Performance Assessment, Regulatory Compliance and the Use of Weighting to Enhance Standard or Rule Based Licensing Systems

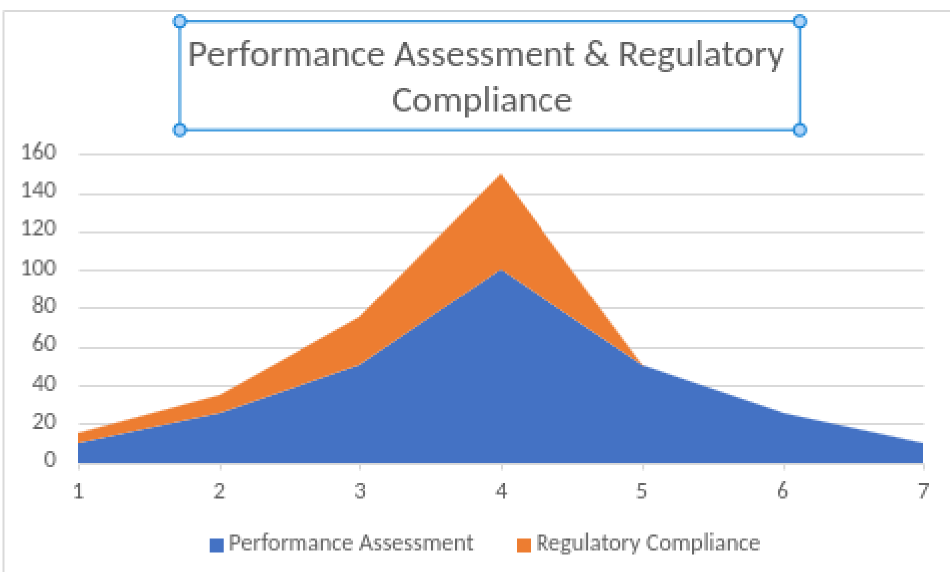
Richard Fiene, Ph.D.

May 2021

The purpose of this short paper is to delineate the commonalities and differences between performance assessment and regulatory compliance. In presenting performance assessments and regulatory compliance side by side it has the potential of introducing a new licensing measurement paradigm which goes beyond basic compliance with rules and standards. This paper builds upon previous technical research notes that are available at http://rikinstitute.com/blog/ which deal with the measurement issues related to licensing and regulatory compliance.

Whenever we think about performance assessments in the Environmental Rating Scales, CLASS, Accreditation Programs, or Quality Rating and Improvement Systems (QRIS), we find more normally distributed curves or distributions where skewness and kurtosis being very low. With regulatory compliance, the same type of normally distributed scores is not the case; the data are very skewed in a positive fashion which means that the majority of the programs are in full compliance (100%) with all the rules or standards. The resulting skewness and kurtosis are much higher which clearly indicates the non-parametric characteristics of the distribution. See the following Table.

Table 1: Data Distributions for Performance Assessment and Regulatory Compliance



Let's walk through Table 1 and discuss the commonalities and differences between performance assessment and regulatory compliance. The vertical axis is a frequency count, the number of programs meeting the particular scores on the horizontal axis. The horizontal axis runs from 1 = Deficient to 7 =

Exemplary. Four (4) = Compliant or Average. These scores represent how well a program meets the rules or standards that are being applied. Anything measured at a 4 or lower would be measured as a risk mitigation while anything above a 4 would measure performance above a specific compliance with the rule standard which is a score of 4. It is suggested that in order to increase the variance in the scoring protocol, weights be applied which measure relative risk or relative performance above or below the average score of 4. In the licensing research literature these would equate to a Risk Assessment Matrix (RAM) or a Performance Assessment Matrix (PAM).

An important discerning characteristic of the two distributions is the continuous nature of the performance assessment scores and the truncated nature of the regulatory compliance scores. The regulatory compliance scores essentially go up to a score of 4 on the Table 1 graphic which indicates full compliance with the rule/standard. It does not continue on as the performance assessment scores do.

The above graphic depiction is presented as a potential licensing measurement paradigm shift in how to think about the relationship between regulatory compliance and performance assessments. Generally, in the past, these two measurement systems have had their own silos and have not been looked at side by side. This paper is suggesting that we alter our vantage point and begin to see these two measurement systems along a continuum one building on the other in a stepped type of model.

RIKI Technical Research Note on the Licensing Key Indicator Predictor Methodology Threshold Updates, Regulatory Compliance, False Positives & Negatives, Data Dichotomization, and Licensing Measurement

April 2021

The purpose of this technical research note is to provide the latest updates to the Key Indicator Predictor Methodology and associated measurement issues, such as eliminating or reducing false positives and negatives, the use of data dichotomization with regulatory compliance frequency distributions.

It has always been recommended that a data dichotomization model be employed in distinguishing between the highly regulatory compliant from the low levels of regulatory compliance. The suggested model was 25/50/25 in which the top 25% constituted the highly compliant group, the middle 50% constituted the substantial – mid range compliant group, and the bottom 25% constituted the low compliant group. This was different from what had been done in the past in which fully compliant (100%) facilities were compared with those facilities who had any violations of regulatory compliance. It was found that by utilizing the 25/50/25 model a clearer distinction could be made between the high and low compliant groups. Generally, the top 25% are those facilities that are in full (100%) compliance, with the middle 50% are those facilities that have regulatory non-compliance ranging from 1 – 10 violations. The bottom 25% are those facilities that have regulatory non-compliance of greater than 10 violations. In this dichotomization model, the middle 50% are not used in the calculations, only the top and bottom 25%.

The dichotomization model described in the above paragraph has worked very well in producing licensing key indicator predictor rules by eliminating false negatives and decreasing false positives in the resultant 2 x 2 Key Indicator Predictor Matrix. The Fiene Coefficients for the licensing key indicator predictor rules have been more stable and robust by utilizing this model. It was made possible because of the increasing sample sizes selected for analyses and in some cases where population data were available. Also, the overall level of full compliance in states/provinces has increased over time and that has been a contributing factor as well in eliminating false negatives. False positives have been decreased because of the same factors but will never be eliminated because of the nature of the data distribution being highly positive skewed. Because of this distribution, there will always be false positives identified in the analyses. But that is the lesser of two evils: a rule being in compliance although it is present in the low regulatory compliant group.

However, are there ways to mitigate the impact of false positives. Based upon results from the *Early Childhood Program Quality Improvement & Indicator Model Data Base (ECPQI2MDB)* maintained at the Research Institute for Key Indicators/Penn State, there appears to be several adjustments that can be made so that the impact of false positives is not as pronounce as it has been in the past. The first adjustment that can be made is to increase the sample size so that additional non-compliance is identified. This is difficult at times because the nature of licensing or regulatory compliance data trends towards very high compliance for most facilities with little non-compliant facilities. It is the nature of a regulatory compliance or licensing program; these are basic health and safety rules which have had a history of substantial to full compliance with the majority of the rules. The data are extremely positively skewed. There is little variance in the data. So, increasing the sample size should help on all these accounts. In addition to increasing the sample size, an additional methodology was developed in order to increase the variance in licensing/regulatory compliance data by weighting rules/regulations based upon the risk children are placed in because of non-compliance. This proposal makes a great deal of sense but its application in reality hasn't played out as intended. What most jurisdictions do in implementing the risk assessment methodology is to identify the most heavily weighted rules but then to deal with these rules as high risk rules and not using the weights assigned to them for aggregating regulatory compliance scores. The use of the methodology in this way is very effective in identifying the specific rules based upon risk, but does little to nothing in increasing the variance in the regulatory compliance data distribution. The data distribution remains severely positively skewed.

Another way to mitigate the impact of false positives is to increase the data dichotomization of the data distribution but this is recommended only with the increase sample size. If it is done without an increased sample size, the resultant Fiene Coefficients for the licensing key indicator predictor rules will be less robust and stable. For example, the data dichotomization model of 25/50/25 could be increased to a 10/80/10 model which should help in decreasing the false positives in the analyses. But this is cautionary, for example, in going to a 5/90/5 model could again make the resultant Fiene Coefficients for the licensing key indicator predictor rules less robust and stable. The sample size needs to be very large or the full population needs to be measured in order to do these analyses and co-balance the increased data dichotomization because the cell sizes will be decreasing significantly. The following 2 x 2 matrix will depict these relationships for generating the Licensing Key Indicator Predictor Fiene Coefficients (FC).

Licensing Key Indicator Predictor Fiene Coefficient (FC) Table

Individual Rules/Groups ->	High Compliant (Top 25%)	Low Compliant (Bottom 25%)
Rule In Compliance	FC (++)	FP (+)
Rule Out of Compliance	FN (-)	FC (--)

((FC (++) + (FC (--)) > ((FN (-)) + (FP (+))

where FC = Fiene Coefficient which results in Licensing Key Indicator Predictor Rules (FC = .25 or >);

FN (-) = False Negative; FP (+) = False Positive

The cells represented by the Fiene Coefficients should always be larger than the False Positive and Negative results in the above table. With the above dichotomization 25/50/25 model and high levels of full 100% regulatory compliance, false negatives can be eliminated and by increasing the sample size, false positives will be decreased but never fully eliminated. Full 100% regulatory compliance increased levels will help to eliminate false negatives, but it will also increase the chances of false positives. There is a delicate balance with confounding the increased sample sizes (false positives will decrease) and increased levels of full 100% regulatory compliance (false positives will increase). This will take a bit of adjusting to get this balancing just right.

By utilizing the *ECPQI2MDB* it has demonstrated that the above-mentioned dichotomization models may be difficult to hit the percentages exactly. The actual models may be more heavily weighted in the percent for the high group as versus the low because of the regulatory compliance data distribution being highly positive skewed as mentioned earlier. This may have an impact on the Fiene Coefficients (FC) for licensing key indicator predictor rules but it will not impact the actual selection of the licensing key indicators – they will remain the same, just the FCs will change.

One last footnote on the relationship between regulatory compliance and program quality. This relationship has been addressed several times over the past four decades in the regulatory science and human services regulatory administration fields; but it needs to be re-emphasized as it relates to this discussion about licensing measurement. Regulatory compliance and program quality are linear and non-random in moving from low regulatory compliance to mid-substantial regulatory compliance as with low program quality to mid program quality. However, when one moves from substantial regulatory compliance to full 100% regulatory compliance the relationship with program quality is more non-linear and random.

Regulatory Compliance, Licensing, and Monitoring Measurement Principles: Rule Compliance Versus Rule Performance

Richard Fiene, Ph.D.

January 2021

The purpose of this short paper is to delineate the parameters of regulatory compliance, licensing and monitoring measurement principles (throughout this paper the term "regulatory compliance" will be used to encompass these principles). Regulatory compliance is very unique when it comes to measuring it because it is very different from other measurement systems and this impacts how one uses various statistical analyses. In this paper, the limitations of the measurement system will be highlighted with potential solutions that have been devised over the past several decades. Hopefully this paper will add to the measurement and statistical analysis licensing research literature. It is meant for those agency staff who are responsible for designing regulatory compliance, licensing and monitoring systems. Its focus is the human services but the basic principles can be applied to any standards-based system that is based upon a compliance or performance model.

The organization of this paper is as follows. First, let's introduce what is included when we talk about measurement principles for regulatory compliance, licensing and monitoring systems. Second, provide examples that should be familiar to most individuals who have been involved in the human services, in particular the early care and education field. Third, what are the limitations of these various systems that have been identified in the research literature. Fourth, what are some potential solutions to these limitations. And, fifth, what are the next steps and where do we go to build reliable and valid measurement systems dealing with regulatory compliance, licensing, and program monitoring as these relate to the human services delivery system.

So, what is included in this approach. I can be any rule, regulation, or standard based measurement system. Generally, these systems are focused on a nominally based system, sometimes they will be ordinally based. By a nominally based system, either the facility being assessed is in compliance with a particular set of rules, regulations, or standards or it is not. In an ordinally based system, a facility may attain a score on a Likert scale, such as 1 through 5 where 1 is non-optimal and 5 is excellent. These types of measurement scales involve a performance component and are not limited to more of a compliance focus as is the case with a nominally based system. These distinctions are important as one will see later in this paper when it comes to the selection of the appropriate statistics to measure data distributions and the subsequent analyses that can be undertaken.

What are examples of these types of systems? For nominally based systems, just about all the licensing systems in the USA, Canada and beyond employ this type of measurement strategy. As has been said in the previous paragraph, either there is compliance or there is not. It is very black or white, there are not shades of gray. For ordinally based systems, these systems are a bit more diverse. Accreditation, Quality Rating and Improvement Systems (QRIS), the new Head Start Grantee Performance Management System (GPMS), the Environmental Rating Scales, and the CLASS are all examples of ordinally based systems based upon a Likert type measurement system. There are many others, but as a research psychologist whose total career (50 years) has been spent in early care and education, this has been the focus of my research.

The limitations of the above systems are numerous and, in some ways, are difficult to find solutions. In the past, these measurement systems have focused more on the descriptive aspects of data distributions rather than attempting to be predictive or inferential. The first major limitation of the data from regulatory compliance systems is the fact that the data distribution is markedly skewed. What does skew data mean? Most data distributions are normally distributed with very few occurrences at the extremes with the majority of the cases in the middle section of the measurement scale. IQ is an example of a normally distributed data distribution. In a skew data distribution, the majority of data are at one end of the data distribution, either at the positive end or the negative end of the distribution. With regulatory compliance data, it is at the positive end with the majority of facilities being in full or 100% compliance with the rules. Very few of the facilities are at the negative end of the distribution.

What is the big deal? The big deal is that statistically we are limited in what we can do with the data analyses because the data are not normally distributed which is an assumption when selecting certain statistical tests. Basically, we need to employ non-parametric statistical analyses to deal with the data. The other real limitation is in the data distribution itself. It is very difficult to distinguish between high and mediocre facilities. It is very easy to distinguish between high and low performing facilities because of the variance between the high performing facilities and the low performing facilities. However, that is not the case between high and mediocre preforming facilities. Since the majority of facilities are either in full or substantial compliance with the rules, they are all co-mingled in a very tight band with little data variance. This makes it very difficult to distinguish differences in the facilities. And this only occurs with regulatory compliance data distributions. As will be pointed later in this paper, this is not the case with the second measurement system to be addressed dealing with ordinal measurement systems.

There is also a confounding factor in the regulatory compliance data distributions which has been termed the theory of regulatory compliance or the law of regulatory compliance diminishing returns. In this theory/law, when regulatory compliance data are compared to program quality data, a non-linear relationship occurs where either the facilities scoring at the substantial compliance level score better than the fully compliant facilities or there is a plateau effect and there is no significant difference between the two groups: substantial or fully compliant facilities when they are measured on a program quality scale. From a public policy stand point, this result really complicates how best to promulgate compliance with rules. This result has been found repeatedly in early care and education programs as well as in other human service delivery systems. It is conjectured that the same result will be found in any regulatory compliance system.

Another limitation of regulatory compliance data is the fact that it is measured at a nominal level. There is no interval scale of measurement and usually not even an ordinal level of measurement. As mentioned above, either a facility is in compliance or not. From a statistical analytical view, again this limits what can be done with the data. In fact, it is probably one of the barriers for researchers who would like to conduct analyses on these data but are concerned about the robustness of the data and their resulting distributions.

Let's turn our attention to potential solutions to the above limitations in dealing with regulatory compliance data.

One potential solution and this is based upon the theory of regulatory compliance in which substantial compliance is the threshold for a facility to be issued a license or certificate of compliance. When this public policy determination is allowed, it opens up a couple of alternate strategies for program monitoring and licensing reviews. Because of the theory of regulatory compliance/law of regulatory compliance diminishing returns, abbreviated or targeted monitoring reviews are possible, differential monitoring or inferential monitoring as it has been documented in the literature. This research literature on differential monitoring has been dominated by two approaches: licensing key indicators and weighted risk assessments.

A second solution to the above limitations deals with how we handle the data distribution. Generally, it is not suggested to dichotomize data distributions. However, when the data distribution is significantly skewed as it is with regulatory compliance, it is an appropriate adjustment to the data. By essentially having two groups, those facilities that are in full compliance and those facilities that are not in full compliance with the rules. In some cases, the fully compliant group can be combined with those facilities that are in substantial compliance but this should only be employed when there are not sufficient fully compliant facilities which is hardly never the case since population data and not sampled data are available from most jurisdictions. When data samples were drawn and the total number of facilities were much smaller, substantial compliant facilities were used as part of the grouping strategy. The problem in including them was that it increased the false negative results. With them not being included, it is possible to decrease and eliminate false negatives. An additional methodological twist is also to eliminate and not use the substantial compliant facilities at all in the subsequent analyses which again helps to accentuate the difference scores between the two groups of highly compliant and low compliant scoring facilities.

The next steps for building valid and reliable regulatory compliance systems are drawing upon what has been learned from more ordinally based measurement systems and applying this measurement structure to regulatory compliance systems. As such, the move would be away from a strict nominally based measurement to more ordinal in which more of a program quality element is built into each rule. By utilizing this paradigm shift, additional variance should be built into the measurement structure. So rather than having a Yes/No result, there would be a gradual Likert type (1-5) scale built in to measure "rule performance" rather than "rule compliance" where a "1" indicates non-compliance or a violation of the specific rule. A "5" would indicate excellent performance as it relates to the specific rule. A "3" would indicate compliance with the specific rule meeting the specifics of the rule but not exceeding it in any way.

This paradigm shift has led to the creation of Quality Rating and Improvement Systems (QRIS) throughout the USA because of a frustration to move licensing systems to more quality focused. The suggestion being made here is to make this movement based upon the very recent developments in designing such systems as is the case with Head Start monitoring. Head Start GPMS is developing an innovative Likert based ordinal system which incorporates compliance and performance into their monitoring system. Other jurisdictions can learn from this development. It is not being suggested as a replacement for QRIS or accreditation or ERS/CLASS assessments but as a more seamless transition from licensing to these various assessments. As indicated by the theory of regulatory compliance and the law of regulatory compliance diminishing returns, this relationship between licensing and program quality is not linear. By having this monitoring system approach in place, it may be able to reintroduce more of a linear relationship between licensing and program quality.

Regulatory Compliance & Program Quality Grid Model: Technical Research Note

Richard Fiene, Ph.D.

December 2020

Depicted below if a regulatory compliance grid model showing the relationship between regulatory compliance (RC) and program quality (PQ).

An explanation of the below chart will demonstrate how regulatory compliance and program quality in human service facilities interact. The horizontal blue axis depicts the various levels of regulatory compliance while the vertical green axis depicts the various levels of program quality of facilities. It ranges from 1-5 or low to high for each axis. The red "**X's"** represent the relationship that has been identified in the research literature based upon the theory of regulatory compliance in which there is either a plateau effect or a downturn in quality as regulatory compliance increases. The one italized "***X***" is an outlier that has also been identified in the research literature in which sometimes (it does not happen often) low compliant programs really are at a high quality level.

It is proposed in order to mitigate the plateau effect with regulatory compliance and program quality standards because regulatory compliance data distributions are severely skewed which means that many programs that have questionable quality are being included in the full (100%) compliance domain. When regulatory compliance standards are increased in their quality components this will lead to a higher level of overall quality as depicted in the "**XX**" cell all the way on the lower right. It also helps to mitigate the severe skewness in the regulatory compliance data distribution. The data distribution does not approximate a normally distributed curve which is the case with the program quality data distribution.

Regulatory Compliance x Program Quality Grid Model

PQ/RC ->	1 Low	2 Med	3 Substantial	4 Full 100%	5QualityAddons
1 Low	XXX				
2		XX			
3 Med			XX	XXX	
4			XX	X	
5 High	***X***				**XX**

By utilizing this model, it helps to deal more directly in taking a non-linear relationship and making it linear again when comparing regulatory compliance with program quality. This model provides a theoretical approach supporting what many state licensing administrators are thinking from a policy standpoint: add more quality to health and safety rules/regulations. This grid/matrix also depicts the three regulatory compliance models: Linear, Non-linear, and Stepped.

Theory of Regulatory Compliance Models

Richard Fiene, Ph.D.

August 2018

Three models are presented here which depict the theory of regulatory compliance as it has evolved over the past four decades. Initially, it was thought that there was a linear relationship between regulatory compliance and program quality as depicted in the first line graph below (see Figure 1). As compliance increased a corresponding increase in quality would be seen in the respective programs.

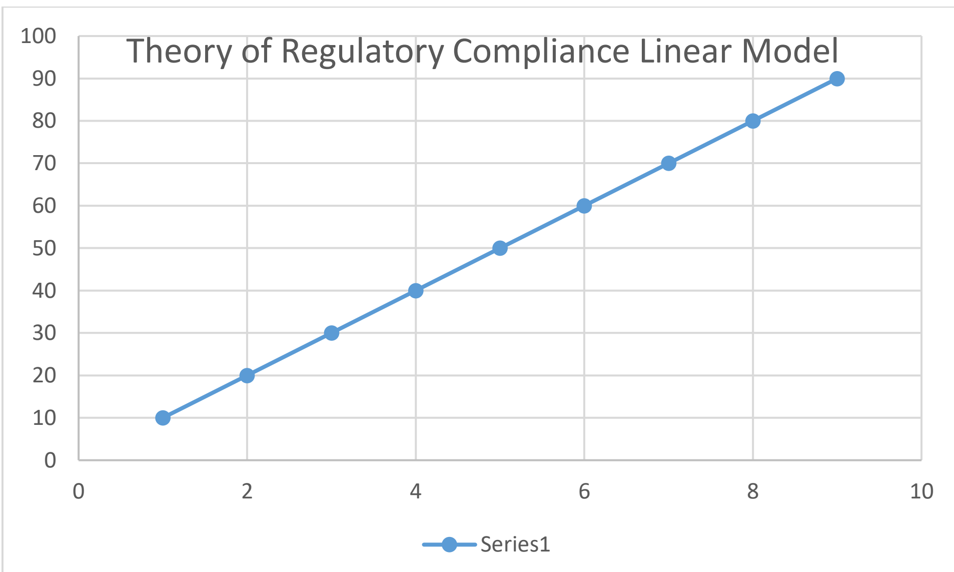


Figure 1

This initial graphic needed to be modified because of various studies conducted in order to confirm this regulatory compliance theory. It was discovered that at the lower ends of regulatory compliance there still was a linear relationship between compliance and quality. However, as the compliance scores continued to increase to a substantial level of compliance and then finally to full (100%) compliance with all rules, there was a corresponding drop off in quality as depicted in the second line graph below (see Figure 2).

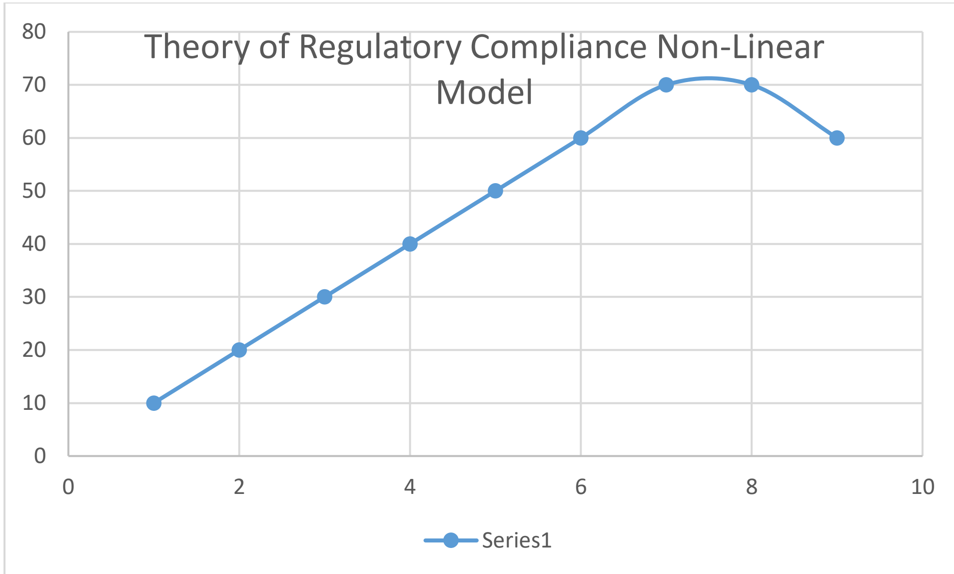


Figure 2

This Non-Linear Model has worked well in explaining the Theory of Regulatory Compliance and the studies conducted for the past three decades. However, the most recent studies related to the theory appear to be better explained by the latest proposed model in Figure 3 which suggests using a Stepped Model rather than a Non-Linear Model. The Stepped Model appears to explain more fully how certain less important rules can be significant predictors of overall compliance and quality.

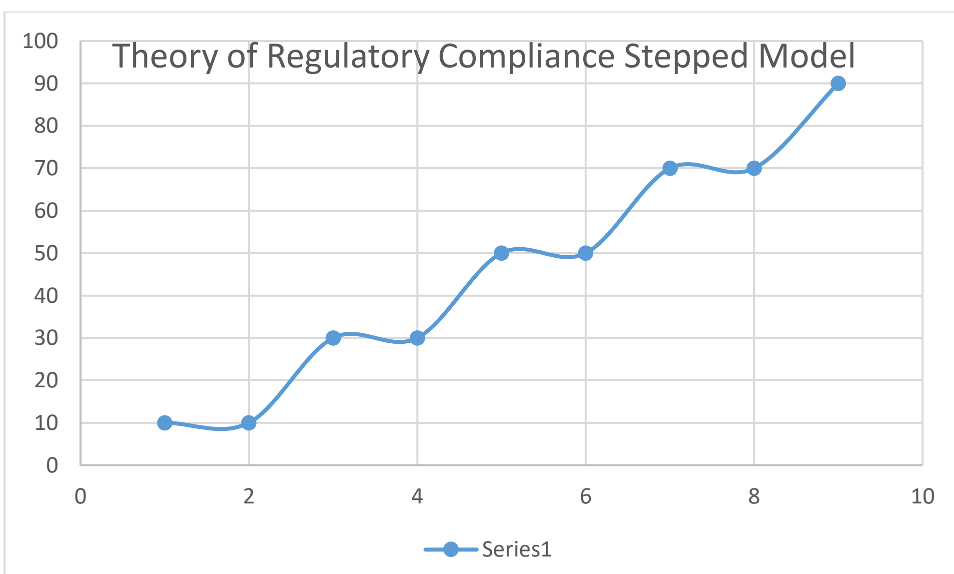


Figure 3

This last model has more flexibility in looking at the full regulatory field in attempting to find the "predictor" or right rules that should be selected as key indicators. It is about identifying those key indicator rules that move the needle from one step to the next rather than focusing on the plateau. So rather than having just one plateau, this model suggests that there are several plateaus.

Richard Fiene, Ph.D., Research Psychologist, Research Institute for Key Indicators (RIKILLC); Senior Research Consultant, National Association for Regulatory Administration (NARA); and Professor of Psychology (ret), Penn State University.

The Implications in Regulatory Compliance Measurement When Moving from Nominal to Ordinal Scaling

Richard Fiene, Ph.D.

May 2018

The purpose of this paper is to provide an alternate paradigm for regulatory compliance measurement in moving from a nominal to an ordinal scale measurement strategy. Regulatory compliance measurement is dominated by a nominal scale measurement system in which rules are either in compliance or out of compliance. There are no gradients for measurement within the present licensing measurement paradigm. It is very absolute. Either a rule is in full compliance to the letter of the law or the essence of the regulation or it is not. An alternate paradigm borrowing from accreditation and other program quality systems is to establish an ordinal scale measurement system which takes various gradients of compliance into account. With this alternate paradigm, it offers an opportunity to begin to introduce a quality element into the measurement schema. It also allows to take into consideration both risk and prevalence data which are important in rank ordering specific rules.

So how would this look from a licensing decision making vantage point. Presently, in licensing measurement, licensing decisions are made at the rule level in which each rule is either in or out of compliance in the prevailing paradigm. Licensing summaries with corrective actions are generated from the regulatory compliance review. It is a nominal measurement system being based upon Yes/No responses. The alternate measurement paradigm I am suggesting in this paper is one that is more ordinal in nature where we expand the Yes/No response to include gradients of the particular rule. In the next paragraph, I provide an example of a rule that could be measured in moving from a nominal to ordinal scale measurement schema.

Rather than only measuring a rule in an all or none fashion, this alternate paradigm provides a more relative mode of measurement at an ordinal level. For example, with a professional development or training rule in a particular state which requires, let's say, 6 hours of training for each staff person. Rather than having this only be 6 hours in compliance and anything less than this is out of compliance, let's have this rule be on a relative gradient in which any amount of hours above the 6 hours falls into a program quality level and anything less than the 6 hours falls out of compliance but at a more severe level depending on how far below the 6 hours and how many staff do not meet the requirement (prevalence). Also throw in a specific weight which adds in a risk factor and we have a paradigm that is more relative rather than absolute in nature.

From a math modeling perspective, the 1 or 0 format for a Yes or No response becomes -2, -1, 0, +1, +2 format. This is more similar to what is used in accreditation systems where 0 equals Compliance and -1 and -2 equals various levels of Non-Compliance in terms of severity and/or prevalence. The +1 and +2 levels equal value added to the Compliance level by introducing a Quality Indicator. This new formatting builds upon the compliance vs non-compliance dichotomy (C/NC) but now adds a quality indicator (QI) element. By adding this quality element, we may be able to eliminate or at least lessen the non-linear relationship between regulatory compliance with rules and program quality scores as measured by the

Environmental Rating Scales (ERS) and CLASS which is the essence of the Theory of Regulatory Compliance (TRC). It could potentially make this a more linear relationship by not having the data as skewed as it has been in the past.

By employing this alternate paradigm, it is a first demonstration of the use of the Key Indicator Methodology in both licensing and quality domains. The Key Indicator Methodology has been utilized a great deal in licensing but in few instances in the program quality domain. For example, over the past five years, I have worked with approximately 10 states in designing Licensing Key Indicators but only one state with Quality Key Indicators from their QRIS – Quality Rating and Improvement System. This new paradigm would combine the use in both. It also takes advantage of the full ECPQI2M – Early Childhood Program Quality Improvement and Indicator Model by blending regulatory compliance with program quality standards.

A major implication in moving from a nominal to an ordinal regulatory compliance measurement system is that it presents the possibility of combining licensing and quality rating and improvement systems into one system via the Key Indicator Methodology. By having licensing indicators and now quality indicators that could be both measured by licensing inspectors, there would be no need to have two separate systems but rather one that applies to everyone and becomes mandated rather than voluntary. It could help to balance both effectiveness and efficiency by only including those standards and rules that statistically predict regulatory compliance and quality and balancing risk assessment by adding high risk rules.

I will continue to develop this scale measurement paradigm shift in future papers but wanted to get this idea out to the regulatory administration field for consideration and debate. This will be a very controversial proposal since state regulatory agencies have spent a great deal of resources on developing free standing QRIS which build upon licensing systems. This alternate paradigm builds off my Theory of Regulatory Compliance's key element of relative vs absolute measurement and linear vs non-linear relationships. Look for additional information about this on my website RIKI Institute Blog - https://rikinstitute.com/blog/.

Richard Fiene, Ph.D., Senior Research Psychologist, Research Institute for Key Indicators; Professor of Psychology (retired), Penn State University; and NARA Senior Research Consultant. Rjf8@psu.edu. http://RIKInstitute.com.