

The Case Against Testing Young Children to Evaluate Teacher Effectiveness:

*A Position Statement from the Marsico Institute
for Early Learning and Literacy*

2010

Introduction

The Marsico Institute for Early Learning and Literacy (MIELL) is primarily concerned with educational environments and outcomes as they pertain to very young children (birth through age eight), and “accountability for effectiveness” initiatives for licensed public school teachers overlap with this age group. Although to our knowledge no states have yet tied teachers’ performance evaluations or employability to their kindergarteners’ or first graders’ test scores (a few have done so for second grade), legislation such as that passed in Colorado in May of 2010 has not made any distinction between “early childhood” and later grades. Any teacher evaluation system that is based, even in part, on the test scores of children below third grade constitutes an even greater experiment than the systems for higher grades that are already causing great controversy throughout the country. In this position statement, we present arguments for why deducing teacher effectiveness from student test performance is neither developmentally appropriate nor scientifically viable in the early childhood grades. Furthermore, we recommend alternative methods that could be used to assess and enhance the effectiveness of teachers who work with our youngest children.

The core of MIELL’s vision is to improve the quality of educational experiences that young children receive in child care, preschool, and the early grades, primarily in terms of their interactions with their teachers. The quality of teacher-child interactions is not a “soft” indicator, but rather is the hallmark of effective early educational experiences. Importantly, it is also something for which we can hold teachers accountable, given adequate and appropriate systemic and programmatic supports.

Policymakers and the public alike seem increasingly and unquestioningly to accept as truth that student achievement test score growth is the only objective indicator of teacher effectiveness, and a necessary part of any “tough” teacher evaluation system. The high degree of error associated with even the best versions of calculations of student growth (i.e., value-added modeling using the most corrections for pre-existing student differences), in the well-tested grades (3rd-12th) and subjects (reading and math), has been discussed extensively elsewhere.^{1,2,3} The conclusion that such methods are “flawed and inaccurate” has led many experts and credible organizations, such as the National Academy of Sciences and Economic Policy Institute, to recommend that student test performance be considered only a small part of teacher evaluation systems.⁴ We and many others^{5,6,7,8} are deeply concerned about the lack of evidence documenting that those who produce achievement gains are also those who demonstrate higher-quality teaching practices in the classroom^{9,10}. Surely, such evidence is needed since reasonableness would not allow a “by any means necessary” approach to education. For our youngest learners, the consequences of such an approach are even graver, since young children do not have the agency to evaluate the relative merits and personal usefulness of the various teachings inputs they receive.

Colorado’s legislation (SB 191) specifies that all licensed teachers employed in a school district or charter school (which therefore includes preschool and K-2 teachers with teaching licenses who are employed to teach in school districts or charter schools) will have evaluations, upon which decisions

ⁱ One study, not peer-reviewed, has documented this correlation between principals’ observations, achievement gains, and teacher education and experience.⁹ Principals’ observations were better predictors of student achievement gains than teacher education and experience. A second study, also not peer-reviewed, did find exploratory associations between test score gains and blind observers’ ratings of teachers.¹⁰ This study was exploratory because it compared only five effective teachers to six ineffective teachers, and therefore no statistical comparisons were conducted. In the University of Pennsylvania’s “Operation Public Education,” system, although the value-added modeling and observational components are separately well-researched, we find no evidence documenting the proportion of cases for which these two methods of evaluation agree.

about probationary status (lack of tenure) will be made, and at least half of which must be based on the “academic growth of the teacher’s students.” As difficult as it will be to articulate this system for the higher grades and well-tested subjects, we are particularly concerned about the potential negative consequences that may result from using this system to evaluate the effectiveness of teachers serving very young children. Preschool and grades K-2 are considered the “untested grades” because, despite the use of some standardized testing in these grades, such tests have never been used in statewide evaluations of district, school, or teacher effectiveness. We believe that forcing a definition of academic growth in the early grades will be either virtually impossible, or harmful to children. Below, the primary arguments for this claim are summarized.

For tests of student growth to be used in high-stakes decisions, they must be credible, meaningful, and accurate assessments of children’s learning. However, early assessments do not reliably predict the kinds of student outcomes that are the target of teacher effectiveness evaluations.

There have been two major meta-analyses (analyses of multiple studies) conducted to examine the question of whether early assessments are meaningfully related to later achievement. Using different methodologies, these studies came to the same conclusion^{11,12,13}: “Tests used to make predictions [in early childhood]-even relatively short-term predictions-are insufficiently stable to justify assigning stakes based on them.”¹⁴ This is thought to be true for multiple reasons including the fact that the younger children are, the more susceptible they are to the subtle variations in test conditions, instructions or test format, and their own particular mood or condition.^{15,16} The meta-analysts (who were not evaluating their own instruments) note that this does not mean that there are no valid assessments of early learning, but rather, validity and predictability are dependent on the type of test, content of the test, and testing conditions, and that changes to any one of these features results in a different conclusion about children’s level of learning.¹⁷

No matter the quality of the test or testing conditions, however, substantial unreliability will occur also because young children’s learning and development progresses in fits and starts¹⁸ - both within individual children and across groups of children. This finding is strongly supported by brain research.¹⁹ Because a young child’s ability to learn academic skills is highly dependent on his or her stage of biological maturation, this developmental variability means that at any given assessment point, a large proportion of young children may have been “caught” just prior to or after a developmental shift, making it a poor reflection of that child’s learning. If the test at the beginning of kindergarten, for example, did not accurately capture children’s learning, the test at the end of kindergarten or beginning of first grade cannot possibly accurately capture whether children have attained a year’s worth of growth. While this is true across grades, there is a substantially higher degree of error in early childhood since “instability or change may be the rule rather than the exception during this period.”²⁰

For these reasons among others, there does not exist a “resource bank” of assessments that are rigorous and credible, that can be employed in educational settings to assess “achievement growth” in the early grades. The fields of early childhood and developmental psychology are beginning to develop innovative and appropriate educational assessments for very young children, including those designed to assess key foundations of lifelong learning such as executive function.²¹ However, these assessments are designed to provide accurate evaluations of children, not their teachers. In short, accountability initiatives that are relying on existing tools will find there is no arsenal of adequate assessments of young children that have been shown to be valid and reliable, as well as sensitive to the influence of teachers. However, this is not a matter of the “state of the science” or the quality of

the tests. Higher error will always be present in early childhood because of the increased instability of development that is characteristic of young children.

The only accurate method for capturing young children’s learning - authentic assessments that are administered by a child’s own teacher, ongoing and observation-based - is prohibited by the purposes of teacher effectiveness evaluations.

Beyond the lack of generalizability, one of the most striking findings from the meta-analysis discussed above confirmed what early childhood researchers have been saying for a long time (e.g., Shepard, 1994) – that of the methods of assessment, those that involve the subjective ratings of individuals who know the child are the most accurate and therefore also the most predictive of later achievement.^{22,23} Only someone with ongoing, subjective knowledge of the child, who has witnessed the child’s activities across multiple contexts, can accurately assess which skills are a legitimate part of a child’s repertoire. Not just based on what seems right for children, but on decades of research that show the poor performance, predictive, and functional value of traditional or standardized methods, the superiority of teacher-administered authentic methodologies stands as one of the few definitive conclusions of the field of early childhood research.^{24,25,26,27,28,29}

In Colorado, other legislation (SB 212, known as Colorado Achievement Plan for Kids, CAP4K) has mandated the design of a new statewide assessment system, which will now include a school readiness assessment to be administered during the kindergarten year. Although the assessment recommendations are not complete, a recent update to stakeholders by the Colorado Department of Education (Colorado Early Childhood Progress and Possibilities event, November 18, 2010) revealed that the first recommendation of the school readiness assessment committee will be for that assessment to be “authentic, ongoing, and observation-based” and for the purpose of developing “individual readiness plans” for all children in Colorado public schools. Not surprisingly, the primary conclusion this committee has made is based on one of the most established givens in the field of early childhood research. This also means, however, that if the resultant assessment follows these appropriate recommendations, it cannot be used in the system designed for SB 191 to evaluate the effectiveness of individual teachers, due to the obvious conflict of interest. Thus, Colorado’s Council for Educator Effectiveness (CCEE), which will provide the recommendations for implementing the SB 191 statute, is now in a position of considering recommending testing in the early grades that is either A) diametrically opposed to research, the early childhood field at large, and similar conclusions of Colorado’s education stakeholders, or, B) scientifically unviable because it is administered by the same individuals it is intended to evaluate. For this reason, MIELL’s position is that SB 191 be amended to reflect the different timelines and learning profiles that exist in the early grades. In short, we believe that the CCEE will find that SB 191 is untenable in its current, universally applied form.

The concept of “one year of academic growth” does not have a meaningful anchor in early childhood since the processes of biological maturation and the acquisition of academic skills are intertwined and unfold at highly variable rates within and across children. Historical efforts to apply similar concepts to children eight and under have failed - and more importantly, they have harmed children.

Even more important than the methodology used to conduct assessment on young children is the unintended harm caused by picking and choosing specific learning tasks that a teacher must

“confer” to a young child in order to be deemed effective. While such processes are widely known to lead to “teaching to the test” across all ages of students,^{30,31,32,33,34,35} this set of circumstances is particularly dire in early childhood for a number of related reasons.

Consider the fact that the younger a child, the more unspecified, global, and dynamic learning is, and the less content-specific it is. Development and academic growth are inextricably linked at early ages, but development is controlled by maturation, genetics, and home environment to a much larger degree than it is by teachers.³⁶ The effects of poverty and deprived home environments are evident in children by 9 months of age, and continue to widen through age two and beyond.^{37,38,39,40} The small amount of variability left over for teachers to influence by the time a child enters their classroom, although very important, is near-impossible to parse out on an assessment.

In addition to the global nature of early development, its highly variable rate of unfolding – even within normal and “universal” developmental achievements - makes it impossible to hold teachers in a single grade accountable for an equal amount of growth. Although neural development is known to slow down between the ages of three and five, cognitive reorganizations using existing structures, that may have an even more significant behavioral effect, continue to occur beyond this age. Take for example the well-known “five-to-seven year shift.”⁴¹ Children have been shown to undergo a radical transformation on a number of dimensions critical to academic functioning including attention, memory, and social understanding sometime between the ages of five and seven.^{42,43,44} This change has been found to be universal – not subject to differences in culture or schooling system. However, given the variety of ages at which children enter school in this country, and the three-year range of the 5-to-7 year shift itself, this critical change - which essentially makes formal schooling possible - could occur in a normally developing child anywhere between preschool and third grade. Sir Ken Robinson, Professor Emeritus of Education at the University of Warwick in the United Kingdom, is gravely concerned about the effects of educating children in “batches” and asks, “Why is there this assumption that the most important thing children have in common is how old they are?” By extension, how can we hold early grade teachers accountable for the same amount of academic growthⁱⁱ of their students when A) this growth is complexly tied to development, B) development is largely not under teacher control, C) the same cognitive reorganization - which is critical to functioning in a school setting - can and regularly does occur across multiple grades?

While it might seem like a good idea to parse out limited pieces of content for which to hold teachers responsible (e.g., letter naming), the unintended consequences of this type of curriculum-narrowing are well-documented.^{45,46} Moreover, the most important reasons for having an education system in the first place are “sleeper effects.”⁴⁷ In other words, letter-naming is not an end, but a means to an end. It is said that prior to age eight children are learning to read, but after this age they are reading to learn. This latter achievement is the “end game”. Research has shown that de-coding skills predict early reading achievement (first and second grade), but that the oral language skills learned in preschool (complexity of syntax, discourse, and vocabulary) are better predictors of reading achievement in third and fourth grade.⁴⁸ By holding teachers accountable for the “precursors” that can be attained within one grade-level, we take teachers’ eye off the prize, reward superficiality, and characterize them as ineffective when it may be that their approach is “slower but deeper,” that is, more appropriately focused on the long-term benefits.

ⁱⁱ We note that using achievement growth, rather than achievement level, does not cure the problem of vastly different starting levels. This is true for multiple reasons including the fact that growth does not occur in equal amounts from all starting levels. In early childhood, there are other complex reasons such as the qualitative shifts in cognition discussed above. A year’s worth of growth therefore becomes reified by the test itself, rather than legitimately accounting for each child’s qualitatively different starting point.

Recent documents commissioned by the National Council for Accreditation of Teacher Education (NCATE) have explained how this parsing out of narrow skills harms children.⁴⁹ Consider the following example:

“Motor coordination is important for penmanship, but need not hinder opportunities in literacy development by preventing expository writing. The child who tears his penmanship paper with frustrated erasures might be better served with fine motor coordination exercises and alternative routes of expression, rather than further exercise in the physically impossible. Such physical limitations are frequently allowed to interfere with literary expression, an example of a limitation in one domain of development influencing growth in another.” (p. 16)

Said differently by Alexander (2010):⁵⁰

“I do not wish to argue that there are no merits to the examination of more short-term outcomes; there are. However, these more ‘immediate’ gains must be considered in relation to what they suggest about long-term growth and development. For example, do we want to risk...turning ‘struggling readers’ into ‘struggling thinkers’ by failing to teach them to think critically and intensively about what they read or hear?” (p. 636)

Such negative, unintended consequences have already occurred in early childhood settings. Meisels (2006) documents the history of the Head Start National Reporting System (NRS) which, to-date, may still be the largest ever standardized test administration in U.S. History (450,000 4-year-olds tested twice per year from 2003 to 2006).⁵¹ The NRS was discontinued in 2007 after a report from the U.S. General Accountability Office determined that nearly 1 out of 5 Head Start grantees openly admitted to changing instruction to fit the NRS in the first year it was administered. As Meisels notes, the low level of formal education characteristic of the Head Start work force means that there is high risk for tailoring pedagogy to accountability tests. Note that the NRS was for programmatic-level accountability, not teacher-level; yet, the inappropriate curriculum narrowing occurred anyway.

A somewhat different vulnerability due to a lack of training exists even in licensed and degreed teachers in preschool and the early grades. That is, teachers deployed to teach in the early grades often have little or no knowledge of, or skills applying, modern child development principles.^{52,53} Thus, with a lack of experience and high-quality models to apply developmental principles, along with national and local emphasis on a set of necessarily narrow instructional outcomes, even educated and experienced teachers are vulnerable to the possibility of thwarting individual children’s optimal paths of development in service of what are believed to be precursor skills. So grave is this concern that several organizations have gone so far as to recommend that all large-scale testing prior to third grade should cease.⁵⁴

Conclusion and Recommendations

In sum, MIELL takes the position that using young children’s scores on standardized tests to deduce the effectiveness of individual early childhood teachers is inaccurate, untenable, and harmful to children. In contrast, teachers can and should be accountable for their words and actions towards children, and the early childhood field is increasingly better equipped to make valid and reliable assessments of these practices. Years of research have led to the development of observational systems that evaluate teachers for conducting themselves in ways that lead to higher achievement for students. This is not an indirect assessment of student achievement, but rather a direct assessment

of teacher effectiveness, which is what education reform legislation such as that in Colorado is charged with evaluating. For example, Pianta and colleagues have developed the Classroom Assessment Scoring System (CLASS), which has been modified and developed for all grades K-12, and extensively tested and validated thus far for grades K-5.⁵⁵ Based on prior research on what teacher behaviors predict child outcomes including academic achievement, and philosophically similar to other systems, the CLASS includes evaluation of emotional supports (e.g., positive climate, respect for student perspectives), classroom organization (e.g., effective use of time, engaging learning formats), and instructional supports (e.g., concept development, quality of feedback, and quality of language use). The CLASS or a similar system could be applied with trained peer observers from other districts (perhaps who teach in adjacent grades but not the exact same grade), to evaluate the effectiveness of teachers in the early grades. Using blind peer observers increases the credibility of the system and answers the criticism that we currently have no “objective” evaluations of teacher effectiveness.

Besides accuracy and fairness, the other critical reason for focusing on blind observations in the early childhood grades is that they provide a more direct, and likely effective route to enhancing instruction.⁵⁶ Secretary of Education Arne Duncan said, “Under the best of circumstances, this [value-added] information would be thoughtfully discussed among teachers and principals with the goal of identifying the strongest teachers so we can learn from them and better support those who are struggling.” (Remarks at the Statehouse Convention Dinner in Little Rock, AR, August 25, 2010, <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-statehouse-convention-center-little-rock-arkansas>). However, a consistent process for how teachers might “learn from” their effectiveness ratings that result from their students’ performance on tests (or, as implied from Secretary Duncan’s quote, the effectiveness ratings of *other* teachers) has yet to be identified, much less validated. In contrast, observations of teacher practices do not have a complicated relationship with teacher effectiveness, and can therefore be used more readily and

Have you considered this?

We firmly believe that because using young children’s test scores to evaluate teacher effectiveness is both inaccurate and potentially harmful (due to unintended consequences such as holding children back from other paths or in other areas because they have not achieved the average level in “precursor” skills), this practice will not pass the credibility test. However, for stakeholders who insist that a student test score component is non-negotiable, alternative models for K-2 could be considered.

Specifically, although we believe that defining a year’s worth of academic growth in any of the early childhood grades will not be possible, we can agree that the overall purpose of the early grades is to set a foundation for success in the later grades. Thus, we suggest that the following idea be explored: Students could be assessed at the end of the second grade for their readiness to participate successfully in the third grade. This would be in addition to the ongoing, diagnostic testing that teachers do throughout K-2 to monitor progress and individualize instruction. All K-2 teachers in a school (or region, or district) could be held accountable for the second grade scores, thus, reinforcing the notion that all early grades teachers are responsible for readying students to become academic achievers. A principal- and/or peer-review system would allow informed judgment to assess which individuals were most likely contributing to any higher- or lower-than-expected performance.

To correct for student differences and minimize disincentives to work with challenging students (which is a challenge to be faced in all models of evaluation systems), benchmarks for expected second grade performance could be established, in part, based on the scores of ongoing assessments used throughout K-2. It would not be in teachers’ best interest to either over- or under-rate their students in this case. It would be in teachers’ best interests to accurately assess their students, since accurate data would put them in the best position for a second grade benchmark that is both achievable, but rigorous enough so as to be a viable basis for promotions.

effectively to coach teachers and enhance their teaching. Such systems of improvement have been used with success in early childhood education in terms of student outcomes including literacy, math, and social-emotional development.^{57,58,59,60}

In short, we appeal to the legislators and legislative councils in Colorado and other states to design or amend teacher effectiveness legislation with these critical differences between younger and older children in mind. A uniform evaluation system applied to all licensed teachers is untenable, in part, for the reasons we have discussed here. Experts in the untested subjects (e.g., history, science) will likely have different but equally valid reasons for a non-universal application of a teacher evaluation system via “student academic growth” on tests.

Suggested citation for this position statement:

Moreno, A.J., Maloney, V.R., & Brown, K.L. (2010). The case against testing young children to evaluate teacher effectiveness: A position statement from the Marsico Institute for Early Learning and Literacy. Issue Brief No. 4, University of Denver.

References

- ¹ Economic Policy Institute. (2010, August). *Problems with the use of student test scores to evaluate teachers*. (Issue Brief No. 278). Washington DC. Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. S., & Shepard, L. A. Retrieved from <http://epi-data.org/education/>
- ² Braun, H. I. (2005). Using student progress to evaluate teachers: A primer on value-added models. Educational Testing Service.
- ³ Schochet, P. Z., & Chiang, H. S. U.S. Department of Education, Institute of Education Sciences, Center for Education Evaluation and Regional Assistance. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. (NCEE Publication No. 2010-4004).
- ⁴ See endnote ¹
- ⁵ Alexander, P. A. (2010). Through myth to reality: Reframing education as academic development. *Early Education and Development*, 21, 633-651.
- ⁶ Bettelheim, R. (2010). Outdated teaching is failing our children. *USA Today*. Retrieved from http://www.usatoday.com/news/opinion/forum/2010-11-10-column10_ST1_N.htm
- ⁷ Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109-119.
- ⁸ Rothstein, R. (2008). The corruption of school accountability. *School Administrator*, 65, 14-15.
- ⁹ Jacob, B. A., & Lefgren, L. (2007). *Can principals identify effective teachers? Evidence on subjective performance evaluation in education*. Retrieved from econ.byu.edu/faculty/Lefgren/Assets/papers/principals.pdf
- ¹⁰ Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, 20, 165-184.
- ¹¹ Kim, J., & Suen, H. K. (2003). Predicting children's academic achievement from early assessment scores: A validity generation study. *Early Childhood Research Quarterly*, 18, 547-566.
- ¹² La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70, 443-484.
- ¹³ Rapin, I. (2003). Value and limitations of preschool cognitive tests, with an emphasis on longitudinal study of children on the autistic spectrum. *Brain and Development*, 25, 546-548.
- ¹⁴ Meisels, S. J. (2006, March). *Accountability in early childhood: No easy answers* (Report No. 6). Herr Research Center for Children and Social Policy at Erikson Institute, 9.
- ¹⁵ Kagan, S. L., & Scott-Little, C. (2004). Early learning standards: Changing the parlance and practice of early childhood education? *Phi Delta Kappan*, 85, 388-396.
- ¹⁶ National Association of School Psychologists. (2005). *Position statement on early childhood assessment*. Bethesda, MD.
- ¹⁷ See endnote ¹¹
- ¹⁸ Kagan, S., L., & Kauerz, K. (2007). Reaching for the whole: Integration and alignment in early education policy. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 11-30). Baltimore, MD: Paul H. Brookes.

¹⁹ Fischer, K. W. (2008). Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education. In A. M. Battro, K. W. Fischer & P. Léna (Eds.), *The educated brain* (pp. 127-150). Cambridge U.K.: Cambridge University Press.

²⁰ La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70, 443-484.

²¹ Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychometric Assessment*, 22(2), 306-317.

²² See endnote ¹¹

²³ Shepard, L. A. (1994). The challenge of assessing young children appropriately. *Phi Delta Kappan* (76).

²⁴ See endnote ¹¹

²⁵ Neisworth, J.T. & Bagnato, S.T. (2004). The mismeasure of young children: The authentic assessment alternative. *Infants and Young Children*, 17, 198-212.

²⁶ Copple, C., & Bredekamp, S. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.

²⁷ Grisham-Brown, J., Hallam, R. A., & Pretti-Frontczak, K. (2008). Preparing Head Start personnel to use a curriculum-based assessment: An innovative practice in the "age of accountability." *Journal of Early Intervention*, 30, 271-281.

²⁸ Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic assessment for infants and toddlers: Exploring the reliability and validity of the Ounce Scale. *Applied Developmental Science*, 14(2), 55-71.

²⁹ National Association for the Education of Young Children. (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8*. Washington, DC.

³⁰ See endnote ⁵

³¹ Amrein-Beardsley, A. (2009). The unintended, pernicious consequences of "staying the course" on the United States' No Child Left Behind policy. *International Journal of Education Policy and Leadership*, 4, 1-13.

³² Carlson, D. L., & Clay, T. (2010). Evoking a spirit of play: M & M's [R] stories and (un)real possibilities for teaching secondary literacy. *Kappa Delta Pi Record*, 46, 164-169.

³³ Gallagher, K. (2010). Reversing readicide. *Educational Leadership*, 67, 36-41.

³⁴ Longo, C. (2010). Fostering creativity or teaching to the test? Implications of state testing on the delivery of science instruction. *Clearing House, A Journal of Educational Strategies, Issues and Ideas*, 83, 54-57.

³⁵ Rothstein, R. (2008). The corruption of school accountability. *School Administrator*, 65, 14-15.

³⁶ National Institute of Child Health and Development, Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, 74, 976-1005.

³⁷ Belsky J. Early day care and infant-mother attachment security. In: Tremblay, R. E., Barr, R. G., Peters, R. D., & Boivin, M. (Eds.). (2009). *Encyclopedia on Early Childhood Development*. Montreal, Quebec: Centre of Excellence for Early Childhood Development; 1-6.

³⁸ Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371-399.

³⁹ Children Now. (January, 2010). *California Report Card 2010: Setting the agenda for children*. Oakland, CA.

⁴⁰ Hawley, T., & Gunner, M. (2000). *Starting smart: How early experiences affect brain development* (2nd ed.). Washington, DC: Ounce of Prevention Fund and Zero to Three.

⁴¹ Sameroff, A. J., & Haith, M. M. (1996). *The five to seven year shift: The age of reason and responsibility*. Chicago, IL: University of Chicago Press.

⁴² Bartgis, J., Thomas, D. G., Lefler, E. K., & Hartung, C. M. (2008). The development of attention and response inhibition in early childhood. *Infant and Child Development*, *17*, 491-502.

⁴³ Jordan, J. A., Mulhern, G., & Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year olds. *Journal of Experimental Child Psychology*, *103*, 455-468.

⁴⁴ Marti, E., Garcia-Mila, M., & Teberosky, A. (2005). Notational strategies for problem solving in 5- to 7-year-olds. *European Journal of Developmental Psychology*, *2*, 364-384.

⁴⁵ See endnote ¹⁴

⁴⁶ Miller, E., & Almon, J. (2009, March). *Crisis in the kindergarten: Why children need to play in school*. Alliance for Childhood.

⁴⁷ See endnote ⁵

⁴⁸ Collins, M. F., & Dennis, S. F. (2009). Targeting oral language development in high-risk preschoolers. *A Research-to-Practice Journal for the Early Intervention Field*, *12*, 245-256.

⁴⁹ Snyder, J., & Lit, I. (2010). *Principals and exemplars for integrating developmental sciences knowledge into educator preparation*. National Council for Accreditation of Teacher Education.

⁵⁰ See endnote ⁵

⁵¹ See endnote ⁴⁵

⁵² Pianta, R. C., Hitz, R., & West, B. (2010). *Increasing the application of developmental sciences knowledge in educator preparation: Policy issues and recommendations*. National Council for Accreditation of Teacher Education.

⁵³ Gasbarro, M. R., Maloney, V. R., & Moreno, A. J. (2009). *Early childhood teacher preparation in Colorado: Connects and disconnects*. Retrieved from <http://www.du.edu/media/documents/marsicoinstitute/ECTeacherPrepConnectsandDisconnects.pdf>

⁵⁴ See endnote ⁴⁶

⁵⁵ See endnote ⁷

⁵⁶ See endnote ⁷

⁵⁷ Cusumano, D. L., Armstrong, K., Cohen, R., & Todd, M. (2006). Indirect impact: How early childhood educator training and coaching impacted the acquisition of literacy skills in preschool students. *Journal of Early Childhood Teacher Education*, *27*, 363-377.

⁵⁸ Hsieh, W.-Y., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009). Using coaching to increase preschool

teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly*, 29, 229-247.

⁵⁹ Mohler, G. M., Yun, K. A., Carter, A., & Kasak, D. (2009). The effect of curriculum, coaching, and professional development on prekindergarten children's literacy development. *Journal of Early Childhood Teacher Education*, 30, 49-68.

⁶⁰ Rudd, L. C., Lambert, M. C., Satterwhite, M., & Smith, C. H. (2009). Professional development + coaching = enhanced teaching: Increasing usage of math mediated language in preschool classrooms. *Early Childhood Education Journal*, 37, 63-69.