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.

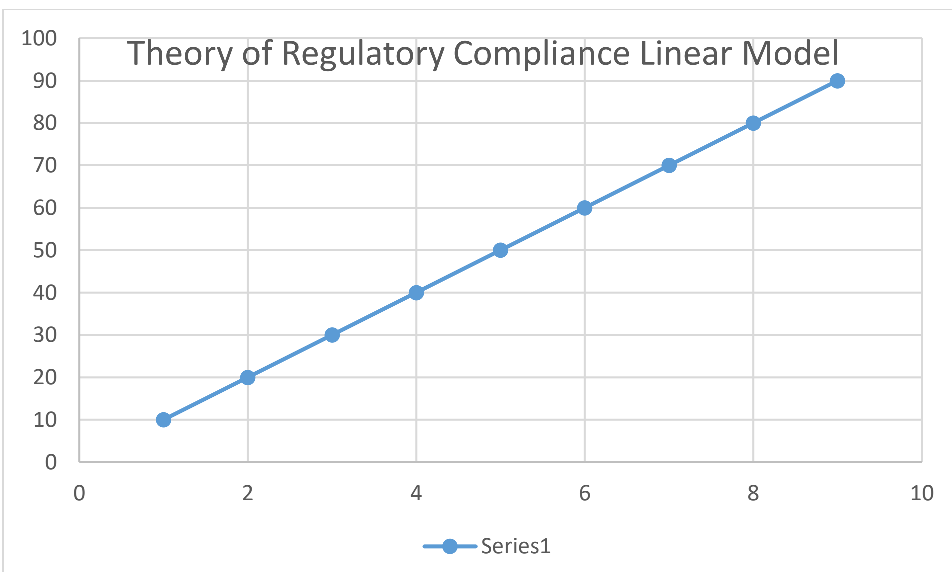


Figure 1

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).

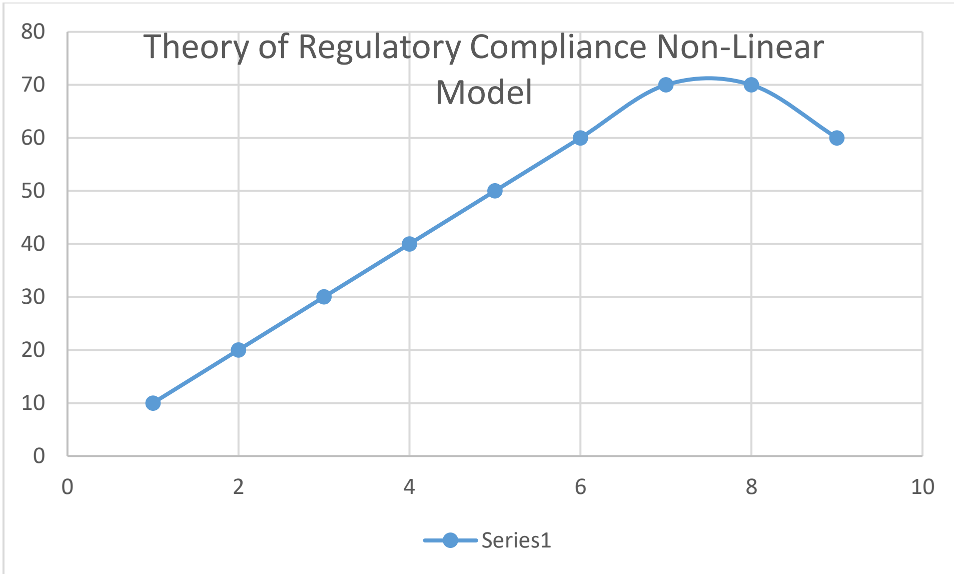


Figure 2

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.

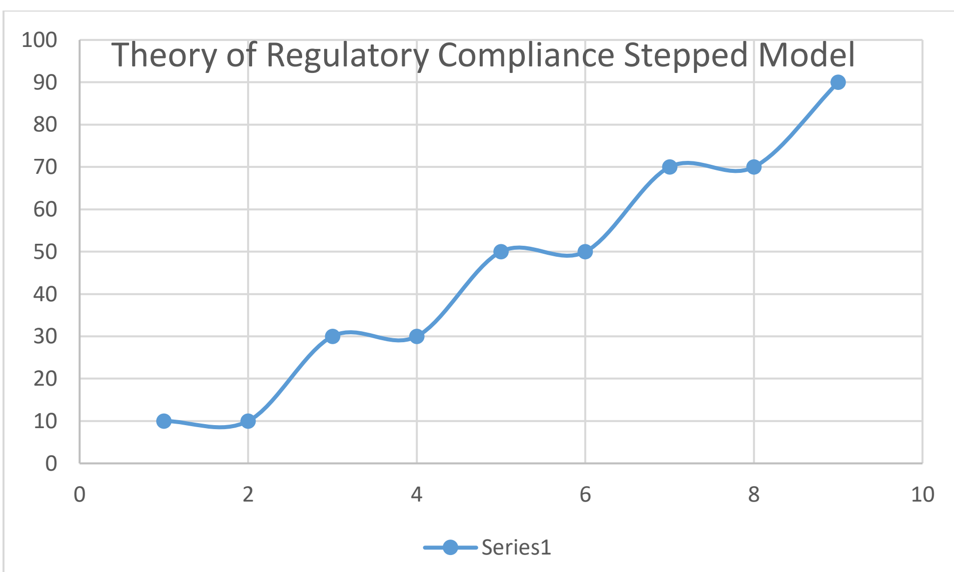


Figure 3

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.

GPMS Performance Assessment Matrix Scoring Algorithm

Measuring Compliance and Performance Together in One System

This slide deck will present a proposed scoring algorithm for the Office of Head Start's Grantee Performance Management System (GPMS). It is a unique application of risk assessment matrices by measuring both compliance and performance in one system.

Performance Assessment Matrix (PAM) Measuring Compliance/Performance

Prevalence -> Severity or Importance (below)	Often (+/-)	Sometimes (+/-)	Rarely (+/-)	Performance Measure Weights (below)
High	**-9,0,+9**	**-8,0,+8**	**-7,0,+7**	**Top PMs**
Mid	**-6,0,+6**	**-3,0,+3**	**-4,0,+4**	**Mid PMs**
Low	**-3,0,+3**	**-2,0,+2**	**-1,0,+1**	**Low PMs**

Provides the basic 3 x 3 Matrix used in many risk assessment matrices which assess potential risk as a combination of prevalence, how likely is something to happen and if it does happen what are the potential consequences or risks. What is different about the Performance Assessment Matrix is that it measures this combination of likelihood & severity in both a positive approach (importance/performance) and a negative (severity/compliance) approach. The negative approach has been described, the positive approach measures the 3 x 3 matrix in terms of importance.

The Performance Measures (PMs) are then placed along this severity or importance scale based upon their relative median weights as determined by a stakeholders ranking. The prevalence data will be drawn from the EAS measurement protocol for each PM.

Key Elements of GPMS:PAM

- **The GPMS:PAM is based upon Risk Assessment Matrices but it takes into account both importance as well as severity (+/-).**
- **The GPMS:PAM accounts for weighting of the PMs by ranking them from High to Low in importance based upon their Medians. Severity is measured by the inverse of Importance: the more it is not present, the worse it is.**
- **It builds off of the EAS protocol (Often, Sometimes, Rarely) in order to measure prevalence.**

There are three key elements to the Grantee Performance Measures System: Performance Assessment Matrix - (GPMS:PAM):
1) It builds upon the risk assessment matrices research that is very pervasive in the regulatory science field in which relative risk as measured by prevalence or scope of occurrence and the severity or actual risk to a group are measured together. The bottom line with risk assessment matrices is what risk will the event pose for a group and what is the likelihood that the event will occur. An event would be rated much higher if it poses severe risk and it is likely to happen than an event which has low risk and is unlikely to happen. And of course there are middle ground risk assessment scores where the risk may be high but the likelihood is extremely low.
2) What is different about the GPMS:PAM is that it measures not only "severity" but it also measures "importance". So two matrices need to be built to measure these two concepts. Severity is measured on a typical "compliance" risk assessment matrix, while importance will be measured on a "performance" matrix.
3) The PMs to be measured within this system have been rank ordered from high to low by a representative group of stakeholders involved in Head Start. There are 23 PMs in total with 9 PMs ranked high, 7 PMs ranked mid-range, and 7 PMs ranked at a low-range. Each of the PMs will be measured using the Evidence Assessment System - EAS protocol.

GPMS Scoring Algorithm (SA)

- **GPMS:SA = (PAM(HSPS(EAS:ΣPM1+PM23))) + (QRIS + ACC + PD) + (LIC)**
- **<u>Where:</u>**
- **PAM = -144 thru +144 (Core Score); HSPS = Standards; EAS = Evidence Assessment System (-1, 0, +1)**
- **Σ PM1:PM23 = the 23 PMs to determine compliance & performance**
- **QRIS = 0 thru 5 (Quality Initiative Add on)**
- **ACC = 0 or 5 (Quality Initiative Add on)**
- **PD = 0 thru 5 (Quality Initiative Add on)**
- **LIC = 1 for full license**
- **CLASS = Dependent Variable and should positively correlate with GPMS Score.**
- **Risk Indicators = Precursor to GPMS Score.**

This slide provides the overview to the GPMS scoring algorithm and the components that make up its composition. The GPMS scoring algorithm is made up of the 23 PMs along with potential other inputs from external systems, such as: Quality Rating and Improvement Systems (QRIS), Accreditation (ACC), and Professional Development (PD). All these systems are voluntary systems and are value-added to the overall scoring algorithm. A grantee will not be penalized if they do not participate in any of these other quality initiatives but if they do, they will have bonus points added to their overall score. Also, licensing data should be addressed.

Risk indicators are potential flags which are measured as precursors to an actual review. They may provide guidance in how a review will be done and what to target or focus upon.

The last component is the CLASS which is being proposed as a dependent variable not to be included within the scoring algorithm at this point. If it is included then the latest OHS Regulatory Guidance for the Designated Renewal System needs to be used in order to rank order the specific scoring from the CLASS.

GPMS Scoring Algorithm Sequencing

- **EAS: -1, 0, +1; Rarely = -1; Sometimes = 0; Often = +1.**
- **EAS revised: -2, -1, 0, +1, +2.**
- **EAS revised scoring: None = -2, Rarely = -1, Sometimes = 0, Often = +1, All = +2.**
- **PM: 1 = Deficiency (None/Rarely); 2 = Non-Compliance (Rarely); 3 = Issues (Sometimes); 4 = Compliance (Often); 5 = Exemplary (All).**
- **PM revised: 1, 2, 3, & 5 stay the same; 4 = 4a = Compliance, 4b = Compliance+.**
- **PAM: -9, -8 = PM1, 2; -7 = PM3; +7 = PM4a; +8 = PM4b; +9 = PM5.**

This slide provides the sequencing in moving from one measurement level to the next and provides two revisions in the scoring protocols in order to increase the variance and sensitivity of the scoring algorithm. This revision needs to especially occur at the Compliance level since the majority of grantees are scored at this level and skews the data dramatically based upon a historical review of data from 2020.

With the revised scoring at both the EAS and PM levels, it should provide additional data points to accomplish this mathematically.

GPMS Scoring Algorithm Detail

- **EAS: (-1, 0, +1) --> (-2, -1, 0, +1, +2)**
- **Standard: Compliance or Non-Compliance (1/0)**
- **PM: 1 = Def, 2 = NC, 3 = Issues, 4 = Compliance, 5 = Exemplary**
- **High PMs Score: (-9, -8, -7, +7, +8, +9) x (9PMs) = ((-81) -- (+81))**
- **Mid PMs Score: (-6, -5, -4, +4, +5, +6) x (7PMs) = ((-42) -- (+42))**
- **Low PMs Score: (-3, -2, -1, +1, +2, +3) x (7PMs) = ((-21) -- (+21))**
- **PAM Total Score Range (HighPMs+MidPMs+LowPMs): ((-144) -- (+144))**
- **Final Score: A=144-94; B= 93-43; C=42-(-42); D=(-43)-(-93); F=(-94)-(-144)**
- **OtherQI: Accreditation(5); QRIS(1-5); Professional Development(1-5)**

This slide provides the GPMS scoring algorithm details on how each level of measurement will be assessed.

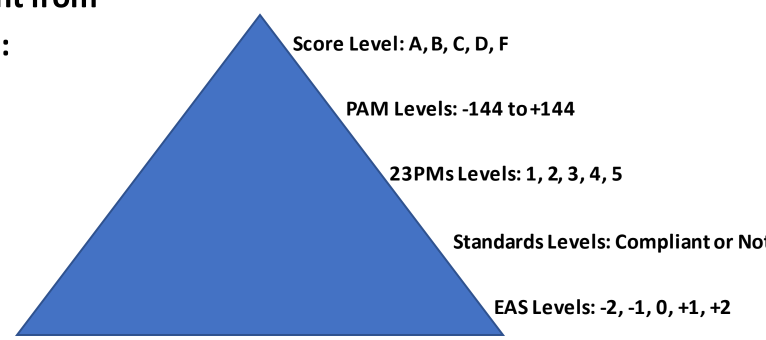
It starts with the EAS protocol, where it is being proposed that the present scoring protocol be expanded from a 3 point scale to a 5 point scale. The three point scale has the following basic characteristics: "rarely, sometimes, often" are used to measure each dimension within a standard/PM/Quality Indicator. It is being suggested that the "rarely, sometimes, and often" scale be transposed to a more mathematical scale as such: -1, 0, +1. Once this is done, it is suggested that this mathematical scale be expanded to the following when scoring each standard/PM: -2, -1, 0, +1, +2; where –2 = None and +2 = All. -1, 0 , +1 stay the same.

QI = Quality Initiatives such as: accreditation, quality rating and improvement systems, and professional development systems are potential Add Ons to the Scoring Algorithm.

Still need to determine how best to build in licensing.

PAM Measurement Hierarchy

- **Measurement from Bottom to Top:**



This pyramid provides a graphical display of how the various measurement levels relate to each other in a logical fashion.

The EAS Level is the basic, beginning level where all measurement begins. It is the most detailed and granular level, and has the most data points.

From the EAS level, these data can be aggregated upwards to the standard, PM, & Quality Indicator levels. There are more standards than PMs but there will be mapping that can occur.

From the PM level, aggregation occurs to the PAM level where all the PMs are added together to come up with a total score with a range of +144 thru to -144.

This PAM score can then be transposed to a specific score level: A, B, C, D, F. This last score level is what is shared with the grantee. All the other math/scoring goes on within the software behind the scenes.

Performance Assessment Matrix: Compliance

Prevalence -> Severity Weights (below)	Often	Sometimes	Rarely
High	**-9**	**-8**	**-7**
Mid	**-6**	**-5**	**-4**
Low	**-3**	**-2**	**-1**

This matrix depicts the relationship between prevalence and severity which is typical in risk assessment matrices.

Performance Assessment Matrix: Importance

Prevalence -> Importance Weights (below)	Often	Sometimes	Rarely
High	+9	+8	+7
Mid	+6	+5	+4
Low	+3	+2	+1

This matrix builds off of the risk assessment matrix as depicted in the previous slide where prevalence cells stay the same but the weights are more positive than negative as in the previous slide.

A higher score (+9) is a good result where in the previous slide a higher score (-9) is not a good result. This matrix needs to be used in conjunction with the previous slide in order to determine the overall PAM Score.

This matrix clearly builds upon the risk assessment matrix research but expands it to account for both positive (+) and negative (-) results. In this matrix, a high importance rating = a high performance rating.

GPMS Scoring Algorithm Levels

Measurement Level	Deficient	Non-Compliance	Compliant with Issues	Compliant	Exemplary
EAS	None -2	Rarely -1	Sometimes 0	Often +1	All +2
PM	1	2	3	4	5
PAM	-9	-8	-7	+7/+8	+9
Final Score	F (-144) - (-94)	D (-93) - (-43)	C (-42) - (+42)	B (+43) - (+93)	A (+94) -(+144)

This graphic depicts the relationship amongst the measurement levels demonstrating how the measurement proceeds from one level to the next level. This matrix provides the detail to the pyramid that is depicted on Slide #7.

The final score measurement level has a preponderance of scores in the central "C" score level because of the supposition that the scaling will be closer to normally distributed given the revised EAS and PAM scoring. However, only with additional data collection and testing will we be able to improve upon this supposition.

Performance Assessment Matrix (PAM) Simulation Examples

Simulation	PM Rating = 1 (-9) Deficient	PM Rating = 2 (-8) Non-Compliance	PM Rating = 3 (-7) Issues Compliant	PM Rating = 4 (0 or +7) Full Compliance	PM Rating = 5 (+8/+9) Exemplary	Total PAM Score ((-81) -- (+81)) or 9-45 Raw Score
Compliance = +7	0	0	0	9	0	+63
Compliance = 0	0	0	0	9	0	0
Raw Scores	0	0	0	9	0	36

This slide and table provides the results of three simulations where certain assumptions were made from a scoring protocol:

The first, bottom row, Raw Scores, contains no weights or prevalence scoring as depicted in the PAM. It is a sum of the PM x the number of PMs. So in the example provided, there were 9 PMs assessed with all of them rated at a "4" level = 36.

The second, middle row, using the PAM but the "Full Compliance Rating" = 0 and has no specific weight/prevalence result. Again, 9 PMs were rated but in this case a "Full Compliance Rating" = 0. Nine PMs x "0" = "0".

The last, top row, uses the PAM but with "Full Compliance Rating" = +7. In this case, the (9PMs) x (+7) = +63. This last simulation provides the greatest variance in the data covering 85% of the full PAM score range. This is significantly better than the other two simulations where only about 50% of the score ranges. These simulations were applied to historical data taken from FA2 2020 reviews.

Performance Assessment Matrix (PAM) Simulation for High PMs (9PMs = High)

Grantee Number in Data Base	PM Rating = 1 (-9) Deficient	PM Rating = 2 (-8) Non-Compliance	PM Rating = 3 (-7) Issues Compliant	PM Rating = 4 (+7) Full Compliance	PM Rating = 5 (+8/+9) Exemplary	Total PAM Score ((-81) -- (+81)) (A - F)
001	0	0	0	9	0	+63 (B)
011	0	1	4	4	0	-8 (C)
020	3	5	1	0	0	-74 (D-)
051	0	3	5	1	0	-52 (D)

Using the simulation with Full Compliance = +7 from the previous slide #11, four examples are provided to demonstrate how the scoring would actually play out with PAM and final scores (A thru F) that are provided to the grantees.

This simulation is based upon reviews from FA2 2020 data related to only the high PMs. The scoring algorithm may need additional tweaking once new data are entered starting in 2021 and all the PMs are included. There will be the need to constantly test the model with new actual data as we move forward to make sure the scoring logic holds.