3 Psychometric Properties of the Preschool PQA

Overview

The Preschool PQA has been used extensively as a research tool by trained independent raters in over 800 preschool classrooms and child care centers. The earliest version, the Program Implementation Profile (PIP), was used in Training for Quality, a national study of 366 diverse early childhood settings (Epstein, 1993, 1999). The second version, or "old" PQA, expanded the content areas assessed from four to the current seven and was used (sometimes with minor variations) in five studies: the 1995 and 1996 cohorts of the Michigan School Readiness Program (MSRP) evaluation with 49 and 32 classrooms, respectively (High/Scope Educational Research Foundation, 1997); the 1998 and 1999 staff development studies of the High/Scope Head Start Quality Research Center (QRC) with 41 and 30 classrooms, respectively (Schweinhart, Epstein, Okoloko, & Oden, 1998; Schweinhart, Oden, Okoloko, Epstein, & Markley, 2000), and the evaluation of the Hansel Head Start Program with 14 classrooms (Schweinhart, 2000).

The third and current version, or "new" PQA, revised the scoring system. It was field-tested in two research projects. The first was the 2000 cohort of Phase 2 of the MSRP evaluation with a sample of 19 classrooms and 2,000 children (Smith, Jurkiewicz, & Xiang, 2002). The study investigated program effects on child development and the relationship of program quality to outcomes. The second project was the Michigan Full-Day Preschool Comparison Study, which compared the effectiveness of part-day and full-day programs in MSRP, Head Start, and private child care settings (Jurkiewicz, 2003). The fall 2000 and spring 2001 cohorts comprised 121 and 132 classrooms, respectively. The robust samples in these studies permitted rigorous testing of the psychometric properties

of the new PQA. A broad range of public and private early childhood settings were represented and the data sets were large enough to conduct a factor analysis and examine the construct validity of the revised PQA.

In addition to looking at the scoring distribution, interrater reliability, internal consistency, and factor structure of the PQA itself, the instrument's validity was assessed by examining its relationship to other measures in several of these studies. In three investigations (Epstein, 1993; Schweinhart et al., 1998; and Smith et al., 2002), scores on the PQA were analyzed relative to one or more other commonly used program quality measures, notably the Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980; Harms, Clifford, & Cryer, 1998), the Caregiver Interaction Scale (CIS; Arnett, 1989), and the Teacher Beliefs Scale (Burts, Hart, Charlesworth, & Kirk, 1990). In addition, PQA scores were examined in relation to children's development using the DIAL-R (Mardell-Czudnowski & Goldenberg, 1990) and/or the High/Scope Child Observation Record (COR; High/Scope Educational Research Foundation, 1992) in three studies with sample sizes of 201, 229, and 155 children, respectively (Epstein, 1993; High/Scope, 1997; Schweinhart et al., 2000).

The psychometric properties of the PQA, based on this series of studies, are summarized below. For more detail on the reliability and validity of the PQA, see the research reports cited above, particularly Jurkiewicz, 2003.

Score Distributions
Conclusion

The Preschool PQA adequately measures the full range of quality along each quality construct.

Evidence

Score distributions for Preschool PQA quality constructs must express sufficient variance to test the entire scale associated with each of them. A construct that has many cases distributed across the entire length of a scale demonstrates it is sensitive across the range it intends to assess. Frequency distributions of scores and descriptive statistics such as means and standard deviations were used to analyze PQA quality constructs. In successful testing, score curves should be spread out across the entire scale associated with each construct.

Table 1 presents the distribution of responses for each quality construct and total scores on the revised PQA. In the old PQA, mean scores ranged from 4.3 to 4.6, with an average of only 3% below level 3 ("unacceptable"). By contrast, in the revised PQA, means ranged from 2.7 to 3.6, with quality ratings below 3 almost always exceeding 10% and averaging 27% across PQA constructs.

Reliability

Conclusion

The Preschool PQA is a reliable instrument with high interrater reliability and internal consistency.

Table 1. Score Distributions of Revised Preschool PQA

PQA Quality Construct[a]	Project	Descriptives			Level of Quality		
		N	Mean	SD	1 to 2	3	4 to 5
I. Learning Environment	Full-Day: Fall 2000[b]	121	3.63	0.69	29.7%	16.2%	54.1%
	Full-Day: Spring 2001[c]	132	3.68	0.79	28.2%	15.9%	55.9%
	MSRP Phase 2[d]	19	3.33	0.46	19.3%	37.4%	43.3%
II. Daily Routine	Full-Day: Fall 2000[b]	103	3.33	0.72	27.9%	26.2%	45.9%
	Full-Day: Spring 2001[c]	114	3.43	0.76	21.9%	27.9%	50.2%
	MSRP Phase 2[d]	19	2.80	0.56	36.2%	50.3%	13.5%
III. Adult-Child Interaction	Full-Day: Fall 2000[b]	99	4.00	0.73	29.2%	10.9%	59.9%
	Full-Day: Spring 2001[c]	101	4.05	0.84	22.6%	11.8%	65.6%
	MSRP Phase 2[d]	18	3.11	0.43	23.5%	49.5%	27.0%
IV. Curriculum Planning & Assessment	Full-Day: Fall 2000[b]	117	3.65	0.86	20.1%	20.5%	59.4%
	Full-Day: Spring 2001[c]	128	3.76	0.96	15.3%	20.9%	63.8%
	MSRP Phase 2[d]	19	2.67	0.47	47.4%	27.4%	25.2%
V. Parent Involvement & Family Services	Full-Day: Fall 2000[b]	111	4.13	0.86	21.5%	10.0%	68.5%
	Full-Day: Spring 2001[c]	121	4.38	0.76	12.9%	7.4%	79.7%
	MSRP Phase 2[d]	17	3.62	0.55	15.4%	27.7%	56.9%
Total PQA Score[e]	Full-Day: Fall 2000[b]	86	3.74	0.63	29.0%	18.0%	53.9%
	Full-Day: Spring 2001[c]	88	3.85	0.68	24.0%	19.3%	56.7%
	MSRP Phase 2[d]	11	3.10	0.46	27.0%	42.7%	30.3%

[a] The sample size was not sufficient to analyze sections VI and VII.

[b] Michigan Full-Day Preschool Comparison Study: Fall 2000 cohort

[c] Michigan Full-Day Preschool Comparison Study: Spring 2001 cohort

[d] MSRP Phase 2 Evaluation: 2000 cohort

[e] Total of constructs I through III

Evidence

Evidence for the Preschool PQA's reliability comes from analyses to determine its interrater reliability and internal consistency.

Interrater reliability. Training to acceptable levels of interrater reliability on the PQA takes 3 days. The first 2 days are devoted to review and practice with the PQA, using raw-footage videotapes and live observations. The third day is used to conduct actual observations and determine interrater reliability. Interrater reliability on the two older versions of the PQA was computed as the percentage of agreement between pairs of raters. Trained data collectors averaged exact agreement (same scores) of 79% and 90%, and close agreement (same or adjacent scores) of 97% and 99%, across categories for these two versions, respectively. To assess interrater reliability on the new PQA, pairs of raters were sent to 10 of the 19 classrooms in the MSRP Phase 2 sample to observe the learning environment, daily routine, and adult-child interaction. Researchers computed Pearson's correlations on their scores. Table 2 presents the results. Although interrater reliability was lower than ideal on the first construct (learning environment), all correlations were significant[1] and ranged from .57 to .75.

Internal consistency. To assess internal consistency, Cronbach's alpha was calculated on five quality constructs (learning environment, daily routine, adult-child interaction, curriculum planning and assessment, parent involvement and family services) and total PQA scores. There were insufficient data to determine internal consistency on the other two constructs (staff qualifications and development, and program management) since these were only rated once at the agency level rather than for each classroom. Internal consistency of the old PQA was based on 107 classroom observations collected across four studies. Cronbach's alpha averaged .94. Internal consistency of the new version was calculated with 185 classrooms in three samples: the fall and spring settings in the Michigan Full-Day Preschool Comparison Study and the MSRP Phase 2 cohort. Table 3 presents the results. Cronbach's alpha

Table 2. Interrater Agreement of Revised Preschool PQA

PQA Quality Construct[a, b]	Paired Raters	
	N	Correlation
I. Learning Environment	9	.57*
II. Daily Routine	9	.75**
III. Adult-Child Interaction	10	.74**

[a] The sample size was not sufficient to analyze sections IV through VII.

[b] The following items were not included in the analysis due to lack of variance: outdoor space, meeting basic physical needs, and support for non–English-speaking students.

* $p < .10$

** $p < .05$

averaged .93 with all but two of the results within the acceptable range of .7 to .9.

Validity

Conclusion

The Preschool PQA is a valid instrument, with an empirically based factor structure and significant relationships with other measures of program quality and child outcomes.

Evidence

The validity of the Preschool PQA was assessed in three ways: confirmatory factor analysis, relationships with other program quality measures, and relationships with child outcomes.

Confirmatory factor analysis. A confirmatory factor analysis was conducted with sections I through V of the revised PQA, using a sample of approximately 150 classrooms. (The sample size was not sufficient to analyze sections VI and VII. These sections are only completed once per agency, with directors who are

---

[1] Significance throughout this manual is defined as $p < .10$ or better, although most significance levels exceed $p < .05$. Exact significance levels are indicated in the tables.

Table 3. Internal Consistency of Revised Preschool PQA

PQA Quality Construct[a, b]	Project	N	Cronbach's Alpha
I. Learning Environment	Full-Day: Fall 2000[c]	121	0.83
	Full-Day: Spring 2001[d]	132	0.87
	MSRP Phase 2[e]	19	0.65
II. Daily Routine	Full-Day: Fall 2000[c]	103	0.86
	Full-Day: Spring 2001[d]	114	0.85
	MSRP Phase 2[e]	19	0.84
III. Adult-Child Interaction	Full-Day: Fall 2000[c]	99	0.90
	Full-Day: Spring 2001[d]	101	0.94
	MSRP Phase 2[e]	18	0.69
IV. Curriculum Planning & Assessment	Full-Day: Fall 2000[c]	117	0.72
	Full-Day: Spring 2001[d]	128	0.79
	MSRP Phase 2[e]	19	0.37
V. Parent Involvement & Family Services	Full-Day: Fall 2000[c]	111	0.91
	Full-Day: Spring 2001[d]	121	0.90
	MSRP Phase 2[e]	17	0.74
Total PQA Score[f]	Full-Day: Fall 2000[c]	86	0.94
	Full-Day: Spring 2001[d]	88	0.95
	MSRP Phase 2[e]	11	0.89

[a] The sample size was not sufficient to analyze sections VI and VII.

[b] The following items were not included in the analysis due to lack of variance: outdoor space, meeting basic physical needs, and support for non–English-speaking students.

[c] Michigan Full-Day Preschool Comparison Study: Fall 2000 cohort

[d] Michigan Full-Day Preschool Comparison Study: Spring 2001 cohort

[e] MSRP Phase 2 Evaluation: 2000 cohort

[f] Total of constructs I through III

responsible for multiple classrooms, rather than with each classroom.) The results are presented in Table 4. Five factors emerged, accounting for 58% of the variance, and their content aligned with the five corresponding PQA sections: Learning Environment, Daily Routine, Adult-Child Interaction, Curriculum Planning and Assessment, and Parent Involvement and Family Services. Factor loadings ranged from .43 to .82, with the majority (64%) at .60 or higher. However, several daily routine items, notably those related to group times (e.g., small- and large-group time), loaded on the adult-child factor. These items were modified in the final version of the PQA. The interaction indicators were taken out and a new item, "Adults support and extend children's learning at group times" was added to the Adult-Child Interaction section.

Relationship to other program quality measures. The first version of the PQA was positively and significantly correlated with the ECERS and the CIS (.86 and .48 overall, respectively). The associations were strongest in those subscales most similar across instruments, for example, PQA Learning Environment and ECERS Furnishings were correlated at .73 and PQA Adult-Child Interaction and CIS Sensitivity were correlated at .77. Teachers' qualifications were also differentially related to scores on the first and second versions of the PQA. Among Head Start teachers for example, the PQA correlated .44, .47, and .47, respectively, with teachers' formal education, hours of inservice training, and years of experience working with young children.

The validity of quality constructs in sections I through V of the revised PQA was assessed in relationship to the Teacher Beliefs Scale. Table 5 presents the results. The PQA was significantly correlated, in the expected positive or negative direction, with appropriate and inappropriate teacher beliefs and practices. With one exception, all correlations were significant and ranged in magnitude from .28 to .49.

Relationship to child outcomes. Program quality measured with the PQA is significantly related to children's developmental outcomes. These results obtain for concurrent outcomes while children are in preschool and longitudinal outcomes when they are in kindergarten. For example, in the national Training for Quality study, the PQA total score and all the subscales were positively and significantly associated with concurrent measures on the language scale of the DIAL-R. Correlations between program quality and children's development in preschool ranged from .35 to .56. Program quality was also related to several subscales on the High/Scope COR. The strongest associations were between the program's daily routine and adult-

Table 4. Confirmatory Factor Analysis of Revised Preschool PQA

PQA Quality Construct and Items[a, b]	Factors (58%)				
	1	2	3	4	5
I. Learning Environment (*N* = 151–152)					
Safe and healthy environment			0.82		
Defined interest areas			0.52		
Logically located interest areas			0.74		
Organization and labeling of materials			0.52		
Varied and open-ended materials			0.56		
Plentiful materials			0.53		
Diversity-related materials			0.61		
Displays of child-initiated work				0.43	
II. Daily Routine[c] (*N* = 134–152)					
Consistent daily routine	0.77				
Parts of the day				0.44	
Appropriate time for each part of day	0.68				
Time for child planning					0.77
Time for child-initiated activities	0.57				
Time for child recall					0.79
Small-group time			0.56		
Large-group time	0.73				
Choices during transition times			0.43		
Cleanup time with reasonable choices	0.74				
Snack or meal time	0.53				
Outside time	0.48				
III. Adult-Child Interaction (*N* = 135–151)					
Handling separation from home	0.60				
Warm and caring atmosphere	0.64				
Support for child communication	0.78				
Adults as partners in play	0.72				
Encouragement of child initiatives	0.78				
Opportunities for child explorations	0.69				
Acknowledgment of child efforts	0.65				
Encouragement for peer interactions	0.70				
Independent problem solving	0.75				
Conflict resolution	0.66				
IV. Curriculum Planning and Assessment (*N* = 148–149)					
Curriculum model				0.60	
Team teaching				0.64	
Comprehensive child records		0.69			
Anecdotal note taking by staff				0.58	
Use of child observation measure				0.62	

(continued on the next page)

Table 4. Confirmatory Factor Analysis of Revised Preschool PQA *(cont.)*

PQA Quality Construct and Items[a, b]	Factors (58%)				
	1	2	3	4	5
V. Parent Involvement and Family Services (*N* = 141–150)					
Opportunities for involvement		0.75			
Parents on policy-making committees				0.43	
Parent participation in child activities		0.74			
Sharing of curriculum information		0.65			
Staff-parent informal interactions		0.61			
Extending learning at home		0.58			
Formal meetings with parents		0.70			
Diagnostic/special education services		0.73			
Service referrals as needed		0.57			
Transition to kindergarten		0.74			

[a] The sample size was not sufficient to analyze sections VI and VII.

[b] The following items in sections I–V were not included in the analysis due to lack of variance: outdoor space, meeting basic physical needs, and support for non–English-speaking students.

[c] Daily routine items that loaded on the adult-child factor were modified in the final version of the PQA. Interaction indicators were taken out and a new item, "Adults support and extend children's learning at group times," was added to the Adult-Child Interaction section.

Table 5. Relationship of Preschool PQA to Teacher Beliefs Scale

PQA Quality Construct[a, b]	Correlation With Teacher Beliefs Scale			
	N	Appropriate Practices	*N*	Inappropriate Practices
I. Learning Environment	51	0.16	50	-.25*
II. Daily Routine	44	0.28*	43	-.32**
III. Adult-Child Interaction	46	0.31**	45	-.43***
IV. Curriculum Planning & Assessment	48	0.35**	47	-.49***
V. Parent Involvement & Family Services	41	0.28*	40	-.43***
Total PQA Score[c]	50	0.43***	49	-.46***

[a] The sample size was not sufficient to analyze sections VI and VII.

[b] The following items were not included in the analysis due to lack of variance: outdoor space, meeting basic physical needs, and support for non–English-speaking students.

[c] Total of constructs I through III.

*p < .10

**p < .05

***p < .01

child interactions, on the one hand, and children's initiative and creative representation on the other. These significant correlations between program quality and child development ranged from .33 to .52.

Similar relationships between the PQA at preschool and children's development in kindergarten appeared in the first cohort of the Michigan School Readiness Program Evaluation. Nearly one-quarter (18 out of 80) of the correlations computed between the PQA and the COR proved significant, ranging from .33 to .65. Key quality factors were the extent to which programs had a consistent and articulated philosophy and curriculum model, the use of funding to support staff development, and the consistency of administrative support and supervision for staff. These PQA variables were in turn related to children's initiative, social relations, language and literacy, logic and mathematics, and total COR scores. There were also significant correlations between these same dimensions of preschool program quality and children's development as rated by their kindergarten teachers. Correlations between the PQA and teachers' ratings ranged from .40 to .48.

Summary

In sum, this series of national and statewide evaluation studies shows the Preschool PQA to be a highly reliable and valid instrument for measuring program quality and determining its relationship to staff qualifications, staff development initiatives, and young children's developmental outcomes. Moreover, the revised PQA corrects the limitations of the earlier versions. Quality constructs have demonstrated their capacity to generate scores across the intended range of targeted programmatic features in early childhood settings. Compared to earlier versions, new PQA score distributions are less positively skewed, averages have dropped, and standard deviations have increased. Overall, interrater agreement is significant and internal consistency is strong. Finally, there is substantial evidence for the validity of the quality constructs based on the confirmatory factor analysis and the relationship of the PQA to other quality measures.