Qualistar Rating Key Indicator Study

Richard Fiene, Ph.D.

June 17, 2014

ABSTRACT

This report provides an analysis of Colorado's quality rating system, the Qualistar Rating, for generating key indicators. Key indicators have been used a great deal in the licensing literature but this is a first time analysis in utilizing this methodology in a QRS (Quality Rating System) or a QRIS (Quality Rating and Improvement System). The key indicator methodology is described in detail applying it to QRS/QRIS. The results clearly indicate that the strongest key indicators are within the Family Partnerships component of the Qualistar Rating; however there are some major limitations to utilizing this methodology with QRS/QRIS.

INTRODUCTION

The Qualistar Rating, administered by Qualistar Colorado, is one of the longest continuously running QRS in the United States. Presently over 50% of states have QRS/QRIS and the research on these program quality rating & improvement systems has increased over the years. One area of research that has been gaining momentum most recently is ascertaining the most effective and efficient delivery system for a QRS/QRIS as the number of early care and education programs participating in QRS/QRIS continues to increase. This report provides an overview to the topic and introduces an option that has been used in the human services/child care licensing field in identifying key indicators of overall compliance with standards. The purpose of the key indicator methodology is to focus monitoring visits on those standards that have the ability to predict overall compliance with the full set of QRS/QRIS standards. The key indicator methodology is part of a program monitoring approach called Differential Program Monitoring which was developed to help streamline the program monitoring of early care and education programs (please see the Appendix for two graphics which help to depict this relationship (Figures 8/9). It was first applied in child care licensing (Fiene & Nixon, 1985) but has been used in many other service types, such as: Head Start Performance Standards (Fiene,

2013a), National Accreditation (Fiene, 1996), and child and adult residential programs (Kroh & Melusky, 2010). The methodologies are based upon statistical protocols that have been developed in the tests and measurements literature in which an abbreviated set of items is used to statistically predict as if the full test was applied. This methodology has been used in regulatory analysis and is now being proposed for use in Quality Rating and Improvement Systems (Fiene, 2013b). This study and report is the first demonstration of its use with QRS.

TECHNICAL ASPECTS OF THE KEY INDICATOR METHODOLOGY

This section provides the technical and statistical aspects of the key indicator methodology. It will provide the specific methodology for generating the key indicators for the Qualistar Rating.

One of the first steps is to sort the data into high and low groups, generally the highest and lowest ratings can be used for this sorting. In very large states such as Colorado this is done on a sampling basis. Frequency data will be obtained on those programs in the top level (usually top 20-25%) and the bottom level (usually the bottom 20-25%). The middle levels are not used for the purposes of these analyses. These two groups (top level & the bottom level) are then compared to how each program scored on each item within the specific assessment tool (see Figure 1). An example from the Qualistar Rating database is provided in Figure 2 (see Figure 2).

Figure 1	*Providers In Compliance or Top 25%*	*Programs Out Of Compliance or Bottom 25%*	*Row Total*
Highest level (top 20-25%)	*A*	*B*	*Y*
Lowest level (bottom 20-25%)	*C*	*D*	*Z*
Column Total	*W*	*X*	*Grand Total*

Because of the differences in the data distribution for the Qualistar Rating, the above cutoff points had to be more stringent with the respective cutoff points for the high and low groups because the majority of the programs were at the Star 2 and 3 levels. In comparing these data to past licensing distributions (see Fiene, 2013d), it would be expected that the majority of programs would be at a Star 1 level, but that was not the case with this sample. Rather than using a 20-25% cut off point, it was changed to 10% to accommodate this difference. Figure 2 depicts that all programs that were in the top 10% were in the highest rating while the bottom 10% were in the lowest rating. The data depicted in Figure 2 are taken from the *Family*

Engagement Standard 5 – The program provides opportunities for staff and families to get to know one another. The reason for selecting this particular standard is that it demonstrates a perfect Phi Coefficient in discriminating between the highest level and the lowest level[1].

Figure 2: Criterion 5 Family Partnerships	*Providers In Compliance or Top 10%[1]*	*Programs Out Of Compliance or Bottom 10%*	*Row Total*
Highest Star level	*11*	*0*	*11*
Lowest Star level	*0*	*10*	*10*
Column Total	*11*	*10*	*21*

Once the data are sorted in the above matrix, the following formula (Figure 3) is used to determine if the standard is a key indicator or not by calculating its respective Phi Coefficient. Please refer back to Figure 1 for the actual placement within the cells and Figure 2 for the data within the cells. The legend (Figure 4) below the formula shows how the cells are defined.

Figure 3 – Formula for Phi Coefficient

$$\phi = (A)(D) - (B)(C) \div \sqrt{(W)(X)(Y)(Z)}$$

Figure 4 – Legend for the Cells within the Phi Coefficient

A = High Group + Programs in Compliance on Specific Compliance Measure.
B = High Group + Programs out of Compliance on Specific Compliance Measure.
C = Low Group + Programs in Compliance on Specific Compliance Measure.
D = Low Group + Programs out of Compliance on Specific Compliance Measure.

W = Total Number of Programs in Compliance on Specific Compliance Measure.
X = Total Number of Programs out of Compliance on Specific Compliance Measure.
Y = Total Number of Programs in High Group.
Z = Total Number of Programs in Low Group.

Once the data are run through the formula in Figure 3, the following chart (Figure 5) can be used to make the final determination of including or not including the item as a key indicator. Based

upon the chart in Figure 5, it is best to have a Phi Coefficient approaching +1.00 since the data are more normally distributed[2] than is the case with licensing data.

Continuing with the chart in Figure 5, a Phi Coefficient between +.75 and -.25 indicates that the indicator is unpredictable in being able to predict overall compliance with the quality rating assessment tool. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance[3]. This can occur with Phi Coefficients above +.75 but it becomes unlikely as they approach +1.00, although there is always the possibility that other standards/rules/regulations could be found to be out of compliance (this was demonstrated in a study conducted by the author (Fiene, 2013c). Another solution is to increase the number of key indicators to be reviewed but this will cut down on the efficiency which is desirable and the purpose of the key indicators.

The last possible outcome with the Phi Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the desired. The indicator would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously undesirable.

Figure 5 – Thresholds for the Phi Coefficient (Fiene & Nixon, 1983, 1985)(Fiene, 2014)

Phi Coefficient Range	Characteristic of Indicator	Decision
(+1.00) – (+.76)	Good Predictor	Include
(+.75) – (-.25)	Unpredictable	Do not Include
(-.26) – (-1.00)	Terrible Predictor	Do not Include

The key indicators should then only be used with those programs that have attained the highest rating. It is not intended for those programs that have attained lower ratings. However, even with those programs that have attained the highest rating, periodically a full, comprehensive review using the full set of standards for Qualistar Colorado should occur (see Figure 6 for a graphical depiction). It is intended that a re-validation of the key indicators occur on a periodic basis to make certain that the key indicators have not changed because of differences in compliance with standards history. This is an important and necessary step for the program to engage in to ascertain the overall validity and reliability of the assessment system. Also there should not have been any major changes in the program while the key indicators are being administered, such as the director leaving or a large percentage of teachers leaving or enrollment increasing significantly, or a change in the licensing or accreditation status of the program.

Figure 6 - Proposed DMLMA System with Key Indicators (KI)

Use of Qualistar Rating Key Indicators (QRKI) for Monitoring with a Full Review every 4th Year for Star 4 Programs.
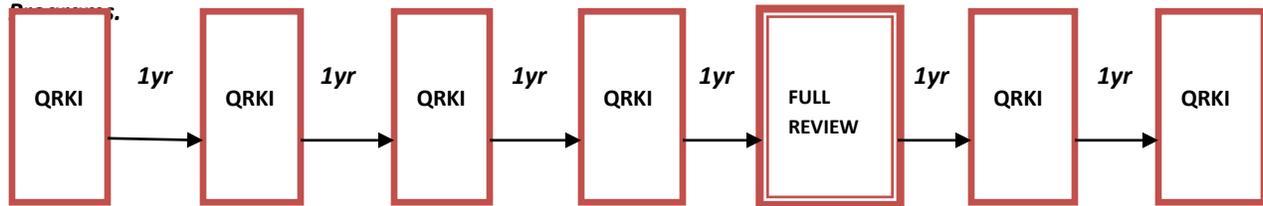
This model is taken from the licensing literature and as will be pointed out in the Limitations and Conclusion Sections may not necessarily be appropriate for QRS/QRIS systems depending on a state's QRS/QRIS data distribution. It is provided for illustrative purposes.

RESULTS

The results reported in this section are based upon a sample selected from the overall Qualistar Rating database from its most recent monitoring reviews (N = 117). This was a representative sample of the program's QRS.

There are five components of the Qualistar Rating: Learning Environment, Family Partnerships, Training and Education, Adult to Child Ratios and Group Size, and Accreditation. See Figures 10-14 in the Appendix for the graphical depictions of the data distributions for the five major criteria. The data distributions are provided because a pre-requisite for calculating the key indicator Phi Coefficients is the dichotomization of data with a skewed data distribution. Figures 10-14 display how much the data are skewed.

The Qualistar Rating is a zero-to-4 star system, with 4 stars indicating the highest level of quality[4]. Eleven programs were rated at the Star 1 level, 19 programs were rated at the Star 2 level, 77 programs were rated at the Star 3 level, and 10 programs were rated at the Star 4 level for a total of 117 programs included in these analyses. There were no programs in the sample that earned less than one star.

Based upon the key indicator methodology described in the previous section, the only Qualistar Rating standards that reached key indicator designation[5] were the following: *Family Partnership Standard/Criterion 5 = The program provides opportunities for staff and families to get to know one another; Family Partnership Standard/Criterion 7 = Families receive information on their child's progress on a regular basis, using a formal mechanism such as a report or parent conference and Family Partnership Standard/Criterion 8 = Families are included in planning and decision making for the program.*

Figure 7 – Key Indicators with Phi Coefficients

	Phi	Significance
Family Partnership Standard/Criterion 5	1.00	.001
Family Partnership Standard/Criterion 7	0.86	.001
Family Partnership Standard/Criterion 8	0.83	.001

There were many other significant correlations (Family Partnerships and Adult-to-Child Ratios and Group Sizes) obtained but none reached the cutoff threshold of .76+ for the Phi calculations. These other correlations are reported in the Appendix after the descriptive graphical displays in Figures 15, 15a, 15b. The Phi Coefficients for the other Criteria (Learning Environment, Training and Education, and Program Accreditation) were not calculated because the data distributions were not skewed as was the case with Family Partnerships and Adult-to-Child Ratios and Group Sizes (see Figures 10-14).

LIMITATIONS

There are two major limitations to this study, 1) the first deals with the statistics being used to generate the key indicators; 2) the second deals with the key indicator methodology.

The first limitation has to do with dichotomization of data which should only be used with very skewed data. Data skewness always occurs with licensing data because of the nature of the data, health and safety protections (the majority of programs are always in compliance with the respective rules). However, this appears to not always be the case with QRS/QRIS data which deals with more program quality aspects of facilities and shows greater variation in the data. If this is the case then dichotomization of data is not appropriate and should not be utilized in order to generate key indicators.

The second limitation of this study is if the key indicator methodology and differential monitoring approaches are appropriate for QRS/QRIS. In Figure 6 above and in the conclusion to this report below, there is a scenario where it can be used but Qualistar Colorado and each state must determine if this is an appropriate approach for their respective program. For example, key indicators will not work in a block model and with a point-system model may generate very limited time savings if the data distribution is normally distributed and there are very few programs at the highest star level. In licensing data base distributions there is always a large number of programs to select from in the highest compliance levels (usually a minimum of 25%).

CONCLUSION/FUTURE RESEARCH/DISCUSSION/RECOMMENDATIONS

This study is the first of its kind in generating key indicators for a QRS based upon the analyses performed with the Qualistar Rating data base. It potentially demonstrates that the use of the key indicator methodology with QRS/QRIS could be feasible and warranted in order to focus limited program monitoring resources in a most efficient and effective manner keeping the above stated limitations in mind as stated in the previous Limitations Section. In the future, Qualistar Colorado may want to pilot an approach utilizing a small group of programs and could focus resources on the Family Partnership/Engagement standards on an ongoing basis between comprehensive reviews as depicted in Figure 6 above for Star 4 programs. The time saved here could then be redistributed to spending more time with the Star 1 programs.

It will be timely to see other states and programs who are interested in generating key indicators if they have Family Partnership/Engagement standards as part of their respective QRS/QRIS to determine if these standards reach the same threshold for key indicator designation as has occurred in this study. It will also be interesting to see if any other state's criteria/standards data distributions are similar to what has been found in the Qualistar Rating or not.

However, as highlighted in the Limitations Section, states and programs need to consider if the key indicator methodology and the resultant differential monitoring model is really warranted and appropriate for their respective QRS/QRIS's. As has been the case with Colorado's Qualistar Rating, only two of the five major criteria: Family Partnerships and Adult-Child Ratio/Group Size were determined to be good candidates for the key indicator Methodology in which the data were skewed[6] enough to warrant dichotomization. The other three major criteria: Learning Environment, Training and Education, and Program Accreditation were determined not to be sufficiently skewed to warrant dichotomization. This sets up a decision making system in which only 40% of the criteria are being used and severely limits the overall predictability of the key indicators selected. Could the other criteria be used to generate key indicators? Of course, but dichotomization of data should not be done when data are not highly skewed (MacCallun, etal, 2002). *Yes, we were successful in generating Key Indicators for the Qualistar Rating but within a limited scenario in how they should be used*. The results are not equivalent to what has been found and utilized in the licensing literature where the licensing data are always highly skewed. If a state or program find that all their standards are skewed in a similar way to licensing data then dichotomization of data and the generation of key indicators is warranted.

A recommendation to Colorado's Qualistar and other programs and states where they find the data from their standards more normally distributed that they not use a key indicator approach. The key indicator approach remains a reliable and valid methodology for licensing but only in very special and limited cases will it be an appropriate monitoring approach for more program quality focused systems, such as QRS/QRIS and accreditation. *For those QRS/QRIS systems where the standards are more normally distributed, the recommendation would be to continue to use the full set of QRS/QRIS standards and not use an abbreviated set of standards.*

NOTES:

1. For analytical purposes, the top 10% of programs received an average score of 8 points or higher on a 10 point scale and the bottom 10% of programs received an average score of 2 points or less on a 10 point scale.

2. The reason for pointing out the need to have a higher Phi Coefficient than what has been reported previously (Fiene & Nixon, 1983, 1985) is the fact that the dichotomization of data should only be used with skewed data and not normally distributed data because it will accentuate differences. However, since the purpose of the dichotomization of data is only for sorting into a high and low group, it would appear to be acceptable for this purpose (MacCallun, etal, 2002. On the practice of dichotomization of quantitative variables, *Psychological Methods, 7, 1,* 19-40.).

3. These results would show an increase in cells B and C in Figure 1 which is undesirable; it should always be the case where A + D > B + C for key indicators to maintain their predictive validity.

4. The following point values equate to the various Star levels in the Qualistar Rating System (for detailed information regarding the QRS system please see the following document: *Qualistar Colorado – Qualistar Rating Criteria Chart*, November 2012):

              Provisional = 0 – 9 points or Learning Environment score of 0
              Star 1 = 10 - 17 points
              Star 2 = 18 - 25 points
              Star 3 = 26 - 33 points
              Star 4 = 34 - 42 points

              Qualistar Rating Criteria Chart:
              Learning Environment = points are awarded based on average classroom scores on the
                    ERS Scales. (Score of component: 1 – 10)
              Family Partnerships = points are awarded based on how well programs communicate
                    with collaborate with, and involve families. Score of component: 1 – 10)
              Training and Education = points are awarded to teachers & center administrators based
                    on their professional development level and amount of experience, with criteria
                    separated by position. Score of component: 1 – 10
              Adult-to-Child Ratios & Group Size = points are awarded based on the average adult-to
                    -child ratio and group size in each classroom. Score of component: 1 – 10
              Program Accreditation = points are awarded for receiving and maintaining national
                    program accreditation through an approved organization. Score of component:
                    0 or 2 points

The reader needs to keep in mind that Qualistar Colorado is not a state agency but rather a private non-profit agency.

5. The three Family Partnership Standards were met at the Star 4 level always or most of the time (see Figure 2).

6. The respective skewness figures are the following: Family Partnership = -1.425; Adult-Child Ratio/Group Size = -1.506; Learning Environment = -0.946; Training and Education = 0.028; Program Accreditation = 7.548. See Figure 16 for basic descriptive statistics for these Criteria.

REFERENCES AND ADDITIONAL RELATED READINGS REGARDING DIFFERENTIAL MONITORING, RISK ASSESSMENT, AND KEY INDICATOR METHODOLOGIES:

☐ Barnard, Smith, Fiene, Swanson (2006). Evaluation of Pennsylvania's Keystone STARS Quality Rating and Improvement System, Pittsburgh: Pennsylvania, Office of Child Development.

☐ Fiene (2013a*).* 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 (2013b)(2014). *Differential monitoring logic model and algorithm.* Middletown: Pennsylvania, Research Institute for Key Indicators.

☐ Fiene (2013c). Head Start Key Indicators. Middletown: Pennsylvania, Research Institute for Key Indicators.

☐ Fiene (2013d). Kansas Child Care Key Indicators. Middletown: Pennsylvania, Research Institute for Key Indicators.

☐ Fiene (2013e). Validation of Georgia's core rule differential monitoring system. Middletown: Pennsylvania, Research Institute for Key Indicators.

☐ Fiene (2007). Child Development Program Evaluation & Caregiver Observation Scale, in T Halle (Ed.), *Early Care and Education Quality Measures Compendium*, Washington, D.C.: Child Trends.

☐ Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, winter 2002-2003, pps 12-13.

☐ Fiene (2002a). *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 (2002b). Improving child care quality through an infant caregiver mentoring project, *Child and Youth Care Forum, 31*(2), 75-83.

☐ Fiene, Iutcovich, Johnson, & Koppel (1998). Child day care quality linked to opportunities for professional development: An applied community psychology example. *Community Psychologist, 31*(1), 10-11.

☐ Fiene (1996). Using a statistical-indicator methodology for accreditation, in *NAEYC Accreditation: A Decade of Learning and the Years Ahead*, S. Bredekamp & B. Willer, editors, Washington, D.C.: National Association for the Education of Young Children.

☐ Fiene (1995). Utilizing a statewide training system to improve child day care quality: The other system in a program quality improvement model. *Child Welfare*, Volume LXXIV, #6, November-December, 1189-1201.

☐ Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services, 5*(2), 38-39.

☐ Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice, 1*(2), 10-11.

☐ Fiene, Greenberg, Bergsten, Carl, Fegley, & Gibbons (2002). *The Pennsylvania early childhood quality settings study*, Harrisburg, Pennsylvania: Governor's Task Force on Early Care and Education.

☐ 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.

☐ Kontos & Fiene (1987). Child care quality, compliance with regulations, and children's development: The Pennsylvania Study, in *Quality in Child Care: What Does Research Tell Us?* Phillips, editor, Washington, D.C.: National Association for the Education of Young Children.

☐ 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

Appendix – Figure 8

DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA©) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model

CI x PQ => RA + KI => DM + *PD => CO*

Definitions of Key Elements:

PC = Program Compliance/Licensing (Health and Safety) (***Caring for Our Children***)
PQ = QRIS/Accreditation/Caregiver/Child Interactions/Classroom Environment Quality (***ERS/CLASS/PAS/BAS***)
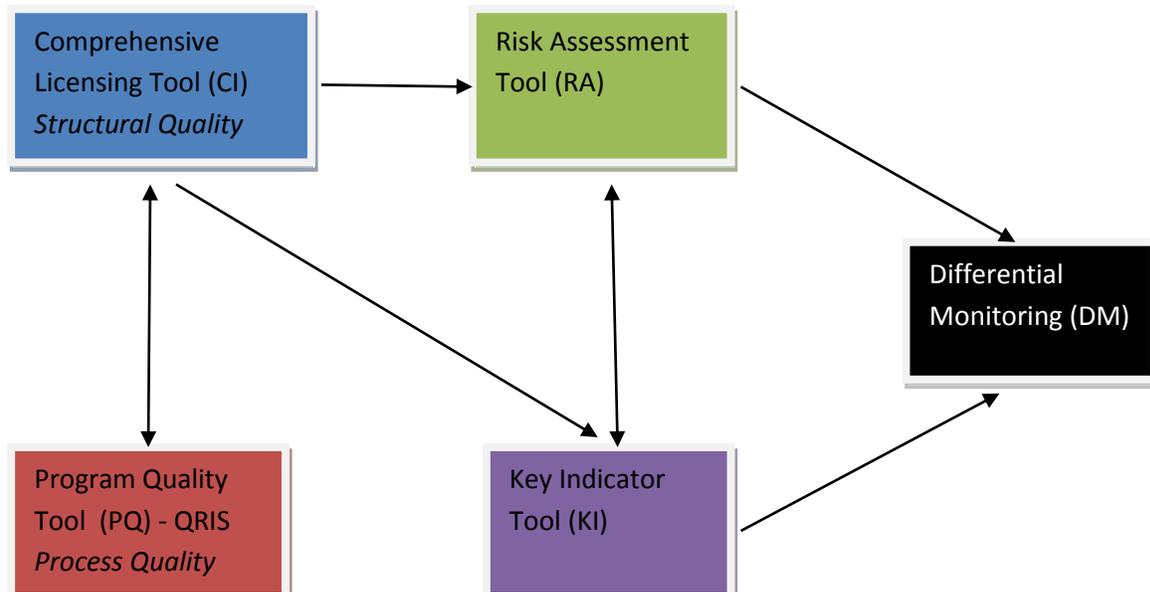**RA** = Risk Assessment, (High Risk Rules) (***Stepping Stones***)
**KI** = Key Indicators (Predictor Rules***) (13 Key Indicators of Quality Child Care***)
DM = Differential Monitoring (How often to visit and what to review)
PD = Professional Development/Technical Assistance/Training (Not pictured but part of Model)
*CO = Child Outcomes (Not pictured but part of Model)*

Comprehensive Licensing Tool (CI) *Structural Quality*	Risk Assessment Tool (RA)	
		Differential Monitoring (DM)
Program Quality Tool (PQ) - QRIS *Process Quality*	Key Indicator Tool (KI)	

Appendix – Figure 9 - Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

All Licensing Rules – Full Compliance Reviews

Differential Monitoring

How Often to Visit?　　　　What is Reviewed?

Frequency

Abbreviated Tool

More Often　　**Less Often**

Risk Assessment Weights　　Key Indicators Predictors
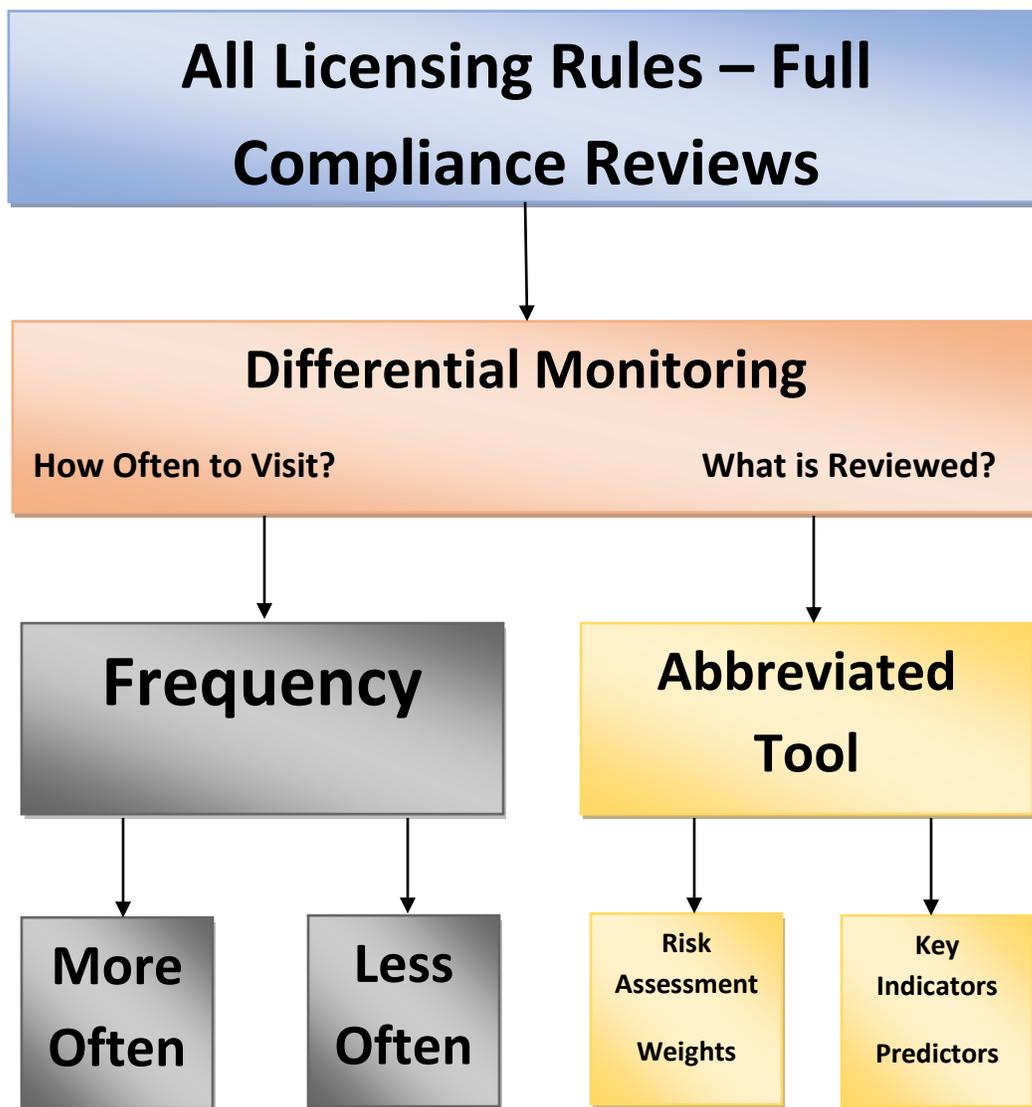
APPENDIX
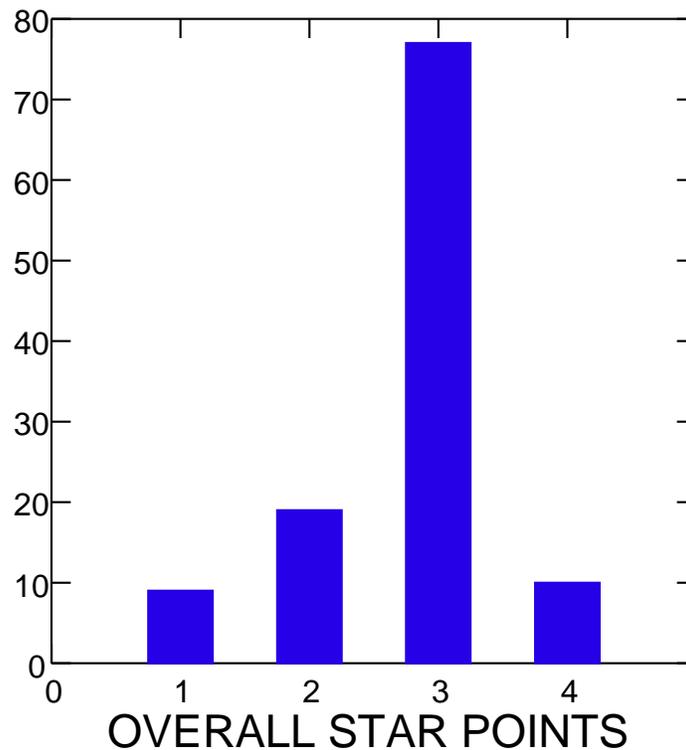
Figure 10

Figures 10-14 depict the data distributions for overall Star points as well as for the major criteria/standards (Training & Education, Learning Environment, Adult-to-Child Ratios & Group Size, and Family Partnerships). Figures 13-14 clearly demonstrate how these respective criteria/standards are extremely skewed data distributions while Figures 10-12 show a more normally distributed data pattern. This is important for which standards can be dichotomized and phi coefficients generated. Dichotomization of data should only be used with skewed data which is the case in figures 13-14. It is not appropriate with the data distributions in figures 10-12. Also see Figure 16 for additional descriptive statistics for the specific criteria.
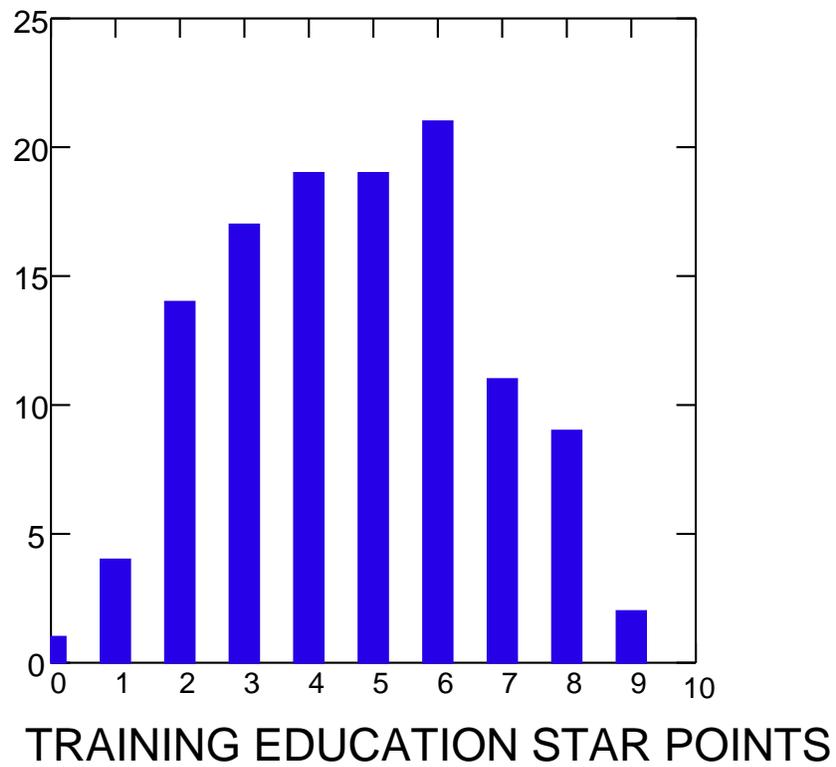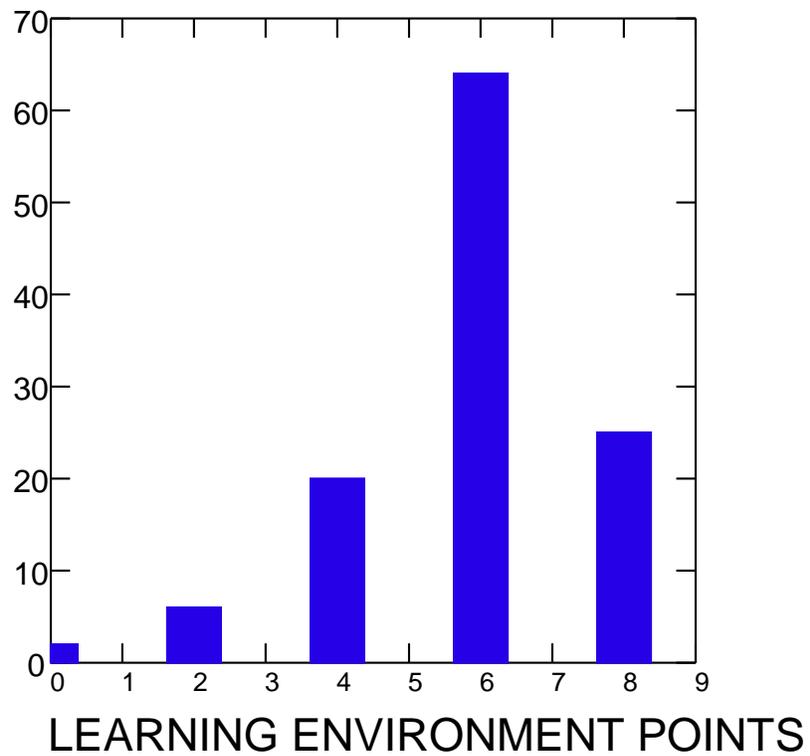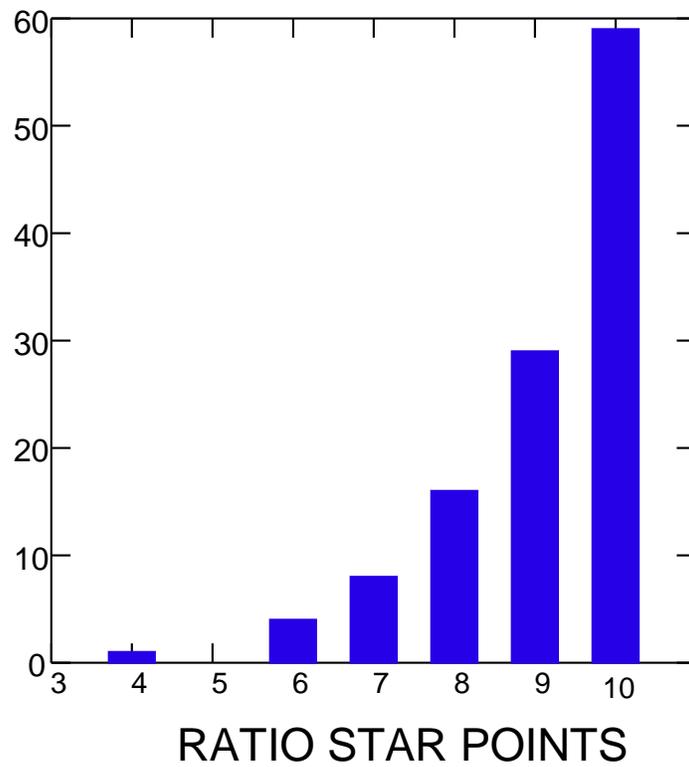
Figure 11

TRAINING EDUCATION STAR POINTS

Figure 12

LEARNING ENVIRONMENT POINTS

Figure 13

RATIO STAR POINTS

Figure 14
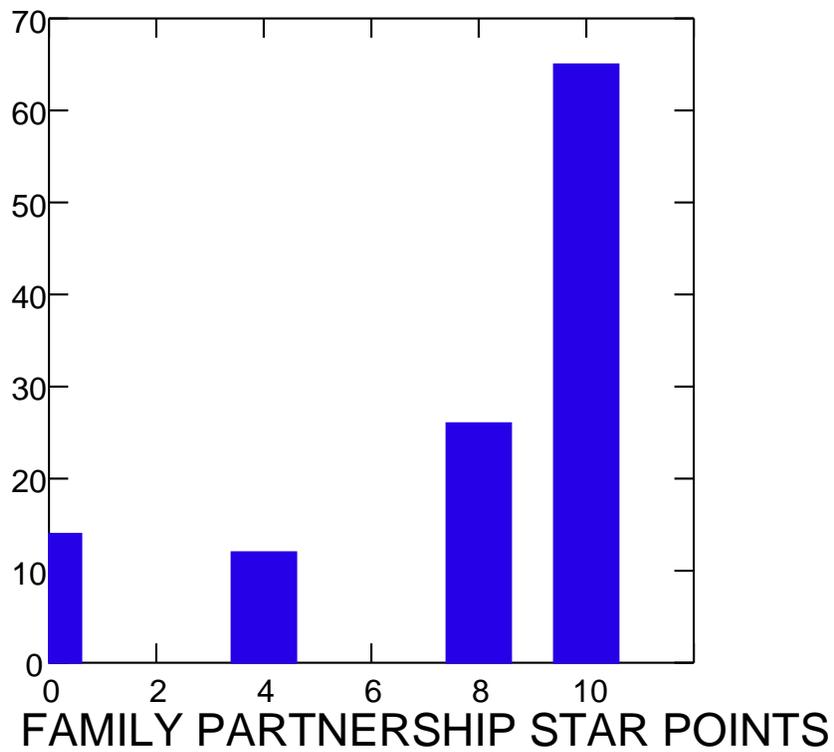

FAMILY PARTNERSHIP STAR POINTS

Figure 15

Selected Relationships amongst the Standards/Criteria and Star Level

Standards/Criteria	Correlation (r)
Family Partnerships x Star Level	.80****
Learning Environment x Star Level	.68***
Training/Education x Star Level	.54**
Adult-Child Ratio/Group Size x Star Level	.46*
Program Accreditation x Star Level	.11

* $p < .05$
** $p < .01$
*** $p < .001$
**** $p < .0001$

Figure 15a

Family Partnership Criteria	Phi	Significance
Criterion 1	.23	ns
Criterion 2	.53	.02
Criterion 3	.46	.04
Criterion 4	.46	.04
Criterion 5	1.00	.001
Criterion 6	.46	.04
Criterion 7	.86	.001
Criterion 8	.83	.001
Criterion 9	.72	.001
Criterion 10	.60	.006
Criterion 11	.46	.04
Criterion 12	.53	.02
Criterion 13	.21	ns
Criterion 14	.46	.04
Criterion 15	.39	ns
Criterion 16	.75	.001
Criterion 17	.60	.006

Legend:
Criteria 1 – 7 involve the program providing information to families.
Criteria 8 – 15 involve families in planning, communicating and decision making for the program.
Criteria 16 – 17 involve a written plan and evaluating the program's family partnerships.

Figure 15b

Adult-Child Ratio/Group Size	Phi	Significance
Adult-Child Ratios	.58	.0001
Group Size	.33	.02

_____

Family Partnerships and Adult-Child Ratio/Group Size standards/criteria phi coefficients were generated because of the skewed data distributions. Phi coefficients were not generated for Learning Environment, Training and Education or Program Accreditation because the data were not sufficiently skewed or showed no variability at all in their respective distributions.

Figure 16

Basic Descriptive Statistics for Criteria

Criteria	Mean	Median	Skewness
Family Partnerships	7.7	10	-1.425
Adult-to-Child Ratios & Group Size	9.1	10	-1.506
Learning Environment	5.8	6	-0.946
Training and Education	4.7	5	0.028
Program Accreditation	0.0	0	7.548
Total Star Level	2.7	3	-1.213

_____