

Emotional quality of early education programs improves language learning: A within-child across context design

Peter Sheldon Rankin¹  | Sally Staton¹  | Azhar Hussain Potia¹  | Sandy Houen¹  | Karen Thorpe² 

¹Institute for Social Science Research, Indooroopilly, Queensland, Australia

²Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

Correspondence

Peter Sheldon Rankin, Institute for Social Science Research, Indooroopilly, QLD, Australia.
Email: p.rankin1@uq.edu.au

Funding information

Australian Research Council's Centre of Excellence for Children and Families over the Life Course, Grant/Award Number: CE200100025; Queensland Government Department of Education and Training; Victorian Government Department of Education and Early Childhood Development; Australian Research Council Linkage Projects Scheme, Grant/Award Number: LP0990200

Abstract

Observational studies comparing child outcomes in early care and education classrooms of differing quality are often confounded by between-child differences. A within-child design, tracking children across contexts, can identify the effects of quality with less confounding. An analysis of Australian children ($N = 1128$, mean age 5 years, 48% female, 2.9% Indigenous, ethnicity data unavailable) tracked across pre-K, K, and year 1 (2010–2012) was conducted to assess how changes in observed quality (Classroom Assessment Scoring System) were associated with changes in cognitive development (Woodcock–Johnson III). Thresholds of quality were also investigated. Increases in Emotional Support were associated with improved language development ($\beta = 0.54$, 95% CI [0.1–0.99], approximating 2.6 weeks development). Results highlight that emotional quality is an integral and potent component of early learning.

Across the last two decades, substantial focus has been placed on early care and education (ECE) programs with the goal of supporting the learning outcomes, well-being, and lifetime achievements of children (Heckman et al., 2018; Psacharopoulos & Patrinos, 2018). Evidence from neuroscience, developmental science, and economics has converged on the conclusion that, when considering where to invest, early education is a prime candidate (Campbell et al., 2014; Heckman et al., 2010; Organisation for Economic Co-operation and Development, 2017). Providing positive educational experiences early in life, at a critical time in shaping neural architecture, can establish optimal within-child learning conditions and avert emergent problems associated with individual child or family circumstance (Yoshikawa et al., 2016). Small-sample, randomized control intervention studies of disadvantaged groups have clearly demonstrated

the potential of ECE to improve children's learning, life chances, and outcomes into adulthood (Campbell et al., 2012; Heckman et al., 2010). These positive effects are most evident for children who are disadvantaged by personal circumstances and/or adversity in their home learning environments (Melhuish et al., 2015). However, evidence pertaining to what constitutes effective experiences within ECE programs available to the general population is less certain. Current evidence directs attention to the qualities of educator-child interactions as key modifiable program features to drive child cognitive development and learning outcomes (Burchinal, 2018), but has yet to converge on what interactional strategies are most potent (Pianta, Hamre, et al., 2020).

Existing research directed to identifying what interactional strategies work to enhance cognitive development in ECE has largely relied on comparing children's

Abbreviations: CLASS, Classroom Assessment Scoring System; CLASS-CO, CLASS-classroom organization; CLASS-ES, CLASS-emotional support; CLASS-IS, CLASS-instructional support; E4kids, Effective Early Education Experiences for Children; ECE, early care and education; ECERS-R, Early Childhood Environment Rating Scale revision; ICC, intra-class correlation; WJIII, Woodcock–Johnson III.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Child Development* published by Wiley Periodicals LLC on behalf of Society for Research in Child Development.

learning within classrooms of different assessed interactional quality (Burchinal, 2018; Fukkink et al., 2017; Perlman et al., 2016). Detailed observational assessments using standard measures have been applied to assess the effect of global interactional quality. Findings have tended to show small and sporadic effects that fall short of theoretical expectations (Burchinal, 2018; Perlman et al., 2016; Pianta et al., 2016). Two key methodological explanations for these weak findings are evident. The first relates to measurement. Studies are potentially confounded with the reliability and validity of the measurement of teacher–child interaction quality (Burchinal, 2018; Gordon et al., 2015; Layzer & Goodson, 2006; Neitzel et al., 2019). Furthermore, many studies assume linear association, yet some identify thresholds of quality as necessary to yield effect (Burchinal et al., 2016). The second relates to design. While intensive randomized control designs yield strong and positive effects, observational studies are necessarily reliant on adequate measurement methods to clarify the contributions of teacher–child interaction to child development, and to identify effective strategies applicable at scale.

In this paper, we offer a robust within-child design to evaluate the effects of different qualities of educator–child interaction on children's cognitive development and learning outcomes. Additionally, we test for quality thresholds to provide indication of potentially meaningful intervention targets. We draw from a cohort study of Australian children ($N = 2606$) recruited from the range of available ECE programs in their pre-K year (age 3–4 years), and subsequently tracked across K, and year 1. In contrast to standard value-add approaches that examine whether classroom-level changes in outcomes for children are associated with assessed interactional quality of the room, our within-child across context design analyses how each child's rate of learning changes as they move between classrooms of different assessed quality. Applying this strategy, variation in unmeasured confounding individual characteristics is attenuated (Hoffman & Stawski, 2009; Maldonado-Carreño & Votruba-Drzal, 2011; Watts et al., 2021). Our study provides an analytically sophisticated and design-controlled approach to identifying ECE program inputs to children's cognitive development. This approach can inform practice interventions that target quality improvement with greater certainty.

Theoretical framework

In this paper, we utilize an individual difference framework (Belsky et al., 2020; Rutter, 2012; Wertz et al., 2020) to evaluate the effects of the quality of ECE experiences on child development. This framework views cognitive development as a hierarchically sequenced, experience dependent, and individually directed process driven by

the interplay of genetic and environmental factors. Thus, we consider child and family characteristics preceding current development as ongoing influences on cognitive developmental and learning outcomes.

The importance of early education and care program quality

At entry to school, all children are not equally prepared for their ongoing educational journey. Rates of learning in the first years of school are not only affected by children's experiences within their first classrooms, but also by the learning experiences that have come before (Landry et al., 2017; Lohndorf et al., 2021; Pianta, Whittaker, et al., 2020). When transitioning to school, individual characteristics (e.g., health, ability, disability, and temperament) and prior experiences within the family weight a child's ability to adapt to school culture and effectively engage in learning within their classroom environments (Lohndorf et al., 2021; Micalizzi et al., 2019; Ramakrishnan & Masten, 2020). For this reason, public investment in ECE targets 'school readiness' to improve children's development, ongoing learning trajectories, and life chances. Yet not every ECE program is successful in delivering positive outcomes for children (Green et al., 2021; Melhuish et al., 2015). Understanding what constitutes a high-quality ECE program and the specific features and thresholds of quality that deliver improved outcomes for children, remains a policy target and research challenge.

In defining what constitutes quality in ECE, research studies have focused on two broad domains: structural and process (Burchinal, 2018). Structural quality comprises the readily measurable features of the ECE environment, notably group size, staff to child ratios, staff qualifications and focus on an educational curriculum. Contemporary research identifies these structural features as important, but not of themselves sufficient to deliver positive child outcomes (Bowne et al., 2017; Burchinal, 2018; Mashburn et al., 2008; Melhuish et al., 2015; Slot et al., 2015; van Huizen & Plantenga, 2018). Their value is in enabling process quality; the amount, content, and inter-personal qualities of interactions in the ECE program as these are more potent predictors of child outcomes (Mashburn et al., 2008). Accordingly, policy actions to enable quality improvement in ECE have regulated structural features of ECE programs. The primary focus of research and intervention strategies has been directed to understanding the domains of process quality that deliver improved child outcomes.

Identifying intervention strategies amenable to delivery at scale, and focused within-classrooms, has been the central focus of contemporary research efforts targeting improved learning outcomes at school entry (Burchinal, 2018; Pianta et al., 2016). Accordingly, the dominant research focus has become qualities of

educator–child interactions and the dominant design strategy comparison of the learning outcomes of children exposed to different levels of process quality. Such observational-level studies are reliant on two key design elements: (1) valid and reliable observational measurement of process quality at scale and (2) control of confounding individual characteristics that drive selection into ECE programs of different quality. Below we consider each of these.

Measuring process quality

Standard observational measurement of educator–child interactions has become the gold-standard method used to assess ECE process quality in research studies and policy settings. Two such measures have come to dominate the assessment of process quality in large-scale, policy-targeted research: The Early Childhood Environment Rating Scale revision (ECERS-R; Harms et al., 1998) and Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). ECERS-R provides global ratings of the ECE environment, both structural and process. ECERS-R applies a hierarchical rating of target behaviors across a total observation period such that higher levels of quality within a domain are not rated if lower thresholds are not attained. Though the measure has been widely used, limitations of the scoring system and psychometric properties have been identified (Bull et al., 2017; Gordon et al., 2015). Furthermore, the global nature of the quality rating may offer less utility in directing intervention strategies (Layzer & Goodson, 2006; Neitzel et al., 2019). The more recently developed CLASS provides a different measurement approach. CLASS focuses specifically on process quality, rating 10 interactional features of ECE quality organized in three quality domains (instructional support [CLASS-IS], emotional support [CLASS-ES], and classroom organization [CLASS-CO]). In contrast to the ECERS rating system, CLASS adopts sequential scoring in which, across a minimum of 2 h, ratings of quality are undertaken for 30-min cycles (20 min of observation and 10 min of scoring). Using this method, records of activity type and teaching formats are made that enable the linkage of assessed quality scores to variations in curriculum and pedagogical formats across the assessment period (Thorpe, Rankin, et al., 2020). Thus, in contrast to ECERS that constructs structural quality, curriculum, and process quality as additive, the CLASS assessment examines these as interactive. CLASS also presents limitations, however, such as concerns about the reliability of ratings (Mantzicopoulos et al., 2018; Styck et al., 2021).

Recent meta-analyses of studies examining the effect of process quality suggest CLASS is a more sensitive measure in detecting effects on child outcomes. Ulferts

et al. (2019) in a meta-analysis of 17 European longitudinal studies found that while CLASS observations had small, statistically significant associations with children's language and literacy across time, these associations were unreliably detected using ECERS as the measure of quality. Similarly, a meta-analysis of six studies conducted in the USA (Hong et al., 2019) found small, statistically significant associations between CLASS-IS and language outcomes, but no such significant effect with ECERS. Taking an Australian sample, in the current study we utilize the CLASS measures to assess effects of three domains of process quality (CLASS-IS, CLASS-ES, CLASS-CO) on child cognitive development and learning attainments.

Disaggregating selection effects and process quality effects

Early care and education is provided through a marketized system in which there is a range of provider types and motives (e.g., for profit vs. non-profit; targeted programs for disadvantaged families). The consequence is a diversity of program offerings. Inevitably, the marketized nature of ECE produces selection effects driven both by parent preference (perceived child need or understanding of quality) and family resource that enable or limit access to their preferred programs. While legislation can specify structural standards, there remains substantial market-driven variation in provision that inevitably influence process quality. Thus, in the Australian context, while there is a national quality standard, there remains considerable variation in ECE quality (Tayler et al., 2013) influenced by the intersection of parent ability to pay and market competition. For example, in low-income communities' provision of food is more likely when there is high market competition, yet no associated change in fees to parents, suggesting cost savings must be attained elsewhere (Thorpe, Searle, et al., 2020). In this case, the effects on process quality are unknown, but savings are most readily made in staffing (United Workers Union, 2021), a key enabler of process quality. Favorable quality rating assessments are also used to attract market share (Thorpe, Rankin, et al., 2020). Yet, higher quality is associated with higher fee structures that, in the absence of targeted subsidy, are only viable for parents who can sustain higher costs (Thorpe et al., 2021). Public policy also introduces systematic bias. Targeted subsidies or provision of intervention programs directed to children living in circumstances of the greatest disadvantage give access to designated groups, but inclusion criteria also serve to exclude others. To identify the independent contribution of dimensions of process quality to children's cognitive development and learning outcomes, therefore, requires effective adjustment

of confounding factors associated with differential selection.

Design and analysis strategies to detect the effects of process quality

Statistical adjustment is the standard approach used to account for confounding and selection effects in between-child observational approaches to analyzing ECE programs (Blanca et al., 2018). Appropriate adjustments, however, are dependent on the availability, knowledge, and understanding of measurements that reliably index confounding and selection bias, along with the correct specification of the analytical and selection model (Rubin, 1974). Typically, family characteristics such as parent education, parent occupation, place of residence, and race are used as indices based on the assumption that these are associated with family values, capacity to pay fees, and ability to access ECE programs. However, many family and child factors that influence the selection of ECE program may not be captured by broad family characteristics. To date, between-child studies have yielded weak and inconsistent results. Meta-analyses of the large body of international studies find, at best, modest effects on children's cognitive development and learning outcomes (Perlman et al., 2016; Ulferts et al., 2019). There is also considerable variation in the domain of process quality identified as effective across studies (Perlman et al., 2016). Variation across observational samples is evident with stronger effects among disadvantaged populations (e.g., Watts et al., 2021). Moreover, associations may not be linear as a number of studies have reported threshold effects in which teacher inputs must exceed a certain quality level to make discernible contributions to child outcomes (Hatfield et al., 2016; Leyva et al., 2015; Li et al., 2019). Unmeasured confounders in correlational studies means between-child designs have limited scope to approximate causality, and consequently provide less certainty in directing policy and practice actions.

A value-add model is the common approach applied to identify strategies that improve children's cognitive development and learning attainments (e.g., Curby et al., 2013). In the value-add method, child outcomes are typically measured at the start and end of an academic period. The comparative difference in the change of child outcomes for each classroom then indicates the value, in terms of academic development, added by the teacher given the starting position of the children. Classroom characteristics, including process quality, are then modeled to evaluate their role in determining this between-classroom variation in added-value. An alternative to the value-add design is to focus on individual children as they move between classrooms

with differing levels of process quality, that is a within-child across context design (Hoffman & Stawski, 2009; Maldonado-Carreño & Votruba-Drzal, 2011; Watts et al., 2021).

Within-child across context design is atypical in the field of ECE effectiveness as they are labor- and cost-intensive, requiring individual children to be tracked as they disperse across classrooms rather than focusing on changes within class groups. When such data are available, however, they present an opportunity to assess the effects of change in quality, both positive and negative, focussed on the individual child as they experience different classroom contexts. Although reliability and validity of measures of program quality and family covariates remain a limitation, this design has the advantages of reducing the effects of unmeasured child and family confounders and selection effects. One recent study (Watts et al., 2021) presents an example. This study examined changes in classroom quality across 3 years of elementary school using CLASS to examine effects on individual children's trajectories of achievement and behavior. This study reports small, but nonsignificant effects of classroom quality across the general population, but significant positive effects on the achievement of children living in poverty when they moved to classrooms with higher organizational quality. To date, although data applicable to a within child-design have been collected (Cash et al., 2019; Pianta et al., 2007; Vernon-Feagans et al., 2019), studies commencing in the ECE setting and utilizing within-child analytical methods have not been reported. Yet, in the marketized ECE sector, selection effects, and variability in quality (e.g., teacher qualification, program type) are more pronounced than in the school sector. Given these circumstances, a within-child across context design offers potential to identify aspects of process quality that can improve child outcomes with greater certainty.

Advancing research to identify early education program effects

Burchinal (2018) in a review of the state of the evidence calls for greater sophistication in design and analytic approach to direct understanding of ECE quality on child development. Here, we provide the first within-child across context analysis of ECE quality effects on child outcome drawing on a diverse sample. We analyze data from the Effective Early Education Experiences for Children (E4kids) study, a sample of 2606 children attending the range of ECE programs in Australia comprising home-based day care, center-based day care, and stand-alone pre-K programs. We assess the effects of child exposure to variation in process quality as they move between classrooms on their cognitive development and learning outcomes.

METHOD

Although informed by a strong design, the current analysis was exploratory to the extent that statistical models were not pre-specified at time of data collection. We report on our data set, determination of our sample size, all data exclusions, imputations and manipulations, and all measures included in our analyses.

Data

Data were analyzed from the E4kids study. E4kids is a study of $N = 2606$ children recruited from Australian Early Childhood Education services with focus on rooms providing for children aged 3–4 years. The protocol for this study has been published previously (Tayler et al., 2013, 2016).

In 2010, the first year of the study, children were recruited from $N = 142$ ECE services, comprising center-based day care ($n = 92$), standalone pre-K ($n = 40$) and home-based day care ($n = 7$), and limited hours care ($n = 3$). Using a sampling approach based on large-scale studies (Adams & Wu, 2003), a three-level stratified sample was drawn. The first stage identified four locations—two major cities in Australia (Melbourne and Brisbane), one inner regional location (Shepparton, Victoria), and one remote location (Mt Isa, Queensland; see Pink, 2011 for location classifications). Within each location, service providers were stratified based on type of program (long-day care, preschool, home-based day care, limited hours care), resulting in 16 strata. Within each stratum, a proportionate quota of services was assigned with the services implicitly stratified and selected by sorting by a weighting of the relative socioeconomic advantage and disadvantage status of the locality (Adhikari, 2006) and service capacity. Programs were randomly selected with replacement in this manner. Specific numbers on strata quotas, intervals, replacements, and weighting are available in Tayler et al. (2016). By this explicit and implicit sampling, E4Kids included a broad range of Australian ECE services across geographic locations and socioeconomic areas.

Within center-based services, all rooms that educated 3- to 4-year-old children, and that had more than 5 children in that age range, were selected. A small number of siblings were captured using this method (see Table S3 for sibling numbers). All children within a selected room were invited to participate. For home-based day care, as a result of lower educator-child ratios, the presence of at least one child 3- to 4-years-old constituted viable selection. The characteristics and frequency of non-consenting children were not detailed at the time of the study. In the years subsequent, the initial sample was tracked into subsequent ECE and school services via supplied email and phone contacts. The new service or school room in which the child was located was then

invited to participate in the study. Thus, the stratification of subsequent education services was indeterminant.

For all classrooms with a study child, observations using the CLASS (Pianta et al., 2008a, 2008b) were made longitudinally each year from 2010 to 2013. Observations of process quality within each child's classrooms were typically pre-K, K, and year 1 although we note that, due to age cut-off or retention in a year level 5.5% of children did not follow this typical sequence. The Supporting Information (Table S1) provides a definition of these programs for international comparison. During classroom visits across 2010–2012, fieldworkers individually tested children using the Woodcock–Johnson III (WJIII) Tests of Cognitive Abilities and Tests of Achievement (Woodcock et al., 2001). Further information about the children was obtained from surveys completed by the primary caregiver (e.g., parent), educators, and ECE directors from 2010 to 2013. The study was approved by the University of Melbourne Human Research Ethics Committee (ID 0932660.2). All participation in the study was voluntary and could be withdrawn at any time.

The sample for this study derives from the years 2010, 2011, and 2012. Of $n = 2606$ children who had at least one academic measure, we only included children with at least partial survey information from the caregiver and academic outcomes for three sampling years ($n = 1128$) and classroom observations for 2 ($n = 68$) or 3 years ($n = 1060$). Thus, the total number of children available for analyses was 1128 (see Table S2 for data availability over time). This analytical sample is compared to the excluded sample and a relevant population in 2011 (the closest national census) in Table S3. This shows slightly higher attrition for migrant children and those with less formally educated caregivers and fewer socioeconomic resources. In addition, in comparison to the national population, the sample captures a broad demographic with some under-representation of Indigenous children, migrant caregivers, and caregivers with lower formal education qualifications. Finally, although not entirely reflecting the initial stratification and weighting, there is general agreement between the proportion of rooms in the analytical sample in 2010 and the types of services in the population (Table S7). We note that beyond Indigenous and migrant status, information on race or ethnicity was not collected.

Measures

Child cognition and achievement

The primary outcomes were five W scores from the WJIII test battery; Cognitive 1-W Verbal Comprehension (measuring lexical knowledge and language development), 5-W Concept Formation (induction), 6-W Visual Matching (perceptual speed), Achievement 4-W Understanding Directions (listening ability, language

development) and 10-W Applied Problems (quantitative reasoning, math achievement, math knowledge) tests (Woodcock et al., 2001). These were administered in person by trained fieldworkers.

Key covariate

Our key covariate was the CLASS observational measure (Pianta et al., 2008a, 2008b). We evaluated the three subscales of *Emotional Support* (4 dimensions), *Classroom Organization* (3 dimensions), and *Instructional Support* (3 dimensions), with each item on a seven-point scale ranging from 1 (lowest quality) to 7 (highest quality). For this study, we considered an observation as valid if there were 2–6 CLASS completed observation cycles of 20 min each (e.g., 1–3 h per room). Most observations (98.2%) applied to our sample had 4 or more cycles. Observations were undertaken by 92 research staff who were trained and certified as reliable both at initial training and through annual re-certification. Following CLASS protocol, certified reliability was expressed as being within one rating of the gold standard coder, with at least 80% agreement across all observations. In-field assessment of fieldworkers against a gold standard CLASS-coder was conducted in 2011 and agreement within one rating was high (96.4%, Cloney et al., 2017). Observers completed their ratings of each scale consistent with the CLASS manual, using Pre-K and K versions of CLASS as appropriate. Each CLASS subscale was calculated as the average of the observation cycles in the classroom at the assessment year. As, in 2011, multiple visits were made to some services for reliability assessment, for that year we used the visit with the most valid cycles. If multiple visits had the same number of cycles, we randomly allocated which CLASS visit was examined for the child in that assessment year. Psychometric results (e.g., measurement invariance across cycles) of the subscales for these observations were excellent and are published elsewhere (Thorpe, Rankin, et al., 2020).

Additional covariates

Several additional covariates were included in the analysis. The time-varying covariates (e.g. home learning environment; child temperament) were from 2010 to avoid adjusting for intermediate outcomes (Ananth & Schisterman, 2017) and to reduce the amount of missing data within the sample. Data from 2010 were also used for time-invariant covariates (e.g., caregiver born overseas) unless missing at this point. If these data were not available in 2010 but subsequently provided, they were backfilled.

Child covariates

Categorical variable covariates regarding the child were gender (0 = female; 1 = male), Indigenous status (0 = no; 1 = yes), child born overseas (0 = no; 1 = yes); child has

a developmental delay (0 = no; 1 = yes); and English-speaking (0 = main and only language; 1 = main, but also speaks other languages; 2 = English not main language).

Continuous child covariates included child age in years (centered at the grand mean) and difficult temperament, which was assessed by the 12-items Short Temperament Scale for Children (Sanson et al., 1994). This scale comprises three subscales (persistence, reactivity, and sociability), derived from aggregation of four items. Caregivers were asked to rate their child's behavior (e.g., "if this child is upset, it is hard to comfort him/her") on a 6-point Likert scale (almost never = 1 to 6 = almost always). Lower persistence, higher reactivity, and lower sociability indicate a more difficult child temperament, so the persistence and sociability scales are reverse coded and averaged with reactivity to derive the difficult temperament scores (Prior et al., 1989).

Caregiver covariates

Categorical caregiver covariates were Indigenous status (0 = no; 1 = yes), caregiver born overseas (0 = no; 1 = yes); caregiver has a low-income health care card (0 = no; 1 = yes); highest level of education (0 = postgraduate degree; 1 = bachelor degree; 2 = diploma; 3 = year 12 or Tafe [technical college] certificate; 4 = year 10 or lower); and presence and effect of 11 stressful life events (Holmes & Rahe, 1967), for example, *death of someone close to you*, on the child (0 = event not experienced, or event experienced but no negative effect on child, or event experienced and negative effect on child; 1 = event experienced and had serious negative effect on child), where the maximum score across the 11 events was retained.

Continuous caregiver covariates were psychological distress as measured by the Kessler-10 (Kessler et al., 2002). This measure is the sum of 10 items that measure psychological distress in the past 30 days. For example, *how often did you feel hopeless?* (1 = none of the time, 5 = all of the time). Additionally, the home learning environment was measured using the average of 12 items that related to learning materials and interaction at home. For each item, caregivers rated the frequency of an activity over the last week, for example, *read to the study child from a book*, on a 0–7 scale (0 = no days; 7 = 7 days).

Analytical plan

In the first step, we calculated the intra-class correlation (ICC) for children for each CLASS dimension by dividing the child random effect variance by the total variance from random intercept linear models. The ICC ranges from 0 to 1 and indicates whether children have stable experiences of early education quality (a high ICC) or fluctuating experiences (a low ICC). Without some within-child variation (e.g., ICC < .7),

it would be impractical to evaluate how changes in quality are linked to changes in child outcomes at the within-child level.

In the second, third, and fourth steps, we used progressively complex hybrid random effect models (Hamaker & Muthén, 2020) to evaluate academic outcome trajectories, the within- and between-person effects of CLASS on academic outcomes, and if there were threshold effect of CLASS on academic outcomes, respectively. The specific formula for a hybrid model is:

$$y_{it} = \beta_0 + (\beta_1 + \eta_{1i})(x_{it} - \bar{x}_i) + \beta_2 c_i + \beta_3 \bar{x}_i + \beta_4 (x_{it} - \bar{x}_i) \bar{x}_i + \mu_{0i} + \omega_{0c} + \varepsilon_{it},$$

where, y is the cognitive outcome for child i at time t . β indicates regression coefficients, x is the score on a time-varying covariate (e.g., CLASS-ES), \bar{x} is the average score for a time-varying covariate (e.g., CLASS-ES), c is a time-invariant covariate (e.g., sex assigned at child's birth), μ is the child random intercept with variance τ_{00}^2 and η represents a random effect on β_1 , for example, with variance τ_{11}^2 distributed as $(\mu_{0i}, \eta_{1i}) \sim \text{Normal}(0, \begin{bmatrix} \tau_{00}^2 & \rho\tau_{00}\tau_{11} \\ \rho\tau_{00}\tau_{11} & \tau_{11}^2 \end{bmatrix})$, ω is the random intercept for unique room year combinations c ($\omega_c \sim \text{Normal}(0, \sigma_c^2)$), and ε is the residual error ($\varepsilon_{it} \sim \text{Normal}(0, \sigma_\varepsilon^2)$). In this model, β_1 is a within-person effect, β_3 is a between-person effect, and β_4 represents the interaction between a within-person and between-person effect, such as the effect of average CLASS-ES on changes in CLASS-ES, that is used to evaluate thresholds.

The models used in step 2 to evaluate the typical trajectories of each academic outcome included a random intercept for each child and between-person and within-person terms for age, as well as their interaction. Random slopes for each child were also estimated for the within-person age term. As all children had three observations of academic outcomes these models used complete data.

Next, the models in step 3, used to evaluate the contribution of each CLASS dimension and subscale in explaining within-person and between-person variation in academic outcomes, extended the previous model to include within- and between-person terms for CLASS, as well as including the additional covariates and a random intercept for unique room by year combinations.

The fourth step included an interaction for the within- and between-person variation in CLASS to evaluate threshold effects, for example, a positive change when average CLASS is high would indicate a child has moved into a very high-quality ECE experience. This approach to estimating quality thresholds is unique, as the data are used to reveal where changes in quality are meaningful based on the average experience of children.

We further investigated each statistically reliable interaction in the fourth step by estimating the regions of significance which show when the effect of one variable becomes significant ($p < .05$) given values of the other variable (Preacher et al., 2006; Roisman et al., 2012). We

considered the interaction reliable if 1) regions of significance were within plus or minus 2 standard deviations of each measure in the interaction, and 2) we interpreted a suitably large number of observations within the regions of significance.

We chose to model each CLASS dimension separately because of high correlations (r up to .86) and consequent multi-collinearity.

As a supplemental analysis comparable to Watts et al. (2021), we estimated step 3 models but categorized CLASS into four groups separating the 25th, 50th, and 75th percentiles of unique room observations, with the lowest 25th percentile forming the baseline. This analysis provides an alternative insight into nonlinear thresholds, though not accounting for how the average experience may moderate the effect of changes in quality.

Covariates

We focus our interpretation on models estimated with (adjusted) covariates. We also ran models without covariates (unadjusted) that are reported in [Supporting Information](#). The unadjusted models included child age to account for cognitive maturation.

Statistical software and model assumptions

All models were run in R (R Core Team, 2020; version 4.0.2) using the lme4 (Bates et al., 2015; version 1.1.23) and lmerTest (Kuznetsova et al., 2017; version 3.1.3) packages. Model diagnostics: Diagnostic plots were examined in R using the qqmath function of the lattice package version 0.20.41 (Sarkar, 2008) to evaluate normality of the residuals and validate assumptions of multilevel regression models (Pinheiro & Bates, 2000).

Missing data

While nearly all children had CLASS scores in three time periods (94%), and complete data on other covariates (80.6%), we aimed to maintain a large sample. Therefore, we used multilevel imputation using mice (version 3.11.0; Van Buuren & Groothuis-Oudshoorn, 2011) and miceadds (version 3.10.28; Robitzsch & Grund, 2020) packages to impute missing covariate information. For covariates that were stable at the child level (e.g., gender) we used two-level-only predictive mean matching, while for time-varying covariates (e.g., CLASS) we used normal multilevel imputation (ml.lmer) that includes group mean centering to accommodate our substantive model of interest. We also included interactions between CLASS subscales and all other covariates in predicting cognitive outcomes using the interactions option. A study evaluating imputation methods for longitudinal

designs showed the normal two-level imputation had superior performance to other multilevel options (Huque et al., 2018). One hundred imputed datasets were created, each being run for 30 iterations and the visit sequence monotonically increasing from least to most missing data. Convergence was visually assessed via trace plots, and imputed data distributions were visually evaluated to non-missing data—these evaluations revealed no major issues with the imputation (Van Buuren, 2012). Results from multiple imputed data sets were combined using the mice, Amelia (version 1.7.6; Honaker et al., 2011), merTools (version 0.5.2; Knowles & Frederick, 2020), and broom. Mixed (version 0.2.6; Bolker & Robinson, 2020) packages.

RESULTS

Summary statistics of child and family covariates are presented in Table 1. These show we obtained a comprehensive analytical sample with expected representation of race, language, and education for a sample recruited with ECE services where employed caregivers are over-represented (Table S3 compares to excluded sample and relevant population). Moreover, the mean and variability of child temperament and proportion of children with developmental delays or experiencing adverse life events suggest a diverse, yet typical, range of developmental typologies are included. Further details on the sample comprising correlations, averages,

and standard deviations of key child variables are presented in Supporting Information (Table S5 imputed; Table S6 observed). These show that CLASS-ES in 2010 (Pre-K) was most reliably correlated with cognitive outcomes in 2010, 2011, and 2012, while CLASS-IS and CLASS-CO had more reliable correlations with cognitive outcomes in later years when more children were in formal school (e.g., 43% at formal school in 2011; 89% in 2012). The average correlation between subsequent cognitive measures was strongest for verbal comprehension ($r = .68$) and weakest for concept formation ($r = .45$). Verbal comprehension and applied problems had the highest between cognitive outcome correlation ($r = .69$, 2010), while the lowest was between concept formation and understanding directions ($r = .41$, 2010). The average correlation between age and cognitive outcomes remained stable across time ($r = .43$, .42, .42). Additional information includes the proportion of children in each grade (Table 2; and proportion rooms in Table S7), number of children per room (Table S4), and grade trajectories of the included (Table S8) and excluded (Table S9) sample. Of note, 52.4% of child observations were during pre-K, 42.7% formal schooling, and 5.9% unknown. Additionally, when known, the most common pre-K settings for children were stand-alone (56%) or center-based services (41%), whilst formal school settings were predominantly Kindergarten (100% 2011; 53% 2012) and year 1 (47% 2012). The excluded sample had comparative grade trajectories but was more likely to be missing room information due to the exclusion criteria.

TABLE 1 Covariates and sample characteristics of the children and caregivers

| | <i>n</i> | <i>M</i> or % | <i>SD</i> | % missing |
|--|----------|---------------|-----------|-----------|
| Child: female | 1128 | 0.48 | | 0 |
| Child: Indigenous | 1121 | 0.03 | | 0.006 |
| Child: born overseas | 1126 | 0.06 | | 0.002 |
| Child: difficult temperament | 947 | 1.98 | 0.66 | 0.16 |
| Child: developmental delay | 937 | 0.05 | | 0.169 |
| Child: English primary language (ref) | 1125 | 0.84 | | 0.003 |
| Child: English primary, speaks another language | 1125 | 0.11 | | 0.003 |
| Child: English not primary language | 1125 | 0.04 | | 0.003 |
| Stressful life event that seriously affected child | 944 | 0.11 | | 0.163 |
| Caregiver Indigenous | 1105 | 0.01 | | 0.02 |
| Caregiver born overseas | 1109 | 0.22 | | 0.017 |
| Caregiver has health care benefits card | 1105 | 0.22 | | 0.02 |
| Caregiver education: post-graduate (ref) | 1110 | 0.19 | | 0.016 |
| Caregiver: bachelor degree | 1110 | 0.3 | | 0.016 |
| Caregiver: diploma | 1110 | 0.13 | | 0.016 |
| Caregiver: year 12 or Tafe | 1110 | 0.29 | | 0.016 |
| Caregiver: year 10 or lower | 1110 | 0.09 | | 0.016 |
| Caregiver: psychological distress | 938 | 14.48 | 4.22 | 0.168 |
| Home learning environment | 941 | 3.25 | 1.14 | 0.166 |

TABLE 2 Classroom grade characteristics reported as the proportion of children (*n* children)

| Year | Grade | | | Formal schooling | | | Missing grade | Missing room |
|------|-------------|--------------|--------------|------------------|-----------|-----------|---------------|--------------|
| | Pre-K | | | K | Y1 | SC/SU | | |
| | Stand alone | Centre based | Family based | | | | | |
| 2010 | 0.47 (529) | 0.44 (496) | 0.03 (37) | | | | 0.04 (43) | 0.02 (23) |
| 2011 | 0.36 (411) | 0.19 (210) | 0.01 (9) | 0.41 (467) | | 0.02 (18) | | 0.01 (13) |
| 2012 | 0.05 (55) | 0.02 (26) | <0.01 (1) | 0.46 (523) | 0.4 (455) | 0.03 (35) | <0.01 (1) | 0.03 (32) |

Note: K = kindergarten, Y1 = year 1, SC = school combined/unknown (K and Y1; K or Y1). Family-based care includes limited hours care. Children without a room were assigned to a unique room by year room number (one room for each year) for modeling.

TABLE 3 Quality information of observed rooms

| Year | CLASS measure | <i>N</i> | <i>M</i> | <i>SD</i> | Min | Max |
|------|------------------------|----------|----------|-----------|------|------|
| 2010 | Emotional support | 219 | 5.13 | 0.92 | 2.58 | 6.9 |
| | Instructional support | 219 | 2.36 | 0.96 | 1 | 5.73 |
| | Classroom organization | 219 | 4.58 | 0.92 | 1.92 | 6.5 |
| 2011 | Emotional support | 599 | 5.47 | 0.64 | 3 | 6.85 |
| | Instructional support | 599 | 2.52 | 0.81 | 1 | 6.33 |
| | Classroom organization | 599 | 5.17 | 0.73 | 2.5 | 6.83 |
| 2012 | Emotional support | 800 | 5.44 | 0.61 | 2.92 | 6.9 |
| | Instructional support | 800 | 2.75 | 0.83 | 1.06 | 5.94 |
| | Classroom organization | 800 | 5.37 | 0.72 | 2.47 | 7 |

Abbreviation: CLASS, Classroom Assessment Scoring System.

Classroom characteristics are in Table 3. Averages of quality, as assessed by CLASS, increased as children moved into formal school grades, and the variability in scores decreased. Average CLASS-ES and CLASS-CO were generally above the middle-quality range (3–5) of the scales, while CLASS-IS was marginally above the low-quality range (1–2). There were high correlations between CLASS scores at each assessment period (e.g., CLASS-ES and CLASS-CO $r = .86$ in 2010; Table S5). However, CLASS scores had low correlations across time (e.g., CLASS-ES 2010 and 2011, $r = .18$; Table S5), reflecting children moving to rooms with different levels of observed quality.

Step 1: Child variation in class scores

Estimating the ICC for the imputed sample (Table S10) showed substantial within-child variability in CLASS-ES (ICC = .08) and CLASS-IS (ICC = .09), while CLASS-CO was particularly variable (ICC = .03). To confirm these results, we ran the same analysis on the sample with CLASS at three time points (n observations = 3180; n children = 1060) and found similar variation (CLASS-ES = 0.08; CLASS-IS = 0.08; CLASS-CO = 0.03; Table S11). Figure 1a plots 20 individual trajectories over

time in comparison to the yearly overall average from 100 imputed datasets.

Step 2: Trajectories of cognitive ability

All measures of the children's cognitive ability increased over time. Figure 1b plots a sample of individual brief intelligence assessment trajectories over time as well as the overall predicted trend from the growth model (Table S12).

Step 3: Effects of CLASS on academic outcomes

Change in quality

Table 4 presents the key coefficients from the adjusted hybrid regression models. This shows that increases in CLASS-ES were the only dimension reliably associated with improved cognitive outcomes, specifically, verbal comprehension ($\beta = 0.54$ [95% CI: 0.1–0.99]). In terms of standardized effect sizes on verbal comprehension, the point estimate for a one standard deviation change in CLASS-ES (0.61 units; Table S13) equates to a 0.02 standard deviation change in verbal comprehension (overall

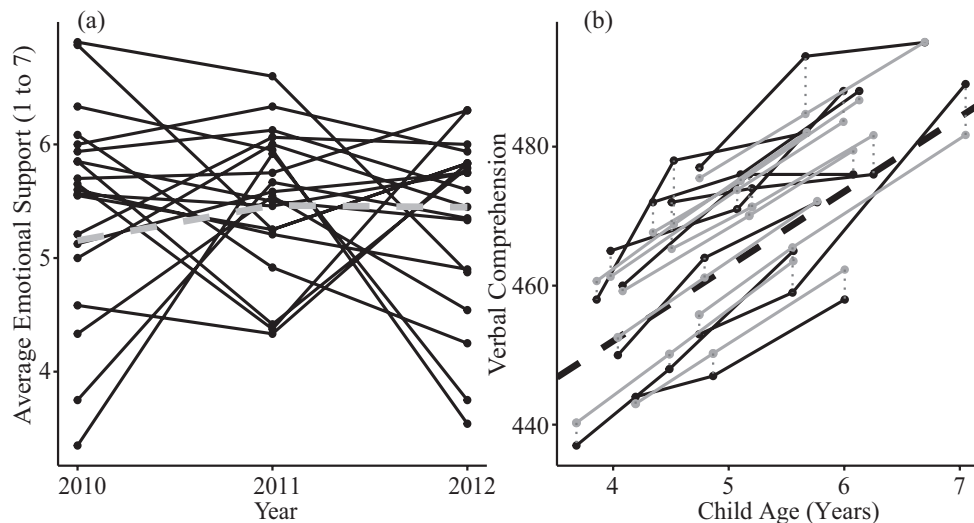


FIGURE 1 (a) Twenty randomly selected child trajectories of emotional support (average in classroom) as measured by Classroom Assessment Scoring System, with dashed line indicating overall average in each year from 100 imputed data sets. (b) Ten randomly selected trajectories of child verbal comprehension scores with predicted values (dotted line connects observation [black] to prediction [grey]) from random intercept and slope growth model. Thick dashed line is the average prediction

$SD = 15.78$). In practice, a 1-point change in CLASS-ES yields a 0.54 change in language scores (5/100 of the effect of aging 1 year; $\beta = 10.72$ [95% CI: 10.37–11.07]).

Higher average instructional support was reliably associated with higher verbal comprehension in the adjusted (Table 3) and unadjusted models (Tables S15–S29). However, the number of reliable associations ($p > .05$) between average CLASS domains and cognitive outcomes in the adjusted model (1 of 15; Table 3) was much lower than that for the unadjusted models (14 of 15; Tables S14–S28).

Step 4: Threshold analyses—Change in quality \times average quality

Two statistically reliable interactions between the average level of CLASS-ES and changes in CLASS-ES predicting cognitive outcomes were identified (Table 4; Tables S14–S28 contain full results, and Tables S29 and S30 for regions of significance tests). Both related to language outcomes: verbal comprehension and understanding directions.

These interactions show whether the effects of changes in quality are dependent on both, or either, the absolute value of quality and the average quality experienced. They identify the threshold at which a change in quality is meaningful and the average dosage of quality required to benefit from these changes. If there were no threshold effects, changes in quality would be the same for children who are in high-quality classrooms on average and children who are in low-quality classrooms on average.

Figure 2a presents the interaction effect for verbal comprehension. This shows, in line with information from the regions of significance testing, that children

with average CLASS-ES across time greater than 5.06 (high-mid range on CLASS; average 5.35, $SD = 0.49$), compared to children with an average below 5.06, experienced a significant improvement in verbal comprehension when there was an increase in CLASS-ES greater than 0.2 (average change ~ 0 , $SD = 0.61$). In absolute terms, there was a significant increase in verbal comprehension for a child with an average CLASS-ES of 5.06 if they moved to a room with CLASS-ES of at least 5.26. On the other hand, a child with an average of 4.06 was not observed to benefit from a shift to 4.26, or 5.26. That is, the benefits of higher emotional support for children's verbal comprehension are dependent on both reaching an exposure threshold across time of CLASS-ES within the high-mid range, and a change in quality into the high-mid range. This positive gain for children when CLASS-ES increased by at least 0.2 was well within two standard deviations of observed changes and informed by 28% of observations that had an average CLASS-ES greater than 5.08, making it reliable (within the gray shaded regions of significance). On the other hand, the negative slope was not significant with the range of the data and is considered unreliable ($p > .05$). Suggesting children who are, on average, in classrooms with CLASS-ES greater than 5.06 did not have a comparative decline in verbal comprehension when quality declined below this level.

Figure 2b presents the interaction effect for understanding directions. The effect was statistically reliable when CLASS-ES increased by 0.84 (within two standard deviations), but the contrasting average groups (average less than 4.65 or greater than 5.83) were only informed by 0.59% and 0.35%, of observations, respectively. Likewise, the left side of the interaction was statistically reliable when CLASS-ES declined by -0.91 (within two standard deviations), but only informed by 2.33% and

TABLE 4 Hybrid regression results, including threshold analysis

| Outcome | Change | Average | Change × average emotional support |
|-------------------------------|---------------------------------|------------------------------|------------------------------------|
| | Emotional support | Emotional support | |
| Concept formation | -0.32 (-1.14 to 0.5) | 1.23 (-0.36 to 2.81) | |
| | -7.28 (-15.54 to 0.98) | 1.22 (-0.36 to 2.8) | 1.36 (-0.25 to 2.97) |
| Visual matching | 0.2 (-0.37 to 0.77) | -0.12 (-1.37 to 1.14) | |
| | 1.94 (-3.77 to 7.65) | -0.11 (-1.37 to 1.14) | -0.34 (-1.45 to 0.77) |
| Verbal comprehension | 0.54 (0.1 to 0.99)* | 0.96 (-0.2 to 2.12) | |
| | -5.65 (-10.15 to -1.15)* | 0.96 (-0.19 to 2.12) | 1.21 (0.33 to 2.08)** |
| Understanding directions | -0.16 (-0.79 to 0.46) | 0.08 (-1.12 to 1.28) | |
| | -9.71 (-16 to -3.43)** | -0.02 (-1.22 to 1.18) | 1.86 (0.64 to 3.09)** |
| Applied problems | -0.1 (-0.87 to 0.67) | 0.5 (-1.31 to 2.3) | |
| | 3.36 (-4.56 to 11.27) | 0.55 (-1.26 to 2.35) | -0.67 (-2.21 to 0.87) |
| Instructional support | | | |
| Concept formation | 0.22 (-0.47 to 0.91) | 0.93 (-0.41 to 2.27) | |
| | -1.36 (-4.92 to 2.2) | 0.93 (-0.41 to 2.27) | 0.56 (-0.67 to 1.8) |
| Visual matching | 0.04 (-0.43 to 0.52) | -0.05 (-1.11 to 1) | |
| | 1.87 (-0.54 to 4.28) | -0.05 (-1.1 to 1.01) | -0.65 (-1.49 to 0.19) |
| Verbal comprehension | 0.25 (-0.12 to 0.63) | 1.46 (0.49 to 2.42)** | |
| | 0.46 (-1.48 to 2.4) | 1.46 (0.49 to 2.42)** | -0.07 (-0.75 to 0.6) |
| Understanding directions | -0.18 (-0.7 to 0.35) | 0.98 (-0.03 to 1.99) | |
| | 1.38 (-1.31 to 4.08) | 0.99 (-0.02 to 2) | -0.55 (-1.49 to 0.38) |
| Applied problems | 0.18 (-0.47 to 0.83) | 0.6 (-0.91 to 2.11) | |
| | 2.07 (-1.23 to 5.37) | 0.62 (-0.89 to 2.13) | -0.67 (-1.82 to 0.48) |
| Classroom organization | | | |
| Concept formation | -0.16 (-0.91 to 0.58) | 1.15 (-0.29 to 2.6) | |
| | -2.53 (-8.78 to 3.72) | 1.15 (-0.29 to 2.59) | 0.49 (-0.79 to 1.77) |
| Visual matching | 0.4 (-0.11 to 0.92) | 0.73 (-0.41 to 1.87) | |
| | 1.71 (-2.59 to 6.02) | 0.74 (-0.4 to 1.88) | -0.27 (-1.15 to 0.61) |
| Verbal comprehension | 0.32 (-0.09 to 0.73) | 1.03 (-0.02 to 2.08) | |
| | 0.25 (-3.21 to 3.7) | 1.03 (-0.02 to 2.08) | 0.01 (-0.69 to 0.72) |
| Understanding directions | -0.03 (-0.6 to 0.54) | 0.41 (-0.68 to 1.5) | |
| | -3.94 (-8.73 to 0.84) | 0.33 (-0.77 to 1.42) | 0.81 (-0.17 to 1.79) |
| Applied problems | 0.03 (-0.67 to 0.73) | 1.39 (-0.24 to 3.02) | |
| | 3.11 (-3.09 to 9.31) | 1.48 (-0.16 to 3.12) | -0.63 (-1.9 to 0.63) |

Note: The interaction term must be significant to interpret the average and change terms in the models containing an interaction. Bold text is used to highlight reliable and interpretable terms.

* $p < .05$; ** $p < .01$.

2.54%, of observations, respectively. Thus, given the seemingly sporadic contrasting interaction shape and small number of observations we considered this interaction unreliable.

Supplementary analysis: Categorical thresholds

Re-estimating step 3 models with categorical cut-points for CLASS supported the results from continuous models. Specifically, when children changed from rooms in the lowest 25th percentile of emotional support to those above the 75th there was an increase in verbal

comprehension ($\beta = 1.25$ [95% CI: 0.33–2.16]). This reflects a threshold of 5.88, above which emotional support influenced verbal comprehension. This is higher than the 5.26 revealed in step 4, highlighting that the continuous interaction revealed thresholds with more nuance. There were also increases in concept formation when instructional support went to the 25th to 50th percentile ($\beta = 1.25$ [95% CI: 0.33–2.16]), and increases in verbal comprehension when classroom organization went up to the 25th to 50th percentile ($\beta = 1.25$ [95% CI: 0.33–2.16]), however, as these cut-points were not along an increasing continuum of effects (e.g., unreliable for children with quality >75th percentile) it suggests they

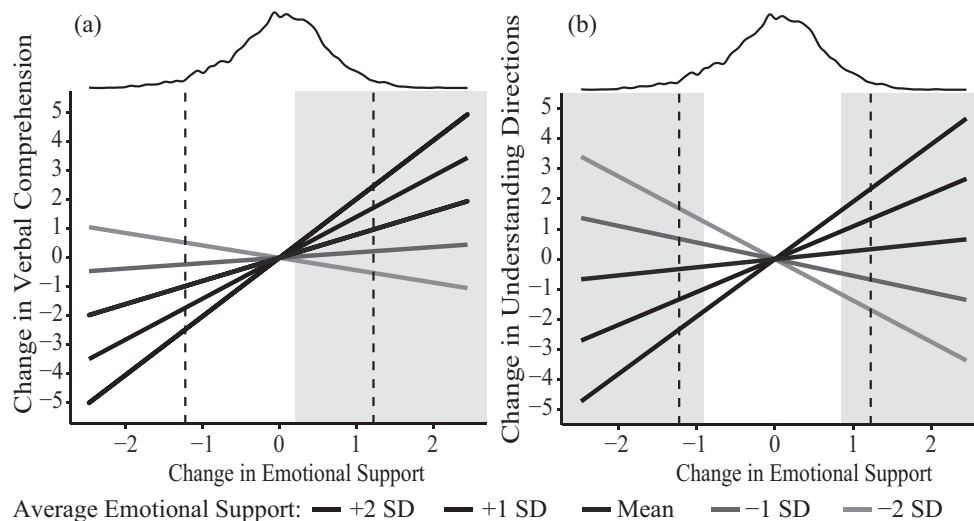


FIGURE 2 Effect of changes in emotional support on (a) verbal comprehension and (b) understanding directions given the average level of emotional support across the years 2010, 2011, and 2012. Regions of significance, where dependent on average emotional support changes in emotional support have a significant influence on the outcome, are indicated by grey rectangle. Dashed line indicates two standard deviations above and below average change in emotional support. Density plot at top of figure is the distribution of changes in emotional support from 100 imputed data sets

are sporadic. See Table S31 for distribution of categorical coding, Table S32 for key results, and Tables S33–S47 for full model results.

DISCUSSION

A key task for research in ECE is to identify program components that are effective in delivering positive learning experiences that optimize school readiness and establish ongoing positive educational trajectories. With that aim we applied a within-child design to assess the effect of change in exposure to interactional quality, measured using CLASS, on children's rates of cognitive development and academic attainment. A within-child design provided a more stringent test of effect by controlling fixed child characteristics. Additionally, we conducted threshold analyses to assess the potential of nonlinear associations. Our sample captured the diversity of ECE provision in Australia, including center-based and family-based day care and stand-alone pre-K programs and captured a socially diverse population of children ($N = 1128$). The children were tracked across 3 years as they moved from their pre-K setting through to school enabling test of the association of variation in quality experienced by each child and their cognitive outcomes. Our within-child approach has yielded more certain findings than those that could be generated by focussing on average classroom quality. While we found the average quality was uniformly associated with higher academic ability in unadjusted models these associations were not robust once adjusting for covariates. This finding highlights the clear distinction between models that are confounded by between-child differences, and those

that apply within-person models to attenuate the effects of pre-existing and stable confounding differences between children (Hoffman & Stawski, 2009; Maldonado-Carreño & Votruba-Drzal, 2011; Watts et al., 2021).

Variation in children's exposure to process quality

Using CLASS domains as our measures, we identified substantial heterogeneity in the quality of early education to which each child was exposed across 3 years of education. Indeed, the variation in CLASS-assessed quality experienced by each child was vastly greater than that experienced between children. This variation in the quality of programs experienced by each child across time is a necessary condition for linking change in quality to differential development. However, the extent of variation is striking given the diversity of the sample that includes not only low-income families but also those with extensive resources who might reasonably be assumed to provide stable access to high-quality ECE experiences. That is the education system is achieving a level of equality in that, regardless of family resource, children on average were not consistently securing higher-quality ECE experiences. One explanation for high variability in quality of teaching, consistent with reports from other countries, is that higher structural resource, while an enabler, is not a guarantee of high-quality interactional experiences (Mashburn et al., 2008; Slot et al., 2015). Thus, although structural aspects of ECE experiences are now often standardized through legislated regulatory mechanisms (Thorpe, Rankin, et al., 2020), variation in process quality persists.

Defining and measuring process quality is a challenging and contested area of scholarship (Burchinal, 2018). Consistency of measurement is difficult to achieve (Pianta et al., 2016). Although, for a variety of reasons (e.g., Pianta et al., 2007), educators will vary in process quality they deliver, measurement error may also amplify variation and may contribute to the findings in this study. Although the assessment of process quality in this study was undertaken by CLASS-certified observers, the CLASS reliability criterion is set at 80% accuracy across ratings within one score of master coder on a scoring system that has a range of 1 to 7. As a score of one is greater than the standard deviation of observed CLASS dimensions, there may be unadjusted measurement error that has inflated the calculated variability of a child's experience across time. In addition, we had two or three observations of CLASS per-child. While two observations per-child are adequate to reliably estimate high ICC's (>.6), more observations are preferable to reliably estimate low ICC's (<.2) and further observations would better clarify the variability (Shoukri et al., 2004). Noting these limitations, the variability in each child's exposure to quality across classrooms was sufficient to proceed with a within-child analysis that linked changes to academic development.

Effects of process quality

Our main finding was that, on average, children who moved from an ECE room with lower to one with higher CLASS-ES had a comparative increase in verbal comprehension. Informed by a within-child design, this finding broadly aligns with prior studies that show an association of CLASS domains and child development and attainment outcomes, and specifically identify language outcomes (Hong et al., 2019; Mashburn et al., 2008; Ulferts et al., 2019). Two recent meta-analyses, one deriving from Europe and the other from the USA, report an association between rates of growth in language outcomes and CLASS in ECE samples, but not with mathematics or social skills (Hong et al., 2019; Ulferts et al., 2019). The association between language and literacy seen in these prior studies and in our Australian study likely reflects the high language focus in assessing quality in ECE classrooms. Our findings specifically identify CLASS-ES, but not CLASS-IS, as associated with verbal comprehension. Given that CLASS-IS has a strong focus on language inputs from teachers we may have anticipated this as a predictor of child language development. One likely explanation is low variability in CLASS-IS in our sample ($av = 2.07$ on a scale of 1–7), and in other international samples (Mashburn et al., 2008; Salminen et al., 2012; Tayler et al., 2013; Von Suchodoletz et al., 2014), reflects either generally poor instruction or, a floor effect in the measure (Thorpe et al., 2022). A further explanation relates to the content of the measure. While CLASS-IS

focuses on teacher-driven language inputs (e.g., extension of vocabulary), child led language interactions are a less dominant focus of this measurement domain. In contrast, CLASS-ES focuses on teacher–child emotional relationship and includes *regard for the child's perspective* and *teacher sensitivity* to the child. The mechanisms connecting emotional support and verbal comprehension are likely associated with rich conversations (“serve and return”) that engage children, afford children opportunity to make inputs, and enable educators to scaffold verbal comprehension (Thorpe et al., 2021).

A converse notable insight from our findings is the limitation of reliable findings to a single CLASS domain and child outcome. Despite an extensive research protocol, empirically robust design, and detailed attention to reliably measuring both ECE quality and outcomes, no reliable associations were found between changes in CLASS-ES and child outcomes beside those for verbal comprehension, nor were there associations between child outcomes and changes in CLASS-IS or CLASS-CO. This finding is consistent with recent meta-analyses that have also observed many unreliable effects (Hong et al., 2019; Ulferts et al., 2019). The results also draw attention to the considerable progress still required to fully understand the aspects of quality that deliver extensive benefits to children (Pianta et al., 2016). Inputs to child development outcomes are complex and convergence in the consistency of quality effects remains elusive. To identify policy and practice actions that achieve significant benefit to children requires that research efforts remain targeted on identifying measures of quality that are theoretically robust, reliable, and predictive of child outcomes. Consideration of broader structural and contextual factors (e.g., Thorpe, Jansen, et al., 2020; Thorpe, Searle, et al., 2020) that may limit capacity to deliver high process quality should also be considered.

Thresholds of process quality

Our threshold analyses identified only one reliable effect. We found that children, who on average, were exposed to mid-high CLASS-ES across time (CLASS scores >5.06) had higher verbal comprehension scores when CLASS-ES increased. Likewise, coding CLASS categorically confirmed this finding, albeit with a higher threshold (CLASS-ES >5.88). These findings reflect reports that suggest the higher thresholds of process quality are required to generate the greatest effect (Burchinal et al., 2016), and that cumulative exposure to high quality provides the most benefit (Vernon-Feagans et al., 2019). The current finding draws attention to the importance of dimensions of classroom quality captured with the CLASS-ES domain across the early years of education, Pre-K through K and year 1. The quality improvement agenda has driven the implementation of educationally focused curricula and interventions focused on teacher

instruction (Organisation for Economic Co-operation and Development, 2021). Such interventions are not intended to remove child inputs or limit inquiry (play-based) learning. Yet, there are ongoing concerns about the intensification of focus on academic readiness, emphasis on more structured learning formats, and reduction of time for child-directed, self-initiated exploration (Yogman et al., 2018). This shift is particularly evident in the transition to formal schooling (K and year 1) where increases in CLASS-IS and CLASS-CO scores are seen with commensurate decline in CLASS-ES (Thorpe, Rankin, et al., 2020; Thorpe et al., 2021, 2022). Our current findings show that CLASS-ES, a quality domain that includes an emphasis on interactional features such as *regard for student perspective* and *teacher sensitivity*, is important in the promotion of language development within the early years of school. One prior study suggested a construct of responsivity underpins CLASS-ES and CLASS-CO (Hamre et al., 2014). While such a construct is potentially underpinning our findings, these were exclusive to Emotional Support and consistent with evidence for three CLASS domains (Li et al., 2020; Sandilos et al., 2014). Our findings flag the dual pedagogical task of improving instructional supports while enabling child inputs into learning. That is, our results show that emotionally supportive environments should not run contrary to educationally focussed instruction but be integral to early years pedagogical approaches that support long-term trajectories of learning (Ansari et al., 2020; Pianta et al., 2021).

Limitations

While the within-child approach provided in this study has the benefits of adjusting for time-invariant differences between children, like all other observational cohort studies our study has limitations that should be considered in interpreting our results (Rosenbaum, 2020). One limitation relates to measurement. There is the potential that variation in quality assessed using CLASS is an artifact of measurement error. While compared to alternative measures, CLASS has been found to be more sensitive to aspects of the classroom environment that predict child development outcomes (Hong et al., 2019; Ulferts et al., 2019) problems of reliability may provide an alternative explanation for fluctuation of quality scores between classrooms (Mantzicopoulos et al., 2018; Styck et al., 2021). Specifically, the criteria for a reliable observation (within one, on the seven-point scale) is equivalent to one standard deviation of observed scores. Thus, instability of quality experiences might be explained by this low reliability standard or by systematic observer differences in applying this standard, given that multiple observers undertook CLASS assessments. Another relates to statistical analyses. While we adjust for time-invariant between-child differences by design,

there may be unmeasured time-variant within-child biases, such as changes in parenting, the home learning environment, and additional educational inputs, we have not accounted for (Rothstein, 2009). In addition, though we assume changes in quality drives cognitive development, our correlational design means we cannot discount the possibility that educators may be able to achieve higher levels of observed quality when teaching children with higher cognitive development (Watts et al., 2021). However, as observations followed individual children into different classrooms with different child compositions, it is unlikely that each child would substantially drive classroom quality across 3 years. Third, to maximize information use we estimated thresholds using the continuous interaction between average quality and changes in quality, as well as analyzing categorical cut-points. Although residuals of the model indicated further exploration unnecessary, we note alternative approaches, such as polynomial nonlinear terms, could be used to investigate thresholds.

Directions for future research

Developmental gains resulting from inputs of ECE have frequently been found to fade out across time, as children enter school environments and classrooms in which children who have not received the benefit of high-quality ECE are also present (Bailey, Duncan, et al., 2020). The composition of a class can have marked effects of child development outcomes (Coley et al., 2019). Fade out would be anticipated when an educator in the new context fails to sustain gains, while sustaining environments are those that enable the continuation or growth of initial gains. Although meta-analyses identify insufficient and limited evidence on sustaining environments (Bailey, Jenkins, & Alvarez-Vargas, 2020), our study provides evidence that maintaining and improving quality across the early years yields ongoing benefits. Our study points to the benefits of within-child approaches and the potential to consider measurement of class composition and other features of ongoing educational environments to understand mechanisms of sustaining versus fade out. Ongoing tracking of samples throughout the school years via data linkage to school records also presents a direction for understanding long-term effects. Regarding the E4kids sample analyzed here, such linkage has been achieved and will be the focus of ongoing analyses.

CONCLUSION

This study aligns with a large body of research finding a limited association between measures of CLASS and children's development and attainments. Consistent with these prior studies, language outcomes are those reliably predicted by CLASS. However, the findings of association

are limited to CLASS-ES and language. The constraint of findings to this one reliable result may reflect measurement limitations, whether of reliability, scope, or specificity. In the case of our Australian study, our results indicate that the emotional quality of the classroom environment matters for children's language attainment and that the higher the support the better. As our measure of emotional quality, CLASS-ES, directs attention to teacher-child relationship and regard for student perspectives we suggest that gains in language development reflect the opportunity for child inputs in the learning environment and the affordance of "serve and return" interactions that scaffold language. An important message from this finding is that the emotional qualities of the environment are not ancillary to instructional support but rather integral to young children's learning.

ACKNOWLEDGMENTS

This research was supported (partially or fully) by the Australian Government through the Australian Research Council's Centre of Excellence for Children and Families over the Life Course (Project ID CE200100025). This study was conducted at the University of Queensland using data from E4Kids under license from the University of Melbourne. E4Kids was funded by the Australian Research Council Linkage Projects Scheme (LP0990200), in collaboration with the Victorian Government Department of Education and Early Childhood Development, and the Queensland Government Department of Education and Training. The authors thank the ECE services, directors, teachers/staff, children, and their families who participated in E4Kids. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.


DATA AVAILABILITY STATEMENT

The data necessary to reproduce the analyses presented here are not publicly accessible. Code and materials to replicate the analyses are publicly accessible and available from the first author. The analyses presented here were not preregistered.

ORCID

Peter Sheldon Rankin  <https://orcid.org/0000-0003-4872-9055>

Sally Staton  <https://orcid.org/0000-0002-9741-8010>

Azhar Hussain Potia  <https://orcid.org/0000-0002-8845-3180>

Sandy Houen  <https://orcid.org/0000-0002-0904-9906>

Karen Thorpe  <https://orcid.org/0000-0001-8927-4064>

REFERENCES

- Adams, R., & Wu, M. (Eds.). (2003). *Programme for international student assessment (PISA): PISA 2000 technical report*. OECD Publishing.
- Adhikari, P. (2006). *Socio-economic indexes for areas: Introduction, use and future directions*. Australian Bureau of Statistics.
- Ananth, C. V., & Schisterman, E. F. (2017). Confounding, causality, and confusion: The role of intermediate variables in interpreting observational studies in obstetrics. *American Journal of Obstetrics and Gynecology*, 217(2), 167–175. <https://doi.org/10.1016/j.ajog.2017.04.016>
- Ansari, A., Hofkens, T. L., & Pianta, R. C. (2020). Teacher-student relationships across the first seven years of education and adolescent outcomes. *Journal of Applied Developmental Psychology*, 71, 101200.
- Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97.
- Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, 56, 100910.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Belsky, J., Caspi, A., Moffitt, T. E., & Poulton, R. (2020). *The origins of you: How childhood shapes later life*. Harvard University Press.
- Blanca, M. J., Alarcón, R., & Bono, R. (2018). Current practices in data analysis procedures in psychology: What has changed? *Frontiers in Psychology*, 9, 2558. <https://doi.org/10.3389/fpsyg.2018.02558>
- Bolker, B., & Robinson, D. (2020). *broom.mixed: Tidying methods for mixed models. R package version 0.2.6 (R package version 0.2.6)*. <https://cran.r-project.org/package=broom.mixed>
- Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407–428.
- Bull, R., Yao, S.-Y., & Ng, E. L. (2017). Assessing quality of kindergarten classrooms in Singapore: Psychometric properties of the Early Childhood Environment Rating Scale—Revised. *International Journal of Early Childhood*, 49(1), 1–20.
- Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3–9. <https://doi.org/10.1111/cdep.12260>
- Burchinal, M., Xue, Y., Auger, A., Tien, H.-C., Mashburn, A., Cavadel, E. W., & Peisner-Feinberg, E. (2016). Quality thresholds, features, and dosage in early care and education: Methods. *Monographs of the Society for Research in Child Development*, 81(2), 27–45. <https://doi.org/10.1111/mono.12237>
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, 343(6178), 1478–1485. <https://doi.org/10.1126/science.1248429>
- Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., Barbarin, O. A., Sparling, J. J., & Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology*, 48(4), 1033–1043. <https://doi.org/10.1037/a0026644>
- Cash, A. H., Ansari, A., Grimm, K. J., & Pianta, R. C. (2019). Power of two: The impact of 2 years of high quality teacher child interactions. *Early Education and Development*, 30(1), 60–81. <https://doi.org/10.1080/10409289.2018.1535153>
- Cloney, D., Nguyen, C., Adams, R. J., Tayler, C., Cleveland, G., & Thorpe, K. (2017). Psychometric properties of the Classroom Assessment Scoring System (Pre-K): Implications for measuring interaction quality in diverse early childhood settings. *Journal of Applied Measurement*, 18(3), 299–318.
- Coley, R. L., Spielvogel, B., & Kull, M. (2019). Concentrated poverty in preschools and children's cognitive skills: The mediational role of peers and teachers. *Journal of School Psychology*, 76, 1–16.

- Curby, T. W., Brock, L. L., & Hamre, B. K. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Education & Development, 24*(3), 292–309. <https://doi.org/10.1080/10409289.2012.665760>
- Fukking, R., Jilink, L., & Oostdam, R. (2017). A meta-analysis of the impact of early childhood interventions on the development of children in the Netherlands: an inconvenient truth? *European Early Childhood Education Research Journal, 25*(5), 656–666. <https://doi.org/10.1080/1350293x.2017.1356579>
- Gordon, R. A., Hofer, K. G., Fujimoto, K. A., Risk, N., Kaestner, R., & Korenman, S. (2015). Identifying high-quality preschool programs: New Evidence on the Validity of the Early Childhood Environment Rating Scale–Revised (ECERS-R) in relation to school readiness goals. *Early Education and Development, 26*(8), 1086–1110.
- Green, M. J., Pearce, A., Parkes, A., Robertson, E., & Katikireddi, S. V. (2021). Pre-school childcare and inequalities in child development. *SSM—Population Health, 14*, 100776. <https://doi.org/10.1016/j.ssmph.2021.100776>
- Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365–379.
- Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257–1274. <https://doi.org/10.1111/cdev.12184>
- Harms, T., Clifford, M. R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale*. Teachers College Press.
- Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly, 36*, 561–571.
- Heckman, J. J., Humphries, J. E., & Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy, 126*(Suppl 1), S197–S246. <https://doi.org/10.1086/698760>
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the high/scope perry preschool program. *Journal of Public Economics, 94*(1–2), 114–128.
- Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6*(2–3), 97–120. <https://doi.org/10.1080/15427600902911189>
- Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research, 11*(2), 213–218.
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*(7), 1–47. <https://doi.org/10.18637/jss.v045.i07>
- Hong, S. L. S., Sabol, T. J., Burchinal, M. R., Tarullo, L., Zaslow, M., & Peisner-Feinberg, E. S. (2019). ECE quality indicators and child outcomes: Analyses of six large child care studies. *Early Childhood Research Quarterly, 49*, 202–217.
- Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology, 18*(1), 168. <https://doi.org/10.1186/s12874-018-0615-6>
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*(6), 959–976.
- Knowles, J. E., & Frederick, C. (2020). *merTools: Tools for analyzing mixed effect regression models (R package version 0.5.2)*. <https://cran.r-project.org/package=merTools>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(1), 1–26.
- Landry, S. H., Zucker, T. A., Williams, J. M., Merz, E. C., Guttentag, C. L., & Taylor, H. B. (2017). Improving school readiness of high-risk preschoolers: Combining high quality instructional strategies with responsive training for teachers and parents. *Early Childhood Research Quarterly, 40*, 38–51. <https://doi.org/10.1016/j.ecresq.2016.12.001>
- Layzer, J. I., & Goodson, B. D. (2006). The “quality” of early care and education settings. *Evaluation Review, 30*(5), 556–576. <https://doi.org/10.1177/0193841X06291524>
- Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher-child interactions in Chile and their associations with prekindergarten outcomes. *Child Development, 86*(3), 781–799.
- Li, H., Liu, J., & Hunter, C. V. (2020). A meta-analysis of the factor structure of the Classroom Assessment Scoring System (CLASS). *The Journal of Experimental Education, 88*(2), 265–287.
- Li, K., Zhang, P., Hu, B. Y., Burchinal, M. R., Fan, X., & Qin, J. (2019). Testing the ‘thresholds’ of preschool education quality on child outcomes in China. *Early Childhood Research Quarterly, 47*, 445–456. <https://doi.org/10.1016/j.ecresq.2018.08.003>
- Lohndorf, R. T., Vermeer, H. J., de la Harpe, C., & Mesman, J. (2021). Socioeconomic status, parental beliefs, and parenting practices as predictors of preschoolers' school readiness and executive functions in Chile. *Early Childhood Research Quarterly, 57*, 61–74.
- Maldonado-Carreño, C., & Votruba-Drzal, E. (2011). Teacher-child relationships and the development of academic and behavioral skills during elementary school: A within- and between-child analysis. *Child Development, 82*(2), 601–616.
- Mantzicopoulos, P., French, B. F., Patrick, H., Watson, J. S., & Ahn, I. (2018). The stability of kindergarten teachers' effectiveness: A generalizability study comparing the framework for teaching and the Classroom Assessment Scoring System. *Educational Assessment, 23*(1), 24–46.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>
- Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Pender, E., Rentzou, K., Tawell, A., Slot, P., Broekhuizen, M., & Leseman, P. (2015). *A review of research on the effects of Early Childhood Education and Care (ECEC) upon child development. EU CARE project*. CARE project, European Early Childhood Education and Care, European Union.
- Micalizzi, L., Brick, L. A., Flom, M., Ganiban, J. M., & Saudino, K. J. (2019). Effects of socioeconomic status and executive function on school readiness across levels of household chaos. *Early Childhood Research Quarterly, 47*, 331–340.
- Neitzel, J., Early, D., Sideris, J., LaForreth, D., Abel, M. B., Soli, M., Davidson, D. L., Haboush-Deloye, A., Hestenes, L. L., Jenson, D., Johnson, C., Kalas, J., Mamrak, A., Masterson, M. L., Mims, S. U., Oya, P., Philson, B., Showalter, M., Warner-Richter, M., & Kortright Wood, J. (2019). A comparative analysis of the Early Childhood Environment Rating Scale–Revised and Early Childhood Environment Rating Scale, third edition. *Journal of Early Childhood Research, 17*(4), 408–422.
- Organisation for Economic Co-operation and Development. (2017). *Starting strong 2017: Key OECD indicators on early childhood education and care*.
- Organisation for Economic Co-operation and Development. (2021). *Starting strong VI: Supporting meaningful interactions in early childhood education and care, starting strong*. <https://doi.org/10.1787/f47a06ae-en>
- Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and

- care settings and child outcomes. *PLoS One*, *11*(12), e0167660. <https://doi.org/10.1371/journal.pone.0167660>
- Pianta, R., Downer, J., & Hamre, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *The Future of Children*, *26*(2), 119–137.
- Pianta, R. C., Belsky, J., Houts, R., Morrison, F., & Network, T. N. I. of C. H. and H. D. (NICHD) E. C. C. R. (2007). Opportunities to Learn in America's Elementary Classrooms. *Science*, *315*(5820), 1795–1796.
- Pianta, R. C., Hamre, B. K., & Nguyen, T. (2020). Measuring and improving quality in early care and education. *Early Childhood Research Quarterly*, *51*, 285–287.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008a). *Classroom Assessment Scoring System™: Manual K-3*. Paul H. Brookes Publishing.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008b). *Classroom Assessment Scoring System (CLASS): Manual Pre-K*. Brookes Publishing Co.
- Pianta, R. C., Lipscomb, D., & Ruzek, E. (2021). Coaching teachers to improve students' school readiness skills: Indirect effects of teacher–student interaction. *Child Development*, *92*, 2509–2528.
- Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's school readiness skills across the pre-K year: Associations with teacher–student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology*, *66*, 101084.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. Springer Science and Business Media.
- Pink, B. (2011). *Australian statistical geography standard (ASGS): Vol. 5—Remoteness structure*. Australian Bureau of Statistics.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multi-level modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448.
- Prior, M. R., Sanson, A. V., & Oberklaid, F. (1989). The Australian temperament project. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 537–554). John Wiley & Sons.
- Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, *26*(5), 445–458. <https://doi.org/10.1080/09645292.2018.1484426>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramakrishnan, J. L., & Masten, A. S. (2020). Mastery motivation and school readiness among young children experiencing homelessness. *American Journal of Orthopsychiatry*, *90*(2), 223–235. <https://doi.org/10.1037/ort0000428>
- Robitzsch, A., & Grund, S. (2020). *miceadds: Some additional multiple imputation functions, especially for mice (R package version 3.10-28)*. <https://cran.r-project.org/package=miceadds>
- Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, *24*(2), 389–409.
- Rosenbaum, P. R. (2020). *Design of observational studies* (2nd ed.). Springer International Publishing.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, *4*(4), 537–571.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
- Rutter, M. (2012). Achievements and challenges in the biology of environmental effects. *Proceedings of the National Academy of Sciences of the United States of America*, *66*(Suppl. 2), 17149–17153.
- Salminen, J., Lerkkanen, M. K., Poikkeus, A. M., Pakarinen, E., Siekkinen, M., Hännikäinen, M., Poikonen, P. L., & Rasku-Puttunen, H. (2012). Observed classroom quality profiles of kindergarten classrooms in Finland. *Early Education & Development*, *23*(5), 654–677.
- Sandilos, L. E., DiPerna, J. C., & The Family Life Project Key Investigators. (2014). Measuring quality in kindergarten classrooms: Structural Analysis of the Classroom Assessment Scoring System (CLASS K–3). *Early Education and Development*, *25*(6), 894–914.
- Sanson, A., Smart, D., Prior, M., Oberklaid, F., & Pedlow, R. (1994). The structure of temperament from age 3 to 7 years: Age, sex, and sociodemographic influences. *Merrill-Palmer Quarterly*, *4*, 232–252.
- Sarkar, D. (2008). *Lattice: Multivariate data visualisation with R*. Springer.
- Shoukri, M. M., Asyali, M. H., & Donner, A. (2004). Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research*, *13*(4), 251–271. <https://doi.org/10.1191/0962280204sm365ra>
- Slot, P. L., Leseman, P. P. M., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, *33*, 64–76.
- Styck, K. M., Anthony, C. J., Sandilos, L. E., & DiPerna, J. C. (2021). Examining rater effects on the Classroom Assessment Scoring System. *Child Development*, *92*(3), 976–993.
- Taylor, C., Cloney, D., Adams, R., Ishimine, K., Thorpe, K., & Nguyen, T. K. C. (2016). Assessing the effectiveness of Australian early childhood education and care experiences: Study protocol. *BMC Public Health*, *16*(1), 1–12.
- Taylor, C., Ishimine, K., Cloney, D., Cleveland, G., & Thorpe, K. (2013). The Quality of Early Childhood Education and Care Services in Australia. *Australasian Journal of Early Childhood*, *38*(2), 13–21. <https://doi.org/10.1177/183693911303800203>
- Thorpe, K., Houen, S., Rankin, P., Pattinson, C., & Staton, S. (2022). Do the numbers add up?—Questioning the measure that places Australian ECEC educators as 'low quality'. *Australian Educational Researcher*, 1–20. <https://doi.org/10.1007/s13384-022-00525-4>
- Thorpe, K., Jansen, E., Sullivan, V., Irvine, S., & McDonald, P. (2020). Identifying predictors of retention and professional wellbeing of the early childhood education workforce in a time of change. *Journal of Educational Change*, *21*(4), 623–647.
- Thorpe, K., Rankin, P., Beatton, T., Houen, S., Sandi, M., Siraj, I., & Staton, S. (2020). The when and what of measuring ECE quality: Analysis of variation in the Classroom Assessment Scoring System (CLASS) across the ECE day. *Early Childhood Research Quarterly*, *53*, 274–286. <https://doi.org/10.1016/j.ecresq.2020.05.003>
- Thorpe, K., Searle, B.-R., Van Halen, O., Lakeman, N., Oakes, C., Harris, H., & Staton, S. (2020). *A market in mealtimes: Social and economic factors associated with meal provision in early care and education (ECE) services*. In SSRN (No. 2020–25; Working Paper Series).
- Thorpe, K., Westwood, E., Jansen, E., Menner, R., Houen, S., & Staton, S. (2021). Working Towards the Australian National Quality Standard for ECEC: What do we know? Where should we go? *The Australian Educational Researcher*, *48*(2), 227–247.
- Ulferts, H., Wolf, K. M., & Anders, Y. (2019). Impact of process quality in early childhood education and care on academic outcomes: Longitudinal meta-analysis. *Child Development*, *90*(5), 1474–1489. <https://doi.org/10.1111/cdev.13296>
- United Workers Union. (2021). *Exhausted, undervalued and leaving: The crisis in early education*. <https://bigsteps.org.au/wp-content/uploads/2021/08/Exhausted-undervalued-and-leaving.pdf>
- Van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press.

- Van Buuren, S., & Groothuis-Oudshoorn, C. G. M. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- van Huizen, T., & Plantenga, J. (2018). Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Economics of Education Review*, 66, 206–222.
- Vernon-Feagans, L., Mokrova, I. L., Carr, R. C., Garrett-Peters, P. T., & Burchinal, M. R. (2019). Cumulative years of classroom quality from kindergarten to third grade: Prediction to children's third grade literacy skills. *Early Childhood Research Quarterly*, 47, 531–540. <https://doi.org/10.1016/j.ecresq.2018.06.005>
- Von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher–child interactions in German preschools. *Early Childhood Research Quarterly*, 29(4), 509–519.
- Watts, T. W., Nguyen, T., Carr, R. C., Vernon-Feagans, L., & Blair, C. (2021). Examining the effects of changes in classroom quality on within-child changes in achievement and behavioral outcomes. *Child Development*, 92(4), e439–e456.
- Wertz, J., Moffitt, T. E., Agnew-Blais, J., Arseneault, L., Belsky, D. W., Corcoran, D. L., Houts, R., Matthews, T., Prinz, J. A., Richmond-Rakerd, L. S., Sugden, K., Williams, B., & Caspi, A. (2020). Using DNA from mothers and children to study parental investment in children's educational attainment. *Child Development*, 91(5), 1745–1761.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Riverside Publishing.
- Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., & Michnick Golinkoff, R. (2018). The power of play: A pediatric role in enhancing development in young children. *Pediatrics*, 142(3), e20182058. <https://doi.org/10.1542/peds.2018-2058>
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, 26(2), 21–35. <https://doi.org/10.1353/foc.2016.0010>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Rankin, P. S., Staton, S., Potia, A. H., Houen, S., & Thorpe, K. (2022). Emotional quality of early education programs improves language learning: A within-child across context design. *Child Development*, 93, 1680–1697. <https://doi.org/10.1111/cdev.13811>