior and parent report of activity level	7
School observed on-task engagement behavior and tests of language development/literacy	7
School observed disruptive/aggressive behavior and teacher report of internalizing problems	7
School observed disruptive/aggressive behavior and teacher report of hyperactivity/impulsivity	7
School observed cooperative prosocial behavior and teacher report of externalizing problems	6
School observed on-task engagement behavior and teacher report of externalizing problems	6
Home observed positive affect behavior and parental report of negative affect	6
Home observed negative affect behavior and parental report of surgency/positive affect	6
Home observed engagement behavior and parental report of surgency/positive affect	6
School observed disruptive/aggressive behavior and tests of intelligence	6
School observed social initiation behavior and teacher report of prosocial skills	5
Home observed affective synchronicity behavior and parental report of negative affect	5
School observed on-task engagement behavior and teacher report of prosocial skills	5
Home observed negative affect behavior and parental report of externalizing problems	5
Home observed negative affect behavior and parental report of approach behavior	5
Home observed engagement behavior and parental report of externalizing problems	5
Home observed non-compliant/ disruptive behavior and parental report of negative affect	5
Home observed non-compliant/ disruptive behavior and parental report of surgency/positive affect	5
Home observed non-compliant/ disruptive behavior and tests of intelligence	5
Home observed non-compliant/ disruptive behavior and tests of language development/literacy	5
Home observed social withdrawal/avoidance behavior and parental report of negative affect	5
School observed off-task behavior and teacher report of hyperactivity/impulsivity	5
School observed disruptive/aggressive behavior and parental report of externalizing problems	5
School observed disruptive/aggressive behavior and tests of social skills	5
School observed cooperative prosocial behavior and teacher report of prosocial skills	4
School observed cooperative prosocial behavior and tests of social skills	4
School observed cooperative prosocial behavior and tests of language development/literacy	4
School observed social initiation behavior and tests of language development/literacy	4
School observed solitary play behavior and teacher report of externalizing problems	4

Home observed affective synchronicity behavior and parental report of surgency/positive affect	4
Home observed affective synchronicity behavior and parent report of activity level	4
Home observed positive affect behavior and parental report of surgency/positive affect	4
Home observed positive affect behavior and tests of motor development	4
Home observed negative affect behavior and parental report of internalizing problems	4
Home observed negative affect behavior and parental report of soothability	4
School observed on-task engagement behavior and tests of intelligence	4
Home observed engagement behavior and parental report of internalizing problems	4
Home observed engagement behavior and parental report of soothability	4
Home observed engagement behavior and parental report of approach behavior	4
Home observed engagement behavior and parent report of activity level	4
Home observed engagement behavior and parental report of attentional control	4
Home observed engagement behavior and tests of intelligence	4
School observed on-task engagement behavior and tests of math skills	4
Home observed compliant/cooperative behavior and structured interactions of attachment security	4
Home observed aggressive behavior and tests of intelligence	4
Home observed aggressive behavior and structured interactions of attachment security	4
Home observed non-compliant/ disruptive behavior and parent report of activity level	4
Home observed non-compliant/ disruptive behavior and parental report of attentional control	4
Home observed prosocial behavior and tests of intelligence	4
School observed off-task behavior and teacher report of prosocial skills	4
School observed negative affectivity behavior and tests of intelligence	4
School observed fidget/grossmotor restlessness behavior and tests of math skills	4
School observed fidget/grossmotor restlessness behavior and tests of language development/literacy	4
School observed disruptive/aggressive behavior and parental report of negative affect	4
School observed disruptive/aggressive behavior and teacher report of inattention	4
School observed conflict with peers behavior and tests of language development/literacy	4
School observed cooperative prosocial behavior and parental report of externalizing problems	3
School observed cooperative prosocial behavior and parental report of negative affect	3
School observed cooperative prosocial behavior and parental report of prosocial skills	3
School observed cooperative prosocial behavior and teacher report of internalizing problems	3
School observed cooperative prosocial behavior and teacher report of hyperactivity/impulsivity	3
School observed cooperative prosocial behavior and teacher report of peer aggression	3
School observed cooperative prosocial behavior and teacher report of self-control	3
School observed cooperative prosocial behavior and tests of intelligence	3
School observed social initiation behavior and parental report of negative affect	3
School observed social initiation behavior and teacher report of negative emotionality	3
School observed positive engagement/proximity with teacher behavior and teacher report of externalizing	3
School observed positive engagement/proximity with teacher behavior and tests of social skills	3
School observed passive onlooking behavior and parental report of negative affect	3
School observed solitary play behavior and parental report of negative affect	3
School observed solitary play behavior and teacher report of prosocial skills	3
School observed on-task engagement behavior and teacher report of fussy/difficult behavior	3

School observed solitary play behavior and tests of social skills	3
Home observed affective synchronicity behavior and parental report of soothability	3
Home observed affective synchronicity behavior and parental report of attentional control	3
Home observed positive affect behavior and parental report of hyperactive/impulsive behavior	3
Home observed positive affect behavior and parent report of activity level	3
Home observed positive affect behavior and tests of language development/literacy	3
Home observed positive affect behavior and tests of positive affectivity	3
Home observed positive affect behavior and tests of effortful control	3
Home observed positive affect behavior and structured interactions of attachment security	3
Home observed negative affect behavior and parental report of fussy/difficult behavior	3
Home observed negative affect behavior and parental report of attentional control	3
Home observed negative affect behavior and parental report of prosocial skills	3
Home observed engagement behavior and parental report of fussy/difficult behavior	3
School observed on-task engagement behavior and tests of attentional control	3
Home observed engagement behavior and tests of language development/literacy	3
School observed on-task engagement behavior and tests of inhibitory control	3
Home observed compliant/cooperative behavior and parental report of negative affect	3
Home observed compliant/cooperative behavior and parent report of activity level	3
Home observed compliant/cooperative behavior and tests of language development/literacy	3
Home observed disengaged/off-task behavior and parent report of activity level	3
Home observed disengaged/off-task behavior and tests of intelligence	3
Home observed disengaged/off-task behavior and tests of language development/literacy	3
Home observed disengaged/off-task behavior and tests of motor development	3
Home observed aggressive behavior and parental report of externalizing problems	3
Home observed aggressive behavior and parental report of internalizing problems	3
Home observed aggressive behavior and parental report of negative affect	3
Home observed aggressive behavior and parent report of activity level	3
Home observed aggressive behavior and tests of language development/literacy	3
Home observed non-compliant/ disruptive behavior and parental report of externalizing problems	3
Home observed non-compliant/ disruptive behavior and parental report of internalizing problems	3
Home observed non-compliant/ disruptive behavior and parental report of fussy/difficult behavior	3
Home observed non-compliant/ disruptive behavior and tests of math skills	3
Home observed non-compliant/ disruptive behavior and tests of motor development	3
Home observed non-compliant/ disruptive behavior and structured interactions of attachment security	3
Home observed responsiveness behavior and tests of intelligence	3
Home observed responsiveness behavior and tests of language development/literacy	3
Home observed responsiveness behavior and tests of motor development	3
Home observed verbal communication behavior and parental report of fussy/difficult behavior	3
School observed off-task behavior and parental report of externalizing problems	3
Home observed prosocial behavior and parent report of activity level	3
Home observed prosocial behavior and structured interactions of attachment security	3
Home observed social withdrawal/avoidance behavior and parental report of surgency/positive affect	3
Home observed social withdrawal/avoidance behavior and parent report of activity level	3

Home observed social withdrawal/avoidance behavior and tests of intelligence	3
School observed off-task behavior and teacher report of internalizing problems	3
School observed off-task behavior and tests of intelligence	3
School observed off-task behavior and tests of attentional control	3
School observed disruptive/aggressive behavior and parental report of hyperactive/impulsive behavior	3
School observed disruptive/aggressive behavior and parental report of effortful control	3
School observed disruptive/aggressive behavior and parental report of attentional control	3
School observed negative affectivity behavior and tests of math skills	3
School observed disruptive/aggressive behavior and teacher report of negative emotionality	3
School observed disruptive/aggressive behavior and teacher report of fussy/difficult behavior	3
School observed disruptive/aggressive behavior and teacher report of language skills	3
School observed negative affectivity behavior and tests of language development/literacy	3
School observed disruptive/aggressive behavior and teacher report of self-control	3
School observed disruptive/aggressive behavior and peer report of sociometric status	3
School observed disruptive/aggressive behavior and tests of attentional control	3
School observed disruptive/aggressive behavior and tests of effortful control	3
School observed conflict with peers behavior and tests of social skills	3
School observed conflict with teacher/non-compliant behavior and teacher report of externalizing problems	3
School observed conflict with teacher/non-compliant behavior and teacher report of prosocial skills	3
School observed negative affectivity behavior and parental report of prosocial skills	2
School observed positive affectivity behavior and teacher report of prosocial skills	2
School observed conflict with teacher/non-compliant behavior and tests of language development/literacy	2
School observed cooperative prosocial behavior and parental report of internalizing problems	2
School observed cooperative prosocial behavior and parental report of hyperactive/impulsive behavior	2
School observed cooperative prosocial behavior and parental report of social problems	2
School observed positive affectivity behavior and tests of intelligence	2
School observed cooperative prosocial behavior and teacher report of negative emotionality	2
School observed cooperative prosocial behavior and teacher report of fussy/difficult behavior	2
School observed cooperative prosocial behavior and teacher report of shyness	2
School observed cooperative prosocial behavior and teacher report of popularity	2
School observed cooperative prosocial behavior and tests of inhibitory control	2
School observed positive affectivity behavior and tests of attentional control	2
School observed cooperative prosocial behavior and tests of math skills	2
School observed cooperative prosocial behavior and tests of motor development	2
School observed cooperative prosocial behavior and tests of compliance	2
School observed social initiation behavior and parental report of shyness	2
School observed social initiation behavior and parental report of attentional control	2
School observed social initiation behavior and parental report of prosocial skills	2
School observed social initiation behavior and teacher report of externalizing problems	2
School observed social initiation behavior and teacher report of internalizing problems	2
School observed social initiation behavior and tests of intelligence	2
School observed elicits positive peer response behavior and tests of language development/literacy	2
School observed positive engagement/proximity with teacher behavior and teacher report of negative	2

School observed positive engagement/proximity with teacher behavior and teacher report of shyness	2
School observed positive engagement/proximity with teacher behavior and teacher report of prosocial skills	2
School observed positive engagement/proximity with teacher behavior and teacher report of self-control	2
School observed on-task engagement behavior and teacher report of internalizing problems	2
School observed positive engagement/proximity with teacher behavior and tests of intelligence	2
School observed positive engagement/proximity with teacher behavior and tests of math skills	2
School observed positive engagement/proximity with teacher behavior and tests of language	2
School observed passive onlooking behavior and parental report of shyness	2
School observed passive onlooking behavior and parental report of attentional control	2
School observed passive onlooking behavior and teacher report of externalizing problems	2
School observed passive onlooking behavior and teacher report of negative emotionality	2
School observed passive onlooking behavior and teacher report of self-control	2
School observed on-task engagement behavior and teacher report of inattention	2
School observed passive onlooking behavior and tests of intelligence	2
School observed solitary play behavior and parental report of shyness	2
School observed solitary play behavior and parental report of attentional control	2
School observed solitary play behavior and teacher report of internalizing problems	2
School observed solitary play behavior and teacher report of negative emotionality	2
School observed solitary play behavior and teacher report of shyness	2
School observed solitary play behavior and teacher report of self-control	2
School observed solitary play behavior and tests of intelligence	2
School observed solitary play behavior and tests of language development/literacy	2
School observed parallel play behavior and parental report of negative affect	2
School observed parallel play behavior and teacher report of prosocial skills	2
School observed parallel play behavior and teacher report of social problems	2
School observed self-reliant behavior and tests of inhibitory control	2
School observed self-reliant behavior and tests of compliance	2
Home observed affective synchronicity behavior and parental report of externalizing problems	2
Home observed affective synchronicity behavior and parental report of internalizing problems	2
School observed on-task engagement behavior and teacher report of language skills	2
Home observed affective synchronicity behavior and parental report of approach behavior	2
Home observed affective synchronicity behavior and teacher report of externalizing problems	2
Home observed affective synchronicity behavior and teacher report of prosocial skills	2
Home observed reciprocity behavior and parental report of negative affect	2
Home observed confrontational behavior and parental report of negative affect	2
School observed on-task engagement behavior and teacher report of math skills	2
Home observed intimacy/proximity seeking behavior and parental report of internalizing problems	2
Home observed positive affect behavior and parental report of externalizing problems	2
Home observed positive affect behavior and parental report of approach behavior	2
Home observed positive affect behavior and parental report of attentional control	2
Home observed positive affect behavior and teacher report of externalizing problems	2
Home observed positive affect behavior and tests of intelligence	2
Home observed positive affect behavior and tests of inhibitory control	2

Home observed positive affect behavior and tests of compliance	2
Home observed positive affect behavior and structured interactions of negative affect	2
School observed negative affectivity behavior and teacher report of externalizing problems	2
School observed on-task engagement behavior and teacher report of emotion regulation	2
Home observed negative affect behavior and parental report of hyperactive/impulsive behavior	2
Home observed negative affect behavior and parental report of effortful control	2
Home observed negative affect behavior and parental report of attention seeking	2
Home observed negative affect behavior and teacher report of externalizing problems	2
Home observed negative affect behavior and tests of intelligence	2
Home observed negative affect behavior and tests of language development/literacy	2
Home observed negative affect behavior and structured interactions of negative affect	2
Home observed negative affect behavior and structured interactions of attachment security	2
Home observed engagement behavior and parental report of effortful control	2
Home observed engagement behavior and structured interactions of negative affect	2
Home observed compliant/cooperative behavior and parental report of internalizing problems	2
Home observed compliant/cooperative behavior and parental report of fussy/difficult behavior	2
Home observed compliant/cooperative behavior and parental report of soothability	2
Home observed compliant/cooperative behavior and parental report of approach behavior	2
Home observed compliant/cooperative behavior and parental report of surgency/positive affect	2
Home observed compliant/cooperative behavior and parental report of attentional control	2
Home observed compliant/cooperative behavior and parental report of attention seeking	2
Home observed compliant/cooperative behavior and tests of intelligence	2
Home observed compliant/cooperative behavior and tests of working memory	2
Home observed compliant/cooperative behavior and tests of math skills	2
Home observed compliant/cooperative behavior and tests of motor development	2
Home observed compliant/cooperative behavior and structured interactions of negative affect	2
Home observed compliant/cooperative behavior and structured interactions of compliance	2
Home observed disengaged/off-task behavior and parental report of negative affect	2
Home observed disengaged/off-task behavior and parental report of fussy/difficult behavior	2
Home observed disengaged/off-task behavior and parental report of attention seeking	2
Home observed disengaged/off-task behavior and tests of working memory	2
Home observed disengaged/off-task behavior and tests of math skills	2
Home observed aggressive behavior and parental report of soothability	2
Home observed aggressive behavior and parental report of approach behavior	2
School observed on-task engagement behavior and tests of compliance	2
Home observed aggressive behavior and parental report of surgency/positive affect	2
Home observed aggressive behavior and parental report of prosocial skills	2
Home observed aggressive behavior and tests of math skills	2
Home observed aggressive behavior and tests of motor development	2
Home observed aggressive behavior and structured interactions of negative affect	2
Home observed aggressive behavior and structured interactions of compliance	2
School observed compliant behavior and teacher report of externalizing problems	2
Home observed non-compliant/ disruptive behavior and parental report of shyness	2

Home observed non-compliant/ disruptive behavior and tests of working memory	2
School observed compliant behavior and teacher report of internalizing problems	2
Home observed non-compliant/ disruptive behavior and structured interactions of negative affect	2
Home observed non-compliant/ disruptive behavior and structured interactions of compliance	2
Home observed clarity of cues behavior and tests of intelligence	2
Home observed clarity of cues behavior and tests of language development/literacy	2
Home observed clarity of cues behavior and tests of motor development	2
Home observed responsiveness behavior and parental report of externalizing problems	2
School observed compliant behavior and teacher report of hyperactivity/impulsivity	2
Home observed responsiveness behavior and parental report of negative affect	2
Home observed responsiveness behavior and parental report of soothability	2
Home observed responsiveness behavior and parent report of activity level	2
Home observed verbal communication behavior and parent report of activity level	2
Home observed verbal communication behavior and parental report of attention seeking	2
Home observed prosocial behavior and parental report of negative affect	2
School observed negative affectivity behavior and teacher report of prosocial skills	2
Home observed prosocial behavior and parental report of soothability	2
Home observed prosocial behavior and parental report of prosocial skills	2
Home observed prosocial behavior and tests of language development/literacy	2
Home observed prosocial behavior and tests of motor development	2
Home observed prosocial behavior and structured interactions of negative affect	2
Home observed prosocial behavior and structured interactions of compliance	2
Home observed approach behavior and parental report of negative affect	2
School observed off-task behavior and parental report of hyperactive/impulsive behavior	2
Home observed approach behavior and parent report of activity level	2
Home observed approach behavior and parental report of prosocial skills	2
Home observed approach behavior and tests of intelligence	2
Home observed affectionate physical contact behavior and parental report of internalizing problems	2
Home observed affectionate physical contact behavior and parental report of soothability	2
Home observed affectionate physical contact behavior and parental report of approach behavior	2
Home observed affectionate physical contact behavior and parent report of activity level	2
Home observed affectionate physical contact behavior and parental report of prosocial skills	2
Home observed affectionate physical contact behavior and structured interactions of negative affect	2
Home observed affectionate physical contact behavior and structured interactions of compliance	2
School observed off-task behavior and parental report of negative affect	2
Home observed affectionate physical contact behavior and structured interactions of attachment security	2
Home observed social withdrawal/avoidance behavior and parental report of externalizing problems	2
Home observed social withdrawal/avoidance behavior and parental report of fussy/difficult behavior	2
Home observed social withdrawal/avoidance behavior and parental report of soothability	2
Home observed social withdrawal/avoidance behavior and parental report of effortful control	2
Home observed social withdrawal/avoidance behavior and parental report of attentional control	2
School observed off-task behavior and parental report of effortful control	2
Home observed social withdrawal/avoidance behavior and tests of language development/literacy	2

Home observed social withdrawal/avoidance behavior and tests of motor development	2
Home observed social withdrawal/avoidance behavior and structured interactions of attachment security	2
Home observed social withdrawal/avoidance behavior and physiological measurement of cortisol	2
Home observed separation anxiety/dependent behavior and tests of language development/literacy	2
School observed off-task behavior and parental report of attentional control	2
School observed off-task behavior and parental report of prosocial skills	2
School observed off-task behavior and teacher report of inattention	2
School observed negative affectivity behavior and teacher report of self-control	2
School observed off-task behavior and teacher report of inhibitory control	2
School observed off-task behavior and teacher report of attentional control	2
School observed off-task behavior and teacher report of negative emotionality	2
School observed off-task behavior and teacher report of popularity	2
School observed off-task behavior and teacher report of self-control	2
School observed off-task behavior and peer report of sociometric status	2
School observed off-task behavior and tests of inhibitory control	2
School observed off-task behavior and tests of effortful control	2
School observed fidget/gross motor restlessness behavior and teacher report of hyperactivity/impulsivity	2
School observed disruptive/aggressive behavior and parental report of internalizing problems	2
School observed disruptive/aggressive behavior and parental report of inattention	2
School observed negative affectivity behavior and tests of attentional control	2
School observed disruptive/aggressive behavior and parental report of prosocial skills	2
School observed disruptive/aggressive behavior and teacher report of inhibitory control	2
School observed disruptive/aggressive behavior and teacher report of attentional control	2
School observed disruptive/aggressive behavior and teacher report of shyness	2
School observed disruptive/aggressive behavior and teacher report of math skills	2
School observed disruptive/aggressive behavior and teacher report of skills for learning	2
School observed disruptive/aggressive behavior and teacher report of social problems	2
School observed disruptive/aggressive behavior and teacher report of teacher-child closeness	2
School observed disruptive/aggressive behavior and teacher report of emotion regulation	2
School observed disruptive/aggressive behavior and tests of inhibitory control	2
School observed disruptive/aggressive behavior and tests of motor development	2
School observed positive affectivity behavior and parental report of prosocial skills	2
School observed conflict with peers behavior and teacher report of externalizing problems	2
School observed conflict with peers behavior and teacher report of negative emotionality	2
School observed conflict with peers behavior and teacher report of prosocial skills	2
School observed conflict with peers behavior and tests of intelligence	2
School observed conflict with teacher/non-compliant behavior and teacher report of internalizing problems	2

Supplementary Table 5.*Associations between primary and secondary outcomes at the instrument/procedure level*

Interactions between primary and secondary outcomes	k
School observation: student teacher classroom interaction observation code STCIO Reinke 2010 and report scale: teacher observation of classroom adaptation checklist TOCA C	2
School observation: student teacher classroom interaction observation code STCIO Reinke 2010 and achievement test: Woodcock Johnson achievement test WJAT	2
Home observation: coding interactive behavior CIB Feldman 2007 and report scale: infant characteristics questionnaire ICQ Bates 1979	2
Home observation: Nursing Child Assessment Teaching Scale NCATS Barnard 1978 and achievement test: Bayley scales of infant development	2
Home observation: CARE Index Crittenden 1988 and achievement test: Bayley scales of infant development	2
Home observation: CARE Index Crittenden 1988 and achievement test: McCarthy scales of children's abilities MSCA	2
Home observation: emotionality Clark Kochanska 2000 and report scale: child behavior questionnaire CBQ	2
Home observation: emotionality Clark Kochanska 2000 and achievement test: Wechsler preschool and primary scale of intelligence WPPSI	2
Home observation: emotionality Clark Kochanska 2000 and achievement test: laboratory temperament assessment battery LabTAB Kochanska 1996	2
School observation: behavioral observations of students in schools BOSS Shapiro 1996 and report scale: Devereux student strengths assessment DESSA	1
School observation: behavioral observations of students in schools BOSS Shapiro 1996 and report scale: strengths and difficulties questionnaire SDQ	1
School observation: CBQ and report scale: child behavior questionnaire CBQ	1
School observation: CBQ and report scale: school liking and avoidance questionnaire	1
School observation: CBQ and report scale: teacher child relationship scale Pianta 1995	1
School observation: CBQ and report scale: teacher rating scale of school adjustment Birch 1997	1
School observation: CBQ and achievement test: continuous performance task NICHD 2003	1
School observation: CBQ and achievement test: Woodcock Johnson achievement test WJAT	1
School observation: classroom observation scale RimmKaufman 2005 and report scale: children self control scale Humphrey 1982	1
School observation: classroom observation scale RimmKaufman 2005 and report scale: mock report card Pierce 1999	1
School observation: classroom observation scale RimmKaufman 2005 and achievement test: preschool self regulation assessment PSRA Smith 2007	1
School observation: play Mize 1988 and report scale: child behavior questionnaire CBQ	1
School observation: play Mize 1988 and report scale: coping styles questionnaire Eisenberg 1993	1
School observation: play Mize 1988 and report scale: Derryberry and Rothbart temperament questionnaire	1
School observation: play Mize 1988 and report scale: internalizing negative emotion Larsen 1987	1
School observation: play Mize 1988 and report scale: peer sociometric evaluation Asher 1979	1
School observation: play Mize 1988 and report scale: popularity Harter 1979	1
School observation: play Mize 1988 and report scale: positive and negative affect schedule PANAS	1
School observation: play Mize 1988 and report scale: self control rating scale Kendall 1979	1
School observation: Fast Track study CPPRG 1999 and report scale: social health profile social competence scale SHP SCS	1
School observation: Fast Track study CPPRG 1999 and report scale: teacher observation of classroom adaptation checklist TOCA C	1
School observation: individualized classroom assessment scoring system inCLASS Downer 2010 and report scale: emotion regulation checklist Shields 1997	1
School observation: individualized classroom assessment scoring system inCLASS Downer 2010 and report scale: teacher child rating scale Hightower 1986	1

School observation: individualized classroom assessment scoring system inCLASS Downer 2010 and structured observation: Test Session Observations Checklist Woodcock Johnson III Woodcock 2001	1
School observation: individualized classroom assessment scoring system inCLASS Downer 2010 and structured observation: Preschool Self Regulation Assessment Assessor report Smith 2007	1
School observation: individualized classroom assessment scoring system inCLASS Downer 2010 and achievement test: preschool self regulation assessment PSRA Smith 2007	1
School observation: individualizing student instruction ISI Connor 2007 and achievement test: head toes knees shoulders task HTKS	1
School observation: individualizing student instruction ISI Connor 2007 and achievement test: Woodcock Johnson achievement test WJAT	1
School observation: Manchester inventory for playground observation MIPO and report scale: social skills rating system SRSS Gresham 1990	1
School observation: Manchester inventory for playground observation MIPO and structured observation: ASEBA observational schedule for school age children Achenbach 2001	1
School observation: observed child engagement scale RimmKaufman 2005 and report scale: children self control scale Humphrey 1982	1
School observation: observed child engagement scale RimmKaufman 2005 and report scale: mock report card Pierce 1999	1
School observation: observed child engagement scale RimmKaufman 2005 and achievement test: preschool self regulation assessment PSRA Smith 2007	1
School observation: on task Rapport 1982 and report scale: conners parent teacher rating scale CPTRS	1
School observation: play observation scale Rubin 1989 and report scale: child social preference scale CSPA	1
School observation: play observation scale Rubin 1989 and report scale: child behavior scale Ladd 1996	1
School observation: play observation scale Rubin 1989 and report scale: colorado child temperament inventory	1
School observation: Pomplun 1996 cooperative group rating scale and report scale: teacher friendship questionnaire	1
School observation: Pomplun 1996 cooperative group rating scale and achievement test: test of word reading efficiency TOWRE Torgeson 1997	1
School observation: Roper questionnaire 1978 and report scale: teacher questionnaire Roper 1979	1
School observation: Sharp behavior identification checklist and report scale: Hahnemann preschool behavior rating scale	1
School observation: Sharp behavior identification checklist and achievement test: Caldwell preschool inventory	1
School observation: Sharp behavior identification checklist and achievement test: preschool interpersonal problem solving test PIPS	1
School observation: Sharp behavior identification checklist and achievement test: what happens next game WHNG	1
School observation: social interaction observation system Neckerman 1994 and report scale: child behavior checklist CBCL	1
School observation: social interaction observation system Neckerman 1994 and report scale: parent child rating scale PCRS	1
School observation: social interaction observation system Neckerman 1994 and report scale: school social behavior scales SSBS	1
School observation: speech intelligibility scale Markides 1986 and report scale: children speech intelligibility scale McConnell 1999	1
School observation: speech intelligibility scale Markides 1986 and report scale: Vineland social emotional early childhood scales Sparrow 1998	1
School observation: speech intelligibility scale Markides 1986 and achievement test: Leiter international performance scale	1
School observation: speech intelligibility scale Markides 1986 and achievement test: Peabody Picture Vocabulary Test	1
School observation: speech intelligibility scale Markides 1986 and achievement test: preschool language scale Zimmerman 2002	1
School observation: play Strayer 1976 and report scale: preschool behavior questionnaire Behar 1977	1
School observation: play Strayer 1976 and achievement test: preschool interpersonal problem solving test PIPS	1
School observation: play Strayer 1976 and achievement test: Wechsler preschool and primary scale of intelligence WPPSI	1
School observation: systematic analysis of language transcripts miller chapman 2002 and report scale: children speech intelligibility scale McConnell 1999	1
School observation: systematic analysis of language transcripts miller chapman 2002 and report scale: Vineland social emotional early childhood scales Sparrow 1998	1

School observation: systematic analysis of language transcripts miller chapman 2002 and achievement test: Leiter international performance scale	1
School observation: systematic analysis of language transcripts miller chapman 2002 and achievement test: Peabody Picture Vocabulary Test	1
School observation: systematic analysis of language transcripts miller chapman 2002 and achievement test: preschool language scale Zimmerman 2002	1
School observation: teacher pupil observation tool TPOT Martin 2010 and report scale: strengths and difficulties questionnaire SDQ	1
School observation: Ghent University classroom coding inventory GUCCI and report scale: disruptive behavior disorder rating scale DBDRS	1
School observation: class Hinshaw 1995 and report scale: child behavior checklist CBCL	1
School observation: class Hinshaw 1995 and report scale: conners abbreviated symptom questionnaire CASQ	1
School observation: class Hinshaw 1995 and report scale: DACI structured interview for parents	1
School observation: class Hinshaw 1995 and achievement test: Wechsler intelligence scale for children WISC	1
School observation: class Hinshaw 1995 and achievement test: Woodcock Johnson achievement test WJAT	1
School observation: play Olson 1992 and report scale: child behavior checklist CBCL	1
School observation: play Olson 1992 and report scale: child behavior questionnaire CBQ	1
School observation: play Olson 1992 and report scale: inventory of peer relations Dodge 1987	1
School observation: play Olson 1992 and achievement test: disappointment task Cole 1994	1
School observation: play Olson 1992 and achievement test: false belief tasks	1
School observation: play Olson 1992 and achievement test: laboratory temperament assessment battery LabTAB Kochanska 1996	1
School observation: class Qi Kaiser 2001 and report scale: child behavior checklist CBCL	1
School observation: class Qi Kaiser 2001 and report scale: social skills rating system SRSS Gresham 1990	1
School observation: class Qi Kaiser 2001 and achievement test: Peabody Picture Vocabulary Test	1
School observation: class Qi Kaiser 2001 and achievement test: preschool language scale Zimmerman 2002	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and report scale: conners parent teacher rating scale CPTRS	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and achievement test: Conners continuous performance test CCPTII	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and achievement test: Wechsler intelligence scale for children WISC	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and achievement test: working memory test battery for children WMTB Pickering 2001	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and achievement test: verbal fluency Benton 1998	1
School observation: Scope classroom observation checklist SCOC Scope 2007 and achievement test: non verbal fluency ideational fluency Barkley 2001	1
School observation: play Weiss 1992 and report scale: child behavior checklist CBCL	1
School observation: interaction Furman 1980 and report scale: anxiety disorders interview schedule for children ADIS	1
School observation: interaction Furman 1980 and report scale: social worries questionnaire SWQ	1
School observation: interaction Furman 1980 and report scale: social competence questionnaire SCPQ	1
School observation: interaction Furman 1980 and achievement test: subjective probability social scale Lucock 1988	1
School observation: interaction Furman 1980 and achievement test: revised behavioral assertiveness test for children BATC Ollendick 1981	1
School observation: system for observing fitness instruction time SOFIT McKenzie 1991 and achievement test: Wechsler individual achievement test WIAT	1
School observation: class Mahar 2006 and achievement test: Wechsler individual achievement test WIAT	1

School observation: multiple option observation system for experimental studies MOOSES Tapp 1995 and report scale: Eyberg child behavior inventory ECBI	1
School observation: multiple option observation system for experimental studies MOOSES Tapp 1995 and report scale: social competence behavior evaluation LaFreniere 1992	1
School observation: multiple option observation system for experimental studies MOOSES Tapp 1995 and achievement test: WALLY problem solving test Webster 1990	1
School observation: multiple option observation system for experimental studies MOOSES Tapp 1995 and achievement test: Wally feelings test	1
School observation: school readiness and conduct problems coder observation of Adaptation COCA and report scale: Eyberg child behavior inventory ECBI	1
School observation: school readiness and conduct problems coder observation of Adaptation COCA and report scale: social competence behavior evaluation LaFreniere 1992	1
School observation: school readiness and conduct problems coder observation of Adaptation COCA and achievement test: WALLY problem solving test Webster 1990	1
School observation: school readiness and conduct problems coder observation of Adaptation COCA and achievement test: Wally feelings test	1
School observation: child behavior checklist direct observation form CBCL Achenbach 1986 and report scale: child behavior checklist CBCL	1
School observation: class Werry 1969 and report scale: hyperactivity rating scale Davids 1971	1
School observation: coding interactive behavior CIB Feldman 1998 and report scale: infant characteristics questionnaire ICQ Bates 1979	1
School observation: coding interactive behavior CIB Feldman 1998 and achievement test: Bayley scales of infant development	1
School observation: social behavior observational coding system McCartney 1996 and report scale: colorado child temperament inventory	1
School observation: social behavior observational coding system McCartney 1996 and structured observation: Strange Situation Ainsworth 1978	1
School observation: social behavior observational coding system McCartney 1996 and achievement test: Bayley scales of infant development	1
School observation: social behavior observation system Early Screening Project Walker 1995 and report scale: adhd rating scale IV DuPaul1998	1
School observation: social behavior observation system Early Screening Project Walker 1995 and report scale: conners parent teacher rating scale CPTRS	1
School observation: social behavior observation system Early Screening Project Walker 1995 and report scale: parenting stress index PSI	1
School observation: social behavior observation system Early Screening Project Walker 1995 and report scale: preschool and kindergarten behavior scales PKBS	1
School observation: social behavior observation system Early Screening Project Walker 1995 and achievement test: Battelle developmental inventory BDI Newborg 1988	1
Home observation: coding interactive behavior CIB Feldman 2007 and report scale: Development and Well Being Assessment DAWBA Goodman 2000	1
Home observation: coding interactive behavior CIB Feldman 2007 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: coding interactive behavior CIB Feldman 2007 and achievement test: Bayley scales of infant development	1
Home observation: coding interactive behavior CIB Feldman 2007 and achievement test: Hurting foot and slightly painful empathy eliciting paradigms	1
Home observation: coding interactive behavior CIB Feldman 2007 and achievement test: Mullen Scales of Early Learning MSEL Mullen 1995	1
Home observation: NICHD Love 2005 and report scale: Emotionality Activity Sociability and Impulsivity Temperament Survey Buss 1984	1
Home observation: NICHD Love 2005 and structured observation: Infant Behavior Record IBR Bayley 1969	1
Home observation: early attention to reading situations EARS Feagans 1994 and report scale: child behavior questionnaire CBQ	1
Home observation: early attention to reading situations EARS Feagans 1994 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: early attention to reading situations EARS Feagans 1994 and report scale: Social Competence Scale SCS Conduct Problems Prevention Research Group 1995	1
Home observation: early attention to reading situations EARS Feagans 1994 and structured observation: Infant Behavior Record IBR Bayley 1969	1
Home observation: early attention to reading situations EARS Feagans 1994 and achievement test: Bayley scales of infant development	1

Home observation: early attention to reading situations EARS Feagans 1994 and achievement test: laboratory temperament assessment battery LabTAB Kochanska 1996	1
Home observation: early attention to reading situations EARS Feagans 1994 and achievement test: EF battery Willoughby 2011	1
Home observation: Kochanska 1997 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: Kochanska 1997 and achievement test: laboratory temperament assessment battery LabTAB Kochanska 1996	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and report scale: child behavior checklist CBCL	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and report scale: Differential Emotions Scale DES	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and report scale: emotion regulation questionnaire ERQ Katz 1986	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and report scale: Child Adaptive Behavior Inventory CABI Cowan 1990	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and report scale: Dodge Peer Aggression Scale DPAS Dodge 1982	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and achievement test: Peabody individual achievement test PIAT	1
Home observation: Rapid Macro R MACRO peer interaction coding system Gottman 1983 and achievement test: Wechsler preschool and primary scale of intelligence WPPSI	1
Home observation: Nursing Child Assessment Teaching Scale NCATS Barnard 1978 and report scale: Eyberg child behavior inventory ECBI	1
Home observation: Nursing Child Assessment Teaching Scale NCATS Barnard 1978 and achievement test: Illinois Standards Achievement Test ISAT	1
Home observation: Nursing Child Assessment Teaching Scale NCATS Barnard 1978 and achievement test: Stanford Binet	1
Home observation: Behavior coding system Baumrind 1967 Clarke 1973 Lytton 1975 White 1975 and report scale: child characteristics questionnaire Bates 1979	1
Home observation: Coding system Clark 2000 and report scale: child behavior checklist CBCL	1
Home observation: Social behavior observational coding system McCartney 1996 and report scale: colorado child temperament inventory	1
Home observation: Social behavior observational coding system McCartney 1996 and structured observation: Strange Situation Ainsworth 1978	1
Home observation: Social behavior observational coding system McCartney 1996 and achievement test: Bayley scales of infant development	1
Home observation: Child Temperament and Behavior Q Set CTBQ and report scale: child behavior checklist CBCL	1
Home observation: Child Temperament and Behavior Q Set CTBQ and report scale: child behavior questionnaire CBQ	1
Home observation: Child Temperament and Behavior Q Set CTBQ and achievement test: Peabody Picture Vocabulary Test	1
Home observation: Observational Scale for Mother Infant Interaction during Feeding Chatoor 1997 Lucarelli 2002 and report scale: Italian Questionnaires on Temperament QUIT Axia 2002	1
Home observation: CARE Index Crittenden 1988 and structured observation: Strange Situation Ainsworth 1978	1
Home observation: Alarm Distress Baby Scale ADBB Guedeney 2001 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: mood Skuse 1992 and report scale: infant characteristics questionnaire ICQ Bates 1979	1
Home observation: Parent Child Interaction Code PACIC APPROACH scheme Caldwell 1969 and achievement test: Peabody Picture Vocabulary Test	1
Home observation: Emotional Availability Scales EAS Biringen 2008 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: Mealtime Observation Schedule MOS Sanders 1993 and report scale: child behavior checklist CBCL	1
Home observation: Mealtime Observation Schedule MOS Sanders 1993 and report scale: Feeding Behavior Checklist FBCL Turner 1996	1
Home observation: Relationship Affect Coding System Peterson 2008 and report scale: child behavior questionnaire CBQ	1
Home observation: Relationship Affect Coding System Peterson 2008 and report scale: Highly Sensitive Person scale Aron 1997	1
Home observation: Relationship Affect Coding System Peterson 2008 and report scale: strengths and difficulties questionnaire SDQ	1

Home observation: Relationship Affect Coding System Peterson 2008 and achievement test: role.play.scenarios.using.puppets Cimpian. 2007	1
Home observation: Relationship Affect Coding System Peterson 2008 and achievement test: Vignettes Rotenberg.and.Eisenberg 1997	1
Home observation: Relationship Affect Coding System Peterson 2008 and achievement test: sharing.task Iannotti 1985	1
Home observation: interaction Erickson 1985 and report scale: California Child Q set CCQ Block 1998	1
Home observation: interaction Erickson 1985 and structured observation: Strange Situation Ainsworth 1978	1
Home observation: play behaviors Johnson 1999 and report scale: preschool and kindergarten behavior scales PKBS	1
Home observation: Magnusson 1983 phonological development and achievement test: McCarthy scales of children abilities MSCA	1
Home observation: Magnusson 1983 phonological development and achievement test: TOLD Newcomer 1982	1
Home observation: Magnusson 1983 phonological development and achievement test: Metropolitan Readiness Test Nurss	1
Home observation: Coding System for Mother Child Interactions Healey 2010 and report scale: child behavior checklist CBCL	1
Home observation: Coding System for Mother Child Interactions Healey 2010 and report scale: child behavior questionnaire CBQ	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and structured observation: Mutually Responsive Orientation scale MRO Aksan Kochanska 2006	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Peabody Picture Vocabulary Test	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: laboratory temperament assessment battery LabTAB Kochanska 1996	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Spin.the.Pots Hughes 2005	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Shape.Stroop Kochanska 2000	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Baby.Stroop Hughes 2005	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Bear.Dragon Reed 1984	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Day Night Gerstad 1994	1
Home observation: Attachment Behavior Q Sort AQS Waters 1995 and achievement test: Dimensional.Change.Card.Sort DCCS Zelazo 2006	1
Home observation: conflict Herrera 1997 and report scale: Attachment Q Set AQS Waters 1985	1
Home observation: conflict Herrera 1997 and report scale: Toddler Behavior Assessment Questionnaire TBAQ Goldsmith 1996	1
Home observation: GLOBAL RATING SCALE FOR MOTHER INFANT INTERACTION Murray 1995 and report scale: infant characteristics questionnaire ICQ Bates 1979	1
Home observation: GLOBAL RATING SCALE FOR MOTHER INFANT INTERACTION Murray 1995 and report scale: parenting stress index PSI	1
Home observation: Mannheim assessment scales for recording maternal Child interaction in infancy MBSMKIS Esser 1989 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: Mannheim assessment scales for recording maternal Child interaction in infancy MBSMKIS Esser 1989 and structured observation: Mannheim assessment scales for recording maternal Child interaction in infancy MBSMKIS Esser 1989	1
Home observation: interaction Grockenberg 1990 Kochanska 1992 Patterson 1982 and report scale: infant behavior questionnaire IBQ Rothbart 1981	1
Home observation: interaction Grockenberg 1990 Kochanska 1992 Patterson 1982 and structured observation: Strange Situation Ainsworth 1978	1

Supplementary Table 6

Characteristics of studies reporting on observed disruptive/aggressive behavior in school settings and teacher report of externalizing problems

Primary outcome	Secondary outcomes	Observed activity	Operationalization of behavior	Observation session duration
Student Teacher Classroom Interaction Observation code (Reinke, Herman, & Newcomer, 2016).	Teacher Observation of Classroom Adaptation-Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009) Woodcock Johnson Achievement Battery, 3rd ed. (WJ-III ACH; Woodcock, McGrew, & Mather, 2007)	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5
Social behavior observation system of the Early Screening Project (Walker et al., 1995)	Preschool and Kindergarten Behavior Scales (PKBS) (Merrell, 1994) Battelle Developmental Inventory (BDI) (Newborg et al., 1988)	Mixed scenarios	Frequency counts/Interval sampling	60
Social behavior during free play (slightly modified from Mize & Ladd, 1988)	Child Behavior Questionnaire (CBQ; see Rothbart, Ahadi, & Hershey, 1994); Kendall and Wilcox's (1979) Self-Control Rating Scale; PANAS (Positive and Negative Affect Schedule; Watson, Clark, & Tellegen, 1988); Popularity Harter's (1979) 4 point response scale	Free play/recess (child-directed)	Frequency counts/Interval sampling	10
Social Interaction Observation System (Neckerman, 1994).	School Social Behavior Scales (SSBS). Achenbach Teacher Report Form (TRF). CBCL. Parent-Child Rating Scale (P-CRS).	Mixed scenarios	Frequency counts/Interval sampling	5
Brief Student-Teacher Classroom Interaction Observation code (ST-CIO; Reinke & Newcomer, 2010).	Woodcock-Johnson III Social Competence Scale (T-COMP; CPPRG, 1995). Teacher Observation of Classroom Adaptation-Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009)	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5
The Teacher-Pupil Observation Tool (TPOT; Martin et al., 2010)	Strengths and Difficulties Questionnaire (TSDQ; Goodman, 1997)	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	15
Fast Track study observation system (CPPRG, 1999; Tapp, Wehby, & Ellis, 1995).	Teacher Observation of Classroom Adaptation-Revised (TOCA-R). Social Health Profile Social Competence Scale (SHPSCS, CPPRG, 1999).	Classroom/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	5

Behavioral Observation of Students in Schools (BOSS; Shapiro & Kratochwill, 2000)	Devereux Student Strengths Assessment—Second Step® Edition (DESSA-SSE; Devereux Center for Resilient Children, 2012). Strengths Difficulties Questionnaire (SDQ; Goodman, 1997).	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	4
Play observation scale (Olson, 1992)	Child Behavior Checklist/2–3 (Achenbach, 1992), Kochanska, Murray, Jacques, Koenig, and Vandegest's (1996) toddler-age behavioral battery Rothbart's Child Behavior Questionnaire (CBQ; Ahadi, Rothbart, & Ye, 1993) Disappointment paradigm developed by Cole, Zahn-Waxler, and Smith (1994). False Belief Prediction and Explanation Tasks—Revised (Bartsch & Wellman, 1989) Caregiver/Teacher Report Form, Ages 2–5 (Achenbach, 1997).	Free play/recess (child-directed)	Frequency counts/Interval sampling	30
Fast track classroom observation system (Wehby & Dodge, 1993)	PLS 3 language (auditory comprehension, Expressive communication) CBCL (internalizing, externalizing) Social skills rating system PPVT3 (vocabulary)	Mixed scenarios	Frequency counts/Interval sampling	20
Sharp Behavior Identification Checklist	Preschool Interpersonal Problem-Solving (PIPS) Test. What Happens Next Game (WHNG). Caldwell Preschool Inventory. Hahnemann Preschool Behavior Rating Scale (Hahnemann).	Free play/recess (child-directed)	Frequency counts/Interval sampling	20
Scope Classroom Observation Checklist (SCOC; Scope, Empson, McHale, & Nabuzoka, 2007).	CTRS:L (Conners, 2001). Wechsler Intelligence Scale for Children (WISC IIIUK; Wechsler, 1992) Conners' Continuous Performance Test (CCPTII; Conners, 2002). Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001).	Classtime/Instruction/Seatwork (teacher-directed)	Frequency counts/Interval sampling	8
Playground behavior coding system developed by Weiss, Dodge, Bates, and Pettit (1992).	Child Behavior Checklist (CBCL; Achenbach, 1991).	Free play/recess (child-directed)	Mixed approach	5
School Readiness and Conduct Problems: Coder Observation of Adaptation-Revised (COCA-R).	Wally Problem Solving and Feelings Tests (Webster-Stratton, 1990)	Mixed scenarios	Mixed approach	30
Achenbach direct observation form CBCL	Achenbach report form CBCL Thai youth check list	Classtime/Instruction/Seatwork (teacher-directed)	Mixed approach	10