

**HUMAN SERVICE LICENSING MEASUREMENT, REGULATORY COMPLIANCE, AND PROGRAM
MONITORING SYSTEMS**

Richard Fiene and Karen Kroh

Preface

This paper was originally written in 2000 as a chapter in the *NARA Licensing Curriculum* dealing with the subject of licensing tools and systems. Since then, many changes and refinements have occurred in the human services regulatory administration field related to measurement and program monitoring. This paper will address many of these updates within the context of the original *NARA Licensing Curriculum* chapter since many of those original concepts are still relevant today.

Since 2000, several original results related to regulatory compliance have been supported and enhancements to what was called inferential inspections and now is called differential monitoring have occurred. We have continued to see that key indicator/predictor rules have remained very constant over the past 15 years. The original 13 key indicators of licensing and quality have not changed a great deal – 10 of the original 13 are still present. An important question is, why is this the case, for example why is compliance with a rule related to proper immunization of children a key predictor rule discriminating between high and low overall compliant programs.

Another result that has held up over the years is the non-linear relationship between program compliance and program quality. This continues to be a controversial and troubling relationship from a public policy point of view.

Other areas that need further development is for validation studies to make certain that differential monitoring, risk assessment and key indicators work as intended. There is substantial anecdotal evidence but a solid empirical base needs to be established for these various licensing and monitoring system methodologies.

Introduction

Measurement within licensing and regulatory administration has changed dramatically over the past 40 years. In the 1970s it was more qualitative in nature rather than its quantitative nature today. The qualitative nature was depicted with long narratives obtained from in-depth observations and interviews that described a facility in detail with a listing of violations with specific rules. The observations used a running record format in which a detailed accounting of the facility was obtained. This qualitative system worked well when there were few facilities to be assessed. However, as the number of human care facilities increased and licensing agency administrators felt a greater need to understand compliance trends, movement to a more quantitative measurement system evolved.

This move to quantification of measurement began in earnest in the 1970's. The notion of an instrument based program monitoring or licensing system began to be examined by licensing agencies. Checklists and rating scales were employed, with checklists being used predominantly because of the nature of regulatory compliance, an all or none phenomenon. However, a few states, provinces and cities have considered the use of rating scales to measure compliance with rules. More will be said about the differences between checklists and rating scales later.

By the early 1980s with severe federal cutbacks in funding, licensing administrators found themselves with an increasing number of facilities to license but fewer funds to perform these program monitoring and investigative functions. In response to this concern, more streamlined or abbreviated program monitoring systems were developed utilizing a differential monitoring format, such as the key indicator or indicator checklist methodology which utilized a shortened version of the comprehensive checklist approach used by many states. Key indicator systems have been developing over the past 40 years and in many states are key components of their monitoring and licensing functions as they form a basis for risk assessment regulatory analyses. The key indicator system along with risk assessment systems are two forms of what is known in the human service licensing/regulatory administration literature as differential monitoring inspections. For the interested reader, two federally produced publications, a white paper on program monitoring innovations published by the Assistant Secretary's Office for Planning and Evaluation and a licensing brief on differential monitoring, risk assessment, and key indicators published by the Office of Child Care in the Administration for Children and Families are highly recommended. These two resources are cited in the resource section at the end of this document.

A related but very different technique that complements key indicator systems is the use of weighting or risk assessment systems to determine the relative risk for morbidity or mortality of specific rules related to non-compliance. The original reason for the development of weighting/risk assessment systems is the nature of regulatory compliance data. Because compliance data measure minimum health, safety and well-being rules, the data are highly skewed with very little variance. The use of risk assessment systems was to help to increase the amount of variance in the regulatory data sets. However, risk assessment systems over the past 40 years have taken on a life of their own and are used by many states as stand-alone abbreviated differential program monitoring systems. It is estimated that more states employ a risk assessment approach rather than a key indicator approach when applying differential monitoring.

The key indicator and risk assessment systems have not been limited to licensing and regulatory administration systems but have also been developed and used for other program quality endeavors, such as accreditation, quality rating and improvement systems, national and international standards setting.

Definitions

Instrument based program monitoring – a movement within licensing and regulatory administration from qualitative measurement to a very quantitative form of measurement that includes the use of checklists. This move to more quantitative has been encouraged as more and more states develop electronic data systems.

Key indicator system – a licensing measurement system utilizing a shortened or abbreviated version of a comprehensive checklist measuring compliance with rules through a statistical methodology. Only key predictor rules are included on an indicator checklist. It is a form of differential monitoring or inferential inspections where only a portion of the full set of rules is measured.

Differential monitoring/inferential inspections – an abbreviated inspection utilizing a select set of rules to be reviewed. A key indicator system or a risk assessment system are two examples of differential monitoring approaches. The use of differential monitoring/inferential inspections by licensing agencies was developed as a time saving technique and a technique to focus regulatory efforts on facilities that required additional inspections or technical assistance.

Checklist is a simple measurement tool that measures compliance with state rules in a yes/no nominal format. Either the facility is in compliance with rules or not in compliance. Generally, there is no partial compliance with checklists. Having regulatory compliance data being at a nominal measurement level creates limitations statistically in the types of tests that can be completed.

Rating scale is a more complex measurement tool in which a Likert type of rating is employed going from more to less or high to low. A rating scale is always used in the development of weighting/risk assessment systems. It is not used in measuring compliance with rules or at least it hasn't been used in the past.

Risk Assessment/Weighting system is a Likert type of measurement that utilizes a modified Delphi technique to determine the relative risk to individuals if there are violations with specific rules. Risk assessment/weighting systems are developed by sending a survey to a selected sample of persons/stakeholders in order for them to rank the relative risk of violation with specific rules.

Outcome based systems are measurement systems based upon outcomes, not processes. A facility would be assessed by the outcomes it produced with individuals. For example, the number of consumers (children or adults) developing normally, free from abuse, not in placement, involved actively in the community, properly immunized, free from injuries, etc. are outcome based measures.

Instrument Based Program Monitoring

Instrument based Program Monitoring (IPM) is a particular approach to measurement and assessment. It is in contrast to a more qualitative type of assessment (case study is an example of this type of assessment). IPM is very quantitative and is characterized by the use of checklists (see the next section for a discussion of checklists). The advantages of instrument based program monitoring are the following: cost savings, improved program performance, improved regulatory climate, improved information for policy and financial decisions and the ability to make state/province comparisons.

IPM was a paradigm shift in conducting licensing inspections and licensing of facilities when it was first introduced in the late 1970s – early 1980s. More recently it has come under scrutiny to see if a more balanced approach employing a combination of quantitative and qualitative tools is more appropriate. With that said, IPM is an approach that lends itself to automation, it is objective and it is generally systems-oriented. The IPM approach came into its own in the 1970s and has been used predominantly since then as the primary licensing measurement approach. As stated, some state administrators have argued that the IPM approach is not as effective as the more qualitative, narrative case study approach although they can't argue with its efficiency. A combination of IPM (quantitative approach) with a qualitative approach is probably most effective; however, this is very time consuming and a luxury that most state/province licensing agencies do not have, with more and more facilities to license and fewer and fewer staff to do the licensing.

Checklists

Checklists are the predominant means of collecting licensing data. It simplifies the process, making it very quantifiable. This is one of its strengths, but along with this simplification, a drawback is that some of the richness of the description of a particular facility is lost.

There are particular steps that need to be followed in the development of the checklist. Licensing administrators need to follow this four step process:

- 1) Make interpretations of the rules part of the overall manual for measurement of the comprehensive set of rules.
- 2) Identify the rules to be included in the checklist.
- 3) Consider the organization of the checklist – the flow of the investigation to the facility.
- 4) Decide what type of record keeping will be used – paper, tablet, laptop, et

Rating Scales

Rating scales will not be discussed in detail because their applicability to licensing measurement is rather limited. Only in cases where a licensing administrator was interested in some form of partial compliance would rating scales make sense. The NAEYC (National Association for the Education of Young Children) accreditation system is one example of the use of a rating scale of full, partial or non-compliance with accreditation standards. Many QRIS (Quality Rating and Improvement Systems) use rating scales through the ERS (Environment Rating Scales). While a partial compliance rating may be useful in accreditation standard measurement, it is generally not appropriate for use in licensing rule measurement.

Most licensing agencies do not use partial compliance, and the movement within the regulatory administration field is to consider partial compliance as being equivalent to non-compliance. Either a facility meets the rule or does not meet the rule. There is no middle ground.

Weighting/Risk Assessment Systems

Weighting/risk assessment systems and licensing indicator systems that are described in the next section are enhancements of the basic checklist (instrument based program monitoring) system. Weighting/risk assessment systems are used to increase the amount of various in licensing compliance data. Because licensing data are nominal data ('yes' or 'no' compliance) and are generally highly in compliance, there is little variance in the data set from any particular set of rules. In order to increase the variance in data, weighting/risk assessment are used so

that each rule does not have an equal weight. If you do not weight rules, by default, you have given an equal weight to each rule.

The remainder of this section describes the process for developing a licensing weighting/risk assessment system for use in the implementation of human care licensing rules, displays data from states that have used this approach and discusses the applicability of weighting/risk assessment system for all types of human service licensing.

A licensing weighting/risk assessment system is a regulatory administration tool designed for use in implementing human care licensing rules. A licensing weighting/risk assessment system assigns a numerical score or weight to each individual licensing rule or section of a rule, based upon the relative health, safety and welfare risk to the consumers if a facility is not in compliance with the rule. The type of license issued is based on the sum of the numerical weights for each rule that is not in compliance.

The specific objectives of a licensing weighting/risk assessment system are:

- a) To standardize decision-making about the type of license to be issued
- b) To take into account the relative importance of each individual rule
- c) To ensure that rules are enforced consistently
- d) To improve the protection of consumers through more equitable and efficient application and enforcement of the licensing rules

A licensing weighting/risk assessment system can and should be developed and implemented only if:

- 1) Regular or full licenses are issued with less than 100% compliance with all rules. If a regular license is not issued unless all violations are corrected at the time of license issuance, a weighting/risk assessment system is not necessary. A weighting/risk assessment system is useful if a facility is issued a license with outstanding violations (and a plan to correct the non-compliance areas) at the time of license issuance.
- 2) There is a large number of licensing rules with a variation of degrees of risk associated with various rules. If there are only a few rules with equal or similar risk associated with each rule, a weighting/risk assessment system is not necessary. A weighting/risk assessment system is useful if there are many rules with varying degrees of risk.
- 3) A standardized measurement system or inspection instrument is used to measure compliance with licensing rules. Before developing a weighting/risk assessment system, a standardized measurement instrument or tool should be developed and implemented.

Development of a Weighting/Risk Assessment System

This section will provide a step-by-step process in the development of a weighting/risk assessment system for licensing agency use.

- 1) The first step in developing a licensing weighting/risk assessment system is the development of a survey instrument. A licensing inspection instrument or measurement tool can be adapted into a survey tool...the survey should contain each rule or section of a rule, according to how it is measure in the inspection instrument. Survey instructions should explain the purpose of the survey and instructions for completing the survey instrument. It is suggested that survey participants rate each rule section from 1-8 based on risk to the health, safety and welfare of the clients if the rule is not met (1 = least risk; 8 = most risk).
The survey participant should be instructed to circle their rating choice of 1, 2, 3, 4, 5, 6, 7, 8. An example of a survey question is:

Interior stairways, outside steps, porches and ramps shall have well-secured handrails.

| | | | | | | | |
|----------|---|---|---|-----------|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Low Risk | | | | High Risk | | | |

- 2) Surveys should be disseminated to at least 100 individuals. If a state has more than 3,000 licensed facilities in the type of service being surveyed, consideration for surveying more than 100 individuals should be given. Individuals surveyed should include providers of service; provider, consumer and advocacy associations; health, sanitation, fire safety, medical, nutrition and program area professionals; licensing agency staff including policy/administrative staff and inspectors; consumers of service; parents; and funding agency staff. In order to assure a higher survey return rate, persons selected as survey participants should be contacted prior to the survey to explain the weighting/risk assessment system and request their willingness to complete the survey.
- 3) Survey results from each survey should be collected and entered into a computer data base spreadsheet software package or an online survey software. After all survey data are recorded, means or average weights for each rule or section of a rule should be calculated. If there is sufficient variation in the means for each rule, the individual rule means can be rounded to the nearest whole number. Generally when comparing mean

weights among the various groups surveyed there should be a similarity in rating among the groups, supporting the use of the weights as a reliable measure of risk.

- 4) The next step is to either (a) pilot test the weights with new licensing data for about six months, or (b) apply the weights to a least 25% of historical data from the previous 12 months. The intent of the pilot application is to collect data to use as the database for determining statistical cut-off points for the issuance of specific types of licenses or for administration of various negative sanctions. A total weighted score for each facility based upon the combined weights of all violations should be calculated. Following is an example of how the scores should be calculated:

| <u>Rule Violations</u> | <u>Weights</u> |
|------------------------|----------------|
| Rule 1 | 7 |
| Rule 2 | 6 |
| Rule 3 | <u>+8</u> |
| Sum of Weights = | 21 |

Under the above example a perfect compliance score with no non-compliance areas would be a score of “0”. The higher the score, the lower the compliance would be. However, this is not congruent with the common usage of scores in which the higher score is associated with better compliance. In order to accommodate our familiarity with higher scores for the better facilities, the weighted score should be deducted from an arbitrary constant score of “100”. Thus a weighted non-compliance score of “20” will convert to a positive score of “80”. This is more intuitive to individuals as they think about scores and measurement.

Using the previous example, the final weighted score would be computed as follows:

| <u>Rule Violations</u> | <u>Weights</u> |
|------------------------|----------------|
| Rule 1 | 7 |
| Rule 2 | 6 |
| Rule 3 | <u>+8</u> |
| Sum of Weights = | 21 |

Final Calculation:

| |
|------------|
| 100 |
| <u>-21</u> |
| +79 |

- 5) The fifth step in the process is to compute and apply the standard deviation to the data. The mean and standard deviation of all weighted scores computed in the pilot application in step #4 should then be calculated. The mean and standard deviation of all final weighted scores computed in the pilot application in step #4 should then be calculated. Based upon experience with implementing licensing weighting/risk assessment systems, it is recommended that if a final weighted score is no more than one standard deviation below the mean, a regular license should be issued. If a score is between one standard deviation below the mean and two standard deviations below the mean, a provisional license should be issued (the length of the provisional license will vary based upon the severity of the non-compliance), or intermediate negative sanctions should be administered. If a score is less than two standard deviations below the mean, no license should be issued or a more severe negative sanction should be administered.

For example, if the standard deviation is 18 and the mean is 88, following is the distribution of the weighted scores used to determine the type of license to be issued:

Score of 100 — 70 = Regular license/no sanction

Score of 69 — 52 = Provisional license/intermediate sanction such as warnings, administrative fines or restriction on admissions

Score of 51 and below = No license/severe sanction such as revocation or administrative closure

- 6) The final weighted scores from the pilot application should be applied to the standard deviation cut-off points to determine the type of license or negative sanction issued. These data should be studied to compare types of licenses or sanctions issued under pre-weighting vs weighting.
- 7) Before implementing the licensing weighting/risk assessment system the following additional licensing factors should be considered and incorporated as necessary into the licensing system.
- a) repeated violations from the previous licensing inspection;
 - b) violation with high risk items (possibly a weight of 8.0);
 - c) discretion of licensing inspector to recommend variance from licensing weighting system.
- 8) Whenever licensing rules are amended, or at least every 5 years, the weights should be recomputed and the weighting system re-evaluated. The licensing weighting system as described here can be used to license any type of human care facility including child care, adult care, residential care and part-day care facilities. Licensing weighting/risk assessment systems have been developed in the majority of states.

Since the concept, development and implementation of weighting/risk assessment systems is relatively new to the field of licensing, the long term impact and benefits of weighting/risk assessment systems have not been fully realized through the use of validation studies (Please see the section on Future Research to address this concern). The potential of using weighting/risk assessment systems and modifications of weighting, to help standardize the implementation and enforcement of licensing rules is an exciting area of research to pursue in the field of regulatory administration.

Licensing Indicator Systems

As mentioned in the weighting/risk assessment system section of this chapter, indicator checklists or licensing indicator systems are used to improve upon instrument based program monitoring (checklist) systems. The licensing indicator system is one method of assuring compliance with licensing rules in a time efficient manner. The concept has been developed and successfully implemented in several states and for different human service types. The licensing indicator system was originally developed in Pennsylvania in 1977 for use in licensing child care centers. The original intent was to develop an abbreviated licensing instrument in order to refocus licensing investigation time to assess and assist in quality enhancement activities.

From 1980-1984, the US Department of Health and Human Services funded a project to study and further develop a licensing indicator system for child day care facilities on a national level. The federally funded project, known as the Children's Services Monitoring Transfer Consortium, organized researchers, state licensing administrators and professional staff from Pennsylvania, Michigan, West Virginia, Texas, New York City and California to review and refine the existing Pennsylvania system for possible use by other states.

The purpose of a licensing indicator system is to increase the efficiency and effectiveness of an existing licensing system by refocusing the emphasis of the licensing process. A licensing indicator system is intended to complement, and not replace, an existing licensing measurement system. Through use of the licensing indicator system, less time is spent conducting annual inspections of facilities with a history of high compliance with the licensing rules, and more time is spent a) providing technical assistance to help facilities comply with licensing rules and b) conducting additional inspections of facilities and agencies with low compliance with licensing rules.

The licensing indicator system is actually a shortened version of a comprehensive licensing inspection instrument. A small number of rules are selected based upon a statistical methodology designed for this specific purpose. The licensing indicator system uses a measurement tool, designed to measure compliance with a small number of rules, that predicts high compliance with all the rules. If a facility is in complete compliance with all of the rules

measured in the licensing indicator system, high compliance with all the rules is statistically predicted. It is critical to understand that the rules for the licensing indicator system are selected statistically (the statistical technique is called the phi-coefficient and generally is set at a p value of .01 or higher) and not based upon value judgement (arbitrary assignment, no basis from research literature), risk assessment or frequent rule violations. The rules are selected based upon an SPSSPC+ computer software package that compares violations of facilities with high compliance versus facilities with low compliance. The rules that are most often out of compliance in low compliance facilities and in compliance in high compliance facilities will be the indicator or predictor rules.

Prerequisites for implementing a licensing indicator system

Before developing and implementing a licensing indicator system it is important that the existing licensing system is comprehensive and well established. The following are prerequisites to implementation of an indicator system:

- 1) Licensing rules must be comprehensive, well written and measurable. Rules are the building blocks for any licensing system. If the rules are not well written and measurable a licensing indicator system should not be pursued. Also, if the total number of rules is small, a shortened inspection tool is not valuable.
- 2) There must be a measurement tool designed to standardize the application and interpretation of the rules. A licensing inspection instrument designed to assure statewide consistency in the application of the rules is essential prior to implementing a licensing indicator system.
- 3) There should be a licensing weighting system designed to assess the relative risk to consumers if the rule is not met. This system may be a formal weighting system or a simple classification system which categorizes rules by degree of risk. An example of a high degree of risk to consumers would be the accessibility of heat sources or toxins. Having a signature in a record is an example of a low degree of risk to consumers.
- 4) At least one year of data on rule violations for individual facilities. These data are needed to enter into the computer software system in order to determine the rules that are the indicators or predictors of high compliance.

How to develop a licensing indicator system

The basic steps to developing a licensing indicator system include:

- 1) Select facilities to be used in determining the indicators. If the total number of licensed facilities is less than 200, all 200 facilities can be used. If the total number of licensed facilities exceeds 200, sampling must be done. Generally, a sample of 100 facilities or 10% is acceptable.

When selecting the sample, variables of size of facilities, geographic area, urban/rural, profit/non-profit, public/private and varied compliance levels or scores must be controlled.

2) Violation data for the sampled facilities is entered into a computer software system designed for this purpose (SPSSPC+ is recommended but other statistical packages will do the same).

3) A list of indicator or predictor rules, based on phi coefficients, that were the best indicators of high compliance will be calculated by the computer software system. These are the rules that are most often out of compliance in low compliance facilities and in compliance in high compliance facilities.

4) A small number of additional rules which are determined based on a licensing weighting system or relative risk are added to the statistically selected indicators. The purpose of this step is to assure face validity of the instrument. By adding a smaller number of carefully selected high risk rules to the instrument, the licensing agency can be assured that critical rules are always measured.

5) In order to assure that full compliance with all the rules is maintained, five items selected at random should also be applied as part of the licensing indicator system. The final licensing indicator system instrument contains the indicator rules, high-risk rules and random rules. The total number of rules on an indicator checklist will vary, but will range from 20-45 items.

6) Specific criteria for use of the licensing indicator system are developed.

Criteria for use of the licensing indicator system

The development of very specific criteria for use of the licensing indicator system is perhaps the most critical step of the design process. This is the step at which the determinations are made as to when the licensing indicator system will be used. The determination of use of the system should be standardized and not based upon licensing inspector discretion.

Each licensing agency must develop its own criteria based upon its own historical licensing data and experience. Following are some criteria that may be useful:

1) The facility has had a full or regular license and no negative sanctions have been administered, within the previous two (2) years.

2) The facility has had a score or percentage of compliance above a specified threshold for the previous year.

3) All previous violations have been corrected according to the facility's plan of correction.

4) No significant validated complaints have been found within the past year.

5) The total number of consumers served has not increased by more than a specified percentage during the past year.

6) There has not been significant staff turnover at the facility/agency within the past year. This may be targeted to certain levels of staff turnover, such as direct care staff or facility directors, depending on which staff are particularly key for program stability.

7) A full inspection using the comprehensive licensing measurement instrument must be done at least every three (3) years.

Revision of the licensing indicator system

The licensing indicator system should be continually reevaluated for its effectiveness. The system should be completely revised at least every three years or upon a revision of the rules. In order to achieve the intended purpose of the licensing indicator system of refocusing the emphasis of licensing effort from facilities with high compliance to facilities with low compliance, constant review, evaluation and revision of the licensing indicator system is essential.

Other types of inferential inspection/differential monitoring systems, of which the licensing indicator system is only one, will not be addressed in this chapter because inferential systems/differential monitoring other than the licensing indicator system have not been determined to be statistically valid or reliable. As licensing administrators may potentially need to defend their actions in a court of law, it is essential that the methodology or technique utilized is scientifically sound. When it comes to inferential inspections only those instruments based upon an indicator or weighting/risk assessment methodology can stand up to this rigorous testing.

Outcome Based Systems

This is a relatively new phenomenon in the licensing and regulatory administration field. The emphasis in this new approach is to examine outcomes rather than processes. What are the ultimate outcomes for individuals? Determine this and the argument goes there is no need to measure processes directly.

Outcome measurement is appealing in many respects. It does focus on results, something the human services field has been short on demonstrating. However, there is a fallacy in this approach. Results are the end product, but we always have a process to get to the end product.

Another issue is that the purpose of licensing is to prevent harm to consumers. A purely outcome-based system would potentially harm consumers who were in the facilities later determined to “fail” the outcome test. Moreover, there are two other problems:

1. Insufficient (political) agreement on what are acceptable outcomes.

2. Some outcomes will not manifest for years and/or are contaminated by other variables related to other influences on later behavior.

What makes more sense is to tie outcomes to specific regulatory processes that appear to be in a causal or at least a correlational relationship. If licensing agencies were able to clearly link specific results (outcomes) to specific rules (processes), there would be the empirical ability to focus only on those rules that produced positive results for consumers and families and eliminate all other unnecessary rules that do not produce positive outcomes for consumers and families. Specific studies could be conducted and in fact have already been conducted by university researchers. In child care, for example, low staff:child ratios, pre-service and in-service training of staff, highly qualified staff and small group size are all examples of regulatory variables that have been identified as surrogates to program quality that produce positive outcomes for children.

Outcome based or results-oriented systems will impact licensing, but the research literature demonstrates how licensing agencies can clearly link outcomes to regulatory processes that produce the outcomes. This becomes a powerful argument to legislators when this roadmap of process to outcome can be provided.

Relationship Between Rules and Instruments

This section is included because this is one area that gets many licensing administrators into trouble. Not enough time is spent on making sure that the instruments developed are the exact reflection of the rules. This is where the interpretive rules that are part of any measurement instrument that accompanies the actual instrument should be placed. This helps to increase the reliability of the instrument and doesn't hurt the overall validity of the tool either (more on reliability and validity in the next section). Readers who are interested in doing additional reading in this area should refer to the *NARA Licensing Curriculum's* chapter on The Formulation of Rules, for additional information on the definition and development of interpretive and substantive rules.

When there is not a close link between instrument development and rule formulation this only leads to headaches for licensing agencies. It may take years and not be evident until you get called into a court of law to defend your licensing system but it will happen.

The analogy of playing Russian Roulette may be useful. As licensing administrators, you are never 100% certain that all your facilities are compliant with all the rules. However, there are certain management procedures and processes that you can put in place to help. A clear link between rules and measurement tools is one of them. Since you are never 100% sure of full compliance (in other words all six chambers of the revolver are not empty if they were, you wouldn't have Russian Roulette), you must make difficult decisions related to increasing or

decreasing your chances in playing Russian Roulette. So you have the choice of having the management and procedural safeguards built in (one or two bullets in the revolver) or you don't build in the procedural safeguards (four or five bullets in the revolver). It is obvious statistically where your chances are greater in surviving a potential mishap in a licensing system.

Reliability and Validity

The two concepts of reliability and validity are so critical to measurement, but are so often overlooked in the development of licensing measurement systems. In fact, it has been estimated that as many as 30 states may be using a type of differential monitoring/inferential inspection. But only a few states have followed the rigorous statistical methodology as outlined in the Licensing Indicator System section.

Very simply, validity deals with content of the particular tool or instrument—does it serve the purpose for which it is to be used? Does it measure the rules accurately? Usually the answer to this question is easier for licensing administrators to answer. Since licensing measurement tools should be directly based upon rules, as explained in the previous section, there should not be much difficulty in establishing validity. When the tools are not based on the rules, that is when validity can be and should be called into question.

Reliability deals with the administration of the tool or instrument. Does it measure the rules consistently and in an objective manner? The answer to this question is much more difficult for licensing administrators to answer affirmatively. This poses real problems if each administration of the licensing tool is not consistent and objective. Facilities will not have the rules applied in an equal and fair manner.

Reliability testing should be done methodologically and scientifically. Inter-rater reliability should be established for the tools/checklists that are to be used in the field by licensing field staff. This is a process that has been well documented in the psychological research. This has not been the case within licensing and regulatory administration. Generally checklists are designed quickly and are never tested for reliability. This creates a problem that many of us have heard—the rules are not applied uniformly across the state/province. The reason is that the tool that is used to measure compliance is not reliable.

In order to establish reliability, licensing inspectors need to go out to facilities in pairs assessing the same facility at the same time. They then need to compare their results. Do they agree on what is in compliance and out of compliance at the particular facility? If there is not at least 90% agreement for each rule then additional interpretation of that specific rule is needed. Establishing reliability is not overly difficult nor overly time consuming; however, it will add a bit

more time before staff are really ready to begin to license facilities (90% agreement on each rule and interpretative rule).

Balance Between Compliance and Program Quality

An interesting development in the past five years has been the emphasis on program quality as a result of pressure from consumers, families, advocates and the general public. Consumers and other interested persons are requesting licensing agencies to ensure not only the health, safety and well-being of individuals served in facilities, but also to be concerned advocates for the overall quality of services provided at these facilities.

This increased emphasis and concern for program quality is a difficult area to address for licensing agencies. The resources to complete program quality reviews and to advocate for quality within government are not commensurate with the expectations. However, there are some strategies that can be employed to assist licensing agencies. The first and foremost will be to save time on doing licensing inspections. The indicator system described in this chapter will provide such a tool for saving time. Studies conducted over the past two decades indicate that utilizing an indicator checklist approach saves up to 50% in the on-site inspection time.

The time saved in doing licensing inspections should be used to either:

- a) Conduct additional licensing inspections in new or problem facilities
- b) Provide technical assistance
- c) Complete program quality reviews

This could be done by utilizing a tool from accreditation in observing classrooms, or utilizing a program quality tool from the research literature (for example, Early Childhood Environment Rating Scale). Licensing administrators need to be certain that they have a plan to utilize this extra time or the worst fears of licensing professionals could occur. Two potential scenarios could play out. One is that the time is used to do more and more licensing inspections utilizing the indicator system on more and more facilities. The worst scenario is that staffs are cut. If a state/province can complete all its inspections in half the time, then doesn't it follow that only half the staff is needed? With a clearly articulated plan on how the licensing and program quality reviews will produce higher quality programs should help to prevent this cost cutting approach. However, this is always a fear that licensing administrators must face.

Regulatory Compliance Theory

A very important discovery which has now been replicated on several occasions is the relationship between regulatory compliance and program quality. The essence of the relationship forms the basic tenet for the use of a key indicator approach which states that

there is a non-linear relationship between regulatory compliance and program quality. In laymen's terms, programs that score low on regulatory compliance also score low on program quality measures but when one looks at the full and substantial levels of regulatory compliance the program quality measure scores plateau out where it is difficult to discriminate between program quality. There appears to be a ceiling effect which makes it difficult to say that those programs that are 100% in compliance with all rules are also the best quality programs. This is not always the case and in fact, many times, those programs that are in substantial compliance have higher program quality scores.

From a public policy perspective this creates problems in requiring a hard line on full (100%) regulatory compliance. A more effective public policy would be to require substantial compliance identifying those rules that predict overall compliance and quality. This is the essence of the key indicator approach in identifying those predictive rules.

From a statistical perspective this is an important consideration because it leads us in the direction that the methods we should be using will be curvi-linear rather than linear. This may help to explain some of the reasons why we haven't been finding significant relationships between program compliance and quality.

The above relationship has been replicated in several regulatory compliance studies and the predictive rules have been replicated in several regulatory compliance studies as well. On the basis of this relationship it makes possible for policymakers to consider more effective and efficient program monitoring and regulatory compliance strategies.

Extreme Skewness of Licensing Data

A very important facet of regulatory compliance data is that it is extremely skewed and not close to being normally distributed. Now, why is this important? It makes it very difficult to analyze skewed data because there is very little variance in the data. It becomes difficult to discriminate between levels of quality and regulatory compliance because only one or two rules that are non-compliant may separate 100s of programs. There is less chance that this will be the case when data are normally distributed, it is generally easier to distinguish between the top performers and the mediocre performers.

It was because of this problem that the concept of weighting of individual rules was introduced in order to enhance the differences amongst the various rules. This weighting of rules has led to the development and implementation of risk assessment systems as stand-alone systems for measuring regulatory compliance and the differential monitoring of programs.

Early Childhood Program Quality Improvement & Indicator Model (ECPQI2M4)

An outgrowth of the above approaches is a comprehensive model that can be used to help improve quality called the Early Childhood Program Quality Improvement and Indicator Model (ECPQI2M4). This model is in its 4th edition utilizing a risk assessment, key indicator and differential monitoring approach (DMLMA4). The reason for introducing these models is to provide an overall structure for states when attempting to develop their program monitoring systems. This model provides crosswalks from licensing and program compliance to more focused program quality improvement systems, such as Quality Rating and Improvement Systems (QRIS).

These models provide a paradigm shift in how we address the balance between program compliance and quality by focusing only on those standards that are proven to make a difference. In 20 years we will have fewer regulations and standards and better outcomes for children because the program monitoring system will be working smarter, not harder. Only those standards that empirically demonstrate a positive impact will be enforced, all others will go by the waste-side.

Methodological Issues

These are some of the key methodological issues that have surfaced over the past 40 years that will need to be addressed if we continue to use these methodologies:

- 1) The need for states to routinely conduct reliability testing is vitally important to make sure that their licensing staff/inspectors are consistently measuring rules.
- 2) The balancing between program compliance and program quality.
- 3) Determining the most effective and efficient threshold is critical because as one becomes more efficient a loss of effectiveness does occur which can lead to an increase in false positives and negatives. False positives are not as critical as false negatives because it would only involve inspecting programs with a comprehensive review that wasn't necessary; but with a false negative a program could be missed that really did need a full comprehensive review rather than an abbreviated review. To solve this threshold issue will take a good deal of sensitivity analyses to determine the correct number of key indicator/predictor rules to be included.

Lessons Learned

Over the past 40 years here are some lessons learned which should help us as we move forward:

- 1) We have learned how to deal more effectively with very skewed data through dichotomization.
- 2) Risk assessment only focuses on compliance and high risk rules which generally are always in compliance. This provides an interesting dilemma when thinking about the efficiency of this approach. One can't argue with its effectiveness because you are monitoring for risk and the prevention of harm which is the essence of regulatory compliance. But is it really efficient to review rules that are generally in compliance. Isn't it more efficient to monitor rules that discriminate between high vs low compliant programs to determine if a more comprehensive review needs to be done. Keep in mind that risk assessment is based upon weighting of rules which was never intended to be used as a stand alone system but rather one to enhance the key indicator methodology by creating more variance in the data. A validation study to determine if risk assessment systems really work in finding additional non-compliance is warranted as is the basic tenet of key indicator systems.
- 3) Key indicators focus on high and low compliance differences with these rules generally being somewhere in the middle range, not in compliance the majority of the time nor out of compliance the majority of the time.
- 4) It continues to be a fact that all rules are not created equal nor are they administered equally. This is very important concept for risk assessment systems as well as a measurement issue in that we are not dealing with equality in measuring each rule. This is where the weighting of each rule comes into play.
- 5) Most recently we have seen that when higher standards are applied, especially with Pre-K initiatives, this goes a long way in helping to discriminate the top performers from the mediocre performers. With licensing data which is very skewed this has been a real problem in discriminating between the best programs and those that are at a lower level. Because the data are so skewed all these high performer and mediocre programs are all grouped together. This is the nature of regulatory compliance data which does not do a good job of distinguishing between the best and the near best, what it is good at is determining the really poor performing providers.

Future Research

Here are some of the key research that needs to be accomplished in the near future related to human service licensing measurement, regulatory compliance, and program monitoring systems.

- 1) The crucial need for future research in the human services licensing and regulatory compliance area is for validation studies of the above approaches, Key Indicators and Risk Assessment methodologies to make certain that they are working as they should. This crucial

need has been pointed out by other authors with the ASPE White Paper as the most poignant example of this suggestion.

2) Another validation study is needed regarding the relationship between program compliance and program quality. This is such an important finding about the plateau of program quality scores with increasing regulatory compliance as one moves from substantial compliance with all rules to full compliance with all rules.

3) A clear delineation needs to occur to establish appropriate thresholds for the number of key indicator/predictor rules that provide a balance between efficiency and effectiveness that can diminish the number of false positives and especially false negatives.

Conclusion

This *NARA Licensing Curriculum* chapter provides a brief overview to the major issues confronting licensing administrators when they consider licensing tools and measurement systems. The emphasis upon quantitative systems was reflected in this chapter because of the need to develop cost effective and efficient licensing systems as the number of facilities continues to grow with shrinking resources. Also there is a compounding effect with higher expectations on licensing agencies to be concerned more about program quality.

The chapter showed the various types of measurement tools that apply to licensing and regulatory administration. It is clear that given the nature of licensing there are certain tools more suited than others, such as checklists versus rating scales. A very detailed description of both licensing weighting /risk assessment and indicator systems was provided. The reason for this emphasis is that these are two very valid and reliable tools that can be used by licensing administrators in making their agencies more effective and efficient. The licensing measurement field is changing constantly as new approaches are introduced. For example, within the program evaluation field there is a move to have a better balance between quantitative and qualitative analyses. It will not be long before this initiative has its impact on the licensing measurement field as well.

This paper has provided a much needed update to what is occurring within the human service licensing and regulatory compliance area. It reviewed what has been in place for several decades but also provided new material to consider in future research, methodological issues, and lessons learned.

To summarize what we know about human service licensing measurement, regulatory compliance, and program monitoring:

- ***The relationship between regulatory compliance and quality is not linear.***

- **Regulatory compliance has difficulty in distinguishing the best programs from the mediocre programs.**
- **Regulatory compliance is very effective at identifying the worse programs.**
- **Balance regulatory compliance with quality indicators.**
- **Need to validate differential monitoring approaches, such as risk assessment and key indicators.**
- **What is the ideal threshold for the number of key indicator/predictor rules.**
- **Risk assessment rules are usually in compliance.**
- **Key indicator/predictor rules are not usually in compliance but are not out of compliance a great deal.**
- **What is it about key indicator/predictor rules that make them so effective in discriminating between high and low performing programs.**
- **Licensing data are very skewed and because of this there is the need to dichotomize the data.**
- **There is very little variance in licensing data with generally only 20 rules separating the top compliant programs from the worst.**
- **The majority of programs (60%+) are in substantial or full compliance with rules.**
- **There is a balance between being effective and efficient that needs to be identified because as the system becomes more efficient it becomes less effective.**
- **As a system becomes more efficient it also can produce additional false positives and negatives which results in lessened effectiveness in program monitoring.**
- **Higher standards (as applied through Pre-K or QRIS) help to distinguish between the best and mediocre programs.**

Resources

For the interested reader, please consult the following excellent publications by the Assistant Secretary's Office for Planning and Evaluation and the Office of Child Care that will provide additional insights into program monitoring in general, differential monitoring in particular, risk assessment and key indicator systems:

ASPE/Monitoring White Paper:

http://aspe.hhs.gov/hsp/15/ece_monitoring/rpt_ece_monitoring.cfm

OCC/Differential Monitoring, Risk Assessment and Key Indicators:

https://childcareta.acf.hhs.gov/sites/default/files/public/1408_differential_monitoring_final_1.pdf

This website provides many additional websites and reports:

RIKI Website for additional publications and websites:

<http://rikinstitute.wikispaces.com>

References

Fiene, (2015). *Early childhood program quality indicators model (Version 4): Differential monitoring, risk assessment and key indicators*, Middletown, PA.: Research Institute for Key Indicators.

Lahti, Elicker, Zellman, & Fiene (2014). Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs, *Early Childhood Research Quarterly*, available online 9 June 2014, doi:10.1016/j.ecresq.2014.04.005.

Zellman & Fiene (2012). *Validation of quality rating and improvement systems for early care and education and school-age care*, Washington, D.C.: OPRE and Child Trends.

Fiene & Carl (2011). Child Care Quality Indicators Scale, in T Halle (Ed.), *Quality Rating and Improvement Systems Tool Kit*, Washington, D.C.: Child Trends.

Manlove, Benson, Strickland, & Fiene (2011). *A comparison of regulated child care in rural and urban Pennsylvania*, Pennsylvania: Center for Rural Pennsylvania.

Johnson, Fiene, McKinnon, & Babu (2010). *A study of early childhood education pre-service teacher education at major universities in 38 pre-k states*. New York, New York: Foundation for Child Development.

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). Measurement tools and systems, in *Licensing Curriculum*, National Association for Regulatory Administration, Minneapolis, Minnesota.

Fiene (1996). Unannounced versus announced licensing visits in monitoring child day care programs. *National Association of Regulatory Administration*, Spring, 5-6.

Fiene (1996). The feasibility of using sampling of agency facilities. *National Association for Regulatory Administration Licensing Newsletter*, fall, 12-14.

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). Improving human services through licensing reform and training: A unique partnership between academia and state government, *Dimensions*, Fall, 1-3.

Kuhns & Fiene (1995). Promoting health and safety in child care programs, *Child Care Bulletin*, January-February (1), 3.

Fiene (1995). *National early childhood program accreditation standards*. Atlanta, Georgia: National Early Childhood Program Accreditation Commission.

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.

Fiene (1994). The case for national early care and education standards: Key indicator/predictor state child care regulations, *National Association of Regulatory Administration*, summer 1994, 6-8.

Fiene (1988). Human services instrument based program monitoring and indicator systems, in *Information Technology and the Human Services*, B. Glastonburg, W. LaMendola, & S. Toole, editors, Chichester, England: John Wiley and Sons.

Fiene & McDonald (1987). *Instrument based program monitoring for child welfare*, Portland, Maine: University of Southern Maine.

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.

Morgan, Stevenson, Fiene, & Stephens (1986). Gaps and excesses in the regulation of child day care, *Reviews of Infectious Diseases--Infectious Diseases in Child Day Care: Management and Prevention*, 8(4), 634-643.

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

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), 10-1.

For additional information about this paper, please contact:

Richard Fiene, Ph.D., Research Psychologist
Research Institute for Key Indicators
National Association for Regulatory Administration
RIKI.Institute@gmail.com or rfiene@naralicensing.org