3

Psychometric Properties of the Infant-Toddler PQA

The Development of the Infant-Toddler PQA

As described earlier, the Infant-Toddler Program Quality Assessment (PQA; Hohmann, Lockhart, & Montie, 2013; Epstein et al., 2013) has existed in various forms since 1999 and has been used over the years in many infant and toddler programs to evaluate the quality of care they are providing for young children as well as to document their implementation of the HighScope Infant-Toddler Curriculum.

The development of the Infant-Toddler PQA has involved numerous resources. The content of the descriptors is based on best practices in the field of early education and care, and research and theory on appropriate ways to support the development of infants and toddlers. The first and second editions of HighScope's infant-toddler curriculum manual, *Tender Care and Early Learning: Supporting Infants and Toddlers in Child Care Settings* (Post & Hohmann, 2000; Post, Hohmann, & Epstein, 2011), were a major source of information; HighScope's Preschool Program Quality Assessment (PQA; HighScope Educational Research Foundation, 2003a, 2003b) was also consulted during the development of the instrument. Also helpful in the development process were resources on best practices from infant-toddler and early childhood professional organizations, such as Zero to Three, Resources for Infant Educarers (RIE), The Program for Infant/Toddler Care (PITC), the National Association for the Education of Young Children (NAEYC), Early Head Start, and Early On Michigan. The following instruments and documents were some of the specific references used to improve the Infant-Toddler PQA: the Infant/Toddler Environment Rating Scale — Revised Edition (Harms, Cryer, & Clifford, 2006); Head Start and Early Head Start Performance Standards (US Dept. of Health and Human Services, 2004, April); *Zero to Three Early Learning Guidelines for Infants and Toddlers: Recommendations for States* (Petersen, Jones, & McGinley 2008); state infant and toddler guidelines, specifically Michigan's Early Childhood Standards of Quality for Infant and Toddler Programs (Michigan State Board of Education, 2006); NAEYC's *Developmentally Appropriate Practice in Early Childhood Programs Serving Children From Birth Through Age 8* (Copple & Bredecamp, 2009); and *NAEYC Early Childhood Program Standards and Accreditation Criteria* (2007).

The Infant-Toddler PQA Validation Study

An initial pilot test involving 75 diverse infant and toddler group care settings was conducted in 2001 to validate the Infant-Toddler PQA. HighScope staff, HighScope field consultants, and trained caregivers completed the rating instrument at these sites. Over the years, some of these field consultants continued to use the Infant-Toddler PQA, and the results of these assessments were added to the 75 in the original data set to increase the sample size to 96. Descriptive statistics and factor analysis of data from the pilot study and of the additional data collected later were used to revise the tool to a field-test version (Hohmann, Lockhart, & Montie, 2011).

At the time of the original pilot study, HighScope early childhood and research staff also tested the reliability of the instrument by conducting dual observations in which pairs of staff members

completed PQAs in various infant and toddler programs. The results of these observations were incorporated in the first round of pilot testing and were also used with the additional data described above in revising the tool to a field-test version.

The last round of testing to confirm the validity and reliability of the Infant-Toddler PQA field-test version (Hohmann, Lockhart, & Montie, 2011 [Form A only]) was conducted from April 2011 to February 2012. Data was collected by trained independent evaluators in various for-profit and not-for-profit infant and toddler classrooms implementing a mixture of program types (e.g., Early Head Start) and curriculum models (e.g., HighScope, Montessori, Creative Curriculum, traditional). These sites were located across two midwestern states (Michigan and Indiana). The sample sizes were as follows:

- 40 inter-rater reliability classrooms
- 30 validity classrooms, scored using the Infant/Toddler Environment Rating Scale — Revised (ITERS-R) (Harms, Cryer, & Clifford, 2006) and the Infant-Toddler PQA Form A — Beta Version (Hohmann, Lockhart, & Montie, 2011)
- 30 test-retest reliability classrooms

Results

Score distributions

The Infant-Toddler PQA is designed to measure the quality of care provided in facility-based infant and toddler programs. It evaluates quality in a broad spectrum of settings that range from low to high quality. For research purposes, a certain amount of variance is necessary for the scores to be useful for data analysis. The score distributions in the validation study indicate that the Infant-Toddler PQA adequately measures the full range of quality along a broad continuum.

Table 1 presents the score distributions for Infant-Toddler PQA Form A: Observation Items (Hohmann, Lockhart, & Montie, 2011). The mean scores across the four subscales indicate that a variety of classroom quality levels were captured. The four mean scores are near the middle of the

range (3.44–3.89), meaning that programs included in the study were of average quality overall. A closer look at the minimums and maximums of each construct reveals minimum distributions ranging from 1.50–2.35 and maximum distributions ranging from 4.75–5.00. Both the lowest and highest scores were on the Curriculum Planning and Child Observation subscale, indicating that at least one setting in the study had low quality in this area while another setting had high quality.

Internal consistency

Table 2 presents the internal consistency of the subscales of Infant and Toddler PQA Form A: Observation Items. Internal consistency describes the extent to which items in the same subscale are related to one another and how well each subscale uniquely captures a particular dimension of quality.

A total of 30 classrooms were included in the analysis. Cronbach's alpha was calculated for each subscale and the results are within acceptable levels of internal consistency. Cronbach's alpha ranged from .79 in Learning Environment to .90 in Curriculum Planning and Child Observation, meeting the criteria of .70 or above.

Reliability

To assess the reliability of Infant-Toddler PQA Form A, analyses were conducted to establish levels of inter-rater reliability on each of the instrument's four subscales. Inter-rater reliability was calculated for the four subscales of Form A using paired raters (N = 40 infant-toddler classrooms). Each observer independently scored the rows using the 5-point scale. The scoring rules were applied to generate item scores, and the item scores were then averaged to produce subscale scores. Trained data collectors averaged exact agreement (same scores) ranging from 62% to 69% of the time and close agreement (same or adjacent scores) from 96% to 98% of the time across the four subscales.

Further examination of the items revealed that a few rows within some items were not at the desired reliability level and were therefore pulling the whole subscale level down. In particular, data

Table 1. Score Distributions for the Infant-Toddler PQA Form A				
Subscales	Score Distributions (*N* = 27)			
	Mean	*SD*	Minimum	Maximum
Learning Environment	3.74	0.65	2.35	4.75
Schedules and Routines	3.89	0.74	2.07	4.96
Adult-Child Interaction	3.68	0.77	2.13	4.80
Curriculum Planning and Child Observation	3.44	1.10	1.50	5.00
Total Score for Subscales	3.73	0.69	2.29	4.80

Note: Scores have a potential range of 1 to 5, with 5 being the highest.

Table 2. Internal Consistency for the Infant-Toddler PQA Form A	
Subscales	(*N* = 30)
Learning Environment	.79
Schedules and Routines	.86
Adult-Child Interaction	.82
Curriculum Planning and Child Observation	.90
Total Score for Subscales	.94

collectors differed in how they defined infant-only versus toddler-only rooms, which affected the number of rows they completed or left blank. Based on this data and feedback from the data collectors, an age-group definition chart was developed and added to the instructions for the instrument, and clearer examples and lists were added to the descriptors to better define each row and item.

After completing these changes to Form A, a small group of data collectors were retrained. Each was paired with a reliable observer and assigned to observe a classroom. Their results were used to assess the inter-rater reliability of the questionable items for exact agreement. Based on these changes the exact agreement increased by 10% to 12% across the four subscales.

Interclass correlation coefficients were calculated to examine the inter-rater reliability at the subscale level. Table 3 presents these findings. All four subscale coefficients are within an acceptable range (levels approaching .70 or above) for inter-rater reliability. Of the four subscales of Form A, Adult-Child Interaction had the lowest reliability. This may have occurred because this subscale is the most difficult to observe; the information is collected in real time and is based on particular events that one or the other of the observers may not always have

the opportunity to see or hear. (Observers may have different locations in the room or may focus on different adults at different times.)

Test-retest reliability

A test-retest reliability study was also conducted to examine the stability of Infant-Toddler PQA Form A results over time. In the 30 sites where inter-rater reliability was assessed, an observer returned for a retest approximately 2 weeks (mean = 15.9 days) after completing the initial observation. Table 4

Table 3. Inter-rater Reliability Correlations for the Infant-Toddler PQA Form A

Subscales	(N = 40)
Learning Environment	.73
Schedules and Routines	.79
Adult-Child Interaction	.70
Curriculum Planning and Child Observation	.79
Total Score for Subscales	.83

Table 4. Test-Retest Correlations for the Infant-Toddler PQA Form A

Subscales	(N = 30)
Learning Environment	.88***
Schedules and Routines	.86***
Adult-Child Interaction	.81***
Curriculum Planning and Child Observation	.80***
Total Score for Subscales	.89***

Note: *p < .05; **p < .01; ***p < .001

shows that there were significant correlations between the initial and second observations, indicating that the Infant-Toddler PQA instrument produces very stable results across two time periods.

Concurrent validity

Concurrent validity for Infant-Toddler PQA Form A was measured using the Infant/Toddler Environment Rating Scale (ITERS–R; Harms, Cryer, & Clifford, 2006), which measures overall environmental quality of infant-toddler programs. Table 5 presents the concurrent validity between the Infant-Toddler PQA and the ITERS–R. Originally there were 30 classrooms in the sample, but because information on some classrooms was missing, the analysis focused on 26 classrooms. Results indicate that the Infant-Toddler PQA

Form A is positively and significantly correlated with the ITERS, at .66 overall. Some strong associations were found in expected areas, where subscales of the two instruments measure similar things. For example, the Infant-Toddler PQA Adult-Child Interaction subscale and the ITERS Listening and Talking subscale were correlated at .67, and the Infant-Toddler PQA Curriculum Planning and Child Observation subscale and the ITERS Activities subscale were correlated at .65.

Form B analysis

In addition to the classroom factors assessed by Infant-Toddler PQA Form A, the full instrument also measures the quality of a program's parent involvement and family services, staff qualifications and staff

Table 5. Concurrent Validity Using the Infant-Toddler Environment Rating Scale (ITERS-R) and the Infant-Toddler PQA Form A

ITERS-R Subscales	Infant-Toddler PQA Form A Subscales				
	Learning Environment	Schedules and Routines	Adult-Child Interaction	Curriculum Planning and Child Observation	Total PQA Score
Space and Furnishings	.43*	.29	.32	.35	.36
Personal Care Routines	.35	.42*	.16	.40*	.37
Listening and Talking	.46*	.59**	.67**	.40*	.61**
Activities	.58**	.45*	.42*	.65**	.51**
Interaction	.48*	.45*	.50**	.29	.50**
Program Structure	.54**	.61**	.55**	.36	.61**
Parents and Staff	.38	.40*	.58**	.39*	.48**
Total ITERS Score	.66**	.62**	.56**	.60**	.66**

Note: *p < .05; **p < .01

Table 6. Internal Consistency for the Infant-Toddler PQA Form B — Agency Items	
Subscales	(*N*=50)
Parent Involvement and Family Services	.67
Staff Qualifications and Staff Development	.52
Program Management	.13
Total Score for Subscales	.44

development, and program management. These components make an important contribution to program quality, but they are generally not observable in classrooms or controlled by the teaching staff; they also tend to be more standard across programs. Because these components are primarily evaluated through factual interviews, they are organized in a separate document, Form B: Agency Items for Infant-Toddler and Preschool Programs (Epstein et al., 2013).

As the title suggests, this is a combined form that agency directors can use for programs serving either or both the infant-toddler and preschool age ranges. The agency items pertaining to infant and toddler programs that are a part of this combined form were included in the original 2001 study, but were not included in the last round of data collection. The original analysis of the Infant-Toddler PQA agency items is presented here.

Table 6 presents the findings on the internal consistency of the agency items. Cronbach's alpha was calculated for each of the three subscales. The program management items proved to be less internally consistent (.13) than those in the other subscales, which affected the overall total score.

Therefore, these items have either been deleted or were combined with other similar agency items from Preschool PQA Form B (HighScope Educational Research Foundation, 2003b) in the new Form B.

Summary

The findings presented here indicate that the Infant-Toddler Program Quality Assessment is a valid and reliable instrument for measuring the quality of infant and toddler programs. Its quality constructs have demonstrated their capacity to generate scores across the intended range of infant and toddler program components. The reliability scores across raters are at acceptable levels, and internal consistency is strong. The items that make up the scales and subscales group together well, supporting a sound theoretical base. Further, another quality assessment was positively and significantly associated with the Infant-Toddler PQA constructs. The Infant-Toddler PQA has gone through many transformations and with the statistical analysis to back up the instrument, programs can trust that by using the Infant-Toddler PQA, they will improve the quality of the services they provide for their infants and toddlers.

References

American Academy of Pediatrics, American Public Health Association, & National Resource Center for Health and Safety in Child Care. (2002). *Caring for our children — National health and safety performance standards: Guidelines for out-of-home child care programs* (2nd Ed.). Elk Grove Village, IL: American Academy of Pediatrics, American Public Health Association, & National Resource Center for Health and Safety in Child Care. Retrieved from http://nrckids.org/CFOC

Copple, C., & Bredekamp, S. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: NAEYC.

Education Week. (2009). Quality counts 2009. Bethesda, MD: Author. Retrieved from http://www.edweek.org

Epstein, A. S., & Hohmann, M. (2012). *The HighScope Preschool Curriculum.* Ypsilanti, MI: HighScope Press.

Epstein, A. S., Gainsley, S., Hohmann, M., Jurkiewicz, T., Lockhart, S., Marshall, B., & Montie, J. (2013). *Infant-Toddler Program Quality Assessment (PQA) Form B: Agency Items for Infant-Toddler and Preschool Programs.* Ypsilanti, MI: HighScope Press.

Harms, T., Cryer, D., & Clifford, R. M. (2006). *Infant/Toddler Environment Rating Scale — Revised edition (ITERS-R).* New York: Teachers College Press.

HighScope Educational Research Foundation. (2003a). *Preschool Program Quality Assessment (PQA).* Ypsilanti, MI: HighScope Press.

HighScope Educational Research Foundation. (2003b). *Preschool Program Quality Assessment (PQA): Form B — Agency Items.* Ypsilanti, MI: HighScope Press.

Hohmann, M., Lockhart, S., & Montie, J. (2011). *Infant-Toddler Program Quality Assessment (PQA) Form A: Observation Items — Beta Version.*

Hohmann, M., Lockhart, S., & Montie, J. (2013). *Infant-Toddler Program Quality Assessment (PQA) Form A: Observation Items.* Ypsilanti, MI: HighScope Press.

Michigan State Board of Education. (2006). *Early childhood standards of quality for infant and toddler programs.* Lansing, MI: Author.

National Association for the Education of Young Children. (2007). *NAEYC early childhood program standards and accreditation criteria.* Retrieved from https://oldweb.naeyc.org/academy/standards/

National Association for the Education of Young Children (NAEYC) & National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8.* Washington, DC: National Association for the Education of Young Children.

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 39*(1), 133–164. doi: 10.3102/00028312039001133

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology, 39*(3), 451–469. doi:10.1037/0012-1649.39.3.451

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2012). *Study overview* [The NICHD Study of Early Child Care and Youth Development]. Retrieved from http://www.nichd.nih.gov/research/supported/seccyd/Pages/overview.aspx#datas

Petersen, S., Jones, L., & McGinley, K. A. (2008). *Early learning guidelines for infants and toddlers: Recommendations for states.* Washington, DC: Zero to Three.

Post, J., & Hohmann, M. (2000). *Tender care and early learning: Supporting infants and toddlers in child care settings.* Ypsilanti, MI: HighScope Press.

Post, J., Hohmann, M., & Epstein, A. S. (2011). *Tender care and early learning: Supporting infants and toddlers in child care settings* (2nd ed.). Ypsilanti, MI: HighScope Press.

US Department of Health and Human Services, Administration for Children, Youth, and Families, Head Start Bureau. (2004, April). *Head Start Program Performance Standards and other regulations.* Washington, DC: US Government Printing Office. Also available online at http://eclkc.ohs.acf.hhs.gov/hslc/Regulations%20&%20Policies

Willer, B., & Richie, S. (Eds.). (2005). *Self-study and program quality improvement through NAEYC early childhood program accreditation.* Washington, DC: NAEYC.

Zero to Three. (2008). *Caring for infants and toddlers in groups: Developmentally appropriate practice* (2nd ed.). Arlington, VA: Author.
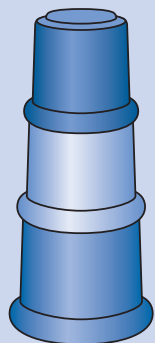
Infant-Toddler
PQA
Administration Manual

The **Infant-Toddler Program Quality Assessment (PQA)** is a validated, authentic instrument designed to measure program quality and identify provider training needs in child care programs serving children aged 6 weeks to 36 months. Based on best practices in the early childhood field, the PQA is much more than an assessment instrument — its detailed indicators provide a guide to implementing high-quality programs and a road map for making program improvements. This **Administration Manual** provides an overview of the instrument, detailed instructions for using it, and a report on the results of the validation study.

The Administration Manual is part of a set that also includes the two main components of the instrument:

- **Form A: Observation Items** is completed mainly through classroom observation and covers four domains of education and care practices in infant and/or toddler rooms: *learning environment; schedules and routines; adult-child interaction;* and *curriculum planning and child observation.*

- **Form B: Agency Items for Infant-Toddler and Preschool Programs** focuses on program-level practices and is completed mainly through interviews. It focuses on three domains: *parent involvement and family services; staff qualifications and staff development;* and *program management.* Form B is a combined form that may be used in either infant-toddler or preschool programs, or in programs serving both age groups.

The instrument may be used in any developmentally based program serving infants and toddlers, including but not limited to those using the HighScope Infant-Toddler Curriculum.

Training results in the most effective use of the Infant-Toddler PQA. To arrange PQA training for providers or independent raters who will be completing this instrument, please contact the Training Coordinator, HighScope Educational Research Foundation, 600 North River Street, Ypsilanti, MI 48198-2898; phone: 734.485.2000, extensions 224, 237; FAX: 734.485.4467; e-mail: **training@highscope.org.** *For more information on HighScope's curriculum or assessment products or training offerings, visit the HighScope website at* **highscope.org.**

HIGHSCOPE
PRESS ®