



Compliance scoring review

Final Report to the Marine Stewardship Council

December 2018

Compliance scoring review.

FINAL Report to the Marine Stewardship Council.

Geir Hønneland

December 2018

The views and opinions expressed in this report are those of the authors and do not necessarily reflect the official policy or position of the Marine Stewardship Council. This work is licensed under Creative Commons BY 4.0. To view a copy of this license, visit (<https://creativecommons.org/licenses/by/4.0>)

How to reference this report:

Hønneland, G (2018). Compliance scoring review. Published by the Marine Stewardship Council [www.msc.org]. 17 pp

Table of contents

1	Introduction.....	3
1.1	Problem statement.....	3
1.2	Objectives.....	3
1.3	The data.....	4
2	Empirical overview	6
2.1	Under which SIs is compliance assessed?.....	6
2.2	How is the SIc score substantiated?.....	7
2.3	What sources are used for scoring compliance?	8
3	Discussion and conclusions	10
3.1	Summary.....	10
3.2	Assessment	12
3.3	Recommendations.....	15

1 Introduction

1.1 Problem statement

The effective management of fisheries is underpinned by the compliance of fishers.

Harmonisation of scoring

- Conditionality between scoring of compliance and the evidence required to demonstrate effectiveness of MCS and sanctions
- Potential ambiguity over whether teams consider the extent to which fishers comply with regulations, versus the confidence that fishers comply with regulations – the current wording for SIc doesn't grade the level compliance itself
- How the amount and/or type of information affects scoring – there is a risk that UoAs with some non-compliance, but high levels of information, may score worse than UoAs with more serious non-compliance, but limited data - so in effect non-compliance is hidden

1.2 Objectives

1. Establish under which PI 3.2.3 SIs compliance is assessed by assessment teams in practice
 - a. Develop a logic map that identifies how compliance should, *in the consultant's opinion*, be considered by assessment teams in PI 3.2.3 scoring issues (SIs). This should consider the nature and extent that SIs a and b should be conditional on the scores and rationales given for SIs c and d.
 - b. Establish how, *in practice*, assessment teams have followed the logic map developed under 1a. Conclude to what extent SIs a and b are conditionally linked to scores given to SIs c and d.
2. Establish to what extent assessment teams consider the *level* of compliance versus the *level of confidence* in compliance in scoring PI 3.2.3 SI c
 - a. Develop a simple typology of information used by teams to assess compliance under SIs c and d. This should consider the type, resolution and source of the information, and also how the information is obtained by the team (e.g. direct request to enforcement agencies, request via the fishery etc.).
 - b. Conclude to what extent teams focus on the level of compliance rather than the level of confidence in compliance. This should consider:
 - i. What level of confidence is generally considered to be ok to meet the different SGs (generally thought, some evidence, high degree of confidence)
 - ii. Whether teams (or enforcement officers) routinely dismiss 'inconsequential' infringements (e.g. administrative errors)
 - iii. What thresholds are generally considered to determine whether fishers 'comply', e.g. do teams conclude compliance when statistics show no issues of non-compliance in 70, 80 or 90% of inspections?

3. Establish to what extent the openness of the enforcement bodies in providing information affects scoring of PI 3.2.3 SIs c and d
 - a. Outline the approach generally taken by teams when there is limited information on compliance. This should consider the apparent weighting of qualitative information, from different sources, versus quantitative data in scoring.
 - b. Conclude to what extent the quantity and quality of *quantitative* information available to teams influences the scoring of SIs c and d

[SIId not discussed]

1.3 The data

The report is based on analysis of Public Certification Reports (PCRs) from 40 MSC certified fisheries, covering the following categories: demersal; small pelagics; tuna and large pelagics; bivalves, molluscs and crustaceans; and freshwater fisheries. Around 25 national enforcement systems are covered. The tuna fisheries are subject to international enforcement regimes – some of the other fisheries also have an international layer in their management system, but enforcement is mainly taken care of by national authorities (although in cooperation with each other).

The fisheries were randomly selected from the overview of assessed fisheries on MSC's 'Track a fishery' website, with the aim of covering all categories listed above, different geographical areas and management regimes, various conformity assessment bodies (CABs) and P3 assessors.

The assessments involved a total of 27 P3 assessors, of which no one is represented with more than three fisheries – this ensures a broad range of expert opinions on how the PIs under P3 should be scored.

The PCR from the latest assessment/re-assessment was used for each fishery. The peer review reports, stakeholder comments and MSC Technical Oversight comments were not investigated, but it should be noted that the scores investigated here are those arrived at *after* these rounds of quality checks of the assessments.

All fisheries in the sample scored in the range 80-100 on PI 3.2.3. 22 fisheries were scored at 80, 19 at 100.¹ Hence, there were no conditions in the selected fisheries related to enforcement, sanctions and compliance. Notably, there were a number of conditions on all other PIs under P3, except one (PI 3.1.3 on overall objectives). There was one condition on PI 3.1.1 (general management framework), four on PI 3.1.2 (consultation rights), three on PI 3.1.4 (incentives, only under v1.3), six on PI 3.2.1 (fishery-specific objectives), twelve on PI 3.2.2 (decision-making processes), five on PI 3.2.4 (research plan, only under v1.3) and four

¹ The total number exceeds 40 since one of the fisheries had multiple UoAs with different scores.

on PI 3.2.5 (reviews, PI 3.2.4 under v2.0) – 35 P3 conditions in total, i.e. nearly one per fishery on average.

2 Empirical overview

2.1 Under which SIs is compliance assessed?

The overarching logic of PI 3.2.3 is that the (comprehensiveness of the) enforcement system is assessed under SIa, the sanctioning mechanisms under SIb and compliance under SIc and SIId (the latter pertaining to *systematic* non-compliance – this will not be pursued in the following). However, by the wording of the guideposts the issue of compliance shall be assessed also under SIa and SIb. Under SIa, the comprehensiveness of the enforcement system shall be assessed as well as its *proven ability to enforce regulations* (i.e., to make fishers comply). Under SIb, the application of sanctions shall be assessed, but also their *proven ability to provide effective deterrence* (again: to make fishers comply). The question, then, is whether teams in effect assess compliance only under SIc and SIId, or also under SIa and SIb. To the extent that the latter is the case, the scoring of SIa and SIb will be a function of the SIc and SIId scores, i.e. the fishery cannot, logically, score better on SIa and SIb than allowed for by the SIc and SIId scores. If the former is the case and compliance is discussed under SIc and SIId only, a low score on these SIs will not prevent a higher score on SIa and SIb.

In the sample of 40 fisheries assessed for this report, only information on the enforcement system is presented under SIa in 16 assessments – hence, the scoring is based on the evaluation of the comprehensiveness of the system alone, not compliance. In 9 of the fisheries, on the other hand, the main presentation of compliance in the fishery comes under SIa (and not SIc, which most often just summarizes what has already been said about compliance under SIa) – here the scores of SIa is contingent upon the assessed compliance in the fishery and not only the comprehensiveness of the enforcement system. Finally, in 15 of the assessed fisheries a middle course is taken: only the enforcement system is presented and discussed (not compliance), but based on that the team explicitly concludes whether the system is able to enforce regulations or not (again: not taking into account actual compliance).

To sum up, in just above 20 % of the assessments, compliance is explicitly assessed under SIa. In 40 % of the assessments, compliance is not mentioned at all, while in nearly 40 % compliance is not documented or discussed, but nevertheless the teams conclude whether the enforcement system has displayed a proven ability to enforce regulations or not (implicitly making a judgement about compliance in the fishery).

The scoring of SIb follows approximately the same pattern as that of SIa, but with some variation. In 11 of the assessed fisheries, only information on the sanctioning mechanisms is presented and the scoring of the system's ability to provide effective deterrence is based on that alone. In 13 of the fisheries, compliance data are presented under SIb (rather than under SIc), and the score of this SI is contingent on proven compliance and not only the existence and application of sanctions. Then finally in 16 fisheries compliance data are not presented, but the team nevertheless concludes whether sanctions provide effective deterrence or not (again: implicitly making a judgement about compliance in the fishery).

Hence, compliance is explicitly assessed somewhat more frequently under SIb than under SIa, in just above 30 % of the fisheries. In just below 30 %, no mention is made of compliance

at all, while in the remaining 40 % compliance is not documented or discussed, but teams still conclude whether sanctions demonstrably provide effective deterrence or not.

Notably, the fisheries where compliance data are presented under SIa and SIb do not overlap to any significant extent. So roughly speaking, compliance is addressed under SIa and SIb in around half of the fisheries (i.e. in relation to the enforcement system's ability to enforce regulations and the sanctions' ability to provide effective deterrence), instead of under SIc. In only around half of the fisheries, then, is compliance primarily discussed under SIc. In turn, data on the comprehensiveness of the enforcement system is used in some fisheries to assess SIc, instead of SIa. The main picture is that *compliance and enforcement are scored haphazardly across SIs a-c*. This is all the more striking since the analysis is based on assessments that have been subjected to scrutiny by peer reviewers, MSC experts and in some fisheries presumably also Accreditation Service International (ASI) assessors.

It should also be mentioned that substantive data on compliance are absent in quite a few fisheries; this is further discussed below. In some assessments, the level of compliance is discussed under SIa while the reliability of the compliance data ('certainty that fishers comply') is the topic of SIc. This is also discussed below.

Finally, is there any support in the material to the hypothesis that the SIc score becomes contingent on the SIa and SIb scores when compliance is assessed under the two latter? The sample is limited, and it isn't always obvious whether compliance is discussed or just concluded on under SIa and SIb. With these reservations: the SIc score is the same as the SIa or SIb score (depending on where compliance is assessed) in 85 % of the assessments where compliance is substantially discussed under SIa or SIb. The score is the same in 60 % of the assessments where there is no mention of compliance under SIa or SIb, and 50 % of the assessments where compliance is concluded on, but not substantially discussed under SIa or SIb. So there is a certain support to the hypothesis that the SIc score follows the SIa and/or SIb score when compliance data are presented under the latter SIs.

2.2 How is the SIc score substantiated?

By the wording of the SIc guidepost, the difference between a 60, 80 and 100 score relates to the team's *confidence* that fishers comply (whether fishers are *generally thought* to comply, whether *some evidence exists* or whether there is a *high degree of confidence* that they do so).² How explicit are assessors that it is the level of *confidence* (that fishers comply) that is at issue here? To what extent do they instead focus on the level of *compliance*, i.e. to what *extent* fishers comply? And in the latter case, what do they accept as good enough for the different scores?

² In addition, there is a requirement that fishers provide information to authorities. This is not pursued any further in this report since there were no issues with this in any of the investigated fisheries (and presumably seldom are).

In 18 of the 40 fisheries investigated, the topic of discussion is clearly the level of compliance in the fishery, not the level of certainty (that fishers comply). Assessors take the information available (see section 2.3 below) as a starting point and discuss whether the reported level of compliance is sufficiently good for a 60, 80 or 100 score. Very seldom is quantitative data presented (again, see below), and in the few cases where that happens, infringements in less than 10 % of inspection is accepted by the teams as good enough for a 100 score. In one of the instances the number of infringements is higher, at 10-15 %, which is also accepted as good enough for a 100 score. One case where the level of infringements is lower, at 5 %, the score is set at 80. (Notably, in most of the assessments where quantitative data is presented – see section 2.3 below – that is done under SIa or SIb instead of under SIc).

In 9 of the fisheries, it appears that the level of certainty is in focus rather than the level of compliance, while in the remaining 13 fisheries there is either a mix of the two or, more common, the wording of the rationale isn't sufficiently clear to say. In one fishery, e.g., the SIc score is substantiated as follows: 'The low incidence of violations and only two minor incidences in the last three years ([...], pers. comm.) indicate a high degree of confidence that fishers comply with the management system.' Here the first part of the sentence addresses the level of compliance (how many infringements have been revealed) while the second part concludes regarding the level of confidence that fishers comply.

In 12-15 cases, there is no information, not even qualitative, on compliance as such in the SIc rationale, only information on the enforcement system. Here the SIc score is made contingent on the SIa score. A typical example is the following: [The enforcement system is comprehensive.] Therefore, there is a high degree of confidence that fishers comply with the management system (100 score).³ In at least three cases, there is no mention of compliance at all throughout SIs a-c.

2.3 What sources are used for scoring compliance?

In the last section, we saw what kind of arguments teams use to substantiate their scores on the compliance indicator. But which sources are used to support the arguments?

Most conspicuous is the lack of quantitative data in the material: only 8 of the 40 assessments have any reference at all to numbers in their rationale for scoring compliance (whether this is done under SIa, SIb or SIc). And nowhere is there any comprehensive quantitative information, e.g. in the form of statistics that break down inspections or infringements on indicators such as year, violation types, vessel groups or ocean areas.³ In some cases, the empirical base of the information is unclear. In other cases, where the source is clear, the number of inspections or infringements is taken from a very limited empirical base – in one fishery, e.g., the compliance rate is reported to be 3.5 %, based on 117 inspections performed during a 10-day *annual* enforcement operation in the fishery.

³ With one possible exception, where infringement numbers are split into recreational and commercial fisheries and two types of infringements are specified.

Reports from Compliance Committees are the main source of information in around 10 of the fisheries, notably the big tuna fisheries where such committees have been set up under international agreement.

In around half of the assessments, personal communication is the only source for scoring compliance. Most such statements are from enforcement authorities, but some are from other regulatory bodies, stakeholders or captains. Some are not specified beyond 'site visit meetings' and 'information collected during the site visit'.

Annual reports from enforcement bodies are used in two fisheries, while the website of the enforcement body is listed as a source in one of the assessments. In one assessment report, a 'list of violations' (of fishing regulations) is mentioned in the reference list but there is no mention of it (or of inspection or infringement incidences) in the rationale. In another fishery, it is stated in the rationale that the client has produced inspection sheets that confirm the enforcement authorities' statement regarding compliance (but the inspections sheets are not in the reference list or specified in any other way). Notably, the only reference to peer reviewed literature about compliance in fisheries is to an 18 years old journal article (i.e. 18 years old at the time of the assessment). In four assessments, no source whatsoever is given to document compliance.

3 Discussion and conclusions

3.1 Summary

1. In the sample of 40 fisheries, there are no conditions on PI 3.2.3 (enforcement, sanctions and compliance), while there are conditions on all other PIs except PI 3.1.3 (objectives). Most conditions (35 %)⁴ are on PI 3.2.2 (decision-making procedures).
2. In 20 % of the assessments, compliance is explicitly addressed and substantially discussed under SIa (enforcement system). In 40 % of the assessments, compliance is not mentioned under this SI, while in 40 % compliance is not documented or discussed, but it is nevertheless concluded on the system's proven ability to enforce regulations, and hence implicitly about compliance as well. The scoring of SIb (sanctions) follows the same pattern, with only little variation: in 30 % of the assessments, the compliance discussion is taken under SIb, while in 30 % no mention is made of compliance here. In the remaining 40 %, compliance is not discussed or documented, but teams conclude on the proven deterrent effect of sanctions and hence implicitly on compliance. In 35 %, there is no information on compliance as such under SIc, only information on the enforcement system. In at least three cases, there is no mention of compliance at all throughout SIs a-c.

To conclude, compliance is scored haphazardly across SIs a-c. More often than not, compliance is discussed and documented elsewhere than under SIc, which is supposed to be the primary place to address compliance. In a few cases, there is no mention of compliance at all throughout SIs a-c.

3. In the scoring of SIc (compliance), some teams score the level of certainty (that fishers comply) and some the level of compliance as such. In 45 % of the fisheries, the topic of discussion is the level of compliance in the fishery, not the level of certainty. Assessors review the information available and discuss whether the reported level of compliance is sufficiently good for a 60, 80 or 100 score. In 25 % of the fisheries, the level of certainty is in focus rather than the level of compliance, while in the remaining 30 % there is either a mix of the two or, more common, the wording of the rationale isn't sufficiently clear to decide.

To conclude, almost half of the teams score the level of compliance rather than the level of certainty under SIc. Some teams score the level of certainty, or a mix, and in quite a few assessments the wording of the rationale is too unclear to conclude which of the two is actually scored.

4. Whether compliance is scored under SIa, SIb or SIc, there is very limited quantitative underpinning of the score. Only 20 % of the assessments have any reference to numbers at all, and (with one exception) no statistics is provided beyond simply stating the percentage of non-compliance throughout a specific period. This period is sometimes not specified, and in some instances it is as short as an annual ten-day

⁴ Approximate percentages are provided in this section, for pedagogical reasons – the sample is too small for exact percentages to make any sense.

enforcement operation. In around half of the assessments, personal communication from enforcement authorities, stakeholders or clients is the only source for scoring compliance. In just a couple of fisheries are annual reports from enforcement authorities or lists of inspections provided. The tuna big tuna fisheries are an exception to this overall picture, where Compliance Committees have been set up under international agreement and information from these reports is used to score fisheries. Notably, in only one assessment is there any reference to peer reviewed literature about compliance, and in that instance it is an 18 year old article that is referenced.

To conclude, personal communication is the primary source used to score compliance. With the exception of the tuna fisheries, where Compliance Reports exist, there is hardly any reference to written material from enforcement authorities. Quantitative data are virtually absent, and there is no reference to relevant peer reviewed literature. Hence, the data are unquantified, unwritten and non-peer reviewed.

3.2 Assessment

1. It is conspicuous that PI 3.2.3 is one of only two PIs under P3 where there are no conditions in the sample of 40 fisheries. Of all the PIs under P3, PI 3.2.3 the most 'outcome-leaning'. While all other PIs are largely procedure and information oriented – focusing on issues such as decision making, dispute resolution, objectives, consultation rights, access to information and review mechanisms – PI 3.2.3 deals with the ability of the management system to actually influence target behaviour.

Is the PI so difficult to measure that teams simply choose the easy way out and not delve into a substantial assessment of fisher compliance? Is it possible to structure the PI, its SIs and guideposts differently in order to find a more feasible way to differentiate between fisheries with good and not so good compliance? In that case, conditions can be introduced for the latter group, which will take the fisheries in a more sustainable direction over the certification period. Today this is a missed opportunity – whether the guideposts are unsuitable or teams unwilling or incompetent: as long as no fisheries score below 80, there is no pull in a more sustainable direction in terms of improved enforcement and increased compliance.⁵

Furthermore, compliance is arguably the issue under P3 (and perhaps across all three Principles) that generates most public attention in matters related to fisheries management. Unsustainable quota levels, inadequate harvest control rules, lacking ETP policies or dispute-resolution mechanisms – nothing can compare with IUU fishing when it comes to media attention. When MSC is not able to address this issue in a way that leads to conditions and hence improved practices, there should be room for improvement.

2. Is it problematic that compliance is scored haphazardly throughout SIs a-c? Does it matter which SI it is scored under as long as it happens under the same PI? Perhaps this is not a big problem, but a more consistent approach would make it easier to harmonize among fisheries, which is not least essential when a fishery receives a condition and there are other fisheries that must be harmonized with. More important substantially is that the order in which compliance is discussed might affect scoring of the different SIs, and in turn of the PI (and Principle) as such. If compliance is assessed under SIa or SIb, the SIc score is more likely to become contingent on the score of the preceding SI(s). When teams score compliance under one of the two first SIs, they tend to simply repeat that information or briefly refer to the preceding discussion. It then becomes more difficult to assign a different score on SIc.

If, on the other hand, compliance is scored under SIc only, there is more room for a differentiated score among the three SIs. As an example, it is perfectly possible for a

⁵ The reason that no fisheries score below 80 since compliance might of course in principle reflect the fact that compliance *is* generally good. The discussion above is based on the assumption that there is every reason to assume that there is as much room for improvement on this PI as on the other PIs under P3.

fishery to achieve a 100 score on SIa if the enforcement system is comprehensive and generally fit to enforce regulations, but only 80 or 60 (or less) on SIc if actual compliance is nevertheless inadequate.

There is reason to believe that the intention behind the current guideposts is such differentiation of scores across the SIs. So why do teams nevertheless score compliance just as often under the two first SIs as under the 'compliance indicator' itself? The most likely reason is arguably that the SIa guidepost speaks not only about the comprehensiveness of the enforcement system, but also about its ability to enforce regulations – for a 100 score, the system must have *demonstrated a consistent ability* to do that. Similarly, SIb by its wording does not only deal with the existence and application of sanctions, but also with the influence of sanctions on fisher behaviour – for a 100 score, sanctions must *demonstrably provide effective deterrence*.

The reason why teams choose to score compliance in different places might be that they interpret the second clauses in the SIa and SIb requirements differently. The requirement under SIa for the enforcement system to have demonstrated an ability to enforce regulations can be interpreted in two ways, either as a requirement i) to actually have induced compliance among fishers (which can only be documented through data about compliance levels), or ii) to have carried out enforcement activities in an appropriate way – i.e. in such a way that, other factors aside, fishers would comply to the extent they behave rationally from an economic point of view (i.e. the risk of being detected in unlawful behaviour is sufficiently high for them to choose a lawful behaviour instead). The same logic applies to SIb, where demonstrable effective deterrence can be interpreted to mean either that fishers have demonstrated a lawful behaviour, or that sanctions are demonstrably severe enough – the sanctioning regime sufficiently robust – to make economically rational fishers choose compliance over non-compliance. The challenge here is that compliance can emanate from other sources than the threat or use of sanctions, which the guidepost does not take into account. It is widely accepted in the literature on compliance that issues such as individual moral, societal norms, communicative efforts by enforcement authorities and the legitimacy of the management system or regulations, can also contribute to compliance. And compliance with regulations might be the natural behaviour of fishers notwithstanding any management efforts, be that deterrence, aims to produce legitimate legislations or communication with the fishing fleet (so-called first-order compliance).

Although the wording of the guideposts arguably makes it more natural to choose the former interpretation – i.e. that actual compliance must be demonstrated under SIs and/or SIb – it is also possible to arrive at the opposite conclusion, i.e. that SIa and SIb should focus on the ability of the management system to enforce regulations in the absence of other sources of compliance (such as norms, legitimacy and first-order compliance). Not least is there reason to view the second interpretation as the most commonsensical as it allows for a differentiated score across SIs a-c. And it takes care of the logical fallacy that deterrence can be proved through documentation of

compliance – deterrence is just one element in the complex fabric of compliance in fisheries.

3. In scoring SIc, nearly half of the assessments in the sample score the level of compliance instead of the level of certainty that fishers comply, which the guidepost asks for. Only one fourth score the level of certainty, while in one in three assessments the wording is so unclear that it is impossible to interpret what the team really scores. The big problem with the guidepost of this SI, however, is that the concept ‘fishers comply’ is not differentiated, i.e. it seems to be a yes/no question whether there is compliance in the fishery (which in reality it’s obviously not – there are degrees of compliance). This is probably the reason why so many teams score the level of compliance instead of the level of certainty – it is simply the natural thing to do in assessing compliance: *to what extent* do fishers comply (a question the guidepost actually doesn’t allow for)? An assessment primarily focused on the level of certainty would have to evaluate the reliability of the data provided to document whatever compliance there is in the fishery. For instance, elaborate statistics on compliance provided by an enforcement body in a well-developed management system would probably be seen as more reliable than a brief, unsubstantiated statement from enforcement authorities situated within a weak management structure.

Finally, it should be noted that the requirement for an 80 score on SIc isn’t particularly strict: for a fishery to pass this SI without condition, the requirement is that ‘*some evidence exists that fishers comply*’. It is not difficult for teams to argue that this is met – if there is little quantitative data on compliance, it can still be argued that ‘some evidence exists’ if only an enforcement officer confirms that this is the case.

4. It is easy to dismiss the empirical base for scoring compliance as it is largely unquantified, unwritten and non-peer reviewed. This is not necessarily problematic, for several reasons: First, hard data do not have to paint a more correct picture of the state of affairs in a fishery than an informed oral statement from the enforcement authority. No fishery is 100 % clean, and there is an inherent challenge in measuring compliance. There is also the risk that the more comprehensive the enforcement system and the more detailed the statistics, the better will non-compliance be reflected in the documentation. So paradoxically, a relatively clean fishery with a sophisticated enforcement and reporting system might come across as having more non-compliance than a fishery with weaker enforcement mechanisms and only rudimentary information on compliance performance. Second, it might be unrealistic to expect much more quantitative information from enforcement authorities than is already provided today. Enforcement bodies in many countries do not share such data for strategic reasons and, increasingly, with reference to personal data protection.⁶

⁶ A bold policy for MSC might, however, be to raise the bar for information on compliance in fisheries and hence aim to improve the situation by laying the ground for conditions that will imply pressure from the fisheries on enforcement authorities to publish more information.

However, it is clearly problematic that the rationales for scoring compliance are in many instances more than inadequate. In some instances, there is no substantiation at all, but simply a repetition of the guidepost. In many instances, only information on the enforcement system is referenced (e.g. legislation), even though the team concludes about compliance as well. (There is hardly any information about actual compliance in legislation.) And in the many references to personal communication, the reference style is often sloppy – only rarely is the source of the information identified or the time and place of the interview specified. ‘Information gathered during the site visit’ is an example from a reference list where much is left to be desired.

More problematic is the absence of reference to peer reviewed literature about compliance. This is striking indeed since there has been a burgeoning social science literature on compliance in fisheries over the last few decades, covering detailed empirical studies of compliance in fisheries across the world, with ever more sophisticated theoretical and methodological approaches and increasing empirical coverage. This is all the more conspicuous since peer reviewed literature is very much present in the rationales for scoring P1 and P2 in MSC assessments. It is the modus operandi of P1 and P2 assessors to search for the most updated peer reviewed literature when they embark on an assessment. That does not seem to be the case for P3 assessors, to put it mildly.⁷

3.3 Recommendations

In general, there is a vagueness about PI 3.2.3 rationales that is arguably not tolerated elsewhere in MSC assessment reports. Guideposts are ambiguous and clearly interpreted in widely different ways by assessment teams. Rationales are inadequate and documentation weak: unquantified, unwritten and non-peer reviewed. This is all the more disturbing since 3.2.3 is the outcome PI per se under P3: it assesses the extent to which management measures – emanating from the entire spectre across P1, P2 and P3 – actually influence the way in which fisheries are conducted. And the vagueness and inconsistencies are striking since the PCRs are the teams’ outputs *after* thorough quality control by peer reviewers, stakeholders, MSC Technical Oversight officers and in some cases also ASI assessors.⁸

Below are two suggestions for improvement:

1. Enforcement, sanctions and compliance should be treated separately under SIs a-c, respectively, instead of randomly across the three SIs, like today. That would

⁷ A likely explanation is that the vast majority of P3 experts are not social scientists (such as political scientists, sociologists and anthropologists, which all have a long tradition for studying compliance in fisheries), but natural scientists and/or practitioners. Many are marine biologists with experience from management and hence considered suitable to assess P3.

⁸ The typical peer reviewer comment to PI 3.2.3 is not related to inconsistencies in scoring (which is often the case in the other PIs under P3), but instead a demand for more quantitative data (which obviously are not available in most fisheries).

improve the possibility for harmonization among fisheries and, importantly, remove the risk that the scoring of SIc becomes contingent on the SIa and SIb scores. There are two possible solutions: i) remove the second clauses of SIa and SIb altogether and focus on the comprehensiveness of the enforcement system under SIa and the robustness of sanctioning mechanisms under SIb; or ii) keep the guideposts as they stand, but specify in the Standard/Guidance that it is the expected ability to enforce/deter that shall be scored, not actual fisher compliance (which shall be assessed under SIc).

2. The framing and wording of SIc is challenging for assessment teams, peer reviewers and stakeholders. Most importantly, the question whether fishers comply is implicitly presented as a yes/no question rather than as a question about degrees of compliance. In fact, the level of compliance isn't actually asked for at all, only the level of certainty *that* fishers comply – are they thought to comply (SG60), does some evidence exist (SG80) or is there a high degree of certainty (SG100) that they comply? In other words, if teams conclude that fishers do *not* comply, the fishery will not only fail to reach a 100 or an 80 score; it will not pass at all. Obviously, this is not how teams have interpreted the guidepost – rather, at least half of the teams have instead scored the level of compliance: is compliance high or low or somewhere in between?

But in an assessment of the level of compliance there is need for a tool to differentiate between good and not so good compliance. Let's take the ongoing debate about how to score PI 3.2.3 in EU fisheries after the introduction of the landing obligation (LO) as an example. Some teams have proposed introducing a condition on PI 3.2.3 in these fisheries (although there is not yet a harmonized decision). Some have proposed a condition on SIa, some on SIc and some on both.⁹ The discussion related to SIa is relatively straightforward since it implies an assessment of whether the enforcement system is sufficiently comprehensive to meet the challenge of a new management measure. It is SIc that is most interesting from a principal point of departure. First, there is the question about information: what kind of documentation exists that fishers do or do not comply with the LO? Some seem to presuppose that fishers will not comply with the LO and demand evidence of compliance in order to score above 60 (i.e. not evidence of non-compliance to score below 80). Others rely on anecdotal evidence, while yet others expect more quantitative documentation. Second, there is the question about the level – or extent – of compliance. Is low compliance with the LO enough to 'score down' the fishery? Or should the LO be considered as just one among many regulations with which compliance must be measured? One might well imagine that a fishery can display perfect compliance with most important regulations, such as quota uptake, bycatch, ETP, gear and area regulations, but not with the LO. Do the fishers 'comply' then according to the guidepost? Is the SG80 requirement that 'some evidence exists that fishers comply' met? A pragmatic take on the scoring of SIc might be to introduce a condition because there is a major new regulation that there is reason to believe is not satisfactorily complied with. Yet it

⁹ This is based on the author's participation in different teams assessing the EU fleet and the introduction of the LO.

might be more correct to argue that some evidence exists that fishers comply with nearly all important regulations (with one exception: the LO) and hence score at 80 or 100.

Based on this, one might consider splitting up the existing SIc into one information and one outcome SI. The information SI could, e.g., ask whether qualitative information exists about compliance in the fishery (SG60), whether there is quantitative information (SG80) or peer reviewed information (SG100), alternatively for SG80 and SG100: aggregate quantitative data (such as total number of inspections and infringements) and comprehensive quantitative data (such as statistics broken down on relevant indicators for the specific fishery, e.g. vessel group, nation, area or gear). The outcome SI, in turn, could require – following the logic of other PIs under P3 – that most important regulations are complied with (SG60), that all important regulations are complied with (SG80) and that all important and other regulations are complied with (SG100). With these SIs and guideposts, the EU fleet might reasonably reach a 60 score on the outcome indicator since some important regulations (such as quota, gear and are regulations) are complied with, but not an 80 score since *all* important regulations (including the LO) are not complied with.

Suggested changes to the standard

- SIa: keep the present guideposts but specify in the Standard/Guidance that it is the expected ability to enforce regulations (based on the robustness of the enforcement system) that shall be scored under this SI, not actual fisher compliance.
- SIb: keep the present guideposts but specify in the Standard/Guidance that it is the expected ability to deter fishers (based on the robustness of the sanctioning regime) that shall be scored under this SI, not actual fisher compliance.
- SIc: new ‘information indicator’ focusing on the existence of knowledge about compliance in the fishery, e.g.: SG60: qualitative information exists; SG80: quantitative information exists; SG100: peer reviewed information exists. Alternatively: SG60: qualitative information exists; SG80: aggregated quantitative information exists; SG100: comprehensive quantitative data exist.
- SI d: new ‘outcome indicator’ focusing on the level of compliance in the fishery, e.g.: SG60: most important regulations are complied with; SG80: all important regulations are complied with; SG100: all important and other regulations are complied with (SG100).
- SIe: same as previous SI d (systematic non-compliance)