

**Assessment in a Continuous Improvement Cycle:
New Jersey's Abbott Preschool Program**

By Ellen Frede, PhD
The College of New Jersey
September, 2005

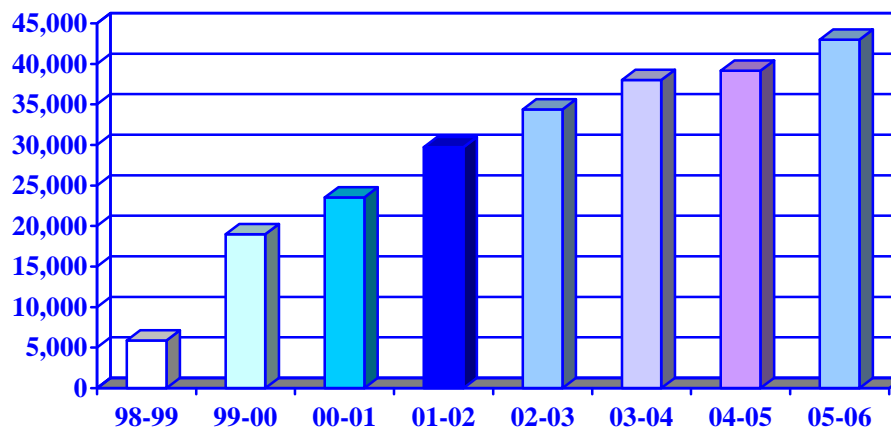
This paper was prepared for the National Early Childhood Accountability Task Force with support from the Pew Charitable Trusts, the Foundation for Child Development and the Joyce Foundation.

Introduction

As part of the landmark New Jersey Supreme Court school funding case, *Abbott v. Burke*, the Court established the Abbott Preschool Program. Beginning in the 1999-2000 school year, 3- and 4- year-old children in the highest poverty districts in the state were designated to receive a high-quality preschool education that would prepare them to enter school with the knowledge and skills necessary to meet the New Jersey Preschool Teaching and Learning Expectations: Standards of Quality (NJ Department of Education, 2004a) and the Kindergarten New Jersey Core Curriculum Content Standards (NJDOE, 2004b). Through a Department of Education (DOE) and Department of Human Services (DHS) partnership, Abbott preschool classrooms combine a DOE-funded six-hour, 180-day component with a DHS-funded wrap-around program that provides daily before- and after-care and summer programs. In total, the full-day, full-year program is available ten hours per day, 245 days a year.

Enrollment in the Abbott preschool program has increased dramatically since its inception in 1999. During the 2004–2005 school year, the sixth year of Abbott preschool implementation, the 31 Abbott districts served over 39,000 3- and 4-year-old children in preschool – 75% of a possible universe of 52,160 children. The projected enrollment for the 2005-06 school year is more than 43,000 children with a DOE budget of more \$450,000,000. Through contracts with the school districts, private child care providers and Head Start agencies offer of Abbott Preschool:

- 37% of children are served in district-run classrooms,
- 7% are served in Head Start classrooms, and
- 56% are in private provider classrooms.

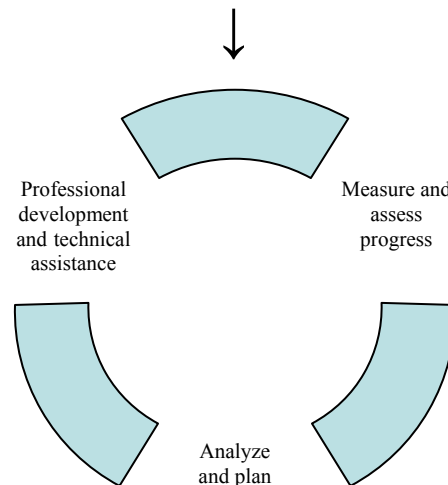


This paper presents an overview of the following efforts over the last three and one half years to develop and implement an assessment system for the New Jersey Abbott Preschool Program and covers the following components:

- The framework of the assessment system,
- The process of defining purposes and setting standards,
- The development and implementation of the system at the state, district and child level,
- Some initial results, and
- Plans for the future.

Program Quality - the Continuous Improvement Cycle

Step 1: Develop Standards and Define Outcomes



High-quality education requires a cycle of continuous improvement. This cycle begins with setting standards and defining outcomes and then proceeds to an iterative process of gathering evidence about progress, analyzing this evidence and making plans for improvement as well as implementing those plans. The process then begins anew with another assessment of progress. This cycle is especially critical in a new initiative since infrastructure and expertise may be lacking.

Defining Purposes, Audiences and Content for Standards & Assessments

Before developing accountability systems in New Jersey, it was important to think about the purpose and audience: Accountable to whom, for what purposes and for what aspects of program services and results? At the state level, purposes could vary in the following ways:

- High stakes decision-making such as determining which programs receive funding;
- Monitoring for public reporting about a program or at the child level, individual report cards;
- Collecting information for program improvement or to inform teaching; or
- A combination of purposes.

The purpose is driven by the audience and this affects to some extent what will be measured and how. For example, the legislature wants to know whether the money is well spent. Therefore accountability may include expenditure analyses as well as child outcome studies. In the Abbott districts, elements of each of the purposes listed above influenced accountability activities. Nevertheless, the over-arching purpose of the accountability system remains collecting and using credible evidence to improve practice and inform decisions at the classroom, site, district and state levels.

Before progress can be assessed, the standards or outcomes desired must be determined. Again, the purpose and audience dictate what types of standards are needed. For example, if the only accountability issue is child outcomes then only child learning standards need be developed. Where the relationship between classroom implementation and child progress is of interest, then standards on curriculum and teaching must be chosen or developed. Program standards that detail criteria for program operations such as administrator credentials or community involvement would be necessary if program implementation is being assessed. Standards at the child, site, district and state level may be necessary. As accountability data are collected the standards may have to be revisited and raised as improvement is seen. This is especially true for new programs. A critical piece of standards development is also ensuring that they are used. That in itself becomes part of the improvement.

Developing Program and Teaching and Learning Standards in New Jersey

In New Jersey, we began with some basic program standards that the Supreme Court set when ordering preschool as a remedy in the *Abbott v. Burke* educational equity case in 1998. These included a maximum class size of 15, certified teachers with early childhood expertise, assistant teachers in every classroom, comprehensive services and a developmentally appropriate curriculum designed to meet learning standards. To ensure high quality and consistency for children across auspice and district and to assist administrators and staff who may have been inadequately prepared in early childhood education, detailed operational standards were necessary. They were also necessary because the Court made clear that funding through an across-the-state per pupil formula would not be adequate and that budget decisions must be based on district contexts. In 2002, only limited regulations that mostly mirrored the basic standards of the Court had been promulgated. Without specific standards for what was part of the program and what was not, inconsistencies in budgets and implementation were becoming evident.

In 2002, the Early Childhood Education Work Group was formed. It consisted of representatives of school districts, private child care centers, Head Start, advocacy groups, community groups, higher education, teacher unions, early childhood professional organizations and state agencies. Subcommittees developed recommendations for guidelines across relevant program component areas. These recommendations were revised and edited by the Department of Education and published as the Abbott Preschool Program Implementation Guidelines (Office of Early Childhood Education, NJDOE, 2002, revised 2005). The guidelines are derived from research

wherever possible, and based on expert opinion where research is not available. In general, the guidelines provide recommendations and not mandates in an effort to accommodate local conditions, contexts and needs. However, many of the provisions included in the guidelines have been incorporated into state regulations as it becomes clear that they are universally warranted. Annual district plans and budgets are driven by the research-based best practices offered in this document.

Learning outcomes for children had been developed in 2000 as a result of a court mandate. However, these expectations were developed mostly with input from practitioners and thus were not always consistent with research. They combined actual learning outcomes such as knowing ten letters with process outcomes such as experiencing cooking activities. To upgrade these standards, work groups were formed with representatives of many of the same groups as listed above but with a heavy emphasis on experts in each of the 8 domains of learning included in the document. To ensure that the learning outcomes did not influence teachers to teach in inappropriate ways, we coupled the child learning outcomes with teaching practices that should assist children in meeting the outcomes. The resulting document, Preschool Teaching and Learning Expectations: Standards of Quality (OECE, DOE, 2004) was adopted by the state board in 2004 but was used for professional development since its completion in 2002. Abbott districts are required to plan or adopt curriculum and teaching practices that will result in children meeting the Expectations. The DOE dedicated intensive annual professional development to this effort and identified five recommended curriculum models that align with the Expectations. An ongoing assessment system discussed below was derived from the Expectations and is now mandated in every classroom.

Another type of standards that is relevant to accountability is criteria for appropriate assessment tools and procedures. Both the program standards and teaching and learning standards have a section setting forth guidelines for child assessment. It was primarily written by Jacqueline Jones of the Educational Testing Service and it reviews appropriate uses of assessment for identifying children with disabilities and describes ongoing observational and portfolio-based assessment as a tool for teaching and learning.

Assessment of Program Quality, Child Outcomes and Fiscal Management

At the state level, the following four types of data collection have been conducted to measure progress and improve practice:

- Conducting randomized structured classroom observations,
- Assessing child progress using standardized instruments on a random sample,
- Measuring program implementation via a self assessment, criterion based system, and
- Conducting randomized fiscal examinations of private providers to determine need for accounting and other fiscal training.

To measure and assess progress statewide, the DOE formed the Early Learning Improvement Consortium (ELIC) in 2002 by bringing together a group of the state's top early childhood education faculty. Drawing on research previously conducted by the Center of Early Education Research (Tarr, Barnett, Lamy and Frede, 2002), the ELIC is responsible for collecting and reporting on data on children and classrooms. Each fall, assessments of kindergartners' skills have been conducted to measure progress toward preparing children to succeed in school. In addition, members of the ELIC conduct classroom observations on randomly selected Abbott preschool classrooms to measure progress in program quality. Findings have been reported yearly and can be found at www.nj.gov/njded/ece.

In addition, districts are required by regulation to conduct annual classroom observations using a structured classroom observation instrument and to implement a state-developed on-going observation and portfolio assessment of children.

Conducting structured classroom observations

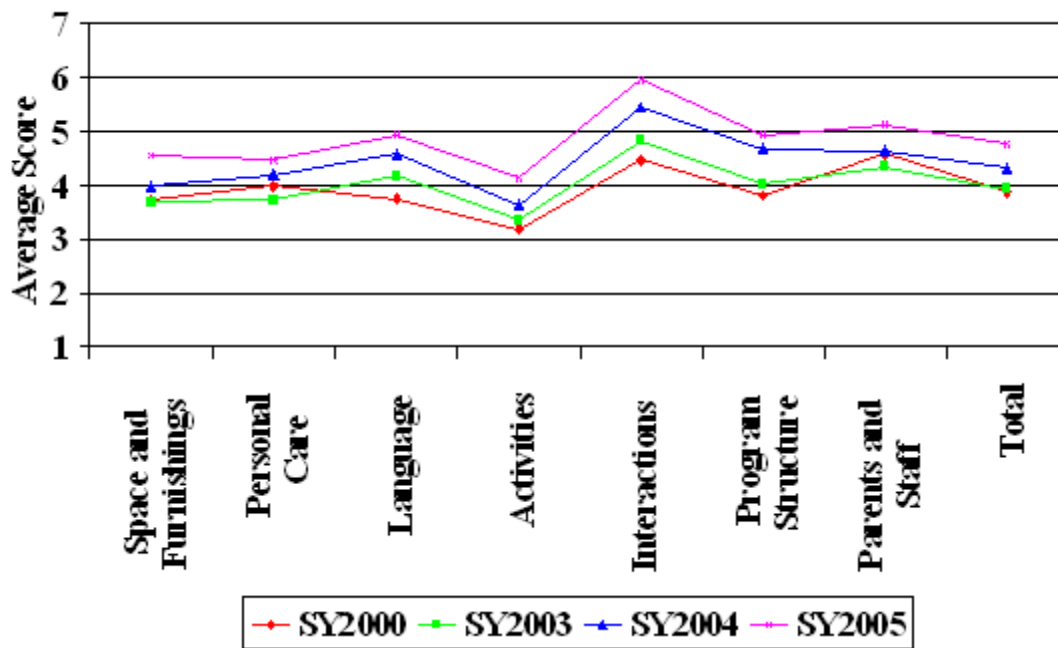
State-level classroom quality was measured using three instruments: the Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms, Clifford & Cryer, 1998), the Support for Early Literacy Assessment (SELA; Smith, Davidson & Weisenfeld, 2001) and the Preschool Classroom Mathematics Inventory (PCMI; Frede, Dessewffy, Hornbeck & Worth, 2001). The ECERS-R, a widely used tool, was chosen to provide a comprehensive look at classroom quality and to allow comparison of New Jersey's scores to scores from other states and other large projects. In addition, statewide results on the ECERS-R in 14 districts since 1999 is available through data previously collected by the Center for Early Education Research (a precursor of the National Institute for Early Education Research) which allows analyses of growth over more years. From that earlier data collection, we were aware that not only were most classrooms providing inadequate educational environments, a substantial number were likely to be unsafe, unhealthy and potentially harmful. An instrument that captured these more basic quality issues was necessary. The design is essentially reporting trend data for the program as a whole via a sample of classrooms using a set of standardized observational tools.

The SELA provides more specific information on classroom practices that support children's early language and literacy skills. The PCMI focuses on the materials and methods used in preschool classrooms to support and enhance children's math skills. Both of these instruments were considered important additions to the ECERS-R since they are designed to more directly assess teaching strategies that relate to academic learning. Results from these instruments are readily translated into professional development and program improvement plans. Teachers and their mentors understand the results. Moreover, they fall within the areas of interest of lawmakers and policy leaders. At the time of this decision, no published observation instruments for the domains of interest were available. The criteria in both the SELA and PCMI are closely related to the teaching practices delineated in the Preschool Teaching and Learning Expectations.

Since the fall of 2003, a random sample of more than 300 preschool classrooms across the 30 Abbott districts, covering 11-14% of classrooms in each district and stratified by public school or private provider, were observed annually. The results presented below are taken from Giant Steps for the Littlest Children: Progress in the Sixth Year of the Abbott Preschool Program (Lamy, Frede, and the ELIC, 2005). Results are reported statewide since in most districts, the sample size is inadequate to represent the district.

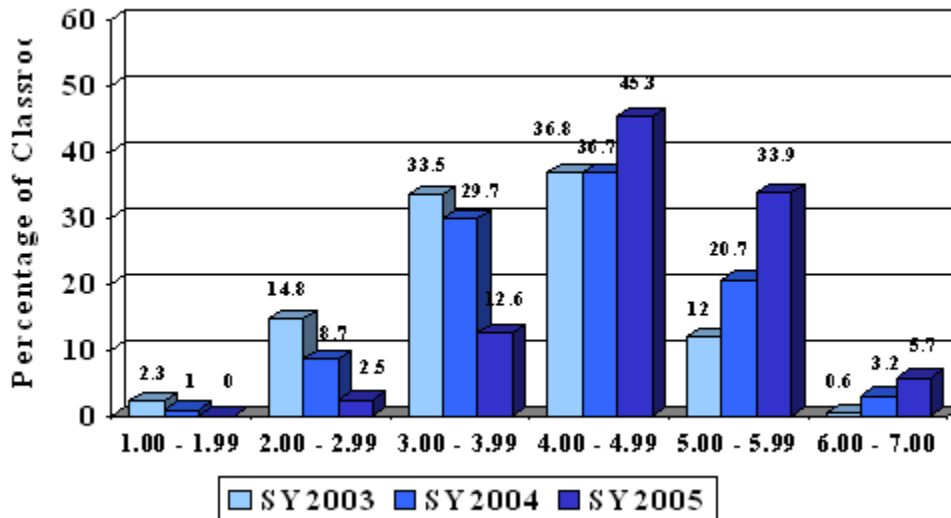
In 2005, average scores for each of the seven ECERS-R subscales increased significantly over previous years. Figure 1 indicates the change in total ECERS-R scores and ECERS-R subscale scores in a sub-sample of 14 Abbott districts in which classroom observations were conducted since the 1999/2000 school year, the beginning of the Abbott effort.

Figure 1. Change in ECERS-R Scores, SY2000 – SY2005



To understand the improvement in all 30 Abbott districts over the past three years, Figure 2 shows the change in the percentage of classrooms scoring 1 through 7 since 2002/2003 (the first year in which observations were conducted in all 30 districts). The percentage of classrooms scoring in the inadequate to minimal range has dropped from over 17 percent in 2003 to 2.5 percent in 2005, while the percentage of classrooms scoring in the good to excellent range has increased from about 13 percent in 2003 to nearly 40 percent in 2005. In 2002-2003, more than 50 percent of the classrooms scored below the midpoint on the seven-point scale. In the current school year 85 percent of the classrooms scored above the midpoint.

Figure 2. Percentage of Classrooms Scoring 1 – 7, SY2003 – SY2005



Results on the SELA also reveal substantial improvements. The percentage of classrooms scoring in the very low quality range has dropped from 12 percent in 2003 to 2 percent in 2005, while the percentage of classrooms scoring in the good to ideal range has increased from 10 percent in 2003 to 24 percent in 2005. In 2002 – 2003, 83 percent of the classrooms scored below the midpoint on this scale. In a dramatic shift, 75 percent of classrooms scored above the midpoint.

Classroom scores on the PCMI, while improving, remain much lower than those of the ECERS-R and the SELA and probably reflect the lack of emphasis given to math in early education generally. However, the percentage of classrooms scoring in the inadequate range dropped from 43 percent in 2004 to 24 percent in 2005, while the percentage of classrooms scoring in the fair to good range increased from 56 percent in 2004 to 75 percent in 2005. Although the percentage of classrooms scoring in the good to ideal range doubled, the percentage remains very small, moving from 0.6 percent in 2004 to 1.2 percent in 2005. However, the percentage scoring above the midpoint in 2004 was only 13 percent. It rose to 22 percent for 2005.

It is reasonable to assume that these dramatic increases in the ECERS-R and similar results on the SELA and PCMI, reported in Giant Steps, are at least partially due to the process of collecting this information, publishing it, helping districts understand it and developing state and local improvement plans based on it. Although results of observations are not published by district due to sample size limitations, DOE and ELIC personnel do discuss district results with the district and use them, along with the district-collected data described below, to develop professional development plans.

Measuring classroom quality to inform professional development at the district level. Since 2003, Abbott districts have been required by regulation to conduct annual structured observations of all preschool classrooms, to aggregate the data and to use the information for program improvement. Preschool professional development coaches,

called Master Teachers, conduct the observations in the beginning of each school year to help them tailor their classroom interventions and other professional development district wide. Typically, districts use the ECERS-R until most or all classrooms are scoring above a 5, then they switch to an instrument that either measures fidelity of implementation of their particular curriculum or that looks at more specific teaching practices.

Districts are required to set a low-end cutoff score for contracting classrooms. Any classroom not meeting that cutoff must have a classroom improvement plan with a timeline for improvement. The district ECE supervisor and the Master Teacher assigned to that classroom meet with the center director and the teacher to develop the plan. The district is required to offer intensive assistance. If adequate progress is not made, an outside, trained and reliable observer repeats the observation and if sufficient improvement does not occur, a center may lose the contract. To date, no center has lost a contract due to low scores on observation instruments. Although trained by the state in the instruments, not all Master Teachers have established reliability on the instruments — which is why the state does not collect this data and why outside observers are required to corroborate the district findings before a contract is terminated.

Measuring Child Outcome to Evaluate Program Effects. In 1999 and in most subsequent years until 2002, CEER collected data on oral language skills at kindergarten entry on a randomized sample of children. Our assumption was that if the scores went up over time and no competing explanation was evident, then it was reasonable to assume that preschool participation was the cause. The ELIC picked up this data collection beginning in 2002 and added a test for early literacy. By 2004, children's oral language scores had increased by one third of a standard deviation. We subscribed to Donald Campbell's directive not to measure a program before it is "proud", meaning that a more meticulous look at effects was not warranted until such time as it is up an running to the satisfaction those responsible for it.

In 2004-2005, with enrollment nearing 80% of the target population and more than 98% of the teachers having at least a bachelor's degree and work toward alternate route certification, combined with rising quality scores, we felt that it was time to conduct a more rigorous analysis of the programs' effects on child outcomes.

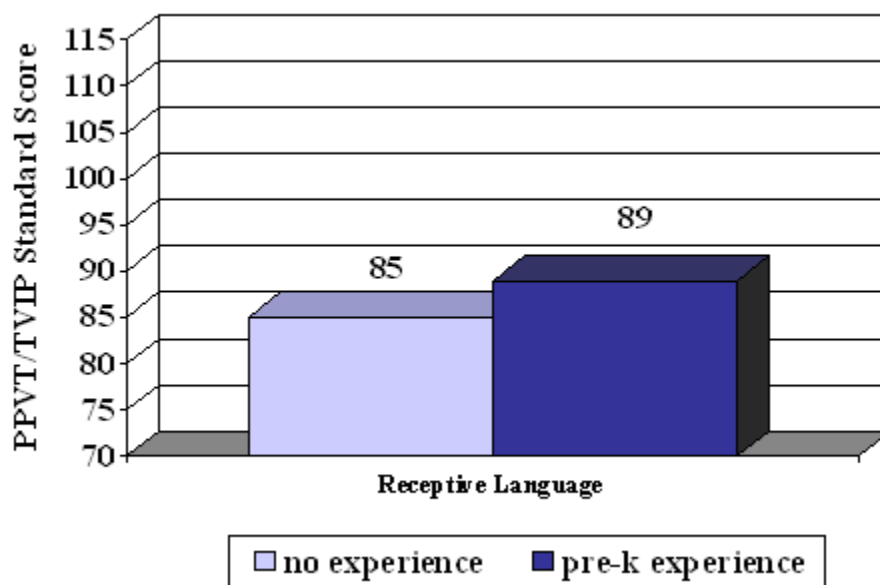
The new research design employed a regression-discontinuity statistical design that was successfully used in Tulsa, Oklahoma to measure the impact of preschool programs on kindergartners' skills (Gormley,). This statistical design enables researchers to compare preschool and kindergarten children.

More than 3,400 children were randomly selected across 3-year old and 4-year-old preschool classrooms as well as in kindergarten. We felt strongly that testing every child in the program as a method of either determining program or child progress is unnecessary, wasteful and potentially harmful because it may result in anemic curriculum, poor teaching methods and incorrect conclusions about the child and the program's success. This belief is also reflected in our program guidelines for districts.

However, we also believe that using random sampling as part of a rigorous research design is an appropriate and useful way to measure program outcomes.

Figure 3 displays the statistically significant ($p < .04$) effect of preschool attendance on average receptive vocabulary scores as measured by the Peabody Picture Vocabulary Test-III (PPVT-3) (Dunn & Dunn, 1997) and the Test de Vocabulario en Imágenes Peabody (TVIP) (Dunn, Padilla, Lugo & Dunn, 1986) for Spanish-speaking children. This four-point difference represents a difference of nearly four months in vocabulary development.

Figure 3. Effects of Abbott Preschool on Oral Language Skills



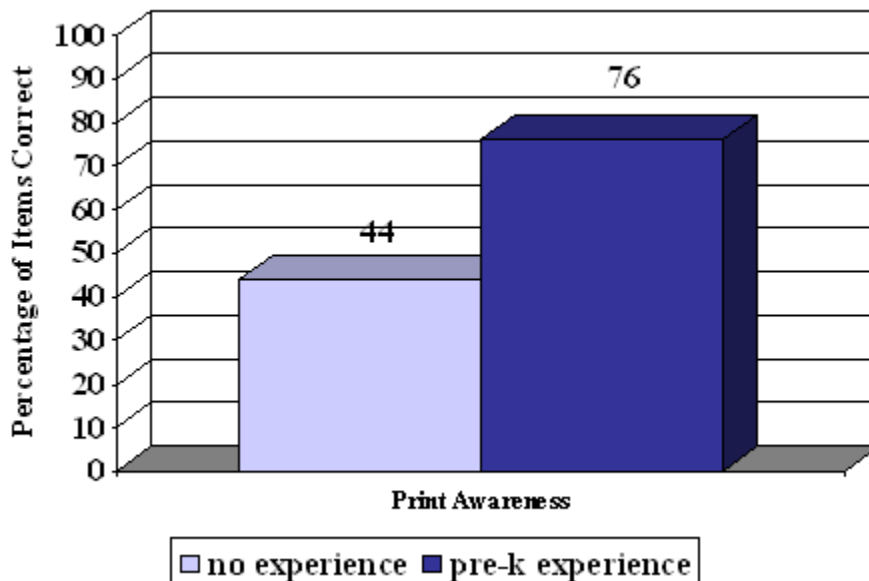
Early literacy skills were measured with the Print Awareness subtest of the Preschool - Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgeson & Rashotte, n.p.). Although not published as of this writing, the Pre-CTOPPP has been extensively validated on middle-class and low-income samples. The Print Awareness subtest measures children's ability to distinguish words and letters from pictures, and measures the extent to which children know that letters have distinct names, shapes and sound associations. As shown in Figure 4, the study found a highly statistically significant ($p < .000$) effect on Print Awareness scores. Kindergartners who attended an Abbott preschool program answered 76 percent of the items on the test correctly, while children who had not had preschool answered only 44 percent correctly.

Mathematical skills were measured with the Woodcock-Johnson Tests of Achievement, 3rd Edition (Woodcock, McGrew & Mather, 2001), Subtest 10, Applied

Problems. Initial analyses reported in Giant Steps did not show any effects on math skills. However, subsequent, as yet unpublished, analyses have. (Lamy, Barnett, Jung, Effects of State Preschool Programs on Young Children’s School Readiness in Five States).

These instruments were selected to measure the skills most directly associated with school success — an area of intense interest for lawmakers and policy leaders. Limited resources dictated that instruments selected be easily and fairly inexpensively administered. The instruments also needed to have Spanish language versions. At the same time, we wanted measures that were widely known and the scores from which would be meaningful to the audience(s). Our purpose in this assessment was not to inform teaching practice or to measure individual child progress but to provide information on the effectiveness of the program.

Figure 4. Effects of Abbott Preschool on Entering Kindergartners’ Early Literacy Skills



Measuring Child Progress to Improve Teaching and Learning. In addition to using child assessment on a random sample of children to evaluate progress in the Abbott program statewide, teachers use on-going child assessment as part of the Continuous Improvement Cycle in their classrooms. To assist teachers in using valid classroom assessment to improve teaching, in 2003 the DOE, with the help of the ELIC, created a language and literacy performance-based assessment. The Early Learning Assessment System (Wollock and the ELIC, 2003) uses teacher observation and portfolio documentation to assess children’s emerging oral language and literacy skills as delineated on the Preschool Teaching and Learning Expectations and the kindergarten Core Curriculum Content Standards. Teachers collect data over three collection periods. The data is analyzed and used to score a rubric on a scale of 1 to 4. Teachers then use

this information to adapt and individualize interactions for specific children and to adjust activities for the entire classroom as well.

We chose to focus on only the language and literacy component because as teachers collect information on oral language they also learn about most content areas. Language and literacy was also the domain of primary interest to lawmakers in New Jersey. (Training more than 5000 teachers, master teachers and administrators in ELAS was daunting enough without adding additional domains.)

As stated above, we believed that using standardized tests on every child is unnecessary. In addition, tests would only render information on a subscribed portion of our content standards. The major purpose of the ELAS is to help teachers teach better. Work done previously in Maryland and research conducted by instrument developers convinced us that performance-based assessment systems can result in quantifiable results. The major impediment to this is establishing reliability and validity of the assessment tool and procedures.

Initially, in 2003-2004, we trained all preschool and kindergarten teachers in 7 of the 30 districts to use the instrument. We conducted extensive meetings to determine what adjustments needed to be made in the protocol and training. In 2004-2005, we selected some districts to receive direct training of teachers and others to receive turnkey training by their Master Teachers. The basis of this selection was mostly the quality of the program already being implemented, whether the district was currently implementing a performance based assessment system, and the expertise of local trainers. In addition, we conducted calibration studies in a few classrooms. As of this writing, we plan to expand the calibration study, compare results on the ELAS with the outcome measures used in the statewide analyses to determine if we have concurrent validity and investigate the psychometric properties of the instrument.

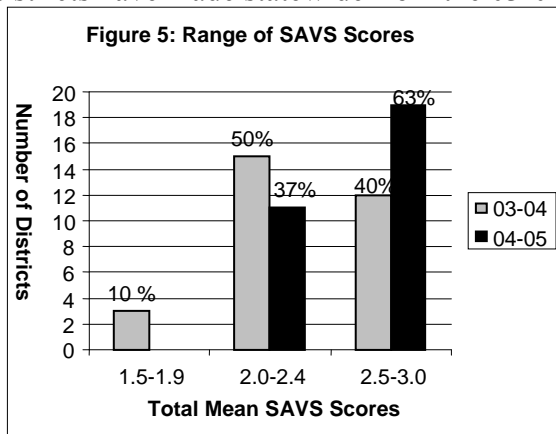
Once we have established the validity of the ELAS and have a statewide data base for individual students, the final score of the ELAS will be collected statewide and used as another measure of program effect, comparing scores from the end of each year to assess children's progress. With these analyses, we are convinced that this method of assessment will merit at least as much confidence as standardized assessment measures which are fraught with problems at this age level. Choosing to pursue this method was costly but the payoff is substantial. Our initial state wide classroom quality evidence indicates that we are improving teaching and learning through the assessment process rather than diminishing them as tests might do.

Measuring Program Quality

The Self Assessment Validation System (SAVS; This section is adapted from Wilkins & Frede, 2005 and Frede and DeMarko-Rowe, 2004), begun in the 2003-2004 school year is an annual process that assists districts and the state in implementing the continuous improvement cycle. The SAVS is designed to guide districts through a systematic self-appraisal of its preschool program to aid in program improvement. The SAVS criteria are

derived from the NJ Abbott Preschool Program Implementation Guidelines, and thus provide an accountability measure for program standards. Each of the 45 items on the SAVS is rated on a three point scale where 1 = not met, 2 = in progress and 3 = fully met. The district then receives an overall SAVS score derived by averaging the item scores. Although instruments exist to assess program administration such as the Program Administration Scale (Talan and Bloom, 2005) and the NAEYC Accreditation Criteria and Procedures (NAEYC, 1998), none of these measure the specific criteria that were relevant to the Abbott program. This is why we chose to develop our own which would be clearly related to guidelines and regulations of the DOE and be of more direct assistance to districts.

During Phase I, district personnel, in collaboration with other relevant parties, such as the local Early Childhood Advisory Council, assess their early childhood program. These initial ratings inform revisions to the operational plan and budget requests submitted to the Department of Education. Since the SAVS is a program improvement tool, districts are encouraged to look critically and honestly at their programs. Phase I is a time for self-assessment and planning for improvement. Phase II allows districts and their DOE liaisons to rescore the SAVS noting growth, areas still in need of development and collecting documentation to justify their scores. These results are then submitted to the department. A validation visit occurs for approximately 1/3 of the districts each year to verify the scores reported in the SAVS. Districts use the results of the SAVS, classroom observations and other data to develop detailed professional development plans which are submitted to the DOE. Figure 5 below shows the progress districts have made statewide from the 03-04 to the 04-05 school year.



Accountability and Fiscal Responsibility

In an effort to ensure that funds are used efficiently and appropriately, the Department instituted fiscal accountability measures which include the following:

- Promulgating clear and specific budget guidance for the more than 500 private providers who contract with districts;
- Quarterly review by district fiscal specialists of expenditures;
- Limited audits of a random sample of approximately 100 agencies;
- District enrollment and expenditure reviews.

Preliminary findings, suggest that providers have difficulty segregating Abbott expenditures, accurately and completely filling out Quarterly Expenditure Reports and record keeping in general. A small number of centers have been suspected of fraudulent expenditures and these have been remanded to the criminal justice system. However, progress is being made. These findings have led to training of fiscal specialists in how to provide technical assistance to providers and a training module is being developed. Providers committing gross negligence or malfeasance have lost their contracts.

Some Examples of Professional Development and Other Technical Assistance

The DOE uses results of all of the data collection efforts to design effective technical assistance for individual districts and to determine appropriate topics for professional development statewide. The following matrix gives some examples of these activities:

SOME FINDINGS OF CONCERN	EXAMPLE INTERVENTION
Original classroom quality	TA to districts on choosing a curriculum and publishing criteria for the five recommended curricula
Supporting English language learners	Require districts to employ bilingual specialists. Provide various professional development opportunities
Classroom practices in mathematics	Specialized PD on Math and begin development of Math ELAS
Inclusion of children with special needs	Require districts to employ Preschool Intervention and Referral Teams and inclusion specialists. Provide intensive training in Positive Behavioral Supports by Lise Fox. Joint training with DOE Office of Special Education.
Continuity and transition	Summit on transition practices followed by specific and comprehensive district transition plans
Family services and community involvement	Require districts to employ Community and Parent Involvement Specialists and provide quarterly state level TA meetings
Fiscal integrity of some private providers	Districts work with the provider to develop a corrective action plan and then monitor its progress. In some cases, contracts were not renewed for providers with egregious findings. Money was recovered from providers where actual expenses were less than approved budgets. To further this effort, a training module on budget and accounting will be developed for district

	fiscal specialists to use with their providers.
Using screening tests incorrectly e.g. as pre and post assessments	Multiple individual TA sessions, presentations on appropriate use of screening tools and other assessments

Plans for the Future

Since 2002, the Office of Early Childhood Education in New Jersey has strived to make policy and intervention decisions based on systematic information in order to improve the programs we provide to preschool children. To do this, we developed standards, collected data at the state, local, and child levels, and used this data to plan for improvement. We then measured our progress and began the process again. We have asked teachers, professional development coaches, and district administrators to follow this same process.

The continuous improvement cycle will inevitably lead to changes in assessment in the future. Last year, the DOE with representatives from DHS, centers and the districts developed the Center Level Assessment System (CLAS), a voluntary self-assessment system for child care centers. Training in the system will be offered this year. As part of the ELIC, colleagues and I at The College of New Jersey are expanding the ELAS into mathematics and are currently piloting the protocol and instrument. Plans call for developing a social/emotional version in the future. As classrooms reach an average score above 5 on the ECERS-R, the state will likely replace it with other domain-specific observation instruments. Each year the SAVS criteria become more stringent, in the process raising the bar.

As part of the ELIC, NIEER will repeat the data collection on children conducted last year, adding data on family background as the first year in a planned longitudinal study of the Abbott Preschool Program. ELIC members will complete analyses of the relationship of classroom quality to child outcome using last year's data and we will look at the predictive validity of the PCMI and the SELA.

New Jersey's Abbott Preschool Program is an ambitious endeavor ordered by the Court to ensure that young children in low income, primarily urban districts start school with chances similar to their suburban peers. It is incumbent on us to gather the best information possible to meet the challenge set forth by the Court and to implement data-based decision making. Operating at the state level with a young universal program and limited evaluation funds, one is never completely confident in the answers revealed — and less confident in those implied. Nevertheless, New Jersey's system of assessment and continuous improvement is superior to the alternative of making decisions on an *ad hoc* basis. By making the system iterative and data collection somewhat redundant while avoiding the pitfalls of testing every child, we can say with confidence that the programs are improving and the children are better prepared to succeed.

Ellen Frede has returned to her faculty position at The College of New Jersey. From February, 2002 through August, 2005, she was Assistant to the Commissioner, Office of Early Childhood Education, New Jersey Department of Education. Although based on the results of research for its conclusions, much of this paper is a narrative of the decision-making process in which she participated while developing an accountability system for the Abbott Preschool Program.

References

- Barnett, W. S., Tarr, J. & Frede, E. (1999). Early Childhood Education Needs in Low Income Communities: Final Report of an Assessment of Young Children's Educational Needs and Community Capacity in New Jersey's Abbott Districts. Rutgers University, New Brunswick, NJ: CEER.
- Barnett, W. S., Tarr, J., Esposito Lamy, C. & Frede, E. (2002). Fragile Lives, Shattered Dreams: A Report on Implementation of Preschool Education in New Jersey's Abbott Districts. Rutgers University, New Brunswick, NJ: CEER.
- Dunn, L. M. & Dunn, L. M. (1997). Peabody Picture Vocabulary Test-Third Edition (PPVT-3). Circle Pines, MN: American Guidance Services.
- Dunn, Padilla, Lugo & Dunn, (1986). Test de Vocabulario en Imágenes Peabody (TVIP). Circle Pines, MN: American Guidance Services.
- Harms, T., Clifford, R. & Cryer, D. (1998). Early Childhood Environment Rating Scale – Revised (ECERS-R). New York, NY: Teachers College Press.
- Frede, E., Dessewffy, M., Hornbeck, A. & Worth, A. (2003). Preschool Classroom Mathematics Inventory. Unpublished instrument.
- Lamy, C. Esposito, Frede, E. & the ELIC (May, 2005) Giant Steps for the Littlest Children: Progress in the Sixth Year of the Abbott Preschool Program. Trenton, NJ: New Jersey Department of Education. www.nj.gov/njded/ece
- New Jersey Preschool Teaching and Learning Expectations: Standards of Quality (2004): Trenton, NJ: New Jersey Department of Education. www.nj.gov/njded/ece
- Kindergarten New Jersey Core Curriculum Content Standards (NJDOE, 2004): Trenton, NJ: New Jersey Department of Education. www.nj.gov/njded/ece
- Smith, S., Davidson, S. & Weisenfeld, G. (2001). Support for Early Literacy Assessment (SELA). Unpublished instrument.
- Talan, T.N. & Bloom, P. Jorde (2004) Program Administration Scale: Measuring Early Childhood Leadership and Management. NY: Teachers College Press
- Whitehurst, G. J. & Lonigan, C. J. (2000). The Get Ready to Read. Washington, DC: National Center for Learning Disabilities.
- Wilkins, R. & Frede, E. (2005) Self-Assessment Validation System (SAVS) 2004-2005: Preliminary Report on Statewide Progress in Abbott Preschool Program Implementation. Trenton, NJ: New Jersey Department of Education. www.nj.gov/njded/ece
- Woodcock, McGrew & Mather (2001). Woodcock-Johnson Tests of Achievement, 3rd Edition. Itasca, IL: Riverside Publishing.