

## Early Childhood Assessment: A Correlational Study of the Relationships Among Student Performance, Student Feelings, and Teacher Perceptions

Julie F. Schappe<sup>1,2</sup>

---

Current preschool assessments lack the ability to capture performance variance and complex thinking skills. This study examined the relationship among student performance, student feelings, and teacher perceptions as they apply to using a formal performance assessment in a preschool setting. A formal performance assessment was administered to 71 preschool-aged children (56–71 months of age). The data were analyzed using descriptive and inferential statistics. As hypothesized, there was a significant and positive correlation between the students' performance and the teachers' perceptions. Furthermore, there was not a significant relationship between either the students' feelings and the teachers' perceptions or the students' feelings and the students' performance. These findings and theoretical study indicate support for the use of formal performance assessment to evaluate preschool students. Limitations and implications of the results are discussed.

---

**KEY WORDS:** early childhood assessment; preschool performance assessment; teacher perceptions; affective assessment; preschool student evaluation; dichotomous checklists; student performance.

Assessment in the preschool classroom typically consists of anecdotal notes based on observations made by the teacher and dichotomous checklists (Eliason & Jenkins, 1990; Neuman & Roskos, 1993; Pretti-Frontczak, Kowalski, & Brown, 2002). The most commonly used dichotomous assessments ignore variance and complexity in performance. Although some preschool programs, notably the Head Start initiative, endeavor to implement varied assessment approaches, unfortunately only a small portion of preschool aged children in the United States are served by these programs (National Center

for Educational Statistics, 2001). The National Research Council Committee on Early Childhood Pedagogy (NRC) specifically notes the failure of existing preschool assessment methods to accurately reflect the breadth and depth of the preschool student's abilities (Bowman, Donovan, & Burns, 2001). The NRC asserts that based on cognitive research findings, the young child's knowledge is more complex than previously considered; instructional and assessment practices should address this complexity (Bowman et al., 2001). Methods of assessment must measure the full range of assisted and unassisted student performance to better inform the teacher-student relationship and subsequent instructional practices. The National Association for the Education of Young Children (NAEYC, 2003) asserts the need to design assessments that are based on a sound understanding of early childhood development and relevant to curricular goals for preschool children. Developmentally Appropriate Practice (DAP)

---

<sup>1</sup>Department of Behavior Sciences and Education, Pennsylvania State University, Camp Hill, PA, USA.

<sup>2</sup>Correspondence should be direct to Julie F. Schappe, Department of Behavior Sciences and Education, Pennsylvania State University, 9 Blackmore Court, Camp Hill, PA 17011, USA; e-mail: jschappe@hotmail.com

supports the use of multiple assessments that involve observation of children while actively engaged in learning (Bredekamp & Copple, 1997). Neuman and Roskos (1993) cite the main difficulties surrounding observation of students are inconsistencies in teachers' decisions involving what to observe, how to observe, how to record observations, and how to make inferences. This study seeks to address this shortcoming in assessment practices by advocating the use of formal performance assessments with preschool students.

Performance assessment is an assessment method that creates a situation designed to actively engage students in a process or task. This active process requires a demonstration of skills or behaviors that are central to developmental and curricular goals. Teachers structure a performance assessment to obtain a picture of the student's learning and rate that demonstration according to a set of scalable criteria. Scalable criteria reflect potential variations in mastery. Popham (2002) identifies seven criteria to consider when designing a performance assessment. He asserts that effective performance tasks should be authentic in context, based on instructional objectives, generalize to comparable tasks, measure multiple instructional outcomes, avoid bias, fit within the resource constraints of the classroom or learning environment, and allow student demonstrations to be reliably and accurately evaluated. Criteria equivalent to each of Popham's criterion for performance assessments appear in the NAEYC position statement (2003) outlining effective assessment practices. The NRC states three clear advantages apply to preschool performance assessment: (a) allows for individual differences in the performance demonstrations; (b) hands-on activities increase student motivation; and (c) performance assessments tend to provide more information in a context that is appealing to and understood by families (Bowman et al., 2001).

Inappropriate assessment practices can waste instructional time, limit the classroom learning experience, alter support services, and incorrectly influence a student's placement for the following academic year. Assessments that are limited in scope misrepresent the child's development because they may ignore variations that result from cultural or contextual perceptions (Fischer, Bullock, Rotenberg, and Raya, 1993; Kagan, Moore, & Bredekamp, 1995; NAEYC, 2003). NAEYC (2003) and Bredekamp and Copple's (1997) assertions that assessments must reflect variation in individual development supports a scalable, rather than dichotomous, measure. Opera-

tionalizing constructs on a scalable performance assessment will reflect the range of the child's progress or developmental need. Through the incorporation of formal performance assessment, this study seeks to arm the decision makers with documented demonstrations of student performance obtained in controlled circumstances and grounded in researched-based evaluative criteria.

The flexibility inherent in the design of a formal performance assessment allows for consistency between the assessment criteria and widely accepted DAP benchmarks for developmental skill levels and behaviors in the preschool classroom. The investigation will address whether or not the use of a specific performance assessment captures performance that will complement the teacher's evaluation of the students' social, cognitive, and fine and gross motor development. In addition, this study measures the students' feelings about the assessment as a gauge to determine whether these feelings correlate with the outcome of the assessment. For the purposes of this study, three measurable criteria provide the basis for evaluation: (1) the teachers' perceptions of achievement measured through survey responses; (2) the students' demonstrated performance measured by a scoring rubric; and, (3) the students' feelings about the activities of the assessment measured by an affective self-assessment. The theoretical framework in Figure 1 illustrates the proposed relationships among the study variables. Specifically, it is hypothesized that:

- H1: There will be a significant and positive relationship between the students' demonstrated performance and the teachers' perceptions of overall student achievement.
- H2: There will not be a significant relationship between the students' feelings about the assessment activities and the students' performance on the assessment.
- H3: There will not be a significant relationship between the students' feelings about the assessment activities and the teachers' perceptions of overall student achievement.

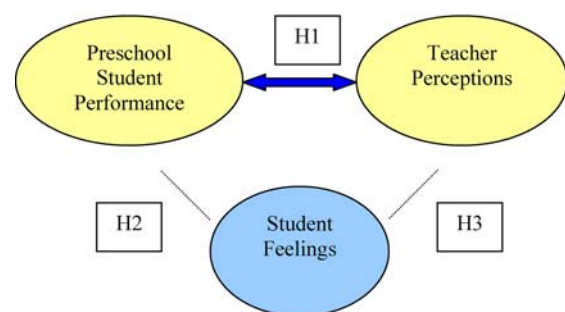


Fig. 1. The Hypothesized Relationship among Study Variables. H1 Hypothesizes a Significant Relationship. H2 and H3 Hypothesize Non-significant Relationships.

## METHODOLOGY

### Subjects

The study subjects were Caucasian, middle socio-economic status, and from a single suburban school. The student subjects ( $n = 71$ ) were pre-school-aged students, 56–71 months of age. There were 41 male and 30 female students in four different self-contained classrooms. The lead teachers ( $n = 4$ ) in these classrooms were the adult teachers participating in the study.

### Data Collection

The data were collected using three measures: a teacher perception survey, a performance assessment rubric, and a student affective self-assessment.

#### *Teacher Perception Survey*

The teacher perception survey contains a defined Likert-type scale with values from 1 (never) to 5 (always) reflecting the extent to which the teacher observed the student demonstrating the 25 identified skills or behaviors. Items 1 through 16 reflected primarily social, emotional, language, and cognitive development; items 17 through 25 specified fine and gross motor behaviors or skills. Sample items included the following: “Speaks in complete sentences of 5–6 words,” “Independently initiates meaningful activity,” and “Follows multiple step directions.” All items included in the survey were based on age-appropriate skills specified in DAP (Bredekamp & Copple, 1997). Given the range of life experience and education of the teachers, clear parameters that specifically described each skill or behavior and the rating scale were established in an effort to reduce personal bias and promote consistency in the responses.

#### *Performance Assessment Rubric*

The performance assessment rubric was based on an authentic performance assessment activity that required each student to build a neighborhood community. The performance assessment rubric measures the student’s demonstrated performance according to four evaluative criteria: organization, content, communication, and self-confidence/self-direction. The first three criteria represent elements from the cognitive taxonomy and the fourth from the affective taxonomy. The structure of the scoring rubric was based on accepted methods for rubric evaluations

(Kubiszyn & Borich, 2000; Popham, 2002). The criteria selected and the language used to describe the ratings to reflect multiple developmental foci and age-appropriate demonstrated behaviors based on DAP (Bredekamp & Copple, 1997). Consideration was given to select four criteria that represented the student’s ability to synthesize previous knowledge and experience to produce a new product. The researcher rated each criterion on a scale ranging from 1 (additional support needed) to 4 (demonstrated independently). Specific demonstrated behaviors anchored each rating choice to minimize error due to bias, therefore reducing the inter-rater variance. Sample scale descriptors for the criteria of content indicated, “The student included more than 3 neighborhood elements representing both homes and community buildings” for a rating of 4, and “The student’s construction did not relate to our topic on neighborhoods” for a rating of 1.

#### *Student Affective Self-Assessment*

The student affective self-assessment measured how students felt about the four activities associated with the assessment: looking outside, reading books, art, and building. The four-item assessment was arranged in horizontal rows. The simple word identifier for each activity was listed next to a picture of that word. Next to each activity picture were three faces: happy, neutral, and sad. For example, from left to right appeared the word “art,” then a picture of markers, and finally, the happy, neutral, and sad faces.

### Procedure

#### *Distribution and Collection of Teacher Perception Surveys*

The teachers were instructed to fill out one survey per child in their class. The teachers placed the completed surveys in a sealed envelope and returned them to the researcher prior to the administration of the performance assessment in their classroom.

#### *Administration of the Performance Assessment*

The researcher spent two 45-minute sessions in each classroom. Given the age of the minor subjects, a standardized script was used to enhance student understanding of the instructions, minimize rater error, and potentially increase the reliability of the assessment. Students were able to work collaboratively during the planning and information gathering

phase, but the neighborhood communities were built independently. The students chose from various classroom materials to construct the neighborhoods. The researcher provided ongoing support by facilitating discussion. Some students required more assistance than others to bring the project to completion, which was reflected in the scoring guidelines. The researcher was the sole administrator of the performance assessment and the sole evaluator for the scoring rubric. The researcher chose to control these two activities to reduce bias and increase consistency.

#### *Administration of the Student Affective Self-Assessment*

Once the scoring rubrics were completed, the researcher administered the student affective self-assessment. The researcher read the form to the students individually and instructed the students to color the face that best reflected how they *felt* about each activity; the study did not ask the students to self-assess their performance.

#### **Data Analyses**

The data were analyzed using descriptive, correlational, and inferential statistics. Twelve surveys were eliminated due to incomplete information or student absence during the performance assessment; consequently, data from 59 subjects were analyzed. The descriptive evaluation involved the calculation of the mean and the standard deviation to determine the variability of the scores. The Pearson Product-Moment Correlation ( $r$ ) was used to evaluate the association between each of the constructs measured. Additionally, ANOVA tests were used with the teacher perception survey and the student affective self-assessment to determine the significance level of mean differences by class. The chi-square test was used to determine the variance among classes for the four performance assessment criteria.

#### **RESULTS**

Table I illustrates the correlations among the study variables. Support was found for the first hypothesis (H1); namely, there is a significant and positive correlation between teachers' perceptions and the students' demonstrated performance ( $r = .35$ ,  $p = .003$ , one-tailed). Support was also found for hypothesis H2: there was not a significant relation-

**Table I.** Correlation Matrix for Study Variables

Measure	1	2	3
1. Teachers' perceptions	–	.35**	.08
2. Students' performance		–	.02
3. Students' feelings			–

Note: Equal variances not assumed. \*\* $p < .01$ , one-tailed.  $N = 59$ .

ship between the students' demonstrated performance and the students' feelings about the assessment activities ( $r = .02$ ). In addition, support was found for hypothesis H3: there was not a significant relationship between the teachers' perceptions and the students' feelings about the assessment activities ( $r = .08$ ).

ANOVA tests were conducted to determine if there were significant mean differences among classes for the student affective self-assessment and the teacher perception survey. Although the mean statistics for the student affective survey differed slightly from one another, the ANOVA results did not indicate a significant mean variance between classes,  $p = .14$ . Descriptive statistics for the teacher perception survey reported for Class 1 ( $M = 4.02$ ) and Class 4 ( $M = 4.69$ ) reflect the greatest difference. The ANOVA results for the teacher perception survey (Table II) indicate a significant mean variance between classes,  $p = .03$ .

The Scheffé post-hoc test was used to further explore the variance between classes (Table III). The Scheffé results indicate a significant mean difference between Classes 1 and 4,  $p = .04$ . There were no significant mean differences between the other classes for the teacher perception survey.

The differences among the classes for the performance assessment rubric were examined using the chi-square analysis. Although the results from the chi-square analysis for the four criteria of the performance assessment rubric (organization, content, oral communication, and self-direction) did not indicate a significance level less than .05 (two-sided), the asymptotic measure of significance for oral

**Table II.** Analysis of Variance Among Classes for the Teacher Perception Survey (TPS)

TPS	SS	df	MS	F	P
Between groups	.46	3	.15	3.10	.03*
Within groups	4.40	55	.08		

Note: \* $p < .05$ .

**Table III.** Scheffe Post-Hoc Test for Significant Mean Differences Teacher Perception Survey

Class	<i>P</i>			
	1	2	3	4
1	–	.89	.79	.04*
2		–	.99	.20
3			–	.33
4				–

Note: \* $p < .05$ .

communication ( $p = .050$ ) did warrant further examination. Neither the likelihood ratio (value = 16.36,  $p = .06$ ) nor the linear-by-linear association (value .82,  $p = .37$ ) indicated significance. Consequently, the chi-square indicated no significant differences among classes and the four criteria of the performance assessment rubric.

## DISCUSSION

The students' performance as measured by the assessment rubric positively and significantly correlated with surveyed teachers' perceptions of overall student achievement. The support for the first hypothesis suggests that the results from the performance assessment might be included in the range of preschool classroom assessments as an example of student achievement when referring to the student's overall progress. Further, the results support Bowman et al.'s (2001) assertion that the young child is capable of complex thinking.

The absence of a significant relationship between student feelings and performance could be attributed to the stage of development of the young child. Although research supports the relationship between attitude and performance in the school-age child, support was not found for the preschool-aged child in this study. The young child lives in the moment (Piaget, 1974). Moods and preferences can be influenced by numerous outside factors that do not involve specific learning activities. It is entirely possible that peer dynamics, family issues, health, or other external factors may have affected the children's moods and, consequently, this result. Given that affective measures are generally accepted as valuable indicators of performance for older students, future research may focus on repeated administration of affective measures to verify consistency, address discrepancies, and the influence of external variables for the preschool student.

Applying the ANOVA to the teachers' perceptions survey and the student affective self-assessment analyzed the data to determine whether significant mean differences between classes affected the results of the study. Although no significant variance was found for the student affective self-assessment, the ANOVA test did show a significant mean difference with regard to the teacher perception survey among Classes 1 and 4. As stated in the methods section, in order to reduce bias and increase inter-rater reliability, the researcher was the sole administrator of the performance assessment and the student affective self-assessment. Although the researcher provided uniform instructions that defined the observed skills and behaviors and the rating scale used for the completion of the teacher perception survey, each individual teacher relied upon personal and professional experience with the children in order to complete each survey. The four teacher participants in this study varied in education and years of teaching experience. It is possible that variance in experience contributed to the significant mean difference with respect to the teacher perception survey. In order to rule out explanations that pointed to the students, the ANOVA results between the classes for the student affective self-assessment and the chi-square test for the performance rubric were examined. No significant mean differences were found between the classes and the other measures. Consequently, the difference for the teacher perception survey between Classes 1 and 4 can be reasonably attributed to the teacher. Theoretical studies indicate the pivotal role of the teacher in the classroom. Bowman et al., (2001) stresses inconsistencies in preschool teacher training and the limited control over qualifications of teachers given that preschools operate outside the sphere of public education. Researchers should continue to consider this variable in future studies.

## Study Limitations

This study utilized convenience sampling. Given that preschool programs vary widely, the researcher selected one school for the study as a means of limiting the influence of variability in program curriculum and teaching practices. The relative power of the statistical analyses of this study would be increased if the study were implemented in multiple schools resulting in a larger sample size. In addition, samples with varied racial, cultural, socio-economic, and geographical characteristics are necessary to enhance the generalizability of the study's findings.

Neither intelligence nor the instance of special needs were measured or statistically controlled in this study. Scripts and defined scale parameters were used to limit bias. However, the subjects were tested on a single occasion; efforts were not made to control for outside factors (health, personal relationship difficulties, or mood) that could have affected the subjects on a given day. Although gross motor skills were included in the teacher perception survey, the study did not incorporate a performance demonstration of gross motor skills. Gross motor skill development is an important part of the early childhood environment and should be a part of classroom activities and assessments (Bredekamp & Copple, 1997; Piaget, 1974). This study does not suggest that a single performance assessment should represent a complete picture of a student's overall achievement.

The study did not attempt to measure the effectiveness of performance assessment as a communication tool. Whether the information from the performance assessment can be easily understood or interpreted by parents cannot be deduced from this study. Nor does this study address the teachers' perceptions regarding their understanding of the methods. Professional development and specific forms of parent-teacher communication for early childhood environments are important topics for future research.

## CONCLUSION

The results of this study support the three propositions put forth by this research. First, the results suggest that administering a formal performance assessment was compatible with the preschool environment. The subject matter of the assessment is a common theme in early childhood environments and one that relates to the student's real life. Given the brevity of the assessment (the ability to fully assess a class of 18 students in two 45-minute sessions) and the compatibility with existing practices (the material needed is commonly found in the classroom), the assessment is readily adaptable to most classrooms.

Second, the students' performance on the assessment positively and significantly correlated with the teachers' perceptions of overall performance. This result is consistent with widely held expectations for older, school-aged students. Research supports the use of performance assessment to address higher-order skills in the elementary, middle, and secondary classroom. This study

supports the use of this type of complex assessment as an effective tool for early childhood teachers. Early learning is complex as the development of social, cognitive, and motor skills interplay and impact overall achievement (Bredekamp & Copple, 1997; Bruner, 1983; Vygotsky, 1978). These skills do not develop in isolation. Simple checklists most often used in early childhood classrooms do not reflect the range of development that occurs in the classroom and tend not to foster complex thinking. Young children do perform at varying levels (Bowman et al., 2001; Bredekamp & Copple, 1997; NAEYC, 2003). As educators seek to answer questions regarding the development of the child, assessments that allow for variation and complexity will best serve the child.

Third, although affective measures are commonly used with school-aged children (Popham, 2002), the results from this study did not support a significant relationship between attitudes and performance for the preschool-aged child. However, the study does question the effect of the variables of age and maturity on the relative impact of affective behaviors.

Consequently, although the results of this study indicate support for the use of performance assessment as a tool to represent overall achievement in a preschool setting, the significant variable in the successful integration of formal performance assessment is the teacher. Theoretically, the classroom teacher is more likely to provide a comfortable, less intimidating assessment environment for the preschool child. Consequently, such an environment should produce a more realistic demonstration of performance. However, the importance of the classroom teacher in the assessment process highlights the inherent weakness in many early childhood classrooms. Enrollment in center-based educational programs indicates that 69.6% of four-year-olds and 76.5% of five-year-olds participate in preschool programs through either Head Start, private preschools, day care schools, or local preschool programs (National Center for Educational Statistics, 2001). Preschools are generally exempt from the rigorous qualifications that apply to elementary schools and teachers. As such, federal guidelines and state requirements for teacher preparation, curricula, certification, and continuing education do not apply to the teachers and schools serving the vast majority of young children at the preschool level in the United States. Professional development of teachers at this level will be an important element when considering appropriate

assessment practices (Bowman et al., 2001; Bredekamp & Copple, 1997; NAEYC, 2003). Researchers and policy makers must work together to improve the quality and consistency of preschool teacher preparation and continuing education. Young children develop at rapid and varied rates. Assessments, such as formal performance assessment, that reflect variation in performance and patterns of complex thinking skills will ultimately provide a more accurate assessment of student achievement.

#### ACKNOWLEDGMENT

The author completed this study in partial fulfillment of the requirements for the degree of Master of Education. The author would like to thank her committee: Dr. Denise Meister, Dr. Steven Melnick, and Dr. Caroline Owens.

#### REFERENCES

- Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.). (2001). *Eager to learn: Educating our preschoolers* (National Research Council Committee on Early Childhood Pedagogy, Commission on Behavioral and Social Sciences and Education). Washington, DC: National Academy Press.
- Bredekamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs* (Rev. ed.). Washington, DC: National Association for the Education of Young Children.
- Bruner, J. (1983). Play, thought and language. *Peabody Journal of Education*, 60(3), 60–69.
- Eliason, C. F., & Jenkins, L. T. (1990). *A practical guide to early childhood curriculum* (4th ed.). Columbus, OH: Merrill Publishing Company.
- Fischer, K. W., Bullock, D. H., Rotenberg, E. J., & Raya, P. (1993). The dynamics of competence: How context contributes directly to skill. In Woziak, R., & Fischer, K. W. (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 93–117). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kagan, L., Moore, E., & Bredekamp, S. (Eds.). (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary*. Washington, DC: National Education Goals Panel, Goal 1 Technical Planning Group.
- Kubiszyn, T., & Borich, G. (2000). *Educational testing and measurement* (6th ed.). New York: John Wiley & Sons, Inc.
- National Association for the Education of Young Children. (2003, November). Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8. Retrieved September 13, 2004, from [http://www.naeyc.org/resources/position\\_statements/pscape.pdf](http://www.naeyc.org/resources/position_statements/pscape.pdf).
- National Center for Educational Statistics. (2001). *National Household Education Survey (NHES), 1991, 1995, 1999* (Table 45). Washington, DC: Department of Education.
- Neuman, S. B., & Roskos, K. A. (1993). *Language and literacy learning in the early years: An integrated approach*. Orlando, FL: Harcourt Brace Jovanovich.
- Piaget, J. (1974). *To understand is to invent*. New York: Wiley.
- Popham, W. J. (2002). *Classroom assessment: What teachers need to know* (3rd ed.). Boston: Allyn and Bacon.
- Pretti-Frontczak, K., Kowalski, K., & Brown, R. D. (2002). Preschool teacher's use of assessments and curricula: A statewide examination. *Exceptional Children*, 69(1), 109–123.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge: Harvard University Press.