Research Notes

Published by the McCormick Tribune Center for Early Childhood Leadership, National-Louis University SPRING 2008

When Quality Counts and Money Matters

Policymakers across the country have begun to implement statewide quality rating systems (QRS) to measure and improve the program quality in early care and education settings. A quality rating system is a systematic approach to assess, improve, and communicate the level of program quality. Currently, there are 15 states implementing statewide quality rating systems; approximately 30 more states are in the design or pilot stages of QRS development. As QRS pilots are taken to scale through statewide implementation, it becomes clear how money matters when quality counts.

Money factors into QRS in multiple ways including: funding to design and pilot QRS; financial incentives to reward the participation of programs and/or practitioners in QRS, financial support for program quality improvements; and funding of an infrastructure to implement, sustain, and evaluate the quality rating system. An important but costly part of the infrastructure is the assessment system for determining the quality of the learning environment in each early care and education program participating in the QRS initiative. While it is most desirable to assess all classrooms in a program to determine a center's true level of quality, financial constraints often only allow for a sample of the classrooms to be examined. These classrooms then serve as a proxy for the quality level of the learning environment for all classrooms participating in the QRS.

The learning environment is a key component within all QRS. In 13 of the 15 states implementing statewide quality rating systems, the quality of the learning environment component is measured using the *Early Childhood Environment Rating Scale-Revised* (or one of the other Environment Rating Scales).[1] Of these 13 statewide quality rating systems, 11 use a random sample of classrooms to determine an overall learning environment component score for the program. Typically, assessments are conducted in one-third of a center's classrooms, including at least one classroom at each age level (infant-toddler, preschool, and school-age). A center's overall score for the learning environment component is then determined by averaging the scores of the sampled classrooms.

There has been limited empirical research looking at the effectiveness of different sampling practices. A study was conducted by the McCormick Tribune Center for Early Childhood Leadership to determine the effectiveness of observing a sample of classrooms in a center rather than observing all the classrooms to determine the level of quality of the learning environment.

Methodology

Data were collected from May 2006 through January 2007 as part of a professional development project for the City of Chicago's Children & Youth Services' Head Start programs. The sample included half-day and full-day programs providing Head Start services. Classroom quality was measured using the *Early Childhood Environment Rating Scale-Revised* (ECERS-R) in all classrooms with preschool-aged children. In January 2008, a subset of the original data was analyzed to determine the difference in average ECERS-R scores when sampling one-third and one-half of the classrooms in a program as compared to the average ECERS-R score for all classrooms in a program. The sample included a total of 468 classrooms in 158 centers. Only centers with a minimum of two classrooms and a maximum of seven classrooms were included.

All possible sampling combinations were calculated. For example in a center with three classrooms, there were six different classroom sampling combinations possible to achieve the 1/3 and 1/2 sampling criteria. The Illinois Quality Counts QRS framework for rating a program on the learning environment component (see Table 1) was then applied to all the possible combinations to determine how frequently the respective sampling combinations matched the accurate component score when using the average ECERS-R score for all classrooms in a program.

Table 1. Illinois Quality Counts QRS: Learning Environment Component

★	★★	★★★	★★★★
Level 1	Level 2	Level 3	Level 4
ERS average: 3.00 - 3.49	ERS average: 3.50 - 4.24	ERS average: 4.25 - 4.99 or national accreditation	ERS average: 5.00 and above

Continued on back page

Findings

This study quantified the difference between sampling one-third and one-half of the classrooms. Table 2 summarizes the data.

Table 2. Percentage of Learning Environment Component Scores Correctly Classified

	Number of centers	Number of classrooms	1/3 of classrooms observed	1/2 of classrooms observed
2 classroom centers	79	158	68%*	68%
3 classroom centers	38	114	58%	72%
4 classroom centers	24	96	76%*	76%
5 classroom centers	8	40	55%	78%
6 classroom centers	3	18	78%	82%
7 classroom centers	6	42	79%	83%
Total	158	468		

* In some instances, using the 1/3 sampling criteria and 1/2 sampling criteria results in the same number of classrooms being observed. Thus the figures are the same for the percentage of learning environment component scores correctly classified.

The results confirmed what common sense would suggest—namely that sampling one-half of the classrooms in a center is a more accurate predictor of the overall quality of a center's learning environment than sampling one-third of the classrooms. Additionally, the results of the data analysis did not find a consistent pattern of under- or over-classifying a program's star level based on a sampling of one-third or one-half of the center's classrooms.

- Sampling one-third of the classrooms in a center resulted in a 67% match in what a star rating of the learning environment component would have been had 100% of the classrooms been assessed.

- Sampling one-half of the classrooms in a center resulted in a 72% match in what a star rating of the learning environment component would have been had 100% of the classrooms been assessed.

- The larger the center, the greater the degree of consistency in learning environment scores across classrooms. However, this result should be interpreted with caution as there were only 9 centers in this sample with more than 5 classrooms.

Policy Implications

Many factors impact the cost associated with implementing statewide quality rating systems. Deciding on the number of classrooms to be observed in each program is clearly one of the most critical decisions impacting cost—the more classrooms assessed the greater the cost. Ultimately such policy decisions involve trade-offs. If cost is controlled by conducting reliable assessments in only a portion of classrooms in each program, the determination of the level of quality will not be as precise and under- or over-classifying the quality level of programs may result.

Important to consider in this study is that directors of the programs assessed had not had training in the specific quality indicators included in the ECERS-R. It is possible that as directors learn more about the specific indicators measured by the ECERS-R and make improvements in their programs prior to an official QRS quality assessment, that the variability between classrooms will decrease. Increasing consistency in quality across classrooms will help ensure that an average ECERS-R score generated from a representative sample of classrooms (one-third or one-half) more closely reflects the average ECERS-R score generated from 100% of the classrooms. This may, in turn, reduce the incidence of either over- or under-rating a center's level of quality in the learning environment component of QRS.

One of the limitations of this study is that the sample pulled exclusively from Head Start and publicly funded preschool programs. These programs may have more access to resources to monitor the quality of programming and may have in place higher quality standards that guide programming. These factors may reduce the variability in quality between classrooms within a program. It may be that pulling from a larger community-based sample, without the resources associated with Head Start, will increase the variability in quality between classrooms and make sampling more classrooms within a program even more important. Similarly, it is interesting to consider why large centers showed less variability in classroom quality. It is possible that as centers increase in size, they may have more access to resources and in-house administrative staff to monitor quality and support curricular implementation. These factors are important to consider in future research in this area.

1. Harms, T., Clifford, R.M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised*. New York: Teachers College Press.

For further information about research conducted by the McCormick Tribune Center for Early Childhood Leadership, call 800-443-5522, ext. 5060. Funding for the Center's Research Notes is provided by the McCormick Tribune Foundation, the Joyce Foundation, and the Illinois Department of Human Services. Individuals may photocopy and disseminate freely.