NARA

The Saskatchewan Early Care and Education Quality Indicators Tool and Validation

Richard Fiene, Ph.D., Senior Research Consultant
National Association for Regulatory Administration (NARA)

The Saskatchewan Early Care and Education Quality Indicators Tool and Validation: The Last Piece of the Puzzle in Creating a Differential Monitoring Approach (SK ECQKI7)

Richard Fiene[1], PhD, Senior Research Consultant/Regulatory Scientist

National Association for Regulatory Administration

Research Institute for Key Indicators/Penn State University

Winter 2023

INTRODUCTION

This report will delineate the development, piloting and validating of the Saskatchewan Early Care and Education Quality Key Indicators (SKECPQKI) Tool. The purpose of the tool is to assess the overall program quality in centered based childcare programs in the Province of Saskatchewan, Canada. The evolution of the tool resulted from a multi-year effort by the Ministry of Education in the Province of Saskatchewan to build an effective and efficient differential monitoring system.

This effort in building a new differential monitoring system started in 2019 and was completed in 2023. The first component of this restructuring was the Saskatchewan Licensing Key Indicator System (2019). This was followed by the Saskatchewan Risk Assessment Rules (2019). Once these were in place and operational, a validation study was conducted to measure that the two methodologies were operating as they should (2020). A work group was initiated in 2019 and completed its work in 2020 on an Early Care and Education Quality Key Indicator Tool (SKECPQKI). The tool was put on hold for 2021 because of the pandemic and a new Canadian Federal initiative to expand childcare services across the province. The tool initiative began again in 2022. The pilot testing and validation occurred in 2023.

The work and these studies in the Province of Saskatchewan by the Ministry of Education is the first demonstration of a full-blown differential monitoring system involving licensing key indicator rules, risk assessment rules, and quality indicators. Besides the development of each tool, each of these tools have been validated as well. All this work was done as a collaborative effort between the Ministry of Education staff and the National Association for Regulatory Administration (NARA) consultant pool. Presently, Saskatchewan's overall system is the best example of a fully developed differential monitoring system for the early care and education field.

This was a monumental effort involving 100's of individuals at the local, provincial, and national levels and many hours of data collection and analysis. All the reports are available on the NARA Website (https://www.naralicensing.org/key-indicators) and the full data set will be available via Mendeley Data Sources (https://data.mendeley.com/datasets/kzk 6xssx4d/1).

BACKGROUND HISTORY

This study and tool grew out of an interest by Saskatchewan Ministry of Education policy makers to establish a balance between regulatory compliance and program quality in the most effective and efficient manner. The Province of Saskatchewan did not have a QRIS (Quality Rating and Improvement System) in place nor plans on developing one. Generally, when a jurisdiction wants to develop a balance between regulatory compliance and program quality with rules/regulations/standards, QRIS's are generally developed and implemented.

In reviewing the research literature on regulatory science, differential monitoring has been a developing approach used by many other jurisdictions in the human service licensing field, especially in the United States and in several other Canadian Provinces. Based upon this review of the research literature and the work of the National Association for Regulatory Administration (NARA) which has been a long-term promoter of this approach and the resulting methodologies of licensing key indicators, risk assessment rules, and most recently quality indicators, a contract was entered into between the Ministry of Education and NARA.

The tool is the direct result of research into identifying licensing and quality key indicators over a 50-year (1970-2022) research effort in which specific methodologies were developed and the differential monitoring approach was tested and implemented in the 1970's. Since that time, a national database which expanded to an international database of common key indicators from jurisdictions' respective key indicator tools. These key indicators resulted in a very similar tool that Saskatchewan is using. In fact, in 2019 when the Saskatchewan work group was established, they started with that specific tool that had been developed (Fiene, 2019). During the 2019-2020 period, the work group made the tool into a more user-friendly tool for Saskatchewan childcare programs.

The big deal with utilizing the key indicator methodology is its ability to statistically predict as if one administered the full tool in question. Therefore, when one administers the first quality indicator in the Saskatchewan Early Care and Education Quality Indicator tool it is as if they have administered a licensing based regulatory compliance instrument since the quality of staff is a statistically predictive rule (Fiene, 2002a). The same is true in administering the curriculum quality indicator because it is a statistically predictive standard when looking at overall program quality (Fiene, 2002b). When it comes to QRIS, having communication between staff and parents and parental involvement is a statistically predictive standard for an overall set of QRIS standards (Fiene, 2014). And finally, when administering the ECERS and ITERS or the CIS quality item indicators these are all statistically predictive items for their respective scales as if you had administered the full scales (Fiene, 2002b).

So, as a state/provincial administrator, I would be interested in focusing my efforts on these indicators which reflect compliance with high quality rules/regulations/standards for early care and education. This would be my starting point. I would make sure that my standards reflected quality teachers with the necessary supports such as coaching/mentoring, an early care and education philosophy based upon an emergent curriculum where children are viewed as competent learners, developmentally appropriate curriculum and child assessments, parental and staff communication and participation, and teacher language based/communicative focus when interacting with children in a give and take manner. All this done within a warm and loving style.

An even more efficient and effective way of using the new program quality tool is to pair it with the National Center for Health and Safety in Child Care's *Parental Guide to Choosing Safe and Healthy Child Care (DHHS: Assistant Secretary's Office for Planning and Evaluation, 2019)*. This is a more aggressive and controversial approach, but it is the most efficient way of conducting monitoring visits in the most abbreviated way. However, as efficiency increases, effectiveness may decrease; so, it is a delicate balancing act. This suggested approach builds off a similar suggestion in which only using *Caring for Our Children: Basics (ACF, 2015)* a DHHS Administration for Children and Families publication would be used as the base for regulatory compliance in the United States.

Differential monitoring grew out of a need for jurisdictions to be more effective and efficient in their oversight and inspection efforts of early care and education programs. This started to occur in the late 1960's and 1970's as many more programs were being established. It was becoming clear that the old one size fits all approach to program monitoring was being overwhelmed by the increasing numbers of programs. Also, from an efficiency standpoint it did not make sense to spend the same amount of time with programs that were performing well as those that really needed additional attention. The birth of differential monitoring which at that time it was called inferential inspections. Different terminology, same concept.

Since then, differential monitoring has two basic methodologies that have been used successfully over the years: risk assessment and key indicators. The two methodologies have the same results, shortened or abbreviated reviews but they differ in their approaches. Risk assessment as the name implies identifies specific standards that place clients/children at greatest risk or morbidity or mortality if not complied with. Key indicators are specific standards that statistically predict overall regulatory compliance with all rules. Each has their place in the differential monitoring approach depending on the jurisdictions' emphasis. Most recently, to balance the emphasis on regulatory compliance has been the introduction of quality indicators which are specific standards drawn from quality initiatives, such as professional development, program quality tools, and quality rating & improvement systems.

It is and always has been recommended that these methodologies be used together and not separately. This final study undertaken in the Province of Saskatchewan completes the cycle of doing just that in developing a fully functional differential monitoring system with key licensing and quality indicators as well as risk assessment rules.

THE STUDY DESIGN AND METHOD

The design of this study was to provide a validation study of the use of the Saskatchewan Early Care and Education Quality Indicators Tool. A convenience sample was selected in which a good variation of overall quality would be present. There were to be three buckets of quality: High, Middle, and Low. These would be defined via ERS scores. Because this was a validation study it was critical to have sufficient variation in the overall quality of programs to test the sensitivity of the new assessment tool.

The below chart (Chart 1) provides guidance to the Ministry of Education policy staff in determining how to collect the program quality data for the research pilot study related to early childhood quality indicators.

Chart 1: Selection Process for Study Programs

Quality	Centers	Classrooms	Ages	Levels	ERS	SKECPQI
High	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3
Middle	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3
Low	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3

Notes:

A = ITERS (Infants) (B-1yr)
B = ITERS (Toddlers) (1yr-2yrs)
C = ECERS (Preschoolers) (3+yrs)
1 = SKECPQI/Infant (QI items 1-5, 7, 9-10)
2 = SKECPQI/Toddler or Preschool (QI items 1-5, 7, 9-10) or (QI items 1-6, 8-10)
3 = SKECPQI/Preschool (QI items 1-6, 8-10)
SKECPQI = Saskatchewan Early Childhood Program Quality Indicators tool

A total of 6 trained data collectors will be needed, 3 for the ERSs and 3 for the SKECPQI. Each observer will collect data from 30 classrooms. Will need a data coordinator who collects all the data, scores the tools and sends them to Dr Fiene. The data collectors should not be aware of which centers are in which group, such as High, Middle, or Low

See Appendix 1 for the Draft of the SKECQKI tool that was used during data collection.

As said earlier this study involves the validation of the Saskatchewan Early Childhood Quality Indicators Tool (SECQIT) and will involve the collection of new data utilizing the new tool and collecting Early Childhood Environmental Rating Scale (ECERS/ITERS) data as well. Independent contract staff will be trained in the use of the SECQIT as well as having had training on the ECERS/ITERS and are proficiently reliable on the ECERS/ITERS.

A sample of 30 childcare programs who volunteer to be part of this study will be selected with 1/3 identified as high quality, 1/3 identified as medium quality, 1/3 identified as low quality. Each program will have both the SECQIT and the ECERS/ITERS administered to them utilizing two independent observers. The data from the SECQIT will be compared to the ECERS/ITERS to determine the relationship between the two/three scales. The research hypothesis is that there will be a positive relationship between the two/three scales in which those programs that score high on the SECQIT will score high on the ECERS/ITERS and those that score low on the SECQIT will score low on the ECERS/ITERS. The ECERS/ITERS will be used as the reference tool for establishing the validity of the SECQIT.

A training program and all necessary revisions to policies and procedures will be conducted as part of this project by NARA Consultants on both phase 1 and 2. It will be determined later if the SECQIT will be

administered on an ongoing basis by contracted staff or by Ministry staff. Reporting templates will be developed as part of this implementation stage. The implementation stage will be evaluated to make certain that all components are in place and working as they should.

Timeline: Phase 1: 6 months; Phase 2: 9 months; Training and Implementation Phase: 12 months, will overlap with phase 1 and 2 and extend beyond both. The total time frame will be 24 months (about 2 years), this will include the final report and final evaluation of the implementation stage

RESULTS

The ECERS and ITERS were used to validate the new Saskatchewan Early Care and Education Quality Indicators Tool. This is standard procedure when conducting a validation study, a recognized empirically based and accepted standard tool is used in correlational analyses to determine if the new tool is measuring the same dimensions as the standardized tool.

The target tool, the Saskatchewan Early Care and Education Quality Indicators, was to be validated against the ECERS and ITERS to determine if there was a quality relationship between the two tools.

The validation analyses involved detailed correlational analyses between the various scales to determine if a relationship existed and how strong that relationship was.

DISCUSSION

Last piece of the puzzle in creating a differential monitoring system, that is how this report is being characterized. The Province of Saskatchewan has undertaken all the other methodologies utilized in a differential monitoring approach. Licensing key indicators and risk assessment rules have been implemented successfully. What remained was the Quality Indicators. This report completes the full cycle of validating these last indicators.

CONCLUSION

I am sure that this report will be read with a certain amount of skepticism in that it suggests using differential monitoring on a much broader scale. The differential monitoring approach has been utilized by many jurisdictions and has been cited in the United States Federal Legislation that reauthorized the Child Care and Development Block Grant. In the legislation, it is suggested but not required that states entertain the use of the approach. Based upon the latest childcare licensing data, it appears that many states have attempted to utilize the approach.

NOTES

Ministry of Education: Kim Taylor, Derek Pardy, Cindy Jeanes, Tanya Mengel, Samantha Ecarnot, Karen Heinrichs, Michelle Vellenoweth, Kristin Jarvis.

NARA Consultant: Rick Fiene.

Data collectors: List who collected the data and key Saskatchewan Ministry of Education staff.

Programs: Ask if the programs would like to be listed.

For additional information regarding this research validation study and report, please contact:

Richard Fiene, Ph.D., Research Psychologist/Regulatory Scientist, Research Institute for Key Indicators/Penn State University. rfiene@rikinstitute.com http://rikinstitute.com

Or

NARA: National Association for Regulatory Administration, Senior Research Consultant.
rfiene@naralicensing.org http://naralicensing.org/key-indicators

REFERENCES

ACF, (2015). Caring for Our Children: Basics, Department of Health and Human Services, Administration for Children and Families, Washington, DC.

Fiene, (2019a). The Saskatchewan Key Indicator System: The First Step in Developing a Differential Monitoring Approach, National Association for Regulatory Administration.

Fiene, (2019b). The Saskatchewan Centre and Home-Based Weighted Risk Assessment Study, National Association for Regulatory Administration.

Fiene, (2020). Validation Research Studies of Key Indicator and Risk Assessment Methodologies in the Province of Saskatchewan, National Association for Regulatory Administration.

Fiene, (2002a). 13 Indicators of Quality Child Care: Research Update, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

Fiene, (2014a). Qualistar Rating Key Indicator Study, National Association for Regulatory Administration.

Lahti, Elicker, Zellman, Fiene, (2014b). Approaches to validating childcare quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs, Early Childhood Research Quarterly, Special Issue.

Fiene, (2019c). Parental Guide to Choosing Safe and Healthy Child Care, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

Fiene, etal (2002b). Pennsylvania Early Childhood Program Quality Study, Harrisburg, Pa: Governor's Office.

APPENDICES

1 SKECQKI: *Saskatchewan Early Care and Education Quality Indicators Tool*

2 Fiene, (2015). *DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL (ECPQIM4©) FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES*, Research Institute for Key Indicators, Penn State University.

3 Fiene, (2022). Regulatory Compliance Monitoring Paradigms and the Relationship of Regulatory Compliance/Licensing with Program Quality: A Policy Commentary, *Journal of Regulatory Science, JRS (2022) Volume 10: Issue 1, pps 1-7.*

4 Fiene, (2019). A Treatise on the Theory of Regulatory Compliance*, Journal of Regulatory Science, JRS (2019) Volume 7, Issue 1, pps 1–3.*

5 Lahti, Elicker, Zellman, Fiene, (2014). Approaches to validating childcare quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs*, Early Childhood Research Quarterly, Special Issue.*

6 Fiene, (2002). Improving Child Care Quality Through an Infant Caregiver Mentoring Project, *Child & Youth Care Forum, 31(2), April 2002, pps 79-87.*

7 Fiene, (2019). *Parental Guide to Choosing Safe and Healthy Child Care*, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

The above papers contain all the key elements, theory, and modeling for the differential monitoring approach and its associated methodologies.

SK ECQKI Report7

---

[1] About the Author:

Dr Richard Fiene, a research psychologist and regulatory scientist, has spent his professional career in improving the quality of childcare in various states, nationally, and internationally. He has done extensive research and publishing on the key components in improving childcare quality through an early childhood program quality indicator model of training, technical assistance, quality rating & improvement systems, professional development, mentoring, regulatory science, licensing, risk assessment, differential program monitoring, key indicators, and accreditation. His research has also made significant contributions in regulatory science related to measurement and monitoring systems.

Dr Fiene is a retired professor of human development & psychology (Penn State University) and founding director of the Capital Area Early Childhood Research and Training Institute with Dr Mark Greenberg. He is presently President & Senior Research Psychologist and Regulatory Scientist for the Research Institute for Key Indicators.

Dr Fiene is regarded as a leading international researcher/scholar on human services licensing measurement and differential monitoring systems. His regulatory compliance law of diminishing returns has altered human services regulatory science and licensing measurement dramatically in thinking about how best to monitor and assess licensing rules and regulations through targeted and abbreviated inspections. The theory has also led to the issuing of human service licenses on the basis of substantial regulatory compliance with all rules rather than full 100% regulatory compliance with all rules. This was a basic licensing and public policy paradigm shift which has impacted regulatory administration.

So what would these targeted and abbreviated inspections look like from a program monitoring perspective? Rather than requiring companies, organizations, or agencies to be in full regulatory compliance, it would focus more on substantial compliance with all rules and full compliance with key indicator rules that statistically predict overall regulatory full compliance. This would be a more effective and efficient allocation of monitoring resources that would lead to increased outcomes for clients and better management for providers. The ultimate goal is to obtain the proper balance of regulatory oversight which is not too stringent nor too lax but rather one that focuses on the right (statistical predictors) rules producing the greatest impact on clients and providers of service.

His research has led to the following developments: identification of herding behavior of two year olds, national early care and education quality indicators, mathematical model (Contact Hours) for determining adult child ratio compliance, solution to the trilemma (quality, affordability, and accessibility) in child care delivery services, Stepping Stones to Caring for Our Children, NECPA: National Early Childhood Program Accreditation, online coaching as a targeted and individualized learning platform, validation framework for early childhood licensing systems and quality rating & improvement systems, an Early Childhood Program Quality Improvement & Indicator Model, Caring for Our Children Basics, Abbreviated Program Monitoring Inspections, Validation Framework for Licensing, Generic Key Indicator Rules, Regulatory Compliance Scoring Scale, RegalMetrics, and has led to the development of statistical techniques for dealing with highly skewed, non-parametric data distributions in human services licensing and regulatory systems, such as data dichotomization.

Dr Fiene had a long career in academia and governmental service. He was a research psychologist and

regulatory scientist during his tenure with the Commonwealth of Pennsylvania's Office of Children, Youth, and Families and the Office of Licensing and Regulatory Administration where he was the research director for both offices. In academia he was a professor of psychology and human development at both the University of North Carolina and the Pennsylvania State University. At Penn State Harrisburg he was Department Head for both the psychology and human development programs during his tenure at the university.

At the national and international levels, Dr Fiene has been a senior research consultant to the National Association for Regulatory Administration, the Federal Office of Child Care, the Administration for Children and Families, and the Federal Department of Health and Human Services. His research has been disseminated to all 50 states and over 120 countries. He received the 2020 Distinguished Career Award from the Pennsylvania Association for the Education of Young Children. Dr Fiene remains active in the regulatory science and early childhood fields through the Prevention Research Center at Penn State where he remains an affiliate professor.

APPENDICES

Saskatchewan Early Care and Education Quality Indicator Tool

Early Childhood Program Quality Improvement and Indicators Model

Background Research Papers:

Fiene, (2022). Regulatory Compliance Monitoring Paradigms and the Relationship of Regulatory Compliance/Licensing with Program Quality: A Policy Commentary*, Journal of Regulatory Science, JRS (2022) Volume 10: Issue 1,* pps 1-7.

Fiene, (2019). A Treatise on the Theory of Regulatory Compliance*, Journal of Regulatory Science, JRS (2019) Volume 7, Issue 1,* pps 1–3.

Lahti, Elicker, Zellman, Fiene, (2014). Approaches to validating childcare quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs*, Early Childhood Research Quarterly*, Special Issue.

Fiene, (2002). Improving Child Care Quality Through an Infant Caregiver Mentoring Project, *Child & Youth Care Forum, 31(2)*, April 2002, pps 79-87.

Fiene, (2019). *Parental Guide to Choosing Safe and Healthy Child Care*, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

Saskatchewan's Early Learning and Child Care Program Quality Key Indicator Instrument (SKECPQI Ver9J)

The Saskatchewan Program Quality Work Group[1]

October 2020 (revised January 2023)

INTRODUCTION and BACKGROUND to SKECPQI

Ten Quality Key Indicators (QKI) make up the Saskatchewan's Early Learning and Child Care Program Quality Key Indicator Instrument (SKECPQI). The details about each of the Quality Indicators and data collection instructions in order to obtain the necessary data to determine if a program meets the Key Quality Indicators are delineated below for each quality key indicator. Part 1 - Quality Key Indicators (QKI) 1 – 5 will be collected via record or document review, interviewing individuals, or observation. Part 2 - Quality Key Indicators (QKI) 6 – 10 will be collected via observations in the classrooms throughout the assessment.

These ten quality key indicators were taken from previous studies conducted over the past 40 years by Dr Richard Fiene utilizing the Regulatory Compliance Key Indicator metric (RCKIm) that he developed in the late 1970's. These QKI have held up over time and have now been coupled together into this tool and being pilot tested in the Province of Saskatchewan. The original tool was reviewed by a Provincial Ministry of Education Work Group who met during 2019-2020 and made some revisions to the original tool. All these changes are reflected in this version of the SKECPQI.

---

1) Saskatchewan Program Quality Work Group: Kim Taylor, Derek Pardy, Cindy Jeanes, Tanya Mengel, Samantha Ecarnot, Karen Heinrichs, Michelle Vellenoweth, Kristin Jarvis, and Rick Fiene.

2) The initial estimated time to complete the assessment is three and a half hours (3.5 hrs)].

PART 1 – Record/Document Review, Interview, Observation Quality Indicators

INDICATOR 1): Number of ECE III Educators (10 minutes)

Assessors will review staff records in order to determine the number of staff who have these credentials in early childhood education. Record the number of ECEs with the appropriate qualifications and divide them by the total number of ECEs in order to come up with a percent for the center.

How to Measure:

Go to the ***Staff Information Summary*** form to obtain the data for this item. There are two particular columns that will do this. Under Certification: *Certification Date and Certification Level* (Highest ECE Level Certified). The certification date should be earlier than the date of the review and the actual level of the certification. In this case, we are interested in the number of (ECEIII's). Record the number of ECEIII working at least 65 hours/month. Then record the number of total teaching staff working at least 65 hours/month below as well. Teaching staff is defined as staff who have a responsibility for working with the children and the programming. Determine the percentage by dividing the total number of staff into the total number of ECEIII Certified teaching staff, ECEIII Certified teaching staff is the numerator, and the total number of teaching staff is the denominator (ECEIII/Total number of teaching staff x 100% = Percent).

Scoring for PQI 1:

The total number of ECEIII Certified teaching staff _____ (1.1)

The total number of teaching staff _____ (1.2)

Total ECEIII teaching staff divided by the total number of teaching staff _____ (%).

Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	*1 = 0 to 25%*	*2= 26 to 50%*	*3 = 51 to 75%*	*4 = 76 to 100%*

INDICATOR 2): Stimulating and Dynamic Environment (10 minutes)

The criteria for measuring this are drawn from ***Play and Exploration Guide.*** The program is child centered. Children are viewed as competent learners, and they have the freedom to access classroom materials independently without adult intervention. The children are provided with meaningful choices through activity/learning centers. There is evidence of the children's interests and their projects in the learning environment.

How to Measure:

Below is the checklist of items that should be present in order to assess if the environment is both stimulating and dynamic for the children. You will want to observe that the following items are occurring in the classroom first. If you do not actually observe it occurring, then check the program plan to find documentation that it normally occurs but you just did not observe today. The checklist items would be found in *Play and Exploration* foundational materials.

Quality Early Learning Environments (Please record all that you observe Y or N):

1. Co-teaching is evident. Y/N _____ (2.1)
2. Children are viewed as competent learners & can access materials independently. Y/N ___ (2.2)
3. Authentic and meaningful materials are used with children. Y/N _____ (2.3)
4. Children are provided with meaningful choices. Y/N _____ (2.4)
5. Children's work, art and photos are displayed respectfully. Y/N _____ (2.5)
6. Family photos are displayed in the early learning program. Y/N _____ (2.6)
7. Documentation of learning is displayed and discusses holistic development. Y/N _____ (2.7)
8. Environment reflects the culture and beliefs of the children, families and staff. Y/N _____ (2.8)

9. Variety of books & other print materials are available throughout the classroom Y/N ____ (2.9)
10. A variety of writing materials are accessible to children most of the time. Y/N _____ (2.10)
11. There is evidence of the children's interests & projects in the classroom. Y/N ___ (2.11)

Scoring for PQI 2:

Total up the number of items where you recorded a "Y" above that you observed (curriculum or in classrooms), divide by 11 x 100% to come up with a percent and record here _____ %. Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	***1 = 0 to 25%***	***2= 26 to 50%***	***3 = 51 to 75%***	***4 = 76 to 100%***

INDICATOR 3): Developmentally Appropriate Curriculum Based on Assessments of Each Child (50-60 minutes)

The key for this quality key indicator is that the program is following an individualized prescribed planning document when it comes to curriculum. It does not mean it is a canned program, in fact, it shouldn't if it is based upon the individual needs of each child's developmental assessment. The assessor will ask to see what is used to guide the curriculum. There should be a written document that clearly delineates the parameters of the philosophy, activities, guidance, and resources needed for the particular curricular approach. There should also be a developmental assessment which is clearly tied to the curriculum. The developmental assessment can be home-grown or a more standardized off-the-shelf type of assessment, the key being its ability to inform the various aspects of the curriculum. The purpose of the assessments is not to compare children but rather to compare the developmental progress of individual children as they experience the activities of the curriculum.

The following key elements should be present when assessing this quality indicator.

- 1) The program practices emergent curriculum, allowing the interests of the children to determine the learning content. The curriculum is informed by individual developmental assessments of each child in the respective classrooms.
- 2) The children and educators are co-learners in the exploration of projects.
- 3) Learning activities of the children are documented, displayed in the learning environment and used to plan further learning activities. This can be assessed developmentally.

How to Measure:

Take a sample of 10 individual children's records and consider the above three elements for EACH record. You should be asking yourself if there is a clear link between an assessment and the developmentally appropriate curriculum so that an individualized learning approach is being undertaken and each child's developmental needs are taken into consideration. These records could be formal, such as portfolios kept for each child or a more informal, anecdotal type of record keeping. The key is that there is a record that can be looked at. It is not adequate if the teacher says they do it from memory – it needs to be written down and documented.

Cross check the child's record to the actual curriculum. Record all the instances (Y's) in which this occurs. All three blocks need to be checked for each record (1-10).

Emergent Curriculum is Practiced (3.1)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 1 +

Children and Educators are Co-learners (3.2)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 2 +

Learning Activities are Documented and Displayed and Used to Plan Future Learning (3.3)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 3 +

All three key elements must have a Y to get an overall score of Y. If all three key elements have a Y for that individual record, then record Y in the corresponding block in the overall score.

1 Ys =	2 Ys =	3 Ys =	4 Ys =	5 Ys =	6 Ys =	7 Ys =	8 Ys =	9 Ys =	10 Ys =

= Total of All Three Key Elements (3.4)

Scoring for PQI 3:

The number of positive records (all Ys for all three elements) where there is a crosswalk from developmental assessment to curriculum _____

Percent of positive records (all Ys) (divide the number of positive records by 10 x 100%) _____ %.

Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	***1 = 0 to 25%***	***2= 26 to 50%***	***3 = 51 to 75%***	***4 = 76 to 100%***

INDICATOR 4): Opportunities for Staff and Families to Get to Know Each Other (10 minutes)

There should be activities both within the center as well as off site where staff and parents have opportunities to meet and greet each other. Communication with family members is documented and enables early childhood providers to assess the need for follow-up. Early childhood providers hold regular office hours when they are available to talk with family members either in person or by phone. Family members are encouraged to lead the conversation and to raise any questions or concerns.

How to Measure:

Look for the following 3 examples in policies developed by the program and determine if they have been carried out with families. It will be necessary to interview staff to complete this indicator if you do not find the three examples in policies:

1. The program provides communication, education, and informational materials & opportunities for families that are delivered in a way that meets their diverse needs. Y/N_____ (4.1)
2. The program communicates with families using different modes of communication, and at least one mode promotes two-way communication. Y/N _____ (4.2)
3. The program demonstrates respect and engages in ongoing two-way communication. The program respects each family's strengths, choices, & goals for their children. Y/N ____ (4.3)

Scoring for PQI 4:

Record the number of Yes's (Y's): _____ (Range: 0 – 3) (Divide by 3 x 100% = _____%). Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	*1 = 0 to 25%*	*2= 26 to 50%*	*3 = 51 to 75%*	*4 = 76 to 100%*

INDICATOR 5): Families Receive Information on Their Child's Progress Regularly Using a Formal Mechanism (Report or Parent Conference) (10 minutes)

Based upon Indicator #3 above, the information gleaned from the developmental assessments should be the focus of the report or parent conference. Parental feedback about the assessment and how it compares to their experiences at home would be an excellent comparison point. All these interactions should be done in a culturally and linguistically appropriate way representing the parents being served.

How to Measure:

Look for the following four examples in policies developed by the program and determine if they have been carried out with families. Record the number of reports completed or parent conferences over the past year. It will be necessary to interview staff to complete this indicator if you cannot determine from records that the conferences or reports were completed.

NOTE: The examples are mutually exclusive and are not additive; the first example is the highest scored, the third example the least scored. After 1-3 are determined, then do the last example.

- 1) The program does have regularly scheduled (at least 2xs/year) parent conferences in which the children's developmental progress is discussed AND provides the family with a report of their child's developmental progress. Y/N _____ (5.1) (Score 3 points). If "Yes" then go to Number 4. If "No", then go to numbers 2 and 3.

- 2) The program has regularly scheduled (at least 2xs/year) parent conferences in which the children's developmental progress is discussed, but it does not provide a report to the parents on their child's developmental progress. Y/N _____ (5.2) (Score 2 points).
- 3) If the program does not have regularly scheduled (at least 2xs/year) parent conferences, does it provide the family with a report of their child's developmental progress. Y/N _____ (5.3) (Score 1 point). Go to Number 4.
- 4) All these interactions are done in a culturally and linguistically appropriate way representing the parents being served. Y/N _____ (5.4) (Score 1 point)

Scoring for PQI5:

Add up the total points based on the Ys; this will range from "0" to "4". The only way a program can receive a "4", is if a program has regularly scheduled parent conferences at least 2xs/year and provides the family with a report of their child's progress; and it is done in a culturally and linguistically appropriate way.

Record the number of points: _____ (Range: 0 - 4)

PART 2 - OBSERVATIONS:

For quality key indicators 6, 7 and 8, it is recommended that the licensing consultant refer to the appropriate Environmental Rating Scale (ERS) tool as a reference tool because these indicators are taken directly from these tools. It is also recommended that these be assessed/observed throughout the assessment and not just during key activity times. Please follow the specific instructions and examples as delineated below and in the appropriate ERS tool: ECERS (Items 12 and 13) or ITERS (Item 12). These specific instructions and examples are provided within this tool for ease of administration and data collection. If there are several preschool aged classrooms randomly select one to do your observations.

INDICATOR 6): Educators Encourage Children to Communicate (20 minutes) (Preschool Class)

Assessors will need to observe this item when they do their classroom observations. Initially you can ask educators or the director how children are encouraged to communicate but in order to gather reliable and valid information regarding this question/standard, it needs to be observed in the various interactions between staff and children. Things to look for would be more back and forth conversations rather than one-way conversations where educators are telling children what to do. Look for opportunities where children can describe what they are doing, how they feel about what they are doing, and why they are doing particular activities. Educators expand upon children's conversations. These opportunities can occur anywhere in the classroom or outside, such as in dramatic play, tabletop activities or on the playground. Materials should be present that encourage communication such as toy telephones, puppets, flannel boards, dolls and dramatic play props, small barns, fire stations, or dollhouses. These create a lot of conversation among children as they assume many different roles.

Children also talk when there is an interested person who listens to them. The staff in a high-quality early childhood classroom will use both activities and materials to encourage growth in communication skills.

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;

Score the classroom a 1 if the following occur:

- No activities used by staff with children to encourage them to communicate, for example: nontalking about drawings, dictating stories, sharing ideas at circle time, finger plays, singing songs. Y/N _____ (6.1)
- Very few materials accessible that encourage children to communicate. Y/N _____ (6.2)

Score the classroom a 2 if the following occur (If the classroom does not have all 3 indicators but has 2 of the indicators then score this item 1+):

- Some activities are used by staff w/children to encourage them to communicate. Y/N _____ (6.3)
- Some materials are accessible to encourage children to communicate. Y/N ____ (6.4)
- Communication activities are generally appropriate for the children in the group. Y/N _____ (6.5)

Score the classroom a 3 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 2+):

- Communication activities take place during both free play and group times, for example: child dictates story about painting; small group discusses trip to store. Y/N _____ (6.6)

- Materials that encourage children to communicate are accessible in a variety of interest centers, for example: small figures and animals in block area; puppets and flannel board pieces in book area; toys for dramatic play outdoors or indoors. Y/N _____ (6.7)

Score the classroom a 4 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff balance listening and talking appropriately for age and abilities of children during communication activities, for example: leave time for children to respond; verbalize for child with limited communication skills. Y/N _____ (6.9)
- Staff link children's spoken communication with written language, for example: write down what children dictate & read it back to them; help them write notes to parents. Y/N _____ (6.10)

Scoring for PQI 6:

Total up the number of "Y's" and record the appropriate level. In order for a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+". If there is a "+" please also mark it in the box.

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 7): Infant Toddler Observation (if applicable) **(20 minutes) (Infant Classroom)**

NOTE: If there is an infant, toddler or combined infant/toddler classroom that needs to be assessed, then use the following ITERS item directly from the ITERS Tool (Item 12), if there is not an infant toddler classroom, then skip to Indicator 8.

Conversations and questions should be used with all children, even young infants. Conversations using verbal and nonverbal turn-taking should be considered when scoring. Most conversations and questions initiated by infants will be nonverbal, such as widening of baby's eyes or waving arms and legs. Observe staff response to such nonverbal communication. For infants and toddlers, the responsibility for starting most conversations and asking questions belongs to the staff. As children become more able to initiate communication, staff should modify their approach in order to allow children to take on a greater role in initiating conversations and asking questions. Staff should provide answers to questions used by children if children cannot answer, and as children become more able to respond, questions should start to include those that the child can answer. If there was not an infant classroom, skip this Indicator and please note that here and on the summary score sheet by marking N/A: _____

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;

Score the classroom a 1 if the following occurs:

- Staff never initiate turn-taking conversations with children, for example: rarely encourage baby to babble back; simple back and forth exchanges with verbal children never observed. Y/N _____ (7.1)

- Staff questions are often not appropriate for children, or no questions are asked, for example: too difficult to answer; carry a negative message. Y/N _____ (7.2)
- Staff respond negatively when children can't answer questions, for example: "You should know this"; "You did not listen". Y/N _____ (7.3)

Score the classroom a 2 if the following occurs (If the classroom does not have all 3 indicators but has 2 of the indicators then score this item 1+):

- Staff sometimes initiate conversations with children, for example: babble back and forth with baby; copy baby's sounds; respond to baby's crying with verbal response; have short back and forth toddler interactions. Y/N _____ (7.4)
- Staff sometimes ask children appropriate questions and wait for the child to respond, for example: ask baby if she likes toy and pay attention as baby smiles; ask toddler what he is eating and wait for him to think of word. Y/N _____ (7.5)
- Staff respond neutrally or positively to children who can't answer questions. Questions asked are sometimes meaningful to children, for example: child responds with interest; does not ignore staff questions. Y/N _____ (7.6)

Score the classroom a 3 if the following occurs (If the classroom does not have all 4 indicators but has 2 or more of the indicators then score this item 2+):

- Staff initiate engaging conversations with children throughout the observation, for example: show enthusiasm; use tone that attracts child's attention. Y/N _____ (7.7)
- Staff often personalize questions and/or conversations for individual children, for example: talk about children's families, preferences, interests; what they are playing with; what they did over weekend; child's mood; use child's name. Y/N _____ (7.8)
- Staff often pay attention to children's questions, verbal or nonverbal, and answer in a satisfying manner for the child. Y/N _____ (7.9)

- Staff ask questions in which children show interest in answering, for example: make the questions funny or mysterious; use attractive tone; meaningful and not too difficult to answer. Y/N _____ (7.10)

Score the classroom a 4 if the following occurs (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff frequently have turn taking conversations with children throughout the observations. Many appropriate questions are used throughout the observation, during both play and routines. Y/N _____ (7.11)
- Staff ask children appropriate questions, wait a reasonable time for child response, and then answer if needed, for example: "Are you hungry? . . . Yes, you are!"; "Where's the ball? . . . These it is! You found the ball". Y/N _____ (7.12)

Scoring for PQI 7:

Total up the number of "Y's" and record the appropriate level. For a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+".

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 8): Educators Use Language to Develop Reasoning Skills (20 minutes) (Preschool)

Assessors will need to observe very carefully as this standard can be difficult to determine because it is tying language and cognition together. Again, this opportunity can occur in any setting in or out of the classroom because it is the basis for problem solving through the use of language. Also look for educators redirecting children's conversations when appropriate. Staff should use language to talk about logical relationships using materials that stimulate reasoning. Through the use of materials, staff can demonstrate concepts such as same/different, classifying, sequencing, one-to-one correspondence, spatial relationships, and cause and effect.

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;

Score the classroom a 1 if the following occur:

- Staff do not talk with children about logical relationships, for example: ignore children's questions and curiosity about why things happen, do not call attention to sequence of daily events, differences and similarity in number, size, shape, cause and effect. Y/N _____ (8.1)
- Concepts are introduced inappropriately, for example: concepts too difficult for age and abilities of children, inappropriate teaching methods used such as worksheets without any concrete experiences; teacher gives answers w/o helping children to figure things out. Y/N _____ (8.2)

Score the classroom a 2 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 1+):

- Staff sometimes talk about logical relationships or concepts, e.g.: explain that outside time comes after snacks, point out differences in sizes of blocks children use. Y/N _____ (8.3)

- Some concepts are introduced appropriately for ages and abilities of children in group, using words and experiences, for example: guide children with questions and words to sort big and little blocks or to figure out why ice melts. Y/N _____ (8.4)

Score the classroom a 3 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 2+):

- Staff talk about logical relationships while children play with materials that stimulate reasoning, for example: sequence cards, same/different games, size and shape toys, sorting games, numbers and math games. Y/N _____ (8.5)
- Children are encouraged to talk through or explain their reasoning when solving problems, for example: why they sorted objects into different groups, in what way two pictures are the same or different. Y/N _____ (8.6)

Score the classroom a 4 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff encourage children to reason throughout the day, using actual events and experiences as a basis for concept development, e.g.: children learn sequence by talking about their experiences in the daily routine or recalling the sequence of a cooking project. Y/N _____ (8.7)
- Concepts are introduced based upon children's interests or needs to solve problems, for example: talk children through balancing a tall block building, help children figure out how many spoons are needed to set a table. Y/N _____ (8.8)

Scoring for PQI 8:

Total up the number of "Y's" and record the appropriate level. In order for a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+".

Circle the Appropriate Level	*1*	*2*	*3*	*4*

For quality key indicators 9 and 10 it is recommended that these be assessed/observed throughout the observation period and not just during key activity times. These two quality key indicators should be observed in two-minute blocks over ten sequences for a total of 20 minutes. These two items should also be used with each age group you are assessing.

Initially it will be necessary to observe these two quality indicators separately, but they could be observed and recorded jointly once you are familiar with the tool and have done sufficient observations.

INDICATOR 9): Educators Listen Attentively When Children Speak (25 minutes)

This quality indicator focuses on the early childhood educator(s) looking directly at the children with nods, rephrases their comments, engages in conversations. Children should have the undivided attention of the specific educator they are addressing. Educators should not be looking away or pre-occupied with others. They should be at the child's level making eye contact. The intent is to observe all children and educators in the room.

How to Measure:

Do this in timed 2-minute observations recording each time you observe this occurring. Record at least 10 different observation periods. These do not need to be consecutive in order to fully observe classrooms and educators. Please use the following scale to assess your recordings: Likert Scale (1-4) where 1 = Never/Not at All; 2 = Somewhat/Few Instances; 3 = Quite a Bit/Many Instances; 4 = Very Much/Consistently):

Make the actual recordings using the Likert Scale (1-4) above for each individual observation and record in each cell below.

10 Observations:

10.1	2	3	4	5	6	7	8	9	10.10

Scoring for PQI 9:

Once all the observations are made, add up the results from the Likert Scale (1-4) and record the total number here: _____ (Range: 10 - 40)(Divide this result by 10) = _____ (1-4)(Round upward or downward to the whole number (3.7 = 4; 2.2 = 2)).

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 10): Educators Speak Warmly to Children (25 minutes)

This quality indicator focuses on the early childhood educator(s) always engaging in a caring voice and body language with every child. Educators do not use harsh language or commands in speaking to children, but rather again are on the child's level making eye contact. Think of the way Fred Rogers would engage his audience where you always felt you were the most important person in the world when he talked to the TV.

How to Measure:

Do this in timed 2-minute observations recording each time you observe this occurring. Record at least 10 different observation periods. Please use the following scale to make your recordings: (This item is on a Likert Scale (1-4) where 1 = Never/Not at All; 2 = Somewhat/Few Instances; 3 = Quite a Bit/Many Instances; 4 = Very Much/Consistently):

Make the actual recordings using the Likert Scale (1-4) above for each individual observation and record in each cell below.

10 Observations:

10.1	2	3	4	5	6	7	8	9	10.10

Scoring for PQI 10:

Once all the observations are made, add up the results from the Likert Scale (1-4) and record the total number here: _____ (Range: 10 - 40) (Divide this result by 10) = _____ (1-4).

(Round upward or downward to the whole number (3.7 = 4; 2.2 = 2)).

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 11): Reflective Supervision Placeholder TBD. The Work Group had a great deal of discussion regarding this Quality Indicator and could not come to an agreeable format for this QI, so for the purposes of the initial pilot testing, the Coaching/Reflective Supervision Quality Indicator is not included. However, this is acknowledged as an important quality indicator and will be addressed in future versions of this tool. ***If you have the time, ask the director if they do any coaching or reflective supervision, please record here if they do coaching/reflective supervision:***

PQI 11: Yes _____ No _____

Notes (record any notations for indicators here by noting the number of the quality indicator and any notes or comments)(attach any additional pages as needed):

After completing your observations, reviewing all documentation, and interviewing staff, when necessary, please transfer all your results to the Summary Table below. If there was not an infant classroom, please note here, no infant classroom: _____. If there was not a toddler classroom, please note here, no toddler classroom: _____. If there was not a preschool classroom, please note here, no preschool classroom: _____.

Key Q Indicator	***Quality Indicator Content***	***Scale Source***	***Potential Score***	***Actual Score***
QKI 1	**Professional Development**	**NAEYC**	**1-4**	**1, 2, 3, 4**
QKI 2	**The Environment**	**Saskatchewan**	**1-4**	**1, 2, 3, 4**
QKI 3	**Curriculum and Assessment**	**NAEYC**	**1-4**	**1, 2, 3, 4**
QKI 4	**Family Engagement I**	**QRIS**	**1-4**	**1, 2, 3, 4**
QKI 5	**Family Engagement II**	**QRIS**	**1-4**	**1, 2, 3, 4**
QKI 6	**Communication (Preschool)**	**ECERS**	**1-4 or NA**	**1, 2, 3, 4, +, NA**
QKI 7	***Infant Classroom***	***ITERS***	***1-4 or NA***	**1, 2, 3, 4, +, NA**
QKI 8	**Reasoning Skills (Preschool)**	**ECERS**	**1-4 or NA**	**1, 2, 3, 4, +, NA**
QKI 9	**Listen Attentively**	**CIS**	**1-4**	**1, 2, 3, 4**
QKI 10	**Speak Warmly**	**CIS**	**1-4**	**1, 2, 3, 4**

Notes:

Use ITERS if: (Infants) (B-1yr)

Use ITERS if: (Toddlers) (1yr-2yr)

Use ECERS if: (Preschoolers) (3yr+)

SKECPQI/Infant (administer QKI items 1-5, 7, 9-10) (Scores 8-32)

SKECPQI/Toddler or Preschool (administer QKI items 1-5, 7, 9-10) (Scores 8-32) or (administer QKI items 1-6, 8-10) (Scores 9-36). Mixed age group (administer QKI items 1-10) (Scores 10-40)

SKECPQI/Preschool (administer QKI items 1-6, 8-10) (Scores 9-36)

All the above 10 quality indicators (SKECPQI) have been taken from other sources having been identified in Quality Indicator Studies conducted by Dr Richard Fiene from 1980 – 2020. Please refer to the source documents for details on their creation: *ECERS, ITERS, QRIS/INQUIRE, CIS/Arnett, NAEYC, SASKATCHEWAN PLAY & EXPLORATION.* For additional information, reports, and publications related to these studies, please go to https://www.naralicensing.org/key-indicators

Or https://rikinstitute.com/publications/

Note:

Members of the Original Saskatchewan Program Quality Work Group are the following:

Ministry of Education: Kim Taylor, Derek Pardy, Cindy Jeanes, Tanya Mengel, Samantha Ecarnot, Karen Heinrichs, Michelle Vellenoweth, Kristin Jarvis, and NARA Consultant: Rick Fiene.

Additional Information contact: Derek Pardy, Government of Saskatchewan, Senior Policy Analyst, Early Years, Ministry of Education, 2-2220 College Ave, Regina, SK, Canada S4P 4V9.

Additional Information regarding the psychometrics of the tool contact: Richard Fiene, Ph.D., Research Psychologist, Research Institute for Key Indicators & Penn State University. RFiene@RIKInstitute.com or RFiene@NARALicensing.org

10/2020; 4/2021; 1/2023 versions

SKECPQI9J

Notes from Dr Fiene to the Assessors/Data Collectors:

It is very important with this pilot testing that you provide me with your candid observations, what worked for you, what did not. Please provide these in the notes section at the end of the tool and if necessary, within the tool narrative as well. Please reference the line numbers (left hand side of the tool) so that it will be easy for us to track your comments.

Record your observations onto the tool but do not worry about scoring the individual items in trying to determine the 1-4 scale. These are the sections that are in **red** type. I will do that once I get the tool back from you. Since this is the first pilot testing of the tool, the 1-4 scaling will need to be assessed if this is the best scale sequence. I will need to run the results through various testing protocols. So, in the data transfer part of the study, pdfs will be made of the tool results and sent to me in total.

However, with that said, if you want to just look over the scoring instructions and provide feedback to me regarding its clarity or lack thereof, that would be great also.

Thank you,

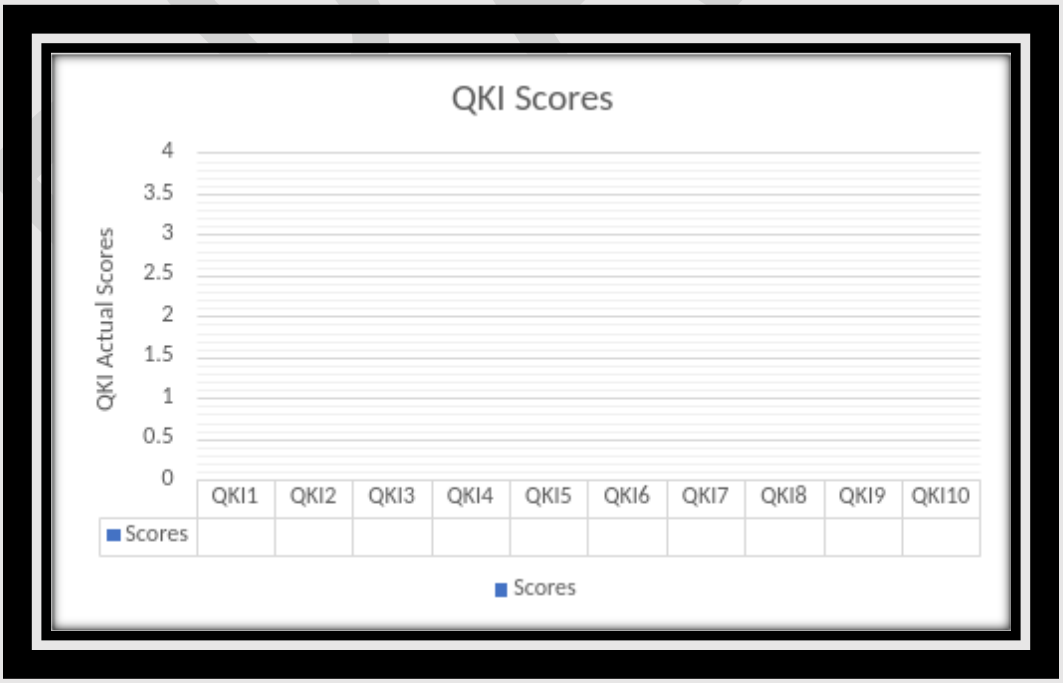
Rick

Richard Fiene, Ph.D., Research Study Principal Investigator, Saskatchewan Early Childhood Program Quality Indicators Project, Ministry of Education, Province of Saskatchewan.

SKECPQI: SASKATCHEWAN EARLY CHILDHOOD PROGRAM QUALITY INDICATORS

CHART/GRAPH

	Scores
QKI1	
QKI2	
QKI3	
QKI4	
QKI5	
QKI6	
QKI7	
QKI8	
QKI9	
QKI10	
TOTAL	



SKECPQI Scoring Protocol

LEVEL	Standardized Scores	Actual Scores
High Quality	Mixed Age: 36+ Preschool: 32+ Infant-Toddler: 28+	Mixed Age: _____ Preschool: _____ Infant-Toddler: _____
High - Mid Quality	Mixed Age: 30 – 35 Preschool: 26 - 31 Infant-Toddler: 22 - 27	Mixed Age: _____ Preschool:_____ Infant-Toddler:_____
Mid – Low Quality	Mixed Age: 20 – 29 Preschool: 16 - 25 Infant-Toddler: 12 - 21	Mixed Age: _____ Preschool: _____ Infant-Toddler: _____
Low Quality	Mixed Ages: 19 or less Preschool: 15 or less Infant-Toddler: 11 or less	Mixed Age: _____ Preschool:_____ Infant-Toddler: _____

QKI and key elements/sub items and comments Scoresheet:

QKI1 _____ 1.1 _____ 1.2 _____ Comments: _____

QKI2 _____ %

2.1 _____ Comments: _____

2.2 _____ Comments: _____

2.3 _____ Comments: _____

2.4 _____ Comments: _____

2.5 _____ Comments: _____

2.6 _____ Comments: _____

2.7 _____ Comments: _____

2.8 _____ Comments: _____

2.9 _____ Comments: _____

2.10 _____ Comments: _____

2.11 _____ Comments: _____

QKI3 _____ %

3.1 _____ Comments: _____

3.2 _____ Comments: _____

3.3 _____ Comments: _____

3.4 _____ Comments: _____

QKI4 _____ %

4.1 _____ Comments: _____

4.2 _____ Comments: _____

4.3 _____ Comments: _____

QKI5 _____ Points

5.1 _____ Comments: _____

5.2 _____ Comments: _____

5.3 _____ Comments: _____

5.4 _____ Comments: _____

QKI6 _____ Level

6.1 _____ Comments: _____

6.2 _____ Comments: _____

6.3 _____ Comments: _____

6.4 _____ Comments: _____

6.5 _____ Comments: _____

6.6 _____ Comments: _____

6.7 _____ Comments: _____

6.8 _____ Comments: _____

6.9 _____ Comments: _____

QKI7 _____ Level

7.1 _____ Comments: _____

7.2 _____ Comments: _____

7.3 _____ Comments: _____

7.4 _____ Comments: _____

7.5 _____ Comments: _____

7.6 _____ Comments: _____

7.7 _____ Comments: _____

7.8 _____ Comments: _____

7.9 _____ Comments: _____

7.10 _____ Comments: _____

7.11 _____ Comments: _____

7.12 _____ Comments: _____

QKI 8 _____ Level

8.1 _____ Comments: _____

8.2 _____ Comments: _____

8.3 _____ Comments: _____

8.4 _____ Comments: _____

8.5 _____ Comments: _____

8.6 _____ Comments: _____

8.7 _____ Comments: _____

8.8 _____ Comments: _____

QKI9 _____ Level

9.1 _____ Comments: _____

9.2 _____ Comments: _____

9.3 _____ Comments: _____

9.4 _____ Comments: _____

9.5 _____ Comments: _____

9.6 _____ Comments: _____

9.7 _____ Comments: _____

9.8 _____ Comments: _____

9.9 _____ Comments: _____

9.10 _____ Comments: _____

QKI10 _____ Level

10.1 _____ Comments: _____

10.2 _____ Comments: _____

10.3 _____ Comments: _____

10.4 _____ Comments: _____

10.5 _____ Comments: _____

10.6 _____ Comments: _____

10.7 _____ Comments: _____

10.8 _____ Comments: _____

10.9 _____ Comments: _____

10.10 _____ Comments: _____

Quality Key Indicators (QKI)	Elements/Items	Data Collection
1	1.	Record Review
2	11	Policy, Records, Interviews
3	4	Policy, Records, Interviews
4	3 . . .	Policy, Records, Interviews
5	4	Policy, Records, Interviews
6	9	Observation
7	12	Observation
8	8	Observation
9	10	Observation
10	10	Observation
TOTAL	**Potential Score = 78**	**Actual Score Obtained = _____**

DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL (ECPQIM^4©) FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES

Richard Fiene, Ph.D.

Professor of Human Development & Family Studies and Psychology (Retired)
Bennett Pierce Prevention Research Center
The Pennsylvania State University

Research Institute for Key Indicators *LLC* (RIKI)
41 Grandview Avenue
Middletown, Pennsylvania USA 17057

717-944-5868 Phone and Fax
DrFiene@gmail.com

Suggestion citation: Fiene (2015). *Differential monitoring logic model (DMLM©): A new early childhood program quality indicator model (ECPQIM4©) for early care and education regulatory agencies*, Middletown, PA: Research Institute for Key Indicators (RIKI).

DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL (ECPQIM^4©) FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES

ABSTRACT

A new Early Childhood Program Quality Indicator Model (ECCPQIM^4©) is described which utilizes targeted program monitoring (Differential Monitoring) via two licensing methodologies: Key Indicators and Risk Assessments. The theoretical and conceptual framework as well as a logic model are presented along with a scoring protocol that can be utilized to compare state/province and national organizations on how they are designing and implementing their program monitoring systems. A state/province/national framework/plan is presented as well as results from five (5) states (Georgia, Kansas, Illinois, Colorado, and New York) and a national organization (Office of Head Start). The five states and national organization are then compared using the Differential Monitoring Scoring Protocol (DMSP©). The Head Start program monitoring system scored a perfect 10 out of 10 in utilizing the DMSP©. Suggestions are made in how the scoring protocol could be used for making comparisons internationally and for future research in comparing various approaches.

Key Words: Program Monitoring, Differential Monitoring, Program Quality, Licensing.

Background

This paper will introduce a Differential Monitoring Logic Model (DMLM©) which provides a new Early Childhood Program Quality Indicator Model (ECPQIM^4©) in which the major monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With this new model, it is now possible to compare results obtained from licensing systems, quality rating and improvement systems (QRIS), risk assessment systems, key indicator systems, technical assistance, and child development/early learning outcome systems (see Figures 1 & 2 for a graphical depiction of the theoretical underpinnings and actual design & logic model for the ECPQIM^4©/DMLM).

The DMLM© can be used by early care and education state/province agencies, Federal agencies, and large provider organizations where an economy of scale is required. This model can be used with state as well as national standards, such as state licensing rules/regulations and *Caring for Our Children (AAP, 2012).* Most states and Federal agencies have either some or all of the key elements of this model in their overall monitoring systems. The purpose of this model is to alter a one-size fits all monitoring system to one that is targeted, spending more time with problem programs who need additional assistance. This is a cost neutral model that is both cost effective and efficient and re-allocates resources from the compliant programs to the non-compliant programs. Presently there is not a measurement rubric for making comparisons within the USA or internationally when it comes to measuring the effectiveness and efficiency of child care and

early care program monitoring systems. This can become a very important tool as the USA begins implementation of the re-authorization of the Child Care and Development Block Grant.

Insert Figure 1

The ECPQIM^4©/DMLM© is based very heavily in translational research and implementation science as a means of building an ongoing program monitoring system based upon the latest empirical demonstrations in the early care and education research literature. It is at the intersection of child care public policy, early care and education interventions, and empirical research. The ECPQIM^4©/DMLM© along with the scoring protocol introduced in this paper could provide a framework for making comparisons amongst states/provinces, national organizations, and countries in how they have designed and implemented their respective program monitoring of child care and early care & education systems similar to how Child Care Aware has developed a reporting format for the USA in comparing states on regulatory and oversight functions. The author reported on such a comparison in a previous study in an earlier edition of this journal (Fiene, 2013). The DMLM© framework and scoring protocol could provide a similar measurement tool for assessing child care and early childhood education program monitoring systems.

DMLM© Key Elements (see Figure 2): **CI** = state or federal child care standards, usually rules or regulations that measure health and safety - *Caring for Our Children* (AAP, 2012) will be applicable here. **PQ** = Quality Rating and Improvement Systems (QRIS) standards at the state level; process quality measures. **RA** = risk assessment tools/systems in which only the most critical rules/standards are measured. *Stepping Stones* (NRC, 2013) is an example of this approach**. KI** = key indicators in which only predictor rules/standards are measured. The *Thirteen Indicators of Quality Child Care* (Fiene, 2002) is an example of this approach. **DM** = differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol**. PD** = technical assistance/training and/or professional development system which provides targeted assistance to the program based upon the **DM** results. **CO** = child outcomes which assesses how well the children are developing which is the ultimate goal of the system.

Insert Figure 2

Once the above key elements are in place, it is then possible to look at the relationships (this is depicted by the arrows that go from one box to another) amongst them to determine if the system is operating as it was intended; in other words, to determine if the DM system is improving the health, safety, program quality and ultimately the overall development of the children it serves.

In the Methodology section, a scoring protocol (DMSP© - Differential Monitoring Scoring Protocol©) is introduced which attempts to quantify these relationships and to give us a means for making measurements and comparisons across various types of organizations.

The DMLM© provides a cross-cutting methodology that can be used in all child care/early care and education delivery systems as well as in other human services. In the past many of these monitoring systems have functioned in silos. The DMLM© integrates all these various monitoring systems together so that the overall monitoring system can be validated as being cost effective and efficient. This can be an important development as available funds become more scarce in the future as international organizations deal with fewer and fewer resources.

Methods

National/State/Provincial Agency Plan for implementing a Differential Monitoring System:

The **first step** in utilizing the DMLM© for a state/province/nation is to take a close look at its Comprehensive Licensing Tool (CI) that it uses to collect violation data on all rules with all facilities in its respective state/province/nation. If the state/province/nation does not utilize a tool or checklist or does not review all violation data than it needs to consider these changes because the DMLM© is based upon an Instrument Based Program Monitoring System (IPM)(Fiene & Nixon,1985) which utilizes tools/checklists to collect data on all rules.

The **second step** for the state/province/nation is to compare their nation's/state's/province's rules

with the National *Health and Safety Performance Standards (Caring for Our Children*)(AAP, 2012) or an equivalent international set of standards to determine the overlap and coverage between the two.

The **third step** for the state/province/nation if it utilizes a Risk Assessment (RA) tool is to assess the relationship between this tool and *Stepping Stones* (NRC, 2013) or an equivalent international set of targeted standards to determine the overlap and coverage between the two.

The **fourth step** for the state/province/nation is to compare the results from the CI with the RA tools.

In the **fifth step**, if a state/province/nation is fortunate enough to have a QRIS – Quality Rating and Improvement System in place and has sufficient program quality (PQ) data available then they will have the ability to compare results from their CI tool with their PQ tool and validate outputs by determining the relationship between compliance with health and safety rules (CI) and program quality (PQ) measures that measure process quality. This is a very important step because very few empirical demonstrations appear in the research literature regarding this relationship.

The **sixth step** is for the state/province/nation to generate a Key Indicator (KI) tool from the CI data base. Please see Fiene & Nixon (1985) and Fiene & Kroh (2000) for a detailed explanation

of the methodology for generating a KI tool. If a state/province/nation did not want to use the KI methodology, a direct comparison could be drawn from The *Thirteen Indicators of Quality Child Care* (Fiene, 2002).

The **seventh step** for the state/nation is to use the RA and KI tools together to determine overall compliance of facilities and how often and which rules will be monitored for future visits. This is the basic component of a Differential Monitoring (DM) approach. Also, this step should drive decisions within the technical assistance/training/professional development (PD) system in what resources are allocated to a particular facility.

The **eighth and final step** for the state/nation is to compare the results from the various monitoring tools (CI, PQ, RA, KI) with any child development outcome (CO) data they collect. This is a relatively new area and few, if any, states/provinces/nations at this point have this capability on a large scale. However, as Early Learning Networks/Systems and Standards (ELS) are developed, this will become more common place.

The ECPQIM^4©DMLM© is presented without two additional items that were present in the 2012/2013 versions which are important to note. The algorithm (Fiene, 2012, 1013) and validation framework (Zellman & Fiene, 2012) are not presented because the author felt that these two components took away from a more direct presentation of differential monitoring. For those interested readers, please refer to my previous abstracts (Fiene, 2012, 2013) which

included the algorithm and validation frameworks.

Just another brief word about the Theoretical Underpinnings for ECPQIM^4. This graphic (Figure 1) attempts to provide the relationships amongst public policy, interventions, and empirical evidence through the lens of translational research, implementation science, and program monitoring. In constructing the ECPQIM^4 concepts were borrowed from each area and integrated them in a model for monitoring early care and education programs. The graphic provides a means for displaying the relationships and potential intersections as well as the content that is important to each scientific/research field.

Figure 3 is provided as additional information regarding differential monitoring conceptually without all the details as in figure 2; and figure 4 is provided to demonstrate the impact that a state's/provincial/national licensing law can have on using the Key Indicators and Risk Assessment methodologies.

Insert Figures 3 & 4

Also, taking Figure 2 and attempting to quantify these relationships, a scoring protocol is proposed as depicted in Table 1. This can provide a numerical means of comparing various

differential monitoring systems and their relative comprehensiveness. This protocol could be a useful tool in future research for determining which combinations work best.

Insert Table 1

The next section provides the results from a national organization and five states who used the above methodology to implement their respective differential monitoring systems.

Results and Discussion

The Early Childhood Program Quality Indicator Model (ECPQIM©) and its latest iteration presented as a logic model: Differential Monitoring Logic Model (DMLM©) have been written about extensively by this author (Fiene & Nixon, 1985; Griffin & Fiene, 1996; Fiene & Kroh, 2000; Fiene, 2013). Several states and Head Start have used the model in order to re-align their program monitoring systems. This paper presents the results of those new program monitoring systems through the lenses of the ECPQIM©/DMLM© logic model display. Each particular approach used various components of the overall comprehensive national model and have been highlighted by connecting arrows. It is proposed that this approach could be applied at an international level as well**.**

The interested reader should obtain a copy of the Office of Child Care's *Licensing Brief on Differential Monitoring, Risk Assessment, and Key Indicators* published by the National Center on Child Care Quality Improvements which gives additional details regarding these approaches and methodologies as well as other state examples. Please go to the following URL website: (https://childcareta.acf.hhs.gov/sites/default/files/1408_differential_monitoring_final_1.pdf). In fact, this paper builds upon that excellent *Licensing Brief*.

Let's start with Figure 5 which provides the Comprehensive National Example that depicts all the possible interconnections and gives national examples from the research literature. As one will see, it is possible for a national organization or a state/provincial agency to select the various components from the model based upon what is available in their particular organization. All do have the program compliance/licensing component (PC) but not all have fully functional program quality initiatives (PQ) or do not have the data to draw from the program quality initiatives.

The next level of components are the key indicator (KI) and risk assessment (RA) approaches or methodologies which organizations or state agencies can use alone or in tandem. One limitation in the key indicator methodology is not to use it with program initiatives if the data are not severely skewed in their data distribution as is the case with licensing data.

The last component is the resulting differential monitoring (DM) approach based upon the results

from using the key indicator and risk assessment methodologies either alone or in tandem. This is the ultimate revision of the program monitoring system in which how often and what is reviewed are answered.

All the components are highlighted (this is indicated by the arrows going from one box to another) in Figure 5 because all are possibilities to be used by a national or state agency. The examples in Figure 5 are drawn from the national research literature so *Caring for Our Children* (AAP, 2012) is the example for Program Compliance, Licensing, and the Health & Safety Comprehensive Instrument (CI). The following examples in Figures 6-11 will show some differences in how national and state agencies have developed their respective differential monitoring systems through their use of key indicator (KI) and risk assessment (RA) methodologies, and linking their licensing/program compliance (PC) and program quality (PQ) initiatives. Tables 1-3 explain the scoring protocol and provide results from the national Head Start program and five states geographically dispersed around the USA (New York, Georgia, Illinois, Kansas, and Colorado). Also see the end of the paper for an explanation of Notes a,b,c in Figure 5.

Insert Figure 5

Figure 6 provides an example from New York (NY) where the state agency is attempting to restructure their early care and education program monitoring system to have a better balance between licensing and key program quality indicators. The plan is to have licensing staff collect data from both areas which means a need to save time in the licensing reviews via key indicators and to only identify indicators of quality through a risk assessment approach. The results from these two methodologies will then be combined into a Quality Indicators Instrument to be used by licensing staff in their annual reviews.

Insert Figure 6

Figure 7 provides an example from Georgia (GA) in which the driving methodology is a risk assessment core rule review system that results in a differential monitoring system called the Annual Compliance Determination Worksheet (ACDW) approach. Key indicators are not used directly but were used as part of the risk assessment core rule development. Please note how the relationship amongst the various components is different from the NY approach delineated in Figure 6. There is a link to their program quality initiatives which proved very significant in the validation studies performed on their Core Rule differential monitoring system.

Insert Figure 7

Figure 8 presents a very different approach from the previous two approaches. In Kansas's (KS) case, the state agency was only interested in developing a key indicator approach and was not interested in risk assessment nor had the capability to tie data together from their program quality initiatives. This is noted by the arrow connections which is more minimal in this depiction. As one can see, this still is a viable option for developing a differential monitoring approach.

Insert Figure 8

Figure 9 depicts the use of both key indicator and risk assessment methodologies in Illinois (IL) with their licensing system but no data interaction with their program quality initiatives. It is proposed that both methodologies will be used together in future licensing reviews of programs which will constitute their differential monitoring system approach.

Insert Figure 9

Figure 10 depicts the new aligned differential monitoring system being employed in Head Start (HS). Head Start has a very comprehensive system that employs various aspects from all the components in their system. The Head Start Performance Standards are very comprehensive, CLASS is used as a major process quality measure and both a key indicator (Head Start Key Indicator – Compliance (HSKI-C)) and risk assessment (Selected Compliance Measures) are utilized in their program monitoring system. The Head Start new Aligned Program Monitoring system comes closest to the comprehensive national model.

Insert Figure 10

In Figure 11 a very different scenario played out in the state of Colorado (CO) in which key indicators were developed for their QRIS system rather than for their licensing system. As mentioned earlier, when applying the key indicator methodology to Quality Initiatives one needs to be very cautious if the data distribution is not exceptionally skewed as is the case with licensing data. Some of the data were sufficiently skewed to be able to be used in generating

quality key indicators but there were limitations noted.

Insert Figure 11

The above results clearly demonstrate how agencies can take very different approaches to designing and implementing their differential monitoring system. The next research question is to determine if agencies that have higher scores (more than 6) if they are more effective and efficient than those agencies that have lower scores (less than 5).

Conclusion

This paper presents the latest examples of national and state agencies differential monitoring approaches. It clearly demonstrates that there are many different approaches to developing and implementing differential monitoring. A key research question for the future as more states utilize the different approaches is to study if one approach is better than the next or a combination works better than most. From 40+ years of experience as a researcher and state policy analyst I would suggest that a more comprehensive approach which employs the full menu of program quality initiatives similar to the Head Start or the New York approaches will be most effective.

As mentioned in the introduction of this paper in describing the Comprehensive National Example of the DMLM© Model Tables 1-3 present a Differential Monitoring Scoring Protocol (DMSP©) that can potentially be used to compare states on how in depth their differential monitoring system is. Table 1 describes the DMSP© in narrative terms delineating the various systems that need to be in place in order to get a particular score. A score of 0 means no systems are in place or do not intersect while a score of 10 means that all of the systems are in place and intersect or are linked. Table 2 gives the points assigned to the specific systems that are part of a differential monitoring system. And Table 3/Figure 12 give the actual points assigned to the state & national examples that have been presented in this paper for ***New York (NY), Georgia (GA), Head Start (HS), Kansas (KS), Illinois (IL), and Colorado (CO)***. The total points assigned to the comprehensive model are also provided as a point of context.

There are a couple of important things to note about the DMSP© in Table 2, such as: if Key Indicators (KI) and Risk Assessment (RA) are linked, it negates KI and RA being scored separately. If KI and RA are developed separately, it is very improbable that they will not be linked but that is always a possibility, so it is listed as so. Linking Program Compliance/Licensing (PC) and Program Quality (PQ) Initiatives is a highly desirable event and is assigned a high score (4 points). Linking KI and RA is also considered a highly desirable event and is assigned a high score (4 points).

Insert Tables 2 & 3 and Figure 12

For future research, it will be interesting to see if this ECPQIM^4©/DMLM© model has applicability from an international perspective. Some of the key elements present in USA state systems are organized very differently in other countries and would have to be adjusted. Also, it will be interesting to see if the DMSP© can be developed as a scoring systems similar to the Child Care Aware Report Card Benchmarks protocol where it will be possible to make comparisons across state and national agencies.

Endnotes a, b, c:

The arrows going from Key Indicators (KI) and Risk Assessment (RA) to Differential Monitoring (DM) can be configured in the following ways: only KI (Kansas); only RA (don't have an example of this as of this writing) or a combination of KI and RA (Illinois) but this configuration could mean all of the KI and RA rules which would be more rules than if only KI or RA rules were selected or only those rules that overlap (KI+RA) which would be a much reduced number of rules. Or a different configuration determined by the state agency.

SENDING00: ECPQIM – DMLM – ICEP1d1 (2)aC RIKI HF

References

AAP (2012). *National health and safety performance standards*, Evanston, IL: American Academy of Pediatrics.

Fiene (2013***).*** A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks, *International Journal of Child Care and Education Policy, 7(1),* 1-15.

Fiene (2012/2013). *Early childhood program quality indicator model (ECPQIM⁴©): Differential monitoring logic model and algorithm (DMLMA©),* Middletown, PA: Research Institute for Key Indicators *LLC* (RIKI).

Fiene (2002). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.

Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.

Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, *14*(3), 198-214.

Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.

NRC (2013). *Stepping stones to caring for our children*. Denver, Colorado: National Resource Center for Health and Safety in Child Care.

Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care,* Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Figure 1

The Theoretical Underpinnings for ECPQIM⁴: Early Childhood Program Quality Indicator Model©

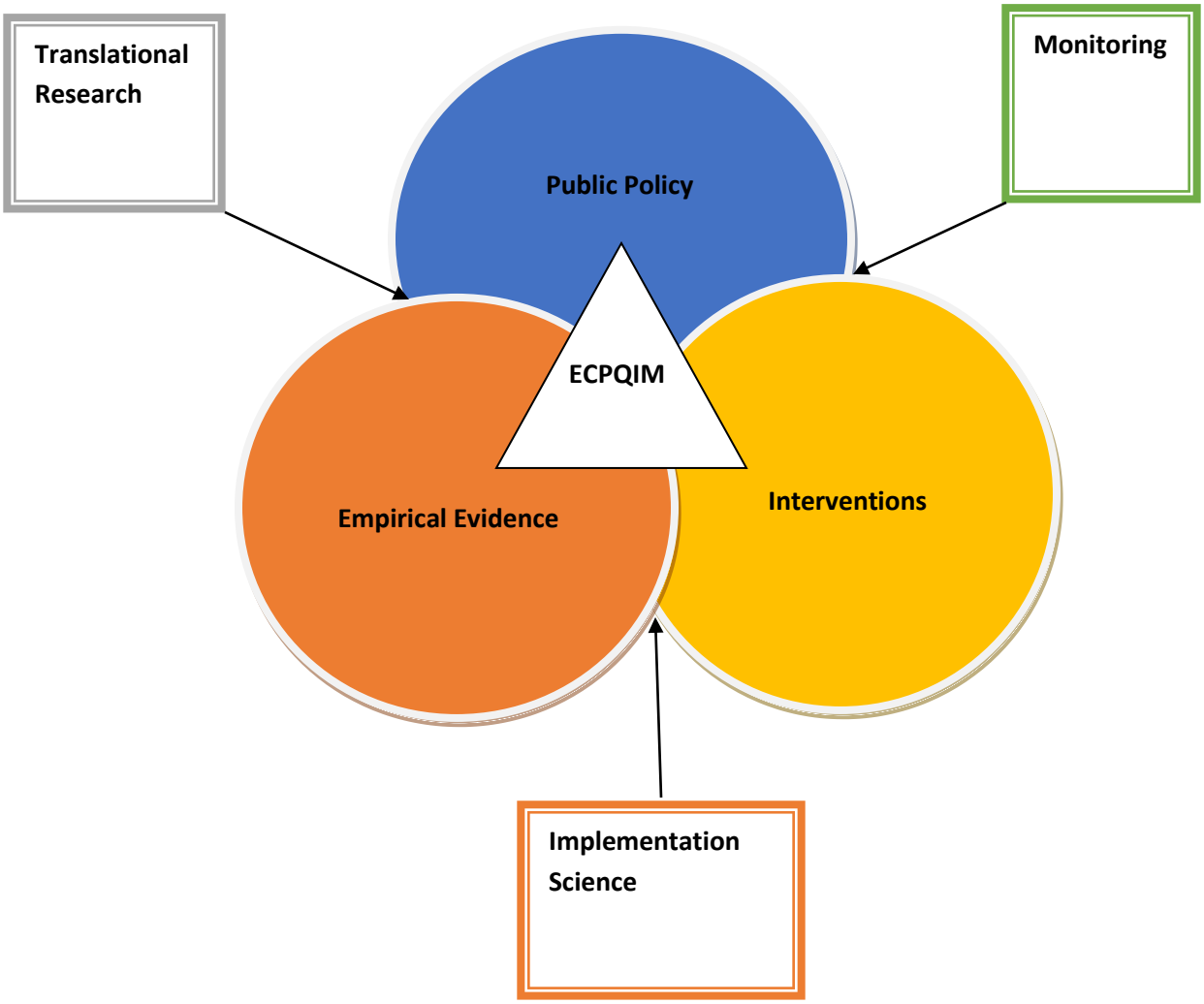


Figure 2

Early Childhood Program Quality Indicator Model (ECPQIM⁴©): Differential Monitoring Logic Model (DMLM©) Comprehensive National Example

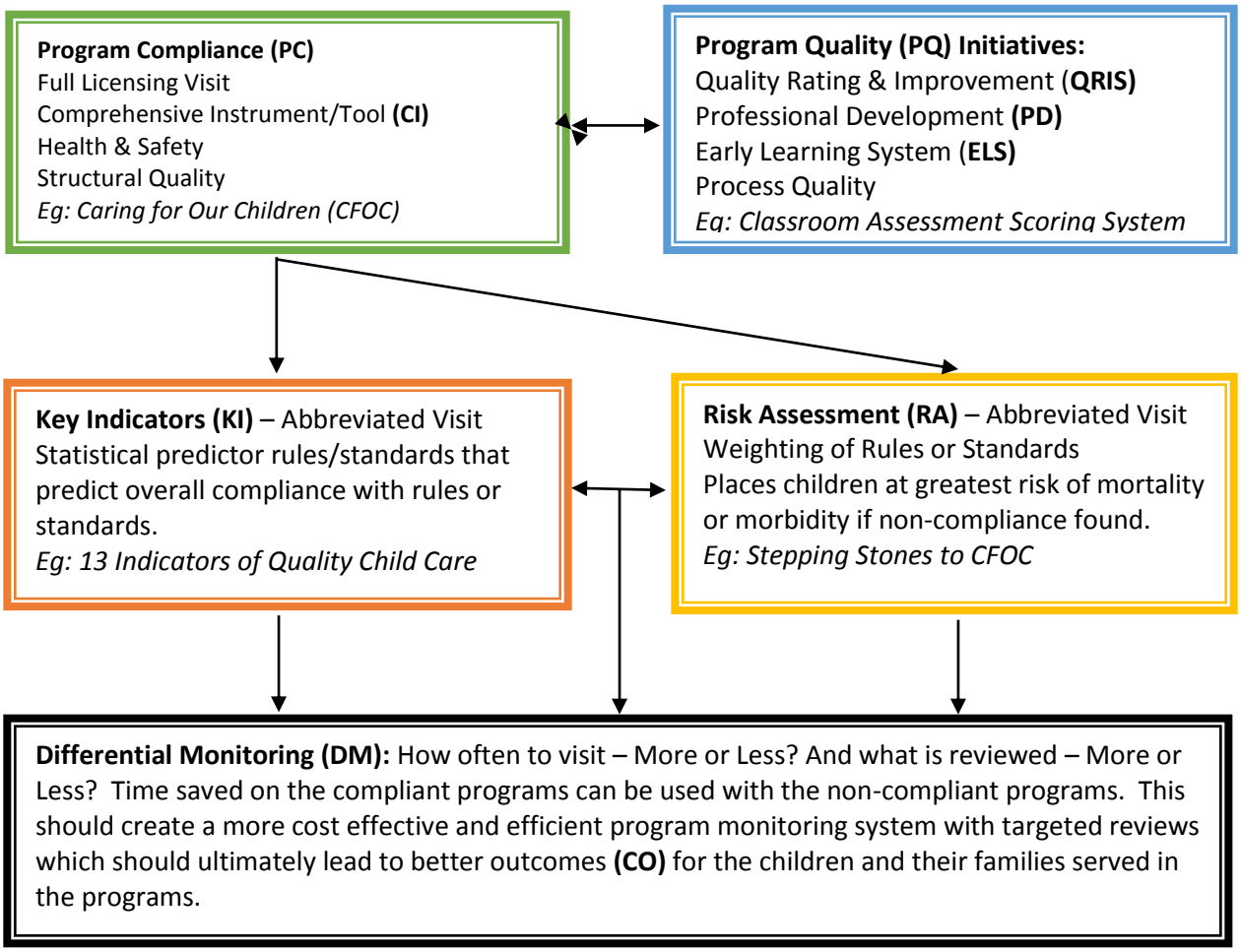


Figure 3

Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

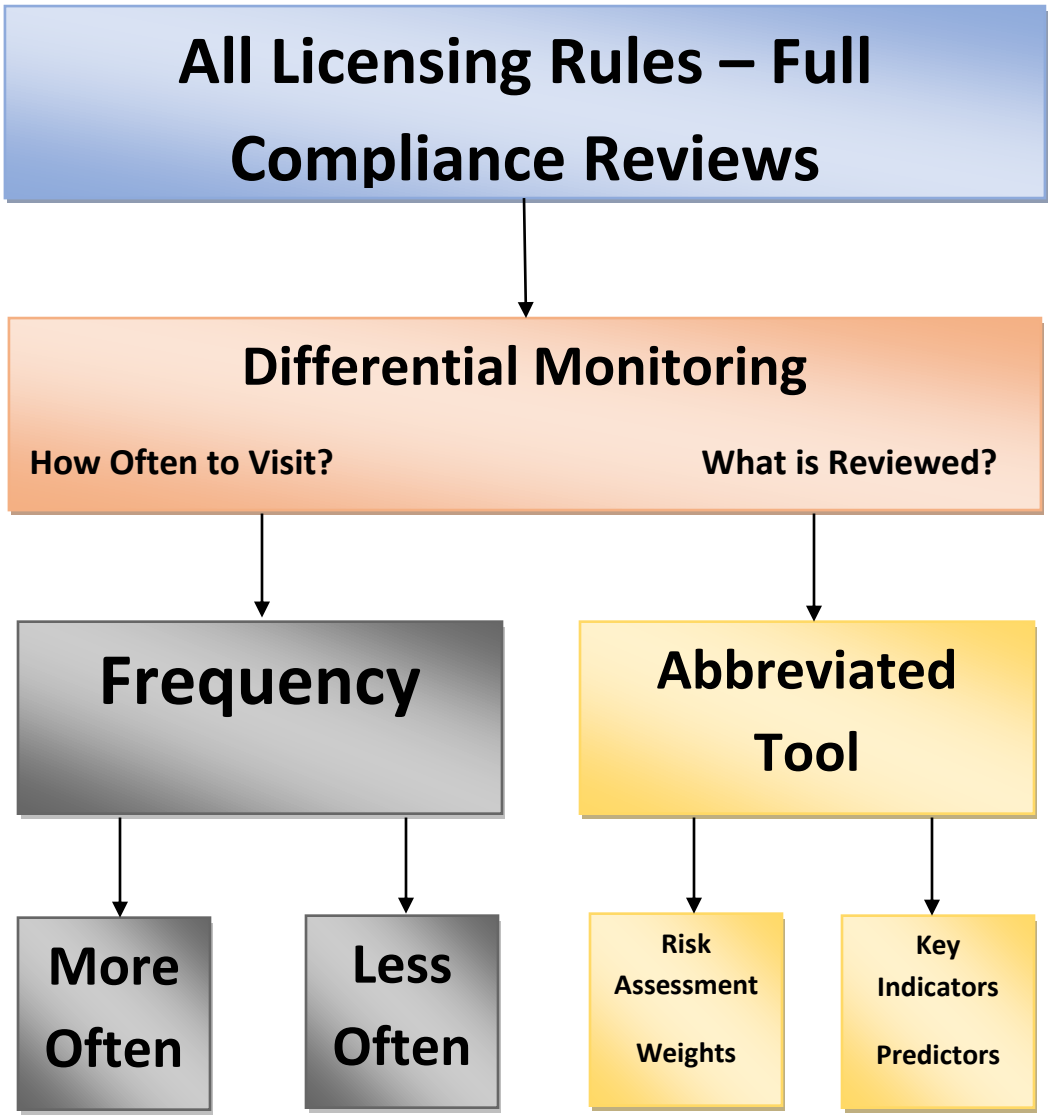
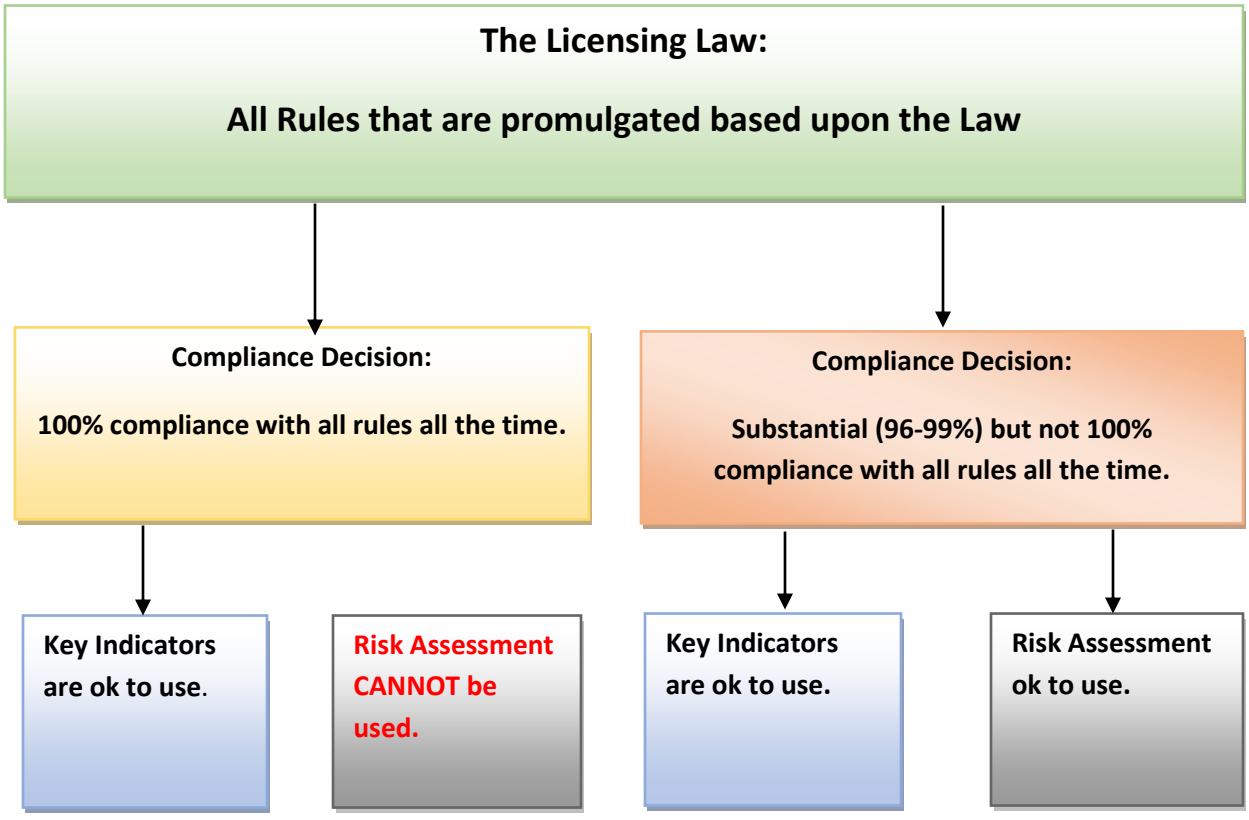
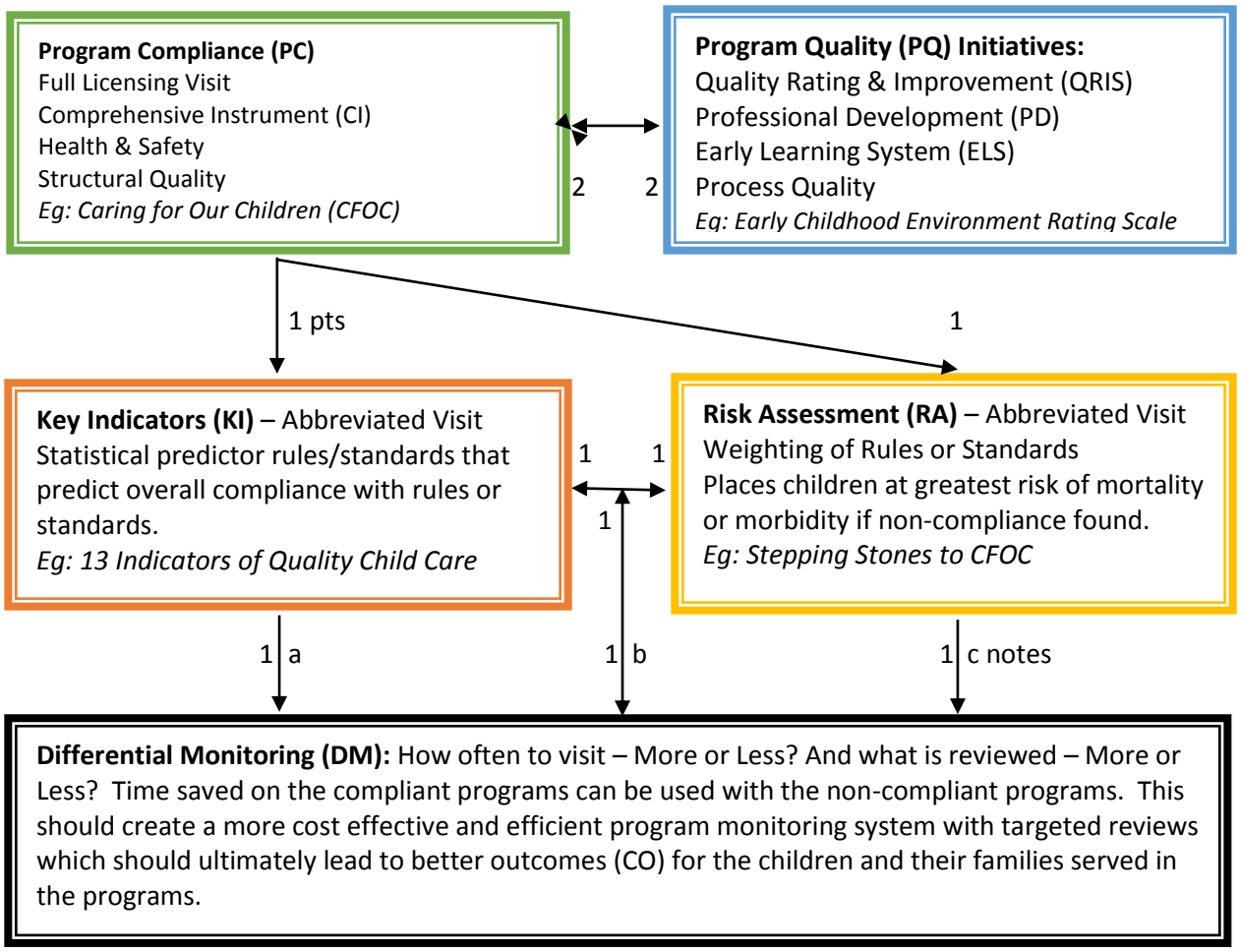


Figure 4

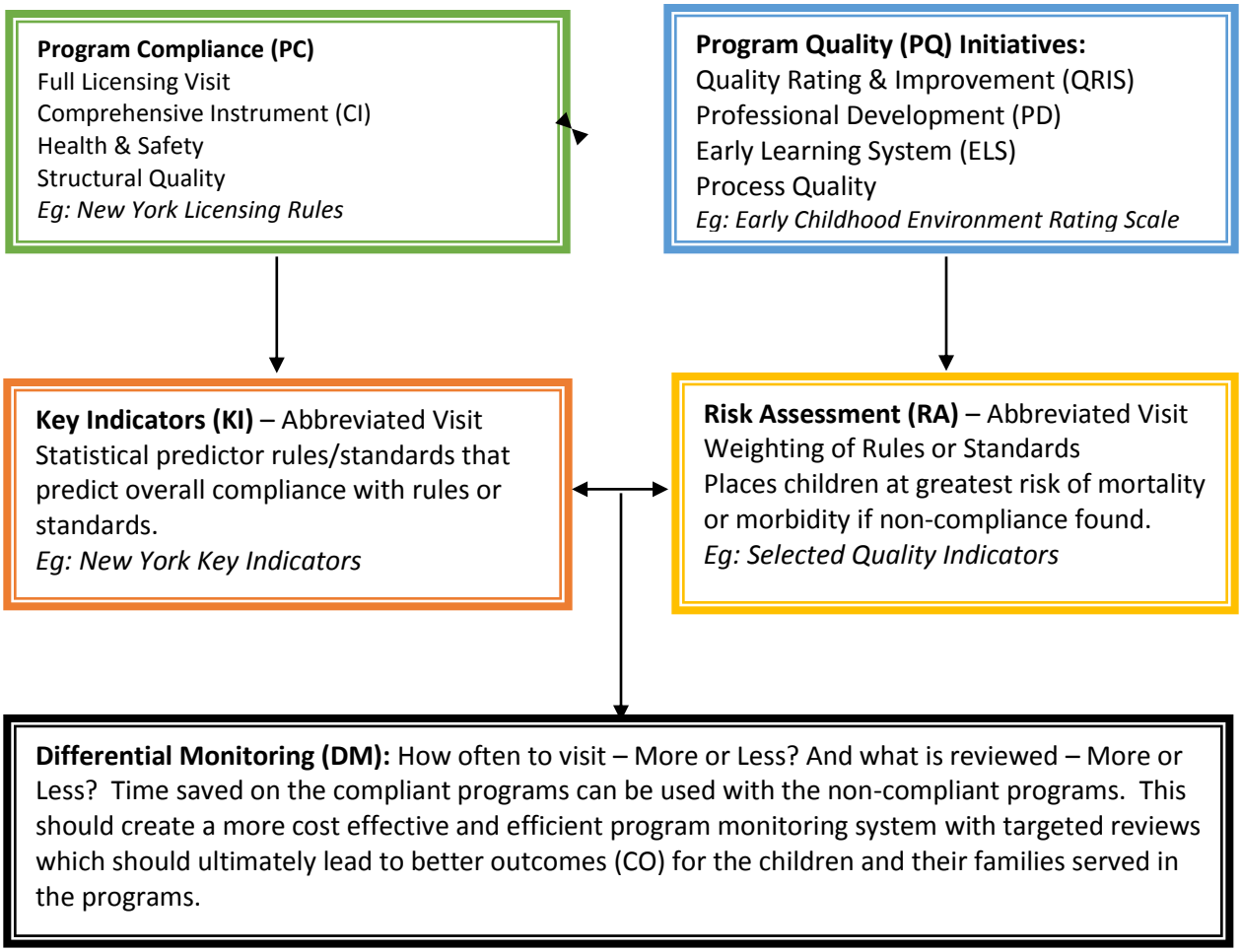
When Key Indicators and Risk Assessments Can Be Used



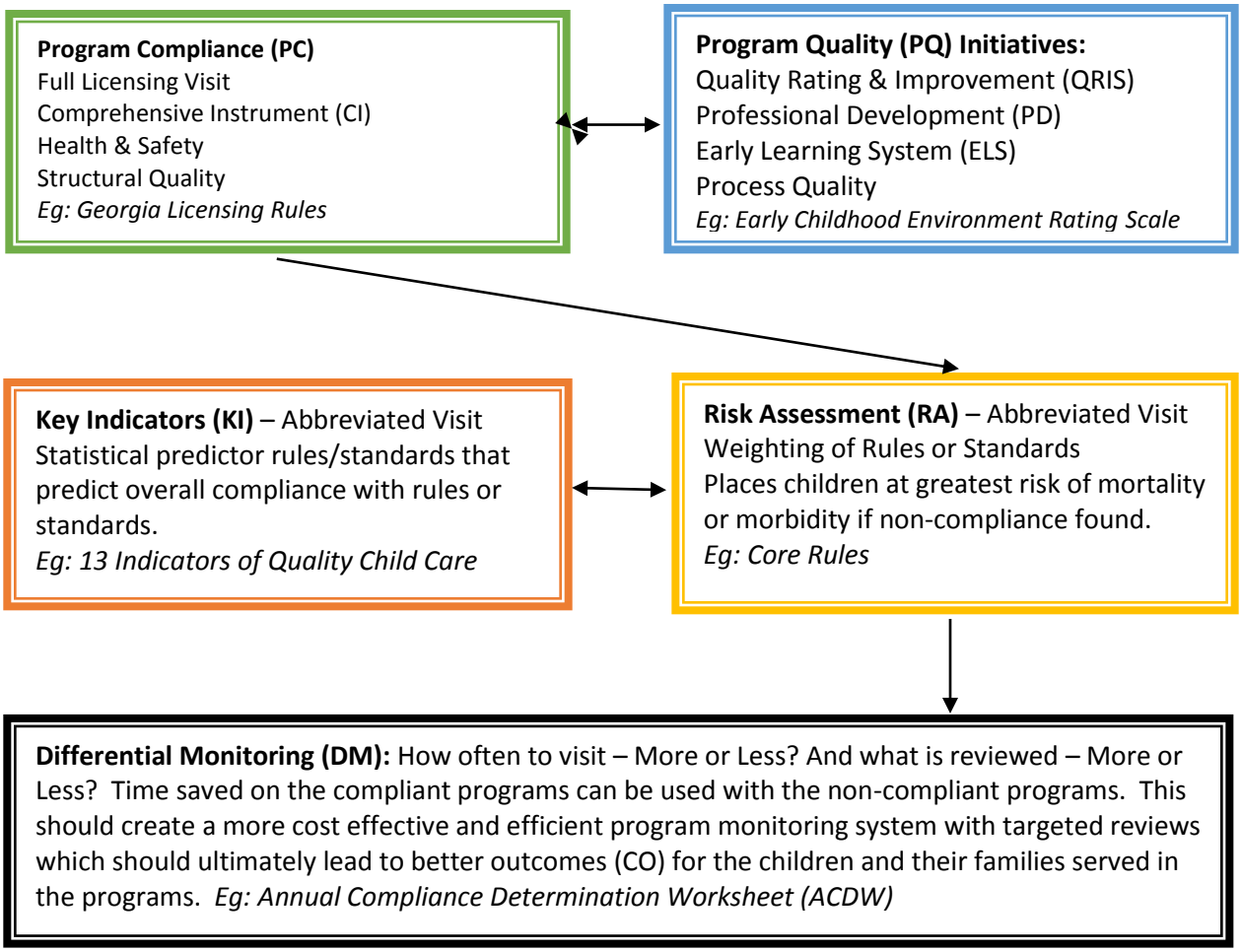
Early Childhood Program Quality Indicator Model (ECPQIM4©): Differential Monitoring Logic Model (DMLM©) Comprehensive National Scoring Protocol Example (Maximum of 10 Points)
Figure 5



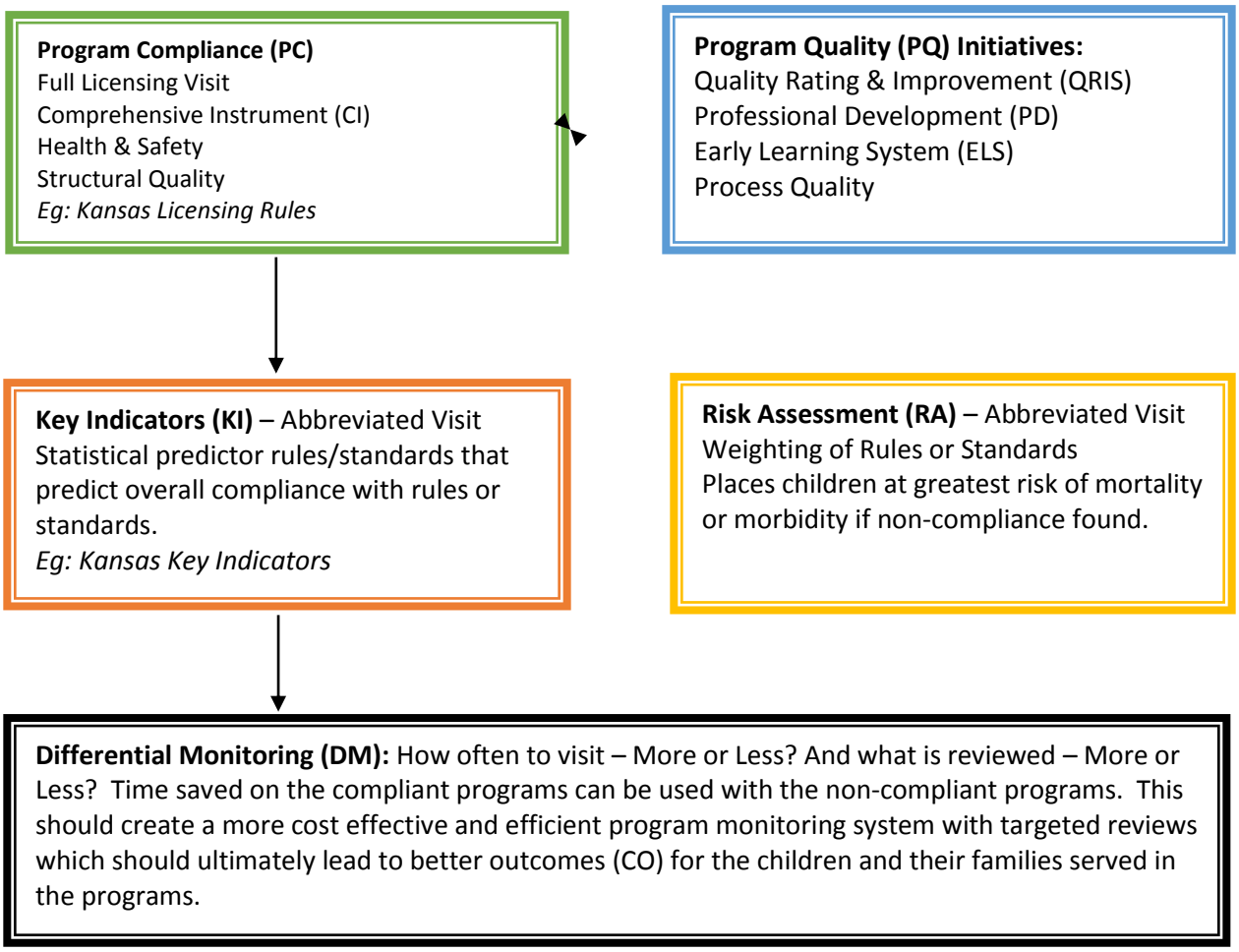
(ECPQIM4©)(DMLM©): New York Example (NY)
Figure 6



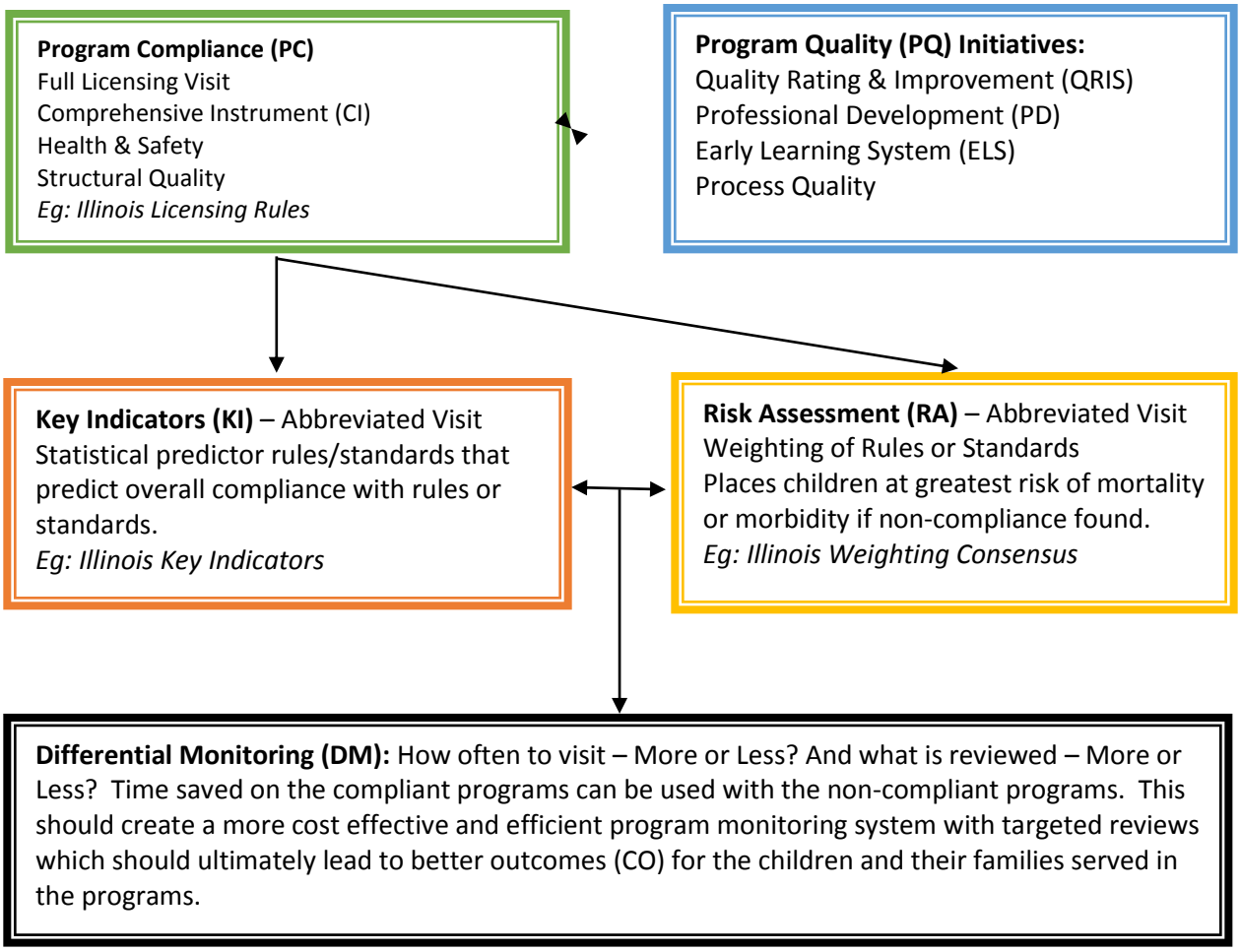
(ECPQIM4©)(DMLM©): Georgia Example (GA)
Figure 7



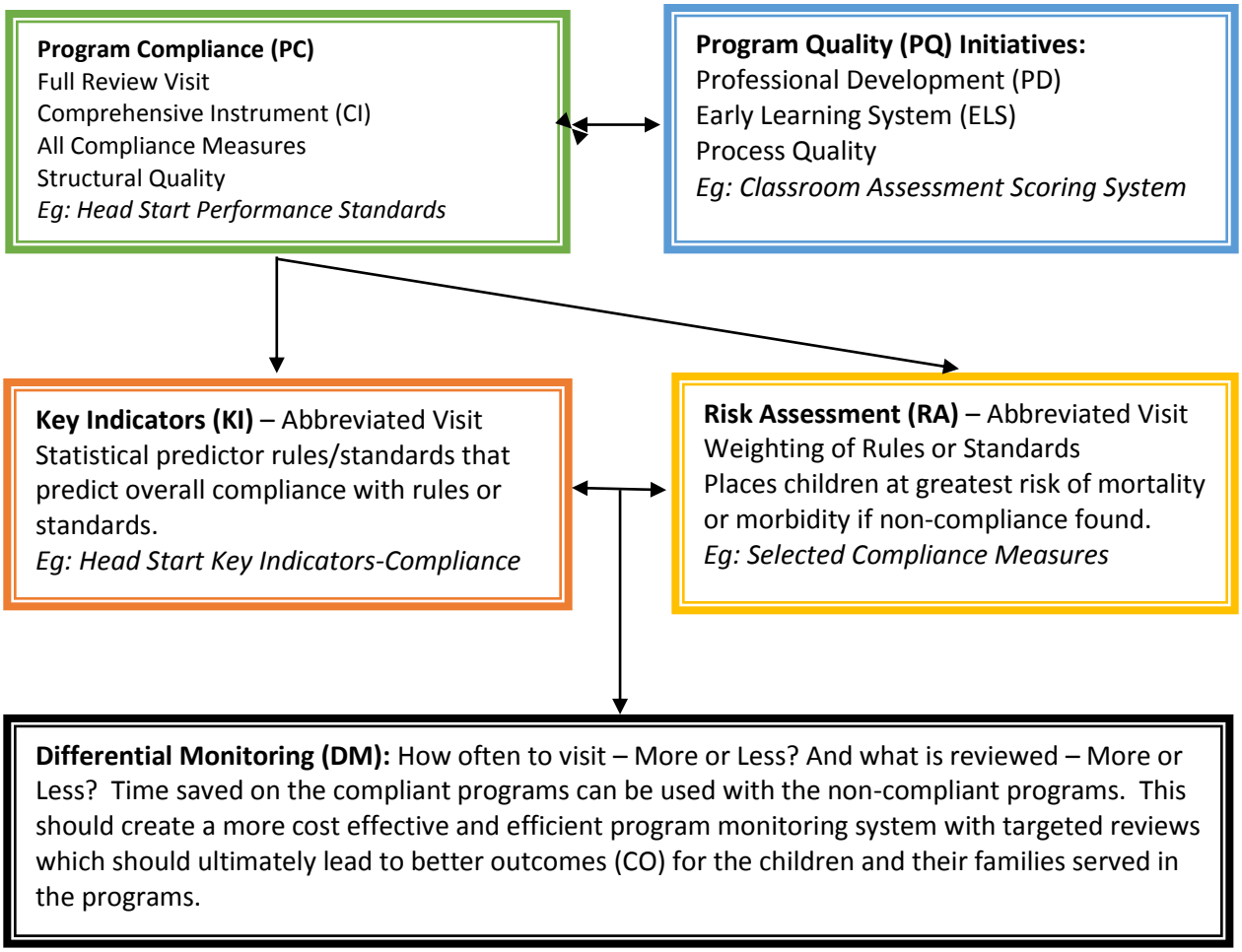
(ECPQIM4©)(DMLM©): Kansas Example (KS)
Figure 8



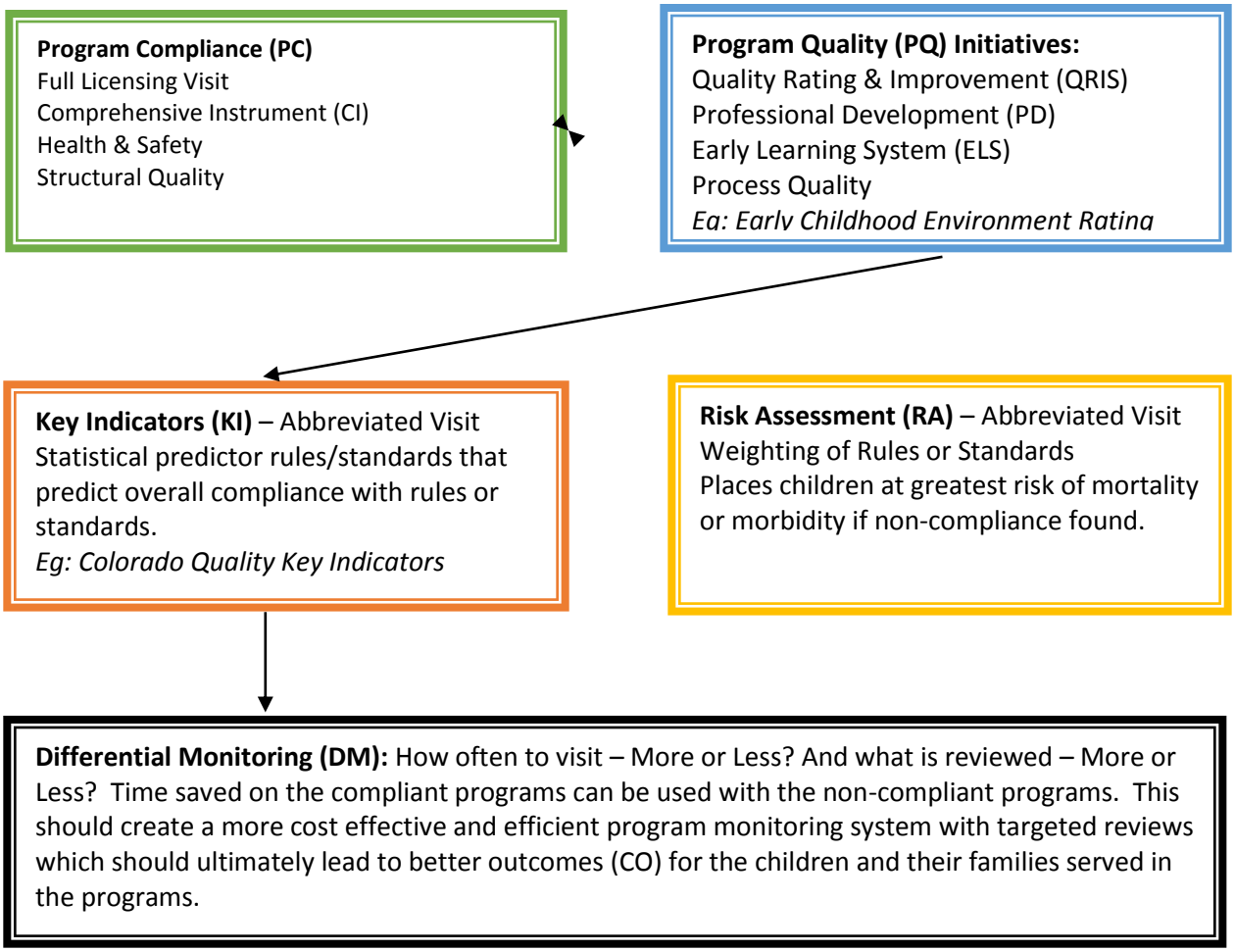
(ECPQIM4©)(DMLM©): Illinois Example (IL)
Figure 9



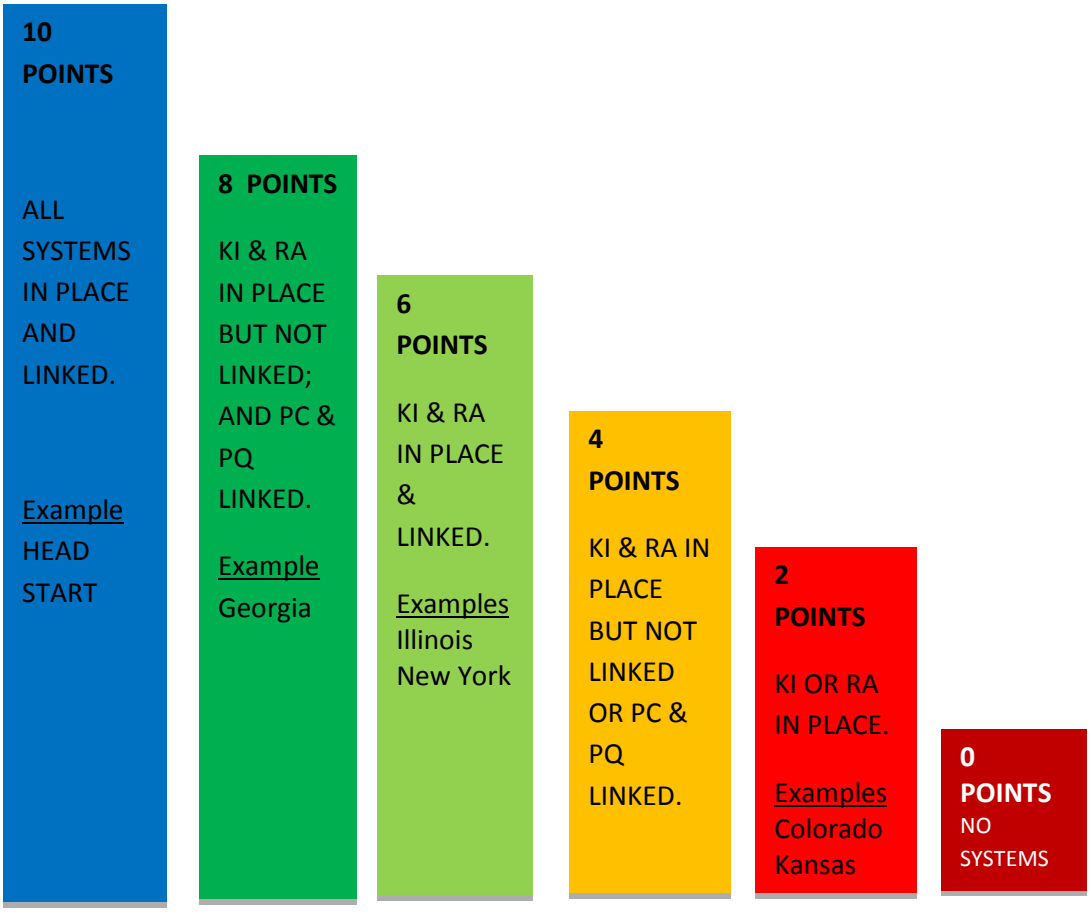
(ECPQIM4©)(DMLM©): Head Start Example (HS)
Figure 10



(ECPQIM4©)(DMLM©): Colorado Example (CO)
Figure 11



DMSP© SCORING PROTOCOL WITH STATE AND NATIONAL AGENCIES AS EXAMPLES
Figure 12



SCORING

KI = Key Indicators; RA = Risk Assessment; PC = Licensing; PQ = Program Quality Initiatives

Table 1: Differential Monitoring Scoring Protocol (DMSP)©

Score	*Systems Present*
0	**No systems in place.**
2	**KI or RA in place and not linked.**
4	**(KI & RA in place but not linked) or (PC + PQ are linked).**
6	**(KI & RA in place) & (KI + RA are linked)**
8	**(KI & RA in place but not linked) & ((PC + PQ) are linked).**
10	**All systems in place and linked.**

KI (Key Indicators); RA (Risk Assessment); PC (Program Compliance/Licensing); PQ (Program Quality Initiatives)

Table 2: Differential Monitoring Scoring Protocol (DMSP)© Point Assignment

Score	*Systems Present and Point Assignment*
0	**No systems in place.**
2	**(KI (1)) & (KI -> DM (1)) or ((RA (1)) & (RA -> DM (1))**
4	**(PC + PQ (4)) or (KI (1) & (KI -> DM (1)) & (RA (1) & (RA -> DM (1))**
6	**(KI + RA -> DM (4)) & (KI (1)) & (RA (1))**
8	**(KI (2) & RA (2)) & (PC + PQ (4)).**
10	**(KI + RA -> DM (4)) & (KI (1)) & (RA (1)) & (PC + PQ (4))**

KI (Key Indicators); RA (Risk Assessment); PC (Program Compliance/Licensing); PQ (Program Quality Initiatives)

Table 3: DMLM© SCORING PROTOCOL WITH STATE EXAMPLES

SYSTEMS (pts)	*MODEL*	*GA*	*NY*	*HS*	*IL*	*KS*	*CO*
KI (1)	*1*	-	1	1	1	1	1
RA (1)	*1*	1	1	1	1	-	-
KI + RA -> DM (4) *KI + RA (2)*	*4*	2	4	4	4	-	-
PC + PQ (4)	*4*	4	-	4	-	-	-
KI -> DM (1)	-	-	-	-	-	1	1
RA -> DM (1)	-	1	-	-	-	-	-
TOTAL (10)	*10*	8	6	10	6	2	2

GA (Georgia); NY (New York); HS (Head Start); IL (Illinois), KS (Kansas); CO (Colorado)

Regulatory Compliance Monitoring Paradigms and the Relationship of Regulatory Compliance/Licensing with Program Quality: A Policy Commentary

Richard Fiene, PhD [a,*]

[a] *Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University*

Abstract

This policy commentary deals with two key issues within regulatory science related to the best methods for measuring regulatory compliance: Program monitoring paradigms and the relationship of regulatory compliance/licensing with program quality. Examples from program monitoring paradigms include: 1) Substantial versus Monolithic. 2) Differential Monitoring versus One size fits all monitoring. 3) "Not all standards are created equal" versus "All standards are created equal". 4) "Do things well" versus "Do no harm". 5) Strength based versus Deficit based. 6) Formative versus Summative. 7) Program Quality versus Program Compliance. 8) 100-0 scoring versus 100 or 0 scoring. 9) QRIS versus Licensing. 10) Non-Linear versus Linear. Examples from the relationship of regulatory compliance/licensing with program quality include: 1) "Do no harm" versus "Do good". 2) Closed system versus Open system. 3) Rules versus Indicators. 4) Nominal versus Ordinal measurement. 5) Full versus Partial compliance. 6) Ceiling effect versus No Ceiling effect. 7) Gatekeeper versus Enabler. 8) Risk versus Performance.

Keywords: regulatory compliance, program monitoring, licensing, program quality.

Introduction

This commentary on policy will deal with two key issues within regulatory science that need to be dealt with by licensing researchers and regulatory scientists as they think through the best methods for measuring regulatory compliance: 1) Program monitoring paradigms; 2) Relationship of regulatory compliance/licensing and program quality. The examples drawn are from early childcare and education but the key elements and implications can be applied to any field of study related to regulatory science that involves rules/regulations/standards. For the purposes of this manuscript "rules" will be used to

* Corresponding author: Richard Fiene; Email: rjf8@psu.edu;
Phone: 717-598-8908; ORCID: http://ORCID: 0000-0001- 6095-5085.

describe or refer to "rules/regulations/standards".

Program Monitoring Paradigms:

This section provides some key elements to two potential regulatory compliance monitoring paradigms (Differential/Relative versus Absolute/Full) for regulatory science based upon the Regulatory Compliance Theory of Diminishing Returns (Fiene, 2019).

As one will see, there is a need within regulatory science to get at the key measurement issues and essence of what is meant by regulatory compliance. There are some general principles that need to be dealt with such as the differences between individual rules and rules in the aggregate. Rules in the aggregate are not equal to the sum of all rules because all rules are not created nor administered equally. And all rules are to be adhered to, but there are certain rules that are more important than others and need to be adhered to all the time. Less important rules can be in substantial compliance most of the time but important rules must be in full compliance all of the time (Fiene, 2019).

Rules are everywhere. They are part of the human services landscape, economics, banking, sports, religion, transportation, housing, etc... Wherever one looks we are governed by rules in one form or another. ***The key is determining an effective and efficient modality for negotiating the path of least resistance in complying with a given set of rules***[2]. It is never about more or less rules, it is about which rules are really productive and which are not. Too many rules stifle creativity, but too few rules lead to chaos. ***Determining the balance of rules is the goal and solution of any regulatory science paradigm.***

Differential/Relative versus Absolute/Full Regulatory Compliance Paradigms: this is an important key organizational element in how rules are viewed when it comes to compliance. For example, in an absolute/full approach to regulatory compliance either a rule is in full compliance or not in full compliance. There is no middle ground. It is black or white, no shades of gray as are the cases in a differential/relative paradigm. It is 100% or zero. In defining and viewing these two paradigms, this dichotomy is the organizational key element for this paper. In a differential/relative regulatory compliance paradigm full compliance is not required and emphasis on substantial regulatory compliance becomes the norm.

Based upon this distinction between differential/relative and absolute/full regulatory compliance paradigms, what are some of the implications in utilizing these two respective approaches. Listed below are the basic implications that occur when selecting either of the two approaches on program monitoring systems: differential/relative versus absolute/full regulatory compliance paradigms.

There are ten basic implications that will be addressed: 1) Substantial versus Monolithic. 2) Differential Monitoring versus One size fits all monitoring. 3) "Not all standards are created equal" versus "All standards are created equal". 4) "Do things well" versus "Do no harm". 5) Strength based versus Deficit based. 6) Formative versus Summative. 7) Program Quality versus Program Compliance. 8) 100-0 scoring versus 100 or 0 scoring. 9) QRIS versus Licensing. 10) Non-Linear versus Linear.

1) Substantial versus Monolithic: in monolithic regulatory compliance monitoring systems, it is one size fits all, everyone gets the same type of review (this is addressed in the next key element below) and is more typical of an absolute paradigm orientation. In a substantial regulatory compliance monitoring system, programs are monitored on the basis of their past compliance history and this is more typical of a relative paradigm orientation. Those with high compliance may have fewer and more abbreviated visits/reviews while those with low compliance have more comprehensive visits/reviews.

2) Differential Monitoring versus One Size Fits All Monitoring: how does this actually look in a program monitoring system. In differential monitoring (Differential/Relative Paradigm), more targeted or focused visits are utilized spending more time and resources with those problem programs and less time and resources with those programs that are exceptional. In the One Size Fits All Monitoring (Absolute/Full Paradigm), all programs get the same type/level of review/visit regardless of past performance.

3) “Not all standards are created equal” versus “All standards are created equal”: when looking at standards/rules/regulations it is clear that certain ones have more of an impact on outcomes than others. For example, not having a form signed versus having proper supervision of clients demonstrates this difference. It could be argued that supervision is much more important to the health and safety of clients than if a form isn’t signed by a loved one. In a differential/relative paradigm, all standards are not created nor administered equally; while in an absolute/full paradigm of regulatory compliance, the standards are considered created equally and administered equally.

4) “Do things well” versus “Do no harm” (this element is dealt with in the second component to this paper below as well): “doing things well” (Differential/Relative Paradigm) focuses on quality of services rather than “doing no harm” (Absolute/Full Paradigm) which focuses on protecting health and safety. Both are important in any regulatory compliance monitoring system but a balance between the two needs to be found. Erring on one side of the equation or the other is not in the best interest of client outcomes. "Doing no harm" focus is on the "least common denominator" – the design and implementation of a monitoring system from the perspective of focusing on only 5% of the non-optimal programs ("doing no harm") rather than the 95% of the programs that are "doing things well".

5) Strength based versus Deficit based: in a strength-based monitoring system, one looks at the glass as “half full” rather than as “half empty” (deficit-based monitoring system). Emphasis is on what the programs are doing correctly rather than their non-compliance with standards. A strength-based system is non-punitive and is not interested in catching programs not doing well. It is about exemplars, about excellent models where everyone is brought up to a new higher level of quality care.

6) Formative versus Summative: differential/relative regulatory compliance monitoring systems are formative in nature where there is an emphasis on constant quality improvement and getting better. In absolute/full regulatory compliance monitoring systems, the emphasis is on being the gate-keeper (more about the gate-keeper function in

the next section on regulatory compliance/licensing and program quality) and making sure that decisions can be made to either grant or deny a license to operate. It is about keeping non-optimal programs from operating.

7) Program Quality versus Program Compliance: (this element is dealt with in greater detail in the second component of this manuscript) differential/relative regulatory compliance monitoring systems focus is on program quality and quality improvement while in absolute/full regulatory compliance monitoring systems the focus in on program compliance with rules/regulations with the emphasis on full, 100% compliance.

8) "100 – 0 scoring" versus "100 or 0 scoring": in a differential/relative regulatory compliance monitoring system, a 100 through zero (0) scoring can be used where there are gradients in the scoring, such as partial compliance scores. In an absolute/full regulatory compliance monitoring system, a 100% or zero (0) scoring is used demonstrating that either the standard/rule/regulation is fully complied with or not complied with at all (the differences between nominal and ordinal measurement is dealt with in the next section on regulatory compliance/licensing and program quality).

9) QRIS versus Licensing: examples of a differential/relative regulatory compliance monitoring system would be QRIS – Quality Rating and Improvement Systems. Absolute/full regulatory compliance systems would be state licensing systems. Many programs talk about the punitive aspects of the present human services licensing and monitoring system and its lack of focus on the program quality aspects in local programs. One should not be surprised by this because in any regulatory compliance system the focus is on "doing no harm" rather than "doing things well". It has been and continues to be the focus of licensing and regulations in the USA. The reason QRIS - Quality Rating and Improvement Systems developed in early care and education was to focus more on "doing things well" rather than "doing no harm". This is not the case in many Canadian Provinces and European countries in which they have incorporated program quality along with specific regulatory requirements.

10) Non-Linear versus Linear: the assumption in both differential/relative and absolute/full regulatory compliance monitoring systems is that the data are linear in nature which means that as compliance with rules increases positive outcomes for clients increases as well. The problem is the empirical data does not support this conclusion. It appears from the data that the relationship is more non-linear where there is a plateau effect with regulatory compliance in which client outcomes increase until substantial compliance is reached but doesn't continue to increase beyond this level. There appears to be a "sweet spot" or balancing of key rules that predict client outcomes more effectively than 100% or full compliance with all rules – this is the essence of the Theory of Regulatory Compliance (Fiene, 2019) – substantial compliance with all standards or full compliance with a select group of standards that predict overall substantial compliance and/or positive client outcomes.

As the regulatory science and administrative fields in general continue to think about the appropriate monitoring systems to be designed and implemented, the above structure should

help in thinking through what these measurement systems' key elements should be. Both paradigms are important, contexts, but a proper balance between the two is probably the best approach in designing regulatory compliance monitoring systems.

Regulatory Compliance/Licensing and Quality

This part of the policy commentary will delineate the differences between regulatory compliance and quality. It will provide the essential principles and elements that clearly demonstrate the differences and their potential impact on program monitoring. Obviously, there is some overlap between this section and the above section dealing with regulatory compliance monitoring paradigms. When we think about regulatory compliance measurement, we are discussing licensing systems. When we think about quality, we are discussing Quality Rating and Improvement Systems (QRIS), accreditation, professional development, or one of the myriad quality assessment tools, such as the Classroom Assessment Scoring System (CLASS) or Environment Rating Scales (ERS's). All these systems have been designed to help improve the health and safety of programs (licensing) to building more environmental quality (ERS), positive interactions amongst teachers and children (CLASS), enhancing quality standards (QRIS, accreditation), or enhancing teacher skills (professional development).

There are eight basic principles or elements to be presented (they are presented in a binary fashion demonstrating differences): 1) "Do no harm" versus "Do good". 2) Closed system versus Open system. 3) Rules versus Indicators. 4) Nominal versus Ordinal measurement. 5) Full versus Partial compliance. 6) Ceiling effect versus No Ceiling effect. 7) Gatekeeper versus Enabler. 8) Risk versus Performance.

1) Let's start with the first principal element building off what was discussed in the above section, "Do No Harm" versus "Do Good". In licensing, the philosophy is to do no harm, its emphasis is on prevention, to reduce risk to children in a particular setting. There is a good deal of emphasis on health and safety and not so much on developmentally appropriate programming. In the quality systems, such as QRIS, accreditation, professional development, Environmental Rating Scales, CLASS, the philosophy is to do good, its emphasis is looking at all the positive aspects of a setting. There is a good deal of emphasis on improving the programming that the children are exposed to or increasing the skill set of teachers or improving the overall environment or interaction that children are exposed to.

2) Closed system versus Open system. Licensing is basically a closed system. It has an upper limit with full compliance (100%) with all rules. The goal is to have all programs fully comply with all rules. However, the value of this assumption has been challenged over the years with the introduction of the Regulatory Compliance Theory of Diminishing Returns (Fiene, 2019). With quality systems, they tend to be more open and far reaching where attaining a perfect score is very difficult to come by. The majority of programs are more normally distributed where with licensing rules

the majority of programs are skewed positively in either substantial or full compliance. It is far more difficult to distinguish between the best programs and the mediocre programs within licensing but more successful in quality systems.

3) Rules versus Indicators/Best Practices. Licensing systems are based around specific standards/rules/regulations that either are in compliance or out of compliance. It is either a program is in compliance or out of compliance with the specific rule. With quality systems, there is more emphasis on indicators or best practices that are measured a bit more broadly and deal more with process than structure which is the case with licensing. It is the difference between hard and soft data as many legal counsels term it. There is greater flexibility in quality systems. With this said, if we can look at other service types, such as adult-residential services, there has been some limited success with blending structural and process elements but it still remains a measurement issue on the process side.

4) Nominal versus Ordinal measurement[3]. Licensing systems are nominally based measurement systems. Either you are in compliance or out of compliance. Nothing in-between. It is either a yes or no response for each rule. No maybe or partial compliance. With quality systems, they are generally measured on an ordinal level or a Likert scale. They may run from 1 to 3, or 1 to 5, or 1 to 7. There are more chances for variability in the data than in licensing which has 1 or 0 response. This increases the robustness of the data distribution with ordinal measurement.

5) Full or None versus Gradients or Gray Area. Building off of the fourth element, licensing scoring is either full or not. As suggested in the above elements, there is no in-between category, no gradient or gray area. This is definitely not the case with quality systems in which there are gradients and substantial gray areas. Each best practice can be measured on a Likert scale with subtle gradients in improving the overall practice.

6) Ceiling effect versus No Ceiling. With licensing there is definitely a ceiling effect because of the emphasis on full 100% compliance with all rules. That is the goal of a licensing program, to have full compliance. With quality systems, it is more open ended in which a ceiling effect is not present. Programs have many ways to attain excellence.

7) Gatekeeper versus Enabler: Licensing has always been called a gatekeeper system. It is the entry way to providing care, to providing services. It is a mandatory system in which all programs need to be licensed to operate. In Quality systems, these are voluntary systems. A program chooses to participate, there is no mandate to participate. It is more enabling for programs building upon successes. There are enhancements in many cases.

8) Risk versus Performance: Licensing systems are based upon mitigating or reducing risks to children when in out of home care. Quality systems are based upon performance and excellence where this is rewarded in their particular scoring by the addition of a new Star level or a Digital Badge or an Accreditation Certificate.

There has been a great deal of discussion in the early care and education field about the relationship between licensing, accreditation, QRIS, professional development, and technical

assistance. It is important as we continue this discussion to pay attention to the key elements and principles in how licensing and these quality systems are the same and different in their emphases and goals, and about the implications of particular program monitoring paradigms and measurement strategies. For other regulatory systems outside the human services field, the same type of model can be applied positioning compliance and quality as a continuum one building off of the other because I feel that with the introduction of more quality into a regulatory context will help to ameliorate the ceiling and plateau effect of diminishing returns on performance and outcomes.

Reference:

Fiene, R. (2019). A Treatise on Regulatory Compliance. *Journal of Regulatory Science*, Volume 7, 2019

Notes:

1. This manuscript should be read along with *A Treatise on Regulatory Compliance* which is referenced above because the two articles build off one another. In the *treatise* description, the specific idiosyncrasies of regulatory compliance data and other key implications of the theory are pointed out that enhance the presentation in this article, such as the extreme nature of skewness that is present in regulatory compliance data, nominal data measurement, the differences between full and substantial regulatory compliance, designing the most cost effective and efficient differential monitoring system, and the need to dichotomize data because of the skewed nature of the data distribution.

2. The ultimate goal is the most cost effective and efficient differential monitoring system for negotiating the path of least resistance in complying with a given set of rules which will provide the proper balance of rules. This should be the goal of any regulatory science paradigm. By using the previous *Treatise* article along with this article should provide a blueprint for the regulatory science field in designing a program monitoring system to measure regulatory compliance where an emphasis on differential monitoring should occur in licensing systems and full-scale monitoring should occur in program quality systems. Another approach is to have both regulatory compliance and program quality built as a continuum in the program monitoring system similar to what Head Start is attempting.

3. There are instances in which this dichotomy is not as clear or straightforward where licensing systems do allow partial compliance as a facility has opportunities to correct non-compliances on their way to achieving full compliance with specific rules. The problem is that this is not necessarily a standardized process and it is difficult to determine if it is used often in licensing agencies' monitoring efforts.

TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http:\\journalofregulatoryscience.org

Regulatory Science

Journal of Regulatory Science 7 (2019) 1–3

A Treatise on the Theory of Regulatory Compliance

Richard J. Fiene[a,*]

[a]*Edna Bennett Pierce Prevention Research Center, Pennsylvania State University, 305 Templar Drive, Elizabethtown, Pennsylvania 17022*

Abstract

This treatise provides some insights into certain assumptions related to regulatory compliance and the implications for regulatory researchers and policy-makers for the future development of rules and regulations. Once regulatory compliance decision making moves from requiring full compliance with all rules to a substantial regulatory compliance decision making approach, the measurement and monitoring systems employed to assess programs and facilities change dramatically.

Keywords: regulatory compliance, risk assessment, key indicators, licensing, monitoring, measurement

1. Introduction

Regulatory compliance is a sub-discipline within regulatory science that focuses on measurement, monitoring systems, risk assessment, and decision making based on regulatory compliance scoring. Regulatory compliance is dominated by nominal scale measurement, that is, either a facility is in or out of compliance with specific rules. There is no middle ground with regulatory compliance as there is with most quality measurements, which are generally made on an ordinal scale. However, some regulators feel that certain regulations are not or should not be subjected to nominal measurement.

A factor with regulatory compliance data is that they generally follow a very skewed frequency distribution, which limits analyses to non-parametric statistics. Because of the skewed data distribution, dichotomization of data is warranted, given the lack of variance in the regulatory compliance frequency distribution - the majority of facilities [1] are either in full or substantial regulatory compliance.

An assumption within regulatory compliance is that full regulatory compliance, that is, 100 percent compliance with all rules [2], is the best (i.e., risk is minimized) possible scenario for the services being delivered and assessed. It is also assumed that all promulgated rules have an equal weight in their relative impact on the desired service delivery model, although this thinking has been changing over time regarding how rules are reviewed and complied with. This short treatise will examine the past 40 years of research delving into regulatory compliance measurement, and will provide some guidance to regulatory researchers and policy-makers as they move forward with both research and policy development related to rules. The data from these research studies have led to a Theory of Regulatory Compliance that demonstrates that substantial regulatory compliance - and not full regulatory compliance - is a more effective and efficient public policy as it relates to decision making on monitoring and licensing.

The results reported herein are drawn from human services delivery systems in the United States and Canada, such as early care and education, as well as child and adult residential services. The results are from state and provincial level licensing systems involving over 10,000 facilities serving over 100,000 clients. All the data are part of an international regulatory compliance database (https://data.mendeley.com/datasets/kzk6xssx4d/1) maintained at the Research Institute for Key Indicators and the Pennsylvania State University.

2. Methods

Alternate methodologies, logic models, and algorithms were developed directly from the Theory of Regulatory Compliance once it was determined that substantial regulatory compliance produced better results than full regulatory compliance. These methodologies created a differential monitoring or targeted monitoring approach based on risk assessment, which measures client morbidity and/or mortality when individual rule

*Corresponding author: Richard J. Fiene, Email: rjf8@psu.edu, Phone: 717-598-8908, ORCID iD: http://ORCID: 0000-0001-6095-5085.

[1] The term “facilities” is used when referring to programs and/or facilities.

[2] The term “rules” is used when referring to rules and/or regulations.

non-compliance is assessed, and the determination of key statistical predictors for overall regulatory compliance [3].

Briefly, the above methodologies provide cost-effective and efficient means for the ongoing monitoring of human service delivery systems by selecting and reviewing only those rules that either have a positive impact on clients, statistically predict overall regulatory compliance, or protect the health and safety of clients [3]. Based on regulatory compliance historical data, decisions could be made as to the frequency and depth of the reviews or inspections. Abbreviated reviews (inspections in which a subset of rules are measured), such as licensing key indicator rules or risk assessment rules, would only be done in those facilities having a history of high regulatory compliance. Those facilities with a history of high regulatory non-compliance would continue to receive full regulatory compliance reviews as they did in the past.

3. Results

Prior to 1979, it was always assumed that there was a linear relationship between regulatory compliance measures and program quality measures of human service facilities. In a study conducted in that year, which compared results from early care and education programs, in particular child care centers, this assumption did hold up when one went from low regulatory compliance to substantial regulatory compliance. However, the results from substantial regulatory compliance to full (100 percent) regulatory compliance did not show the same linear relationship. Rather, it showed that those programs that were in substantial instead of full compliance were actually scoring higher on the program quality measures.

Since 1979, this result has been replicated in many other early care and education delivery system studies, both nationally in the United States (Head Start) [1] and in several states (Georgia, Indiana, Pennsylvania) [2]. In all these studies, one finds a non-linear - rather than a linear - relationship between regulatory compliance and the overall quality of the facilities being assessed.

4. Discussion

Based on the results above, there are several assumptions within regulatory compliance that need to be reconsidered:

1. Public policies that require full (100 percent) compliance with all rules may not be in the best interest of the clients being served, nor an effective use of limited regulatory resources. Potentially, emphasis on substantial regulatory compliance may be a more effective and efficient public policy related to client outcomes when it comes to their health, safety, and quality of life. Note that substantial compliance is still very high regulatory compliance (99-97 percent compliance with all rules) and produces positive client outcomes. As stated above, regulatory compliance data are extremely skewed and not normally distributed. There is very little variance in the data and the majority of programs are in either full or substantial regulatory compliance.

2. If a jurisdiction focuses on a substantial regulatory compliance public policy it opens up many system enhancements, such as differential or targeted monitoring, risk assessment analysis, and statistical key indicator rules that have been demonstrated to be cost effective and efficient approaches to reviewing program performance. In a full regulatory compliance public policy approach, none of these system enhancements can be employed, with the possible exception of the key indicator approach as delineated in number four below.

3. If a jurisdiction takes the position that all rules are not equal, then a risk assessment or weighting approach becomes an alternative based on the assumption that certain rules place clients at greater risk of death, serious injury, or other types of harm.

4. Even if a jurisdiction does not have a licensing law that allows issuing licenses on the basis of substantial compliance, there is the possibility that key indicators could still be used for abbreviated reviews or inspections, if there is no prohibition in statute or regulation that expressly forbids the use of this approach, since key indicators statistically predict full regulatory compliance. In other words, all rules are statistically predicted to be in regulatory compliance based on the results of the key indicators. Therefore, technically, all rules have been reviewed albeit short of a full review or inspection.

5. Based on previous research, utilizing a risk assessment approach along with a key indicator approach is the most cost effective and efficient differential monitoring system model. The reason is that both predictive rules and those rules that place clients at greatest risk are always assessed when a site visit review or inspection is done. Many more jurisdictions use a risk assessment approach at this point, but there is a loss of predictive regulatory compliance by just using it.

6. Based on previous regulatory compliance history, only those facilities in high regulatory compliance would be eligible for abbreviated key indicator and risk assessment reviews, whereas those with a history of high regulatory non-compliance would continue to receive full regulatory compliance reviews. This gets at the essence of the differential monitoring approach, which is cost neutral. Regulatory resources may then be re-allocated from the abbreviated reviews to more in-depth full regulatory compliance reviews.

7. Based on the use of the key indicator and risk assessment methodologies within a differential monitoring approach, it is possible to identify over multiple jurisdictions if there are generic rules that meet the criteria of risk abatement and prediction. Such an application has occurred in the United States with the creation of early care

and education standards entitled *Caring for Our Children Basics*, published by the Administration for Children and Families, US Department of Health and Human Services (2015).

5. Conclusion

Regulatory compliance is relatively new in applying empirical evidence and basic scientific principles to its decision making. In the past, it had been dominated by case studies and long narrative reports that did not lend themselves to quantitative analysis. There is a need to more clearly apply empirical evidence and the scientific method to rule development. Certain assumptions, such as full regulatory compliance as a sound public policy, are lacking in empirical evidence. This treatise on a theory of regulatory compliance is provided for its heuristic value for both regulatory researchers and policymakers in rethinking some basic regulatory compliance assumptions. It is not about more or less, rules but finding the "right rules" that protect clients, predict overall regulatory compliance, and produce positive client outcomes.

6. Declaration of Conflicting Interest

The authors declare no conflicts of interest.

7. Article Information

This article was received March 14, 2019, in revised form April 11, 2019, and made available online May 9, 2019.

8. References

[1] Fiene, R. (2013). *Office of Head Start Key Indicator Project Report.* Retrieved from https://drfiene.files.wordpress.com/2018/04/rheadstart-report.pdf

[2] Fiene, R. (2014). *Georgia Child Care Licensing Study: Validating the Core Rule Differential Monitoring System.* Retrieved from https://drfiene.files.wordpress.com/2018/04/2014-georgia.pdf

[3] Fiene, R. (2016). *Early Childhood Program Quality Indicator and Improvement Model (ECPQIM) and Differential Monitoring Logic Model and Algorithm (DMLMA) Readings*. Retrieved from https://drfiene.files.wordpress.com/2018/03/1ecpqims-book-of-readings.pdf

Early Childhood Research Quarterly xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs

Michel Lahti[a,*], James Elicker[b], Gail Zellman[c], Richard Fiene[a,b,c,1]

[a] University of Southern Maine, Muskie School of Public Service, PO Box 9300, Portland, ME 04104-9300, United States
[b] Purdue University, United States
[c] RAND Corporation, United States

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Child care quality
QRIS
Study design
Methodology

ABSTRACT

In recent years, child care quality rating and improvement systems (QRISs) have become an increasingly popular policy tool to improve quality in early childhood education and care (ECEC) settings and have been adopted in many localities and states. The QRIS proposition is that with higher-quality child care settings, it is more likely that children who attend those high-quality programs will benefit in terms of outcomes like school readiness. However, in order to demonstrate this linkage, QRIS standards and ratings must function as intended, i.e. be valid. This paper presents a framework for validating child care quality improvement standards and processes, along with examples from recent QRIS validation studies in two states. The state examples provide useful data about the strengths and limitations of these validation approaches. We discuss the implications of applying these approaches and provide recommendations to researchers, policy-makers, and program leaders who implement QRIS validation studies.

Introduction

In recent years, child care quality rating and improvement systems (QRISs) have become an increasingly popular policy tool to improve quality in early childhood education and care (ECEC) settings and have been adopted in many localities and states. The QRIS National Learning Network reports that 40 statewide QRISs have launched or piloted, including the District of Columbia (QRIS National Learning Network, 2014). The immediate goal of a QRIS is to raise the quality of care in early learning settings. Existing research suggests that care in higher-quality settings will improve child functioning, including school readiness (Burchinal et al., 2009; Burger, 2010; Howes et al., 2008), especially for children from lower-income families. QRIS logic models that guide these large-scale interventions focus on improving various dimensions of ECEC quality, with the ultimate goal of improving system outcomes, namely; child care program quality, training and technical assistance for child care providers, information and support for families, and, therefore, improvements to children's cognitive, language, social, emotional, and physical development.

The perceived need for QRIS has grown out of documented gaps in quality in existing ECEC programs, especially those serving children from lower-income families (Fuller, Loeb, Kagan, & Carrol, 2004; NICHD ECCRN, 2000) and the inability of the current ECEC system to promote uniformly high quality (Cochran, 2007). QRISs produce program-level quality ratings based on multi-component assessments designed to make ECEC quality transparent and easily understood to parents and other stakeholders. Most also include feedback, technical assistance, and incentives to both motivate and support providers' efforts toward quality improvement (Tout et al., 2010). To make program quality transparent, QRISs typically rely on a multi-tiered rating system with one to five levels of program quality. Therefore, it is important that these ratings show evidence of validity, so that higher-quality programs are rated higher, and lower-quality programs are rated lower.

Recent research has documented the importance of both specificity and thresholds when testing hypotheses about child care quality impacts on children's developmental outcomes (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Howes, Whitebook, & Phillips, 1992; NICHD ECCRN, 2000, 2002). However, common global measures of classroom quality such as the Early Childhood Environment

* Corresponding author. Tel.: +520-326-5154; fax: +520-326-5155.
E-mail addresses: mlahti@usm.maine.edu, Michel.lahti@lecroymilligan.com (M. Lahti).
[1] Professor Emeritus at Pennsylvania State University.

http://dx.doi.org/10.1016/j.ecresq.2014.04.005
0885-2006/

Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 2005) are not always significantly associated with specific child outcomes (Burchinal, Kainz, & Cai, 2011). This may be because these global quality scales do not focus enough on the particular child care quality processes most likely to bring about improved child outcomes (specificity) or they do not provide guidance for the level of quality required to produce improved child outcomes (thresholds). As states implement QRISs, they are using observational measures such as the ECERS-R, and they may also combine other quality measures such as the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) or locally specified quality indicators. Because QRIS quality standards are often complex, including many components and measures at several quality levels, and because they vary from state to state, it is especially important for states to carefully validate their quality rating systems and match measures specifically to the stated outcome goals of the QRIS. For example, if a particular QRIS places more emphasis on the health aspects of children's development, then the ECERS-R and CLASS would not be appropriate tools; but a tool measuring child care health indicators, such as the National Health and Safety Tool being developed by the California Child Care Health Program (Alkon, 2013) would be more appropriate.

Validity data can also enable researchers to test conclusions about whether the quality indicators embedded in QRIS standards lead to adequate quality assessment and whether the methods used to assign quality ratings are working as intended (Cizek, 2007). This paper defines operationally the concept of QRIS validity, presents four general approaches to assessing validity in the context of large-scale QRISs, and critically examines the efforts of two states, Maine and Indiana, to assess the validity of recently implemented QRISs using these approaches.

Validation of a QRIS is a developmental and multi-step process that assesses the degree to which design decisions about program quality standards and measurement strategies are resulting in accurate and meaningful quality ratings. Validation of a QRIS provides designers, administrators, and stakeholders with crucial data about how well the system is functioning. A carefully designed plan for ongoing QRIS validation creates confidence in the system and a climate that supports continuous quality improvement at both the child care provider and system levels (Zellman & Fiene, 2012).

To date, QRIS validation research efforts have been limited, for a number of reasons. First, validation is complex and involves a range of activities, which should include validating standards, measures, and rating protocols. Second, there has been little information available in the field that clarifies the importance and purpose of QRIS validation or identifies recommended strategies. Third, child care quality advocates and policy makers have been extremely busy designing and implementing these statewide systems, often with limited resources. Given these constraints, validation may seem like an abstract luxury that can wait until later. Further, in states with more mature QRISs, there may be some reluctance among stakeholders to assess the validity of an established and accepted quality improvement system. In newer state systems, policymakers may question the need for validation, given arguments recently offered in support of establishing a QRIS system (Zellman & Fiene, 2012; Zellman, Brandon, Boller, & Kreader, 2011). Yet early and ongoing validation research is essential to the long term success of any system.

One challenge is that QRIS validation cannot be determined by a single study. Instead, validation should be viewed as an iterative process with several equally important goals: refining the QRIS quality standards and ratings, improving system functioning, and increasing the credibility and value of rating outcomes and the QRIS system as a whole. A carefully designed validation plan can promote the accumulation of evidence over time that will provide a sound theoretical and empirical basis for the QRIS (AERA, APA, & NCME, 1999; Kane, 2001; Zellman & Fiene, 2012). Ongoing validation activities, carried out in tandem with QRIS monitoring activities (those that examine ongoing implementation processes) and evaluation activities (those that examine specific outcomes) can help a QRIS improve throughout its development, implementation, and maturation (Lugo-Gil et al., 2011; Zellman et al., 2011).

QRIS validation research may produce three important benefits. First, validation evidence can promote increased support for the system among parents, ECEC providers, and other key stakeholders. Ratings that mirror the experiences of parents and providers can build trust and increase the overall credibility of the system. Second, a system that is measuring quality accurately and specifically should better able to target limited quality improvement resources to programs and program elements most in need of improvement. This should result in more targeted and effective supports for programs striving to offer higher-quality services. Third, validation evidence can be used to improve the efficiency of the rating process. If a QRIS is expending resources to measure a component of quality that is not making a unique contribution to a summary quality rating, is not measuring quality accurately, or is not contributing to desired program outcomes, that component can be removed or revised. For example, measures that vary little across providers, whose quality varies substantially in other ways, make little or no contribution to overall quality ratings (Zellman & Fiene, 2012).

Four approaches to validation

A comprehensive QRIS validation plan includes multiple studies that rely on different sources of information and ask different but related questions. We suggest QRIS validation research be organized around four complementary approaches: key quality concepts; quality measurement; ratings outputs; and links to child outcomes (Zellman & Fiene, 2012). Summaries of these approaches are provided in Table 1, which includes the purpose of each validation approach, the types of research that can be undertaken, the questions that are asked, and some limitations of each approach. The four approaches are also elaborated later in the paper, as we summarize results of validation research in Indiana and Maine.

In reviewing the table, and throughout this paper, we use three key QRIS terms: component, standard, and indicator. The term 'quality component' refers to broad quality categories used in QRIS (such as staff qualifications, family engagement, or learning environment). A 'quality standard' is defined as a specific feature of quality, such as specialized training in the use of developmentally appropriate curriculum or developmental assessment training within the staff qualifications component. A set of quality standards comprise each quality component. 'Quality indicators' are the specific metrics used for each quality standard. A given quality standard may have one or more quality indicators. An indicator related to the curriculum/assessment staff training standard may be, for example, "At least 50% of teaching staff have completed the two-course statewide training session on developmentally-appropriate curriculum."

QRIS validation in Indiana and Maine

This section will describe efforts at QRIS validation in two states in order to explore current validation efforts using these four approaches and to identify the successes and challenges experienced in these early QRIS validation studies. In Indiana and Maine, the QRIS designs are similar, but some aspects of the states' child care contexts, specific QRIS quality components, standards, and rating processes employed are somewhat different. Both states launched their QRIS statewide in 2008, and both systems have four quality tiers, referred to as "levels" in Indiana and "steps" in Maine, organized into a "building block" framework, meaning that child

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Table 1
Four related approaches to validating a QRIS.

Approach	Activities and purpose	Typical questions	Issue and limitations
1. Examine the validity of key underlying concepts.	Assess whether basic QRIS quality components and standards are the "right" ones to include by examining levels of empirical and expert support.	*Do the quality components capture the key elements of quality?	*Process subject to interpretation and to political pressure
		*Is there sufficient empirical and expert support for including each standard?	*Limited empirical evidence available; few established links to outcomes of interest.
2. Examine the measurement strategy and psychometric properties of measures used to assess quality.	*Examine properties of key quality measures, e.g., inter-rater reliability on observational measures, scoring of documentation, and inter-item correlations, to determine if measures are psychometrically sound.	*What is the reliability and accuracy of indicators collected using different methods?	*This validation activity is especially important given that some quality component measures were likely developed in low-stakes settings and have not been examined in the high-stakes context of QRIS.
	*Examine relationships among quality measures to assess whether they function as expected.	*Do quality measures perform as expected? (e.g., do expected subscales emerge?) *Do measures of similar standards relate more closely to each other than to other measures? *Do measures relate to each other in theoretically consistent ways?	
3. Assess the outputs of the rating process	*Examine variation and patterns of program-level ratings within and across program types, to assess if QRIS distinguishes levels of quality.	*Do rating distributions vary by program type, e.g., center-based programs vs. home-based programs?	*Measurement error is an important issue that should be examined.
	*Examine relationship of program-level ratings to other validated quality indicators to determine if ratings are assessing quality in expected ways.	*Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?	*These validation activities depend on a reasonable level of confidence about the quality components, standards and indicators as well as the process used to designate ratings.
	*Examine alternate cut points and combining rules to determine how well the ratings distinguish different levels of quality.	*Do levels cut scores and combining rules produce expected rating distributions and meaningful distinctions among programs?	*Comparing QRIS measures to other measures is frequently constrained by the absence of validated alternative measures of the same constructs.
4. Examine how ratings are associated with children's outcomes.	Examine the relationship between program-level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes.	Do children who are attending higher-rated programs have greater levels or gains in skills than children who attend lower-quality programs?	*Appropriate demographic and program level control variables must be included in analyses to account for selection factors.
			*Studies may be done on selected child and program samples to save resources. *Given staff turnover and movement, children may not spend substantial time with quality-rated caregivers (dosage.) *Measurement limitations noted for Approach 3 apply here as well. *Findings from non-experimental studies do not permit attribution of causality about QRIS participation, but preliminary inferences can be made about how quality influences children's outcomes.

care providers must enter at the lowest level and meet all quality standards and indicators at each level in order to advance to the next higher level. The focus on these two states in this paper is to help illustrate the application of these four approaches to operationalizing validation in a QRIS. While the QRIS evaluations in Maine and Indiana have resulted in other kinds of information disseminated for policy makers in these states and publications for other audiences, this paper is unique in that it is only intended to focus on these four concepts of validation.

Both states partnered with university-based researchers to conduct validation research, after piloting aspects of their QRIS design. However, there are also key differences between these two states. For example, the Indiana QRIS standards were developed based on a local community-based model that was then modified by a state stakeholder committee for statewide expansion. The Maine quality standards were developed to align with program-type-specific national accreditation standards. The Maine standards were also vetted through review and comment by many stakeholders and technical assistance was provided by University researchers based on reviews of the scientific literature. Maine QRIS ratings are generated by provider self-report, then verified by state agency staff, while Indiana employs independent raters who directly assess the standards by visiting child care settings. Provider voluntary participation rates are higher among state-licensed providers in Indiana. However, Indiana also has significant numbers of license-exempt child care providers, whereas license exemption is not a prominent feature of the Maine child care system. The key features of each state QRIS are summarized in Table 2. These two states provide useful examples, because while the state child care contexts are different, they each used strategies contained in the four validation approaches discussed above and outlined in Table 1. The successes and limitations of these states' approaches will inform future validation research on QRIS.

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Table 2
Key features of Indiana and Maine QRISs.

QRIS feature	Indiana	Maine
System name	Paths to QUALITY (2008)	Quality for ME (2008)
Eligible child care types (participation rate)	Licensed centers (89%) Licensed homes (62%) Unlicensed registered ministry centers (12%)	Licensed centers (68%) Licensed homes (44%)
Participation rules	Voluntary for all providers (Not required for CCDF participation)	Required for programs participating in federal CCDF/Voluntary for all others
QRIS structure	Building Block; 4 quality levels	Building Block; 4 quality levels
QRIS standards (examples)	Level 1 – Licensed, or completes voluntary certification program	Step 1 – Meets all regulatory standards, in operation for more than one year, and all staff registered in Maine Roads to Quality Registry (MRTQ).
	Level 2 – Learning environment and materials requirements; daily literacy activities; 25% of staff have CDA or equivalent; 15 hrs. in-service training/yr.; etc.	Step 2 – Learning Environment/Developmentally Appropriate Practice requirements; program improvement plan in place; 50% of staff at level 5 on MRTQ career lattice; etc.
	Level 3 – Written curriculum focused on whole child; provision for special needs; 50% of staff have CDA or equivalent; 20 hrs. in-service training/yr.; etc.	Step 3 – Documented use of Early Childhood Learning Guidelines and/or Infant-Toddler Learning Guidelines; Evidence collected at least three times per year on child's development; etc.
	Level 4 – National accreditation; Provide mentoring to other QRIS providers (see www.in.gov/fssa/carefinder/2554.htm#).	Step 4 – National accreditation; written parent involvement plan; etc. (See: https://www.maine.gov/dhhs/ocfs/ec/occhs/qualityforme.htm).
QRIS standards development process	Aligned with national accreditation standards, expansion of community pilot program, modified by state stakeholder committee.	Aligned with national accreditation standards; based on available empirical evidence; and from results of stakeholder reviews and comment.
QRIS rating procedure	Independent ratings contractor, annual site visits, using Paths to QUALITY standards checklist	(1) Provider self-assessment, online system; (2) enrollment system uses linked files from state licensing and registry; and (3) verified by state agency staff.

Method

Indiana

The Indiana QRIS is called "Paths to QUALITY™." The validation research reported here includes a preliminary literature review and an empirical field study including a stratified random sample of 276 child care providers who had voluntarily entered the QRIS during 2008–2009, including 135 classrooms in 95 licensed child care centers, 169 licensed family child care homes, and 14 classrooms in 12 unlicensed registered child care ministry centers. Independent, on-site assessments were completed by university researchers approximately one year after QRIS entry and included: observational global quality assessments of the child care environment using the Environmental Rating Scales (ERS: ITERS-R, ECERS-R, FCCERS-R; Harms et al., 2005); observations of adult–child interaction quality (Caregiver Interaction Scale, CIS; Arnett, 1989); surveys and interviews with child care providers; and interviews with parents whose children had been placed with QRIS providers. Observers were trained to reliability level of 80% exact agreement (*Kappa* = .70) or higher, and maintained reliability during the study. Child development assessments were completed using standardized research-validated measures, with two randomly selected children from each participating child care center classroom or family child care home. For children under three years, measures included the Mullen Scales of Early Learning (Mullen, 1995) for cognitive and language development and the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2002) for social-emotional development. For children three to five years, the measures included the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 1997) and Woodcock–Johnson Applied Problems and Letter–Word Identification subtests (Berry, Bridges, & Zaslow, 2004) for language and cognitive development and the Social Competence and Behavior Evaluation (SCBE; LaFreniere & Dumas, 1997) for social–emotional development. (For a detailed description of the Indiana evaluation methodology, see Elicker et al., 2013; Elicker, Langill, Ruprecht, Lewsader, & Anderson, 2011.)

Maine

The Maine QRIS is called "Quality for ME." The Maine validation research reported in this paper is based on a literature review of quality variables, focus group interviews with providers and parents, and a field study including a stratified random sample of 255 providers who enrolled in the QRIS in 2008 through 2011, including: 153 classrooms in 105 licensed child care centers; 113 licensed family child care homes; and 41 classrooms in 37 Head Start sites. Assessments were completed as soon as possible after a program enrolled into the QRIS, however, this varied based on the length of time required for the state agency to verify enrollment information and schedule on-site observations. Like Indiana, Maine researchers used the ERS global quality assessment scales and conducted surveys with providers and parents. Unlike Indiana, Maine did not collect any child-level outcome data. Assessors were trained to reliability annually by authors of the ERS scale and maintained a 85% inter-rater reliability during the study. (For a detailed description of the evaluation methodology, see Lahti et al., 2011.)

Results

Results of the QRIS validation research in Indiana and Maine are presented in relation to the four approaches to validation recommended by Zellman and Fiene (2012; refer to Table 1).

Approach 1: examine the validity of key underlying concepts

As noted above, the quality components included in a QRIS (e.g., staff qualifications, learning environment, family engagement) essentially define how child care quality will be viewed in each state. Conceptual validation provides justification and support for these chosen elements. This first validation approach asks whether the quality components, standards, and indicators included in a QRIS are the "right" ones; that is, if together they define quality of care. Many state QRISs have adopted similar, though not identical, concepts and program quality standards (Smith, Robbins, Stagman, & Kreader, 2012).

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

One approach that can help to validate the underlying concepts of quality in a QRIS involves assessing the degree to which the quality components used in the QRIS ratings include standards and indicators that are based on empirical evidence that links them to desired program, family and child outcomes. A literature review weighs the existing research evidence and on that basis provides a judgment about whether a particular quality component should be included or excluded from the QRIS. Like many validation activities, such reviews ideally would be updated from time to time to determine if revisions to the QRIS are advisable in light of new research findings. As noted in Table 1, this approach may be limited by available data. Further, available data may be subject to more than one interpretation. Politics can also play a role; supporters of particular elements, e.g., nutrition, accreditation, may want to ensure that such measures are included, regardless of the strength of the research evidence. This literature review approach of conceptual validation was a key method used in developing both Indiana's and Maine's QRIS quality standards.

Indiana: examining the validity of underlying concepts

Standards and indicators for each QRIS level in Indiana were drafted by a state committee of child care providers and stakeholders. The standards were based on an existing community-level Paths to QUALITY model, but also made accommodations for statewide use and integration into the existing state child care licensing and training/technical assistance systems. The highest level quality goal for QRIS in Indiana is national accreditation, so proposed quality standards and criteria at each QRIS level were constructed to help child care providers work toward accreditation in steps.

The Purdue University research team conducted a review of previous evaluations of the Indiana QRIS community-level pilot programs and an in-depth analysis of the proposed QRIS quality standards based on the published child development and child care literature. This literature-based analysis is summarized here. (For a full report, see Elicker, Langill, Ruprecht, & Kwon, 2007; Elicker et al., 2013.)

First, the evaluators looked at each proposed QRIS quality standard and indicator for each type of child care. Ten broad quality components were identified that encompassed all of the proposed quality indicators: regulation; teacher education/training; structural/environmental quality; process quality/interactions; assessment; provisions for children with special needs; program policies; director/owner professional development; parent–teacher communication; and national accreditation. These ten components were then used as key terms to guide an extensive search of the research literature to collect and weigh the available evidence that each component was: (1) generally considered a valid aspect of quality; and (2) empirically associated with children's well-being or positive developmental outcomes. Based on the amount and quality of evidence, each quality component was the rated as follows: (1) some or limited evidence (one or two well-designed studies); (2) moderate evidence (3–5 well-designed studies); or (3) substantial evidence (more than five well-designed studies). The results of this analysis were reported to the state QRIS planning committee, including a conclusion that most of the proposed quality indicators had "substantial evidence" for their validity.

Maine: examining the validity of underlying concepts

Researchers at the University of Southern Maine worked with state agency leaders and other key stakeholders through a process that involved the use of Concept Mapping (The Concept System®, 2012). This process allows for the development of a conceptual framework that can guide planning, and in this case led to the selection of the underlying quality concepts and standards for Maine's QRIS. Similar to what was done in Indiana, University of Southern Maine research staff identified key quality concepts from the literature and national accreditation standards. In addition, concepts emerged from results of eight focus groups with parents and ECEC professionals across the state, including participants from various types of settings, e.g., family child care homes, after school programs, centers, etc. Statements of program quality were developed; these statements were the focus of a mapping process which involved more than twenty-four experts reviewing and rating the statements. The Concept Maps that result from this process allowed participants to visually identify which concepts of program quality were most favored by specific key stakeholder groups. In addition, the mapping software illustrated how closely related the concepts were to each other, based on reviews from the select experts. From this process, a set of components and standards was developed. The final step in selecting program quality standards involved a formal review and comment process that the state agency implemented in various locations across the state (Maine DHHS, 2008).

Approach 2: examine measurement strategies and psychometric properties of quality measures

A second type of validation focuses on the attributes of the individual quality measures used in the QRIS and the way these measures are combined to produce a summary rating of program quality. This approach addresses how well measures are working in the context of the QRIS. These efforts attempt to answer questions such as, "Is there evidence that a given indicator measures what it purports to measure?" "If the QRIS claims to have a specific number of dimensions, do we find those dimensions in the output data?" "Is there sufficient variance in scores on this quality indicator to justify its inclusion in the QRIS?" Addressing these issues involves an examination of the distribution of participating provider quality scores and the internal consistency of multi-item measures.

The research literature provides limited guidance concerning the most appropriate ways to combine measures of quality indicators into summary ratings (Lugo-Gil et al., 2011; Tout, Zaslow, Halle, & Forry, 2009; Zellman, Perlman, Le, & Setodji, 2008). Yet this process is crucial to producing meaningful overall program quality ratings, the key output of the rating assessment process. At minimum, it is important to consider whether certain elements *should* be treated as more important, and if so, how this can be assured in the process of combining them. If this issue is not addressed, unexamined weighting may occur anyway. For example, if measures of individual quality elements are combined without any weighting, then those measures that are longer (e.g., include more items) will count for more in a final rating.

At the time both the Maine and Indiana QRISs were being designed, in the mid-2000s, the predominant global quality measures in use in both states were the Environmental Rating Scales (ERS) (Harms et al., 2005; Harms, Cryer, & Clifford, 2006; Harms, Cryer, & Clifford, 2007). ERS use was predominant in the accreditation quality improvement efforts in both child care centers and child care homes. So there was some familiarity with the measures on the part of providers. This was an important political consideration in terms of developing and promoting the design of the QRIS. In addition, in reviewing emerging QRIS work from other states, it appeared that the ERSs were the predominant global classroom quality measure in use at that time. While ERS was influential in the design of the QRIS quality standards in both Maine and Indiana, it is important to note that the ERS are not used to determine the step or level quality ratings. Many other quality indicators are included in the QRIS standards of both states, including staff qualifications, annual staff training hours, and other indicators that help providers make progress toward the ultimate quality goal of national accreditation.

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Choosing the points at which individual measures (in block design QRISs) and summary ratings are assigned to rating levels is another exercise that has received limited attention. Cut scores can be assessed in a number of ways. One relatively simple one is to use existing data to conduct a "virtual pilot" (Zellman & Karoly, 2012a) in which existing data are used and cut scores are altered and the effects are examined in terms of distributions of summary ratings across programs. A downward limit on cut scores is the need for some variation within each quality component; without it, a component provides no useful information in overall ratings. Designers may compare program distributions using different cut scores, although it is not always clear what an appropriate rating levels distribution should be. However, it is reasonable to assume that an appropriate distribution in the early phase of a QRIS would be one in which there are programs placed at all levels, with decreasing numbers of programs at each succeeding higher level.

Another validation activity might involve an assessment of the relationship of a given indicator to other indicators of quality included in the QRIS. In studies that examine measures to be included together in a QRIS, it is important to look at the degree of correlation found among these measures: ideally, measures will be moderately correlated so that each measure both contributes to an overall assessment of quality yet also provides some non-redundant program quality information (Zellman et al., 2008). Correlation patterns should make sense. For example, two measures of interaction quality should be more closely related to each other than to a measure of adult–child ratios. If such studies reveal for example that the correlation between ratios and interaction processes is very high ($r = .90+$) this result might argue for eliminating one or the other indicator from the QRIS, as they may not be providing unique information (although some QRISs include certain quality elements to ensure that they are paid attention to for other policy related reasons, even if their psychometric properties are not ideal). To date, the Maine and Indiana validation research has not included a comparison of measures internal to the QRIS rating systems, but this is recommended in future research as the systems mature and stabilize.

Measurement error presents another potential challenge in assessing QRIS validity. Most QRISs assume that observational measures are relatively stable over time absent quality improvement efforts. This assumption is consistent with empirical evidence for at least one widely used instrument, the ERS (Clifford, 2005).

A related measurement issue concerns inter-rater reliability. In the twenty systems reviewed by Tout et al. (2010), nearly all QRISs require 80–85% agreement with a master coder (either exact agreement or agreement within one scale point) on ERS; this degree of reliability does not eliminate errors in ERS measurement (Bryant, 2010; Bryant, Burchinal, & Zaslow, 2011). For instance, two raters could be 100% reliable under a standard of 85% agreement within one scale point, but one might give a classroom a score of 3.5 and the other a score of 4.5, a difference that is large enough to affect an overall program rating (Karoly, Zellman, & Perlman, 2013). Based on the range and degree of variability in ERS quality scores at each rated level in both Maine and Indiana QRISs (see Tables 3 and 4), we recommended that program managers strive to increase the reliability of the rating process by clearly defining quality indicators and rating procedures, and conducting regular reliability checks.

Approach 3: assess the outputs of the rating process

A third validation approach focuses on assessing the *outputs* of the rating system: the scores and levels assigned to providers who undergo a rating, and the distributions of those scores within and across different types of providers. Studies conducted under this approach examine the degree to which the quality levels in the QRIS are meaningfully distinct from each other. The results of these studies may provide data that suggest that measures, cut scores, or rules for combining measures need to be changed in order to distinguish the rated quality levels effectively. Because these studies can result in proposals for significant changes to the standards for QRIS levels, it is helpful for these studies to occur prior to studies that examine associations between quality levels and children's development.

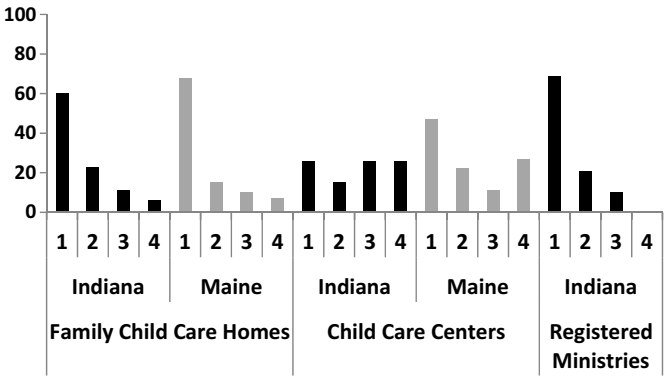


Fig. 1. Percent enrolled providers rated at four quality levels, by state and type of child care.

Output studies may focus on individual indicator scores, such as how providers score on an environmental rating, as well as on the overall quality level that is the final output of the rating process. These studies may also utilize a measure of quality not included in the QRIS rating process to make an evaluation of concurrent validity, by examining whether assessments on both measures co-vary in predictable ways. The following section provides examples of the two states' examinations of the distribution of quality ratings and rating-level advancement patterns for each program type enrolled in the state QRIS.

Examining initial QRIS rating distributions and cut points

While evaluators in Indiana and Maine did not conduct a detailed examination of the weighting or internal consistency of specific quality indicators, they did analyses to reveal the distribution of quality levels. After three years of system implementation, both Indiana and Maine QRIS child care providers were predominately rated at Level 1 or Level 2 (see Fig. 1). It is important to note that in Indiana, all providers enter the system at Level 1, and in Maine, providers can enter the system at any level based on their program rating, and then may advance at will from that level. A recent in-depth study of five state quality rating and improvement systems that were fully implemented found a similar pattern, with four of the five states reporting 40–76% of all programs enrolled in the lower tiers of the system (Mathematica Policy Research, 2011).

In Indiana, licensed child care centers were evenly distributed across the four QRIS levels approximately two years after the program inception. However licensed family child care homes were most frequently found at Level 1, with steeply declining numbers at the other three levels. This higher proportion of Indiana licensed centers rated at Level 3 or Level 4 may have been due to a greater historical emphasis in child care centers than in homes on regulation and attaining national accreditation, greater organizational capacity to complete the requirements of advancement in QRIS, or possibly that QRIS standards more closely reflect center quality than family child care home quality. Unlicensed registered child care ministries, a unique type of child care center in Indiana that is not licensed due to religious affiliation, participated at a much lower rate, and none had yet attained Level 4, reflecting significant challenges facing these unlicensed centers in meeting the Level 1 standards needed to enter the QRIS. These data patterns in Indiana

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Table 3
Indiana QRIS: mean global quality ERS scores[a] as a function of program type and rated quality level.

	Level one (n = 84)	Level two (n = 90)	Level three (n = 74)	Level four (n = 66)
All providers (N = 314)	3.2 (.87)	3.7 (.76)	3.8 (.73)	4.3 (.80)
Family child care homes (n = 167)	2.9 (.64)	3.4 (.75)	3.6 (.67)	4.0 (.89)
Licensed child care centers (n = 133)	4.0 (.77)	4.0 (.68)	4.3 (.66)	4.5 (.67)
Unlicensed registered child care ministries (n = 14)	3.2 (.95)	4.1 (.45)	4.0 (.18)	NA

[a] Possible range = 1–7.

Table 4
Maine QRIS: mean global quality ERS scores[a] as a function of program type and rated quality level.

	Step one (n = 82)	Step two (n = 99)	Step three (n = 79)	Step four (n = 82)
All providers (N = 342)	3.7 (.77)	3.9 (.84)	4.0 (.80)	4.3 (.79)
Family child care homes (n = 129)	3.3 (.67)	3.5 (.80)	3.8 (.91)	4.2 (.83)
Licensed child care centers (n = 165)	3.9 (.72)	4.1 (.80)	4.2 (.68)	4.4 (.84)
Head start centers (n = 48)	NA	NA	4.1 (.75)	4.5 (.71)

[a] Possible range = 1–7.

supported the validity of the QRIS rating system in that they showed variation in quality ratings across participating providers, they reflected the increasing effort necessary to meet quality standards at higher levels, and they were interpretable within the state's child care context.

Another gauge of overall quality rating system utility is the amount of program advancement to higher rated levels. It is reasonable to expect, if the QRIS is viable, that at least some providers will advance in quality level. In the Indiana evaluation, 19% of the licensed centers, 24% of the licensed homes, and 27% of the unlicensed ministry centers had advanced at least one QRIS quality level in a 6-month period between assessments, during which mentoring was provided by local training providers. This advancement pattern, if maintained over time, suggests that even though attaining the highest levels may be challenging, quality improvement is feasible.

For Maine, as Fig. 1 illustrates, center-based programs and family child care type programs are most frequently found at Step or Level One. A disproportionately small number of family child care programs have attained Step Four, the highest quality level, and a disproportionately large number of child care centers and Head Start programs are enrolled at Step Four. This pattern of fewer family child care homes enrolled at higher Step levels has existed throughout QRIS implementation in Maine. Maine family child care home providers argued that some of the program standards were not "a good fit," despite designers' beliefs that standards were well-matched to setting type. The large number of center-based and Head Start programs at the higher Step levels was expected, given that QRIS quality standards closely align with accreditation standards, and center-based programs are more likely to be nationally accredited than family child care homes.

An assumption of the designers of the QRIS in Maine was that programs engaged with QRIS will improve their tier levels consistently over time (Lahti et al., 2011). Approximately 80% of all programs (n = 1118) in the QRIS observed during the study period 2008 through 2011 did not experience a move up from one Step Level to the next. Results indicated that 95 of the 103 events or changes in Step Level from level one to two occurred during the first 23 months of enrollment. Moving from a Step One to Two, center-based care programs had a hazard probability of just .02 while family child care homes stayed virtually flat during this early period of enrollment in the QRIS. For movement from Step Two to Three, neither program type (p = 290) nor regional location (p = 195) appear to be significant in explaining Step level movement. For movement from Step Three to Four, the highest tiers in Maine's QRIS, the analysis indicated that only type of program is a significant covariate explaining advancement. Family child care homes appeared to have a significantly lower probability of advancing a Step Level at this highest quality tier, compared with center-based and Head Start programs. These types of analyses of program movement in the system are relevant to the validation of a QRIS as they illustrate whether or not the way the system as designed is meeting its goals of supporting program advancement, leading to statewide improvement of program quality.

Studies may also be conducted to examine the degree to which given measures relate to other measures that purportedly assess the same concept. Here, strong correlation is desired, as they suggest that measures are measuring the concepts that they purport to measure in ways that are consistent with other measures of the same concepts.

Indiana: assessing the output of the rating process

The Indiana evaluation research included one validation test of state committee-generated quality standards, indicators, and levels by comparing the outputs of the QRIS rating system with independently gathered assessments of quality using validated quality measures, the environmental rating scales (ERS, Harms et al., 2005, 2006, 2007) and the Caregiver Interaction Scale (CIS; Arnett, 1989). The results, originally published by Elicker et al. (2011) and shown in Table 3, indicate that ERS scores co-varied as expected with QRIS level ratings, with a significant mean difference in global scores of 1.1 scale points between Level 1 and Level 4. Table 5 shows that caregiver interaction as observed using the CIS was less related to the rated QRIS quality levels. The overall correlation between the 4-level QRIS ratings and global ERS quality scores was moderate (r = .42, p < .01). The correlation between CIS adult–child positive interaction scores and QRIS level was more modest, but positive and significant (r = .24, p < .01).

Taken together, and looking across all types of providers, these results suggest that the QRIS ratings distinguish levels of quality in somewhat similar ways as two time-tested, validated measures of child care quality. However, mean quality levels at Level 4 were mostly found to be below the "good" rating threshold, suggesting the need to strengthen standards and/or rating procedures at the highest QRIS levels. In addition, finer analysis of the data suggested specific recommendations about quality standards and rating procedures that might be improved for each type of child care. Summaries of individual ERS item means for Level 3- and 4-rated providers led to the identification of a number of ERS items with scores below 4. Program planners are currently improving standards and QRIS rating procedures in light of these findings (Elicker et al., 2013).

In Indiana, patterns of association between QRIS ratings and ERS ratings were not the same for all types of child care. While the

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Table 5
Indiana QRIS: mean total positive interaction (CIS) scores[a] as a function of program type and rated quality level.

	Level one (n = 84)	Level two (n = 87)	Level three (n = 74)	Level four (n = 65)
All providers (N = 310)	3.0 (.44)	3.1 (.46)	3.2 (.37)	3.2 (.52)
Family child care homes (n = 164)	2.9 (.37)	3.1 (.43)	3.2 (.36)	3.3 (.48)
Licensed child care centers (n = 132)	3.2 (.44)	3.1 (.50)	3.2 (.41)	3.2 (.54)
Unlicensed registered child care ministries (n = 14)	2.8 (.58)	3.2 (.48)	3.5 (.11)	NA

[a] Possible range = 1–4.

global ratings were significantly correlated in both licensed centers and licensed family child care homes, the strength of association was stronger for homes, meaning QRIS level ratings in homes more clearly distinguished levels of ERS-related quality, meaning at each QRIS-rated level, the ERS quality differences were generally greater than they were for centers. Second, the overall ERS quality levels for center-based preschool classrooms (using ECECRS-R; M = 4.6 at Level 4) were somewhat higher than for center-based infant classrooms (ITERS-R; M = 4.4 at Level 4) and family child care homes (FCCERS-R; M = 4.0 at Level 4). While the equivalence of quality scores across these three ERS scales is not supported by research evidence, the results taken together suggest the need to strengthen quality standards and assessment procedures for all types of care, so that child care providers at the highest rated levels are providing care that is at or above threshold levels recommended to impact children's developmental outcomes (Zaslow, Martinez-Beck, Tout, & Halle, 2011).

Maine: assessing the output of the rating process

As in Indiana, differences in program quality were measured using Environmental Rating Scales (ERS) mean scores at the classroom level, and these scores were not part of the QRIS standards or ratings. The results presented here are from factorial ANOVAs to examine the effects of Step Level, ERS scale type and child care program type on the dependent variable ERS mean score. Table 4 provides the adjusted mean scores for all 307 classrooms and by each program type by Step Level. Table 4 is extracted from the full report on Maine's QRIS (see Lahti et al., 2011).

The results show an overall significant difference between Step Level and ERS mean score at the classroom/setting level (F = 5.02; df = 3, 307; p = .002). Results of post hoc Bonferroni tests showed a significant difference between Step One and Step Four programs (p = .001) and between Step Two and Step Four programs (p = .001). The total variance of the mean ERS score explained by Step Level was only 5%, indicating weak relationships between the variables. Comparisons of the program type mean ERS quality scores indicated a difference only between the family child care home scores and the center-based scores (p < .001). The family child care home mean scores were lower at each Step Level than the center-based setting scores with the exception of scores at Steps Three and Four. There did not appear to be any significant differences at Step Three or Four between the center-based and Head Start type settings (p = .97). The results provide some evidence for differences in rated quality, with higher ERS means for higher tier or step programs, most distinctly for family child care homes. Overall these mean scores suggest the need for considerable efforts at quality improvement, considering that the majority of settings are scoring below the "5" or "good" level on the ERS measures.

Maine: parent level data on QRIS program quality

In the Maine validation study, parents in programs selected for observation were asked to complete an anonymous survey that focused on services received by the parent and the parent's perceptions of the quality of the program. The belief was that parents served by higher Step level programs should be receiving more supports and services and therefore may rate the program higher in level of quality. The response rate over the three-year study period was approximately 26% (N = 1478). These results are extracted from the full report on Maine's QRIS (see Lahti et al., 2011).

Parental perception of program quality was measured by the 15-item Emlen scale, see Emlen, Koren, and Schultze (2000), and was found not to be correlated to Step Level rating (Pearson's r = .010, p = .68). In terms of services parents should have received according to program standards, a majority of parents reported not receiving: information about other government services for their child; opportunities for parent engagement with the program; daily communication from the program about their child; and being provided an up to date written parent hand-book from their provider. There did not appear to be any difference in step level in terms of parents not consistently receiving these types of services as required by the program quality standards according to parent reports. The use of these data by QRIS administrators was primarily for monitoring purposes focused on services and or supports parents should have received based on requirements in the QRIS standards. While parents were asked about perceptions of program quality, due to a strong desire to reach out to parents as a key stakeholder in the QRIS, that information was not relied upon for program planning or program improvement.

Approach 4: relate ratings to children's development

The fourth approach to validation focuses on children's development. In many respects, this is the final step in validating a QRIS, and one that arguably should be delayed until the questions raised in the earlier approaches are addressed and changes made to the system as necessary. It may even be possible that new data will emerge that makes the costly and difficult effort involved in assessing child outcomes unnecessary. For example, if studies begin to show consistently that certain inputs, e.g., ratings-based coaching lead to substantial improvements in indicators such as instructional support, and if instructional support or other indicators is found to consistently promote improved child outcomes, it may be possible to argue that the inclusion of those inputs and measures of those outputs may suffice.

The logic models that underlie QRISs typically assert that higher quality care will be associated with improved child outcomes. Therefore, one important piece of validation evidence concerns whether children make greater developmental gains in programs with higher program-level QRIS ratings than in programs with lower ratings. While a definitive evaluation of QRIS impact on child outcomes would consist of an experimental study with random assignment of providers and children to QRIS levels, ethical and practical considerations often make experiments impractical, at least on a state-level scale. Instead, current studies evaluating QRIS validity in terms of child outcomes using this approach do not attempt to evaluate causal linkages. Instead, they examine whether the QRIS ratings and the quality components that comprise the ratings are associated in expected ways to measures of children's development. Showing significant associations between QRIS-rated quality would be a first step, a necessary but not sufficient result to demonstrate causal inferences about how QRIS quality influences children's outcomes.

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

To date, few QRIS validation studies have incorporated children's outcomes. Maine did not include this approach to validation. As Elicker and Thornburg (2011) note, results from such studies are mixed, at least in part because of the challenges of conducting them. A primary challenge is the inability to control for all the factors that may confound the quality-outcome correlations for children whose families have selected programs in a non-random way. Additional challenges include the difficulty of recruiting of programs and children across all quality levels; lack of information about the amount of care children received in each setting (dosage); lack of appropriate outcome measures for children of diverse ages, abilities, cultures and linguistic backgrounds; and, lack of variation in the quality of participating QRIS programs. As noted above in the discussion of Approach 3, measurement error remains a problem.

Indiana: examination of ratings associated with children's outcomes

To examine validity-related questions about children's development in the context of the Indiana QRIS, the evaluators assessed the developmental status of 557 children (249 infants/toddlers; 308 preschoolers) who were in the care of QRIS providers. Two children per classroom or home were randomly selected in approximately equal numbers at all four QRIS levels. Data from parent interviews describing annual family income and parents' education levels and participation in the CCDF voucher program were used as control covariates in the analyses. The basic validity question explored was: are children in higher-rated QRIS care functioning at higher levels, socially and cognitively, than children in lower rated care?

It is important to point out that this study of quality and child outcome associations was cross-sectional, with all data collected at one point in time. As mentioned earlier, exploring these correlational relationships does not substitute for longitudinal or experimental designs that can better evaluate the causal impact of the QRIS on child outcomes. However in the implementation phase of QRIS, it is useful to explore the developmental status of participating children, how they are distributed in the child care system, and whether associations between quality measures and measures of children's functioning are occurring in the expected direction (Elicker & Thornburg, 2011).

Bivariate correlations and multiple regression models were used to explore the associations between children's development and the three measures of child care quality: QRIS ratings (4 levels); ERS global quality scores; and a CIS composite rating of positive adult–child interactions. All regression models included parent education, household income, and type of child care. No significant correlations were found between the four-level QRIS quality ratings and either infant/toddler or preschooler developmental status. Some of the researcher-observed quality measures were mildly but significantly correlated with child development measures. For preschoolers, CIS positive interactions were correlated with social competence ($r = .17^{**}$) and receptive language ability ($r = .17$, $p < .01$). For infants and toddlers, ERS global quality scores were associated with social competence ($r = .15$, $p < .01$), and total CIS positive interactions were associated with cognitive/language competence ($r = .17$, $p < .01^{**}$). These significant correlates were entered as predictors in regressions of child outcomes on the quality variables, controlling for the family SES variables (parent education level and household income) that were also significantly correlated with the child outcome variables. As a result, for preschoolers, CIS positive adult–child interactions significantly predicted children's receptive language ability, after controlling for family SES ($b = .12$, $p < .05$). For infants and toddlers, CIS positive adult–child interactions significantly predicted children's cognitive/language competence, after controlling for family SES ($b = .14$, $p < .05$).

Family income was also a significant predictor, $b = .23$, $p = .009$.

Therefore while QRIS rated levels were not significantly associated with any child development measures for either infants/toddlers or for preschoolers, ERS and CIS quality measures were moderately associated with aspects of children's development. Specifically, after controlling for family SES, it was the positive quality of interaction between adults and children that was associated with language and cognitive functioning, for both preschoolers and infants and toddlers.

Therefore it appeared that the specific aspects of child care quality assessed by the ERS and CIS measures are more likely to be associated with children's development than are the composite of quality indicators represented by the 4-level QRIS ratings. This was true even though the QRIS ratings and the ERS and CIS were significantly correlated with each other. As a result, in Indiana, further refinement of QRIS standards and procedures is taking account of these findings, especially by identifying ways to strengthen QRIS standards and ratings to inlcude the quality of adult–child interactions.

Discussion

Limitation to validation study designs

Both of these state studies provide results that describe linear associations among variables. The study designs are limited due to the fact that the investigators have no control over how the QRIS systems are implemented which affects enrollment and therefore sample sizes and selection of measurement strategies were also not in the sole control of the investigator. It will be interesting as additional studies are done and where non-linear associations are found to determine the impact this has on outcomes. These field studies were conducted with all the limitations associated with working in a developing system with multiple stakeholders. While the design presents a limitation in terms of arguing for causality and application of more sophisticated analytic approaches, it should be noted that the state agency program managers and other stakeholders in both states found the information generated from these studies of high value in terms of system planning, program improvements, and resource allocation. Depending upon context and resources, limitations to these two study designs can be remedied in future studies by such design choices as having programs that are on a waiting list be compared to programs already participating in their state's QRIS.

Validation of QRIS is a process that needs attention over time, using more than one approach

The examples from Indiana and Maine illustrate how these validation approaches can work in practice, with tangible benefits for system improvement. These validation activities are specific to the design and implementation of each state QRIS. We believe it is important to stress to QRIS policy leaders that each of the four validation approaches needs to be used appropriately, considering the developmental stage of the QRIS and the unique features of the setting and QRIS. For example, states with QRIS in development can use the four approaches as a framework for planning how to validate their system. Developmentally then, an initial focus of QRIS design would be to validate the key concepts used in the QRIS design. The four approaches highlighted in this paper need to be considered as part of an ongoing process, not a one-time event. As states progress in their implementation of QRISs, more descriptive research is necessary to understand better how these validation approaches work in other settings, for example with point-based QRISs (rather than "building block" QRISs, like those or Indiana and Maine). Use of these approaches enables cross system comparisons,

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

which will allow for the identification of common threats to validity and useful strategies to enhance the validity of a state QRIS.

Validation and early care and education system constraints

A QRIS is not merely a program-level quality-improvement intervention, it is a policy lever for strengthening a state's overall early care and education system that reaches beyond child care (Schaack, Tarrant, Boller, & Tout, 2012). The two state validation efforts highlighted in this paper reflect the challenges and constraints common to other state experiences with validation activities (Lahti, Sabol, Starr, Langill, & Tout, 2013). On-site observations of global program quality and establishing and maintaining inter-rater reliability for QRIS raters is a time consuming and costly endeavor. Keen interest in school readiness may pressure program administrators to collect child outcome level data before a QRIS is well established. Current research on the measures that are in the widest use to predict child outcomes appear to do so consistently, especially for children at-risk, but with modest levels of association with program quality as measured by the ERS's (for example, Burchinal et al., 2011). We recommend to policy makers to always take into account that any validation study is occurring within a dynamic system. System-level constraints such as varying resources available to programs, different type and design of programs, and challenges to measuring quality and reliably collecting information about program quality all influence the design and implementation of state-level QRIS validation studies.

Validation research is critical for performance measurement and improvement for a state QRIS

The 2012 Child Care Development Fund (CCDF) Plan preprint for fiscal years 2014–2015 includes a much larger focus on QRISs (U.S. Department of Health & Human Services, Administration for Children and Families, 2011). In this document, a QRIS is defined as a "...systematic framework for evaluating, improving, and communicating the level of quality in early childhood programs." States are expected to provide a self-assessment based on current program quality initiatives from a set of questions that are also organized according to a "QRIS framework." Validation of program standards or assessment tools is mentioned specifically in relation to information states must provide about data and performance measures on program quality.

The information generated from QRIS validation activities can be used to inform efforts for continuous quality improvement. For example, both Indiana and Maine found that, for at least some types of providers, enrollment patterns in the QRIS, and lack of movement by programs once they are in the QRIS, is resulting in a large proportion of providers at the lower-rated quality levels of the quality tiers. System-level, quality-improvement responses to this information could be to re-assess the design of the system in terms of the ability of programs to meet standards at each tier, or to focus training/technical assistance on specific quality standards that are most challenging for providers to meet. At the same time, care should be taken to ensure that standards reflect current knowledge about the specific indicators and levels of quality most likely to produce the desired child outcomes. Findings from validation studies can be part of the information that state child care administrators use to assess the overall performance of the state early care and education system. The performance data could then be used by program administrators in making decisions about monitoring programs in a differential manner by visiting those programs more often who are having difficulty meeting QRIS standards (Fiene, 2013). We recommend that one focus of future research be learning more about whether and how information from the results of validation studies are used to improve system and program level performance in QRIS.

It will be interesting to determine as more validation studies are completed to analyze the differences between levels and how often lower quality is present in the top level which is the case with measuring compliance with licensing standards (Fiene & Nixon, 1985). Key areas to look at will be the movement of programs from one level to another, how long this takes, and are the increments equal or not in terms of quality improvement.

Another area to be explored which may have an impact on overall QRIS implementation are the fiscal constraints that many states are experiencing due to the recent recession and lower levels of federal funding. It would be interesting to note differences amongst states with large investments in quality improvement initiatives and those states with smaller investments.

It is important to remember that the QRIS is a policy lever and the validation of child care quality standards in a QRIS is a new phenomenon in early care and education policy-making. These four recommended approaches to QRIS validation, illustrated by validation research in these two states, even with their limitations, did provide policy makers and program administrators with information that guided efforts at system quality improvement. The use of these approaches in other studies will create a common nomenclature for better understanding threats to validity in a QRIS and ultimately increase our understanding of how best to design a QRIS that meets the needs of the parents, providers and children it serves.

References

Alkon, A. (2013). *National health and safety tool*. San Francisco, CA: California Child Care Health Program.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*, 541–552. http://dx.doi.org/10.1016/0193-3973(89)90026-9

Berry, D. J., Bridges, L. J., & Zaslow, M. J. (2004). *Early childhood measures profile*. Retrieved from: http://aspe.hhs.gov/hsp/ECMeasures04/report.pdf

Briggs-Gowan, M. J., & Carter, A. S. (2002). *Brief infant-toddler social and emotional assessment (BITSEA) manual, version 2.0*. New Haven, CT: Yale University.

Bryant, D. M. (2010). *Observational measures of quality in center-based early care and education programs (OPRE Research-to-Policy, Research-to-Practice Brief 2011-10c)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Bryant, D. M., Burchinal, M., & Zaslow, M. (2011). Empirical approaches to strengthening the measurement of quality: Issues in the development and use of quality measures in research and applied settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 33–47). Baltimore, MD: Paul H. Brookes.

Burchinal, M. R., Peisner-Feinberg, E. S., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality. Testing for differential associations related to poverty, gender, or ethnicity. *Journal of Applied Developmental Science*, *4*, 149–165. http://dx.doi.org/10.1207/S1532480XADS0403_4

Burchinal, M. R., Vandergrift, N., Pianta, R., & Mashburn, A. J. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in prekindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Burchinal, M. R., Kainz, K., Cai, Y., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2009). *Early care and education quality and child outcomes (Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief #1)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services, and Child Trends.

Burchinal, M. R., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–27). Baltimore, MD: Paul H. Brookes.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, *25*(2), 140–165. http://dx.doi.org/10.1016/j.ecresq.2009.11.001

Cizek, G. J. (2007, August). Introduction to validity. In *Presentation to the National Assessment Governing Board of NAEP* Washington, DC.

Clifford, R. (2005). Structure and stability in the early childhood environment rating scale. In H. Schonfeld, S. O'Brien, & T. Walsh (Eds.), *Question of quality: Proceedings of a conference on defining, assessing and supporting quality in early childhood care and education* (pp. 12–21). Dublin, Ireland: Centre for Early Childhood Development and Education, St. Patrick's College.

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Cochran, M. (2007). *Finding our way: The future of American early care and education*. Washington, DC: Zero to Three.
Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service.
Elicker, J., Langill, C. C., Ruprecht, K., & Kwon, K. (2007). *Paths to QUALITY™: A child care quality rating system for Indiana. What is its scientific basis?* Lafeyette, IN: Purdue University, Center for Families.
Elicker, J., & Thornburg, K. (2011). *Evaluation of quality rating and improvement systems in early childhood programs and school age care: Measuring children's development (Research to Policy, Research to Practice Brief, OPRE 2011-11c)*. Washington, DC: Department of Health and Human Services, Administration of Children and Families, Office of Planning, Research, and Evaluation.
Elicker, J., Langill, C. C., Ruprecht, K., Lewsader, J., & Anderson, T. (2011). *Evaluation of "Paths to QUALITY" Indiana's child care quality rating and improvement system*. West Lafayette, IN: Purdue University Center for Families.
Elicker, J., Langill, C., Ruprecht, K. M., Lewsader, J., Anderson, T., & Brizzi, M. (2013). Paths to QUALITY: Collaborative evaluation of a new child care quality rating and improvement system. *Early Education and Development*, *24*(1), 42–62. http://dx.doi.org/10.1080/10409289.2013.736127
Emlen, A., Koren, P., & Schultze, K. (2000). *A packet of scales for measuring quality from a parent's view*. Portland, OR: Portland State University, Regional Research Institute for Human Services, The Oregon Child Care Research Partnership.
Fiene, R. (2013). *Differential monitoring logic model and algorithm (DMLMA): A new early childhood program quality indicator model (ECPQIM) for early care and education regulatory agencies*. Middletown, PA: Research Institute for Key Indicators.
Fiene, R., & Nixon, M. (1985). Instrument based program monitoring and key indicators in child care. *Child Care Quarterly*, *14*(3), 198–214.
Fuller, B., Loeb, S., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type. *Quality, and Stability, Child Development*, *75*(1), 47–65.
Harms, T., Clifford, R., & Cryer, D. (2005). *Early childhood environmental rating scale–revised*. New York, NY: Columbia University Teachers College Press.
Harms, T., Cryer, D., & Clifford, R. (2006). *Infant toddler environmental rating scale–revised*. New York, NY: Columbia University Teachers College Press.
Harms, T., Cryer, D., & Clifford, R. (2007). *Family day care environmental rating scale–revised*. New York, NY: Columbia University Teachers College Press.
Howes, C., Whitebook, M., & Phillips, D. (1992). Results on the National Child Care Staffing Study concerning effectiveness and characteristics of teachers across the United States. *Child & Youth Care Forum*, *21*(6), 399–414.
Howes, C., Burchinal, M., Pianta, R. C., Bryant, D. M., Early, D. M., Clifford, R. M., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Child & Youth Care Forum*, *37*(1), 27–50. http://dx.doi.org/10.1016/j.ecresq.2007.05.002
Kane, M. T. (2001). Current concerns in validity theory. *Journal of Education Measurement*, *38*, 319–342.
Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly*, *28*, 645–657. http://dx.doi.org/10.1016/j.ecresq.2013.05.001
LaFreniere, P. J., & Dumas, J. E. (1997). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*, *8*(4), 369–377. http://dx.doi.org/10.1037//1040-3590.8.4.369
Lahti, M., Cobo-Lewis, A., Dean, A., Rawlings, S., Sawyer, E., & Zollitsch, B. (2011). *Maine's quality for ME – child care quality rating and improvement system (QRIS): Final evaluation report*. Portland, ME: Cutler Institute for Health and Social Policy, Muskie School of Public Service, University of Southern Maine.
Lahti, M., Sabol, T., Starr, R., Langill, C., & Tout, K. (2013). *Validation of quality rating and improvement systems (QRIS): Examples from four states (Research-to-Policy, Research-to-Practice Brief OPRE 2013-036)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Lugo-Gil, J., Sattar, S., Boss, C., Boller, K., Tout, K., & Kirby, G. (2011). *The quality rating and improvement system (QRIS) evaluation toolkit (OPRE Report #2011-31)*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.
Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service, Inc.
NICHD Early Child Care Research NICHD Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Sciences*, *4*, 116–135.
NICHD Early Child Care Research NICHD Network. (2002). Child care structure → process → outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science*, *13*, 199–206.
Maine, Department of Health and Human Services. (2008). *Quality for ME: Child care quality rating system program standards*. Retrieved from: http://www.maine.gov/dhhs/ocfs/ec/occhs/quality_cbc.pdf
Mathematica Policy Research. (2011). *Defining and measuring quality: An in-depth study of five child care quality rating and improvement systems (OPRE Report 2011-29)*. Retrieved from: http://www.whitehouse.gov/sites/default/files/omb/budget/fy2011/assets/management.pdf
Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual, Pre-K*. Baltimore, MD: Paul H. Brookes.
QRIS National Learning Network (2014). Retrieved from: http://qrisnetwork.org/qris-state-contacts-map
Schaack, D., Tarrant, K., Boller, K., & Tout, K. (2012). Quality rating and improvement systems, frameworks for early care and education systems change. In S. L. Kagan, & K. Kauerz (Eds.), *Early childhood systems, transforming learning* (pp. 71–86). New York, NY: Teachers College Press.
Smith, S., Robbins, T., Stagman, S., & Kreader, J. L. (2012). *Practices for promoting young children's learning in QRIS standards*. Retrieved from: http://www.nccp.org/publications/pdf/text_1070.pdf
The Concept System® (2012). BT computer software. PN Concept Systems, Incorporated, PL Ithaca, NY.
Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the next decade of quality rating and improvement systems, OPRE Issue Brief*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality ratings systems and evaluations*. Washington, DC: Child Trends and Mathematica Policy Research, for Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.
U.S. Department of Health and Human Services, Administration for Children and Families. (2011). *Program instruction (Child Care and Development Block Grant Act of 1990)*. Retrieved from: http://www.acf.hhs.gov/sites/default/files/occ/pi2011_03.pdf
Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes.
Zellman, G. L., & Fiene, R. (2012). *Validation of quality rating and improvement systems for early care and education and school-age care (Research-to-Policy, Research-to-Practice Brief OPRE 2012-29)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Zellman, G. L., Perlman, M., Le, V., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar Early Learning quality rating and improvement system as a tool for improving child-care quality (MG-650-QEL)*. Santa Monica, CA: RAND Corporation.
Zellman, G. L., Brandon, R. N., Boller, K., & Kreader, J. L. (2011). *Effective evaluation of quality rating and improvement systems for early care and education and school-age care (Research-to-Policy, Research-to-Practice Brief OPRE 2011-11a)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Zellman, G. L., & Karoly, L. A. (2012). *Incorporating child assessments into state early childhood quality improvement initiatives*. Santa Monica, CA: RAND Corporation.

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Childhood Research Quarterly* (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005

Improving Child Care Quality Through an Infant Caregiver Mentoring Project

Richard Fiene
The Pennsylvania State University

ABSTRACT: An evaluation of a mentoring training program for infant caregivers is described. Fifty-two infant caregivers from 27 childcare center-based programs were involved in a four month long intervention in which they were paired with an experienced early childhood educator. The focus of the mentoring program was to improve the overall quality of the classroom environment, as well as making the caregivers more sensitive to the needs of the infants. The results clearly indicated that the mentoring program was very effective in improving the overall quality of the classroom, as well as making caregivers more sensitive to infants' needs.

KEY WORDS: infant caregivers; childcare; mentoring; training.

Introduction

This paper describes a child care mentoring project designed to improve the quality of infant and toddler child care programs in south central Pennsylvania. The goal of the mentoring project was to improve the quality of the child care environment and specifically the quality of caregiver-child interactions. As most caregivers in Pennsylvania only receive workshop training, the goal of this project was to compare the mentoring approach to the more typical workshop training. Mentoring is being explored because of its targeted intensive one-on-one nature in delivering training to caregivers based upon needs assessments. The project was conducted during the later half of 2000 and the beginning of 2001. The results presented in this paper are part of the pre- and post-test data collection phase (summer 2000 and winter 2000–2001) of this mentoring project. The actual mentoring intervention occurred from September through December 2000.

Correspondence should be directed to Richard Fiene, Ph.D., Director, Capital Area Early Childhood Training Institute, and Senior Research Associate, Prevention Research Center, The Pennsylvania State University, 2001 North Front Street, Suite 314, Harrisburg, PA 17102; e-mail: rjf8@psu.edu, Web page: http://caecti.org.

Mentoring in childcare has been documented in the literature for the past 10–15 years (Breunig & Bellm, 1996; Fenichel, 1992). It has been demonstrated to be an effective mode of training/technical assistance (Breunig & Bellm, 1996). However, in the majority of studies conducted there are few, if any, demonstrations that utilize a randomized trial design (Breunig & Bellm, 1996). Many studies track the progress of the intervention group, some studies have comparison groups, but few, if any, have employed a randomized design. This research paper will describe the pre- and post-test data collected as part of a study that has employed a randomized design.

The majority of research (Clarke-Stewart, 1987; Goelman & Pence, 1987; Howes, 1987; Phillips, 1987; Kontos & Fiene, 1987; Galinsky, Howes, Kontos, & Shinn, 1994; Scarr, Eisenberg, & Deater-Deckard, 1994; Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997; Helburn, 1995; Fiene, 1995, 1996; Jorde-Bloom, 1988; Love, Schochet & Meckstroth, 1986) completed on early childhood quality has focused on preschool programs, with infant toddler programs rarely as the central focus of the research. The research completed in infant toddler programs has clearly documented the mediocre level of care provided to children in these programs (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997). In the present study, we focus on the first three years of life. All the centers and the classrooms reported upon in this study serve children from birth to less than three years of age.

This report is organized as follows: a methodology section briefly describes the sample selected with basic demographic information on directors, caregivers and the programs. This is followed by a results section that provides pre- and post-test average scores for each of the assessment tools utilized in this study to measure quality, caregiver behaviors, knowledge, and organizational climate of programs. This section is followed with a discussion section and implications regarding this mentoring project.

Methods

Study Design

This study involved 52 caregivers from 27 sites in south central Pennsylvania. All programs were child care centers licensed by the Department of Public Welfare. Seven of the sites were accredited by the National Association for the Education of Young Children.

This study employed a randomized design in which a self-selected group of programs and caregivers were randomly assigned to two groups, either the mentoring group or the comparison non-mentoring

comparison/control group. Intervention model mentoring group received intensive mentoring from a seasoned early childhood professional (minimum of 5–7 years of experience in the early childhood field as both a director and teacher) from September to December 2000. The mentoring model consisted of a problem solving approach in which the mentor spent a good deal of time observing in the beginning weeks in order to develop a trusting relationship with the protégé. Once both the mentor and protégé felt comfortable then suggestions could be entertained by the mentor.

The comparison group did not receive the mentoring intervention and only had the regular workshop type variety training available to them. However, the comparison group did receive mentoring during the Spring 2001 from March to June 2001. What is of interest in this study is to determine how much the two groups have improved from the pre-test data collection because they were essentially equivalent at that point on all measures.

Programs were recruited by the Capital Area Early Childhood Training Institute, a broad based community focused training institute. Program directors were invited to attend a meeting describing the mentoring project. Of those attending, 95% agreed to participate in the project. Fifty two caregivers started the project, 14 caregivers dropped out of the project between pre- and post-test. There was an equal drop out rate from both the mentoring and the control groups.

Data from the four quality measures used for all the programs are presented in Table 1. The four measures of quality were the Infant Toddler Environment Rating Scale (ITERS), the Arnett Caregiver Observation Scale, the Knowledge of Infant Development (KIDI), and the Bloom Scales of Organization Climate.

The program directors' average age is 31 with a range from 24–53

Table 1
ITERS, Arnett, KIDI, Bloom Scale Scores

All Programs (n = 38)	Pre-Test	Post-Test	Change	Significance
ITERS	134	140	+6	ns
Arnett	30	40	+10	ns
KIDI	14	14	-0-	ns
Bloom	78	79	+1	ns

years of age. They are predominantly Caucasian (81%). Eight percent have associate degrees, 78% have bachelor's degrees, and 14% have master's degrees. They had been employed as directors in their program for an average of 31 months with a range from 1 month to 120 months. Their average pay is between $20000–25000 per year. Sixty percent have health insurance and 45% have some form of dental or life insurance. Forty-five percent are in a retirement system.

The average age of caregivers in the programs was 36 with a range from 18–68. They are predominantly Caucasian (77%). Fifty-seven percent have high school diplomas, 16% have some college credits, 5% have CDA's, 16% have associate degrees, 5% have bachelor's degrees, and 2% have master's degrees. They have been employed as caregivers in their program for an average of 34 months with a range from 1 month to 153 months. They have worked in the early childhood field as caregivers for an average of 71 months with a range from 1 month to 312 months. Their average pay is between $10000–15000 per year. Fifty percent have health insurance and 33% have some form of dental or life insurance. Thirty-three percent are in a retirement system.

The average size of the centers is 98 children with 17 staff employed either full time or part time at the program. The average weekly fee for infant care is $137.00 per week and for toddler care is $124.00 per week. The majority of staff are employed at the centers for either less than 1 year or greater than 5 years.

Results

Both the mentoring and comparison groups were tested for equivalence at the beginning of the project in the pre-test data collection phase. There were no statistically significant differences on any of these measures at the pre-test. When the programs and caregivers were measured at the post-test, positive changes occurred although none were found to be statistically significant. In the aggregate, the programs that continued with the mentoring project showed improvements in the overall quality of care.

Tables 2 through 5 present the pre- and post-test data for the intervention and control groups.

These results indicate that the mentoring group showed increases on the program quality scales (ITERS and Arnett). This increase is especially noticeable on the ITERS. Further, there was a decrease in program quality with the control group, going from a score of 137 to 132. On the Arnett scale the mentoring group increased greater than the control group (11 point increase versus a 7 point increase).

Although the above results did not reach statistical significance,

Table 2
ITERS

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	134	141	+7	ns
Control Group	137	132	–5	ns

Table 3
Arnett

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	29	40	+11	ns
Control Group	33	40	+7	ns

Table 4
KIDI

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	14	14	-0-	ns
Control Group	14	15	+1	ns

Table 5
Bloom

	Pre-Test	Post-Test	Change	Significance
Mentoring				
Group	73	74	+1	ns
Control Group	87	91	+4	ns

when specific subscales are analyzed several show significant differences (see tables 6 and 7). Several of the subscales on the ITERS and Arnett reached statistical significance with positive changes in routines (greeting/departing, meals/snacks, nap time, diapering/toileting, health/safety practice/policy) learning activities (eye-hand coordination, active physical play, blocks, pretend play, cultural awareness), sensitivity, and appropriate discipline for the mentoring group. The only statistically significant finding with the control group was in a negative change in interactions in which the scores decreased from pre-test to post-test. Paired t-tests were used in all of these analyses for Tables 6 and 7.

Table 6
Mentoring Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	36	41	.005
Listening activities	8	9	ns
Learning activities	28	31	.05
Interactions	13	13	ns
Adult needs	17	19	ns
Arnett subscales			
Sensitivity	26	31	.001
Appropriate discipline	7	9	.05

Table 7
Control Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	41	42	ns
Listening activities	9	8	ns
Learning activities	29	31	ns
Interactions	15	13	.02
Adult needs	17	17	ns
Arnett subscales			
Sensitivity	28	31	ns
Appropriate discipline	6	7	ns

Discussion

These data demonstrate that the sites that were mentored improved on the ITERS and the Arnett. This is an encouraging result in that the intervention was only 4 months long. It is an important finding because the majority of mentoring projects in the past have utilized anecdotal evidence to demonstrate their effectiveness. Very few programs have conducted randomized trials of their interventions.

It is clear from the data that training/technical assistance interventions are needed in infant toddler programs because of the low scores on various program quality measures. It is also discouraging in that the control programs did not improve in which the ITERS went from 137 (pre-test) to 132 (post-test). This is a finding that will be monitored over time to see if this trend continues. Hopefully this was just an aberration in the data; however there does seem to be support when these data are compared to other studies (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997).

The public policy implications are that an intensive mentoring intervention of only four months can produce positive, although not statistically significant, changes in the overall quality of child care programs both globally and with caregiver interactions. Previous research (Johnson, 1994) has indicated that increasing the number of hours of training produces more developmentally appropriate behaviors in child care staff. Mentoring fits this model because it is an intensive one on one intervention in which the mentor and protégé are engaged in problem

solving activities to improve the overall quality of the interactions and environment of the child care program.

References

Behar, L. & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, 10, 601–610.

Breunig, R. & Bellm, D. (1996). *Early childhood mentoring programs: A survey of community initiatives*. Washington, D.C.: Center for the Child Care Workforce.

Clarke-Stewart, A. (1987). In search of consistencies in child care research. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 105–119). Washington, DC: National Association for the Education of Young Children.

Feagans, L. & Farran, D. (1979). *Adaptive Language Inventory*. Unpublished instrument. University of North Carolina, Chapel Hill.

Fenichel, E. (1992). *Learning through supervision and mentorship.* Washington, D.C.: Zero to Three: National Center for Infants, Toddlers and Families.

Fiene, R. (1995). Utilizing a statewide training system to improve childcare quality. *Child Welfare, 74*(6): 1189–1201.

Fiene, R. (1996). Using a statistical-indicator methodology for accreditation. In S. Bredekamp (Ed.), *A Decade of Accreditation*. Washington, DC: National Association for the Education of Young Children.

Galinsky, E., Howes, C., Kontos, S., & Shinn, A. (1994). *The study of children in family childcare and relative care: Highlights of findings*. New York, NY: Families and Work Institute.

Goelman, H. & Pence, A. (1987). Effects of childcare, family, and individual characteristics on children's language development: The Victoria day care research project. In D. Phillips (Ed.), *Quality in childcare: What does research tell us*? (pp. 89–104). Washington, DC: National Association for the Education of Young Children.

Griffin, A. & Fiene, R. (1995). *Childcare planning and evaluation*. Washington, DC: Zero to Three, National Center for Infants and Toddlers.

Harms, T. & Clifford, R. (1980). *The early childhood environment rating scale*. New York: Columbia University Teacher's College Press.

Helburn, S. (1995). *Cost, quality and child outcomes in child care centers*. Denver, CO: Center for Research in Economics and Social Policy, Department of Economics, University of Colorado.

Howes, C. (1987). Quality indicators in infant and toddler childcare: The Los Angeles study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us*? (pp. 81–88). Washington, DC: National Association for the Education of Young Children.

Hresko, W., Reid, D. K., & Hammill, P. D. (1981). *The Test of Early Language Development*. Austin, TX: Pro-Ed.

Iutcovich, J., Fiene, R., Johnson, J., Koppel, R., & Langan, F. (1997). *Investing in our children's future*. Erie, PA: Keystone University Research Corporation.

Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.

Jorde-Bloom, P. (1988). Assess the climate of your center: Use the early childhood work environment survey. *Day Care and Early Education*, Summer 1988: 9–11.

Katz, L. (1994). Perspectives on the quality of early childhood programs. *Phi Delta Kappan*.

Kontos, S. & Fiene, R. (1987). Childcare quality, compliance with regulations, and children's development: The Pennsylvania study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 57–79). Washington, DC: National Association for the Education of Young Children.

Love, J., Schochet, S., & Meckstroth, R. (1986). *Are they in any real danger: What*

research does—and doesn't—tell us about childcare quality and children's well being. Princeton, NJ: Mathematica Policy Research, Inc.

Maternal and Child Health Bureau (1997). *Stepping stones to Caring for Our Children*, Washington, D.C.: Maternal and Child Health Bureau, Department of Health and Human Services.

Morgan, G., Costley, C., Genser, A., Goodman, M., Lombardi, J., & McGimsey, J. (1993). *Making a career of it: The states of the states report on career development in early care and education*. Boston: The Center for Career Development in Early Care and Education, Wheelock College.

NAEYC (1987). *Accreditation standards*, Washington, D.C.: National Association for the Education of Young Children.

Scarr, S., Eisenberg, E., & Deater-Deckard, R. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9:* 131–151.

National Resource Center for Health and Safety in Child Care and Early Education

A Parent's Guide To

Choosing Safe and Healthy Child Care

Revised July 2019

A Parent's Guide To
Choosing Safe and Healthy Child Care

More and more, research tells us that our children's healthy development depends on safe and positive experiences during the first few years of life. If you are a parent/guardian who works during these early years, choosing good child care is one of the most important decisions you will ever make for your child.

To help you make the right choice for your child, researchers have identified 13 guidelines to think about when choosing a child care program. Research shows that if a program follows guidelines, it is more likely to be a safe and healthy place for your child. Considering these guidelines can help you find a place where you can feel comfortable leaving your child.

You might want to visit several different child care programs, either centers or family child care homes, before you decide which one is best for your family. Call each child care program and schedule an appointment for your visit. Once you are there, stay for at least an hour to watch activities, check the surroundings, and ask questions. Place checkmarks on this form to note which guidelines are met in each of the programs you visit. The checklist provides a place to make notes on up to 3 different child care programs.

Child Care Program Names (fill in): *A* *B* *C*	*Child Care Programs* A	B	C
 **Supervision** *Hint: Discipline should be positive, clear, consistent, and fair.*			
✔ Are children supervised at all times, even when they are sleeping?	☐	☐	☐
✔ How do the caregivers/teachers discipline children?	☐	☐	☐
 **Handwashing and Diapering** ***Hint**: Hands should be scrubbed with soap and water for at least 20 seconds and then rinsed and dried. The water faucet should be turned off with a paper towel.*			
✔ Do all caregivers/teachers and children wash their hands often, especially before eating and after using the bathroom or changing diapers?	☐	☐	☐
✔ Is the place where diapers are changed clean?	☐	☐	☐
✔ Do caregivers/teachers always keep a hand on the child while diapering?	☐	☐	☐
✔ Do caregivers/teachers remove the soiled diaper without dirtying any surface not already in contact with stool or urine?	☐	☐	☐
✔ Do caregivers/teachers clean and disinfect the surface after finishing the changing process?	☐	☐	☐

	Child Care Programs A	B	C
 **Director Qualifications**			
✔ Does the director of a child care center have a bachelor's degree in a child-related field?	☐	☐	☐
✔ Has the director worked in child care for at least 3 years?	☐	☐	☐
✔ Does the director understand what children need to grow and learn?	☐	☐	☐
 **Lead Teacher Qualifications**			
✔ Does the lead teacher in a child care center have a bachelor's degree in a child-related field?	☐	☐	☐
✔ Has the teacher worked in child care for at least 1 year?	☐	☐	☐
✔ Does the teacher give children lessons and toys that are right for their ages?	☐	☐	☐
 **Child: Staff Ratio and Group Size** ***Hint**: Your child will get more attention if each caregiver/teacher cares for fewer children.*			
✔ How many children are being cared for in the child care program?	☐	☐	☐
✔ How many caregivers/teachers are there?	☐	☐	☐
 **Immunizations**			
✔ Is your child up-to-date on all of the required immunizations?	☐	☐	☐
✔ Does the child care program have records proving that the other children in care are up-to-date on all their required immunizations?	☐	☐	☐
✔ Are the caregivers/teachers up-to-date on all of the recommended immunizations?	☐	☐	☐
 **Toxic Substances**			
✔ Are toxic substances like cleaning supplies and pest killers kept away from children?	☐	☐	☐
✔ Has the building been checked for dangerous substances like radon, lead, and asbestos?	☐	☐	☐
✔ Is poison control information posted?	☐	☐	☐
 **Emergency Plan**			
✔ Does the child care program have an emergency plan if a child is injured, sick, or lost?	☐	☐	☐
✔ Does the child care program have first aid kits?	☐	☐	☐
✔ Does the child care program have information about who to contact in an emergency?	☐	☐	☐

	Child Care Programs A	B	C
Disaster Drills			
✔ Does the child care program have a plan in case of a disaster like a fire, tornado, flood, blizzard, earthquake or acts of violence?	☐	☐	☐
✔ Does the child care program do practice drills once every month?	☐	☐	☐

Child Abuse			
✔ Can caregivers/teachers be seen by others at all times, so a child is never alone with one caregiver/teacher?	☐	☐	☐
✔ Have all caregivers/teachers gone through a background check?	☐	☐	☐
✔ Have the caregivers/teachers been trained how to prevent child abuse, how to recognize signs of child abuse, and how to report suspected child abuse?	☐	☐	☐

Medications			
✔ Does the child care program keep medication out of reach from children?	☐	☐	☐
✔ Are the caregivers/teachers trained and the medications labeled to make sure the right child gets the right amount of the right medication at the right time?	☐	☐	☐

Staff Training/First Aid ***Hint****: When infants are sleeping, are they on their backs with no pillows, quilts, stuffed toys, or other bedding in the crib with them?*			
✔ Have caregivers/teachers been trained how to keep children healthy and safe from injury and illness?	☐	☐	☐
✔ Do they know how to do first aid and CPR?	☐	☐	☐
✔ Have they been trained to understand and meet the needs of children of different ages?	☐	☐	☐
✔ Are all child care staff, volunteers, and substitutes trained on and implementing infant back sleeping and safe sleep policies to reduce the risk of SIDS (Sudden Infant Death Syndrome, crib death)?	☐	☐	☐

Playgrounds			
✔ Is the playground inspected for safety often?	☐	☐	☐
✔ Is the playground surrounded by a fence?	☐	☐	☐
✔ If there is a sandbox, is it clean?	☐	☐	☐
✔ Is the playground equipment safe, with no sharp edges, and kept in good shape?	☐	☐	☐
✔ Are the soil and playground surfaces checked often for dangerous substances and hazards?	☐	☐	☐
✔ Is equipment the right size and type for the age of children who use it?	☐	☐	☐

For more information on choosing a safe and healthy child care setting, contact your local child care resource and referral agency by checking Child Care Aware, a national consumer education parent hotline and web delivery system at: https://www.childcareaware.org/.

The 13 areas above are based on the 2002 research paper by R. Fiene, *13 Indicators of Quality Child Care: Research Update* available at: http://aspe.hhs.gov/hsp/ccquality-ind02/.

Produced by:
National Resource Center for Health and Safety in Child Care and Early Education
University of Colorado Anschutz Medical Campus, Aurora, CO
https://nrckids.org

Supported by:
Initial development of this *Parent's Guide* was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Cooperative Agreement #U46MC09810.

Revised July 2019