D R A F T

Evaluation of Proposed Methodology for Measuring Marsh Compliance with the Everglades Phosphorus Criterion

prepared for

U.S. Department of the Interior

by

William W. Walker, Jr., Ph.D., Environmental Engineer

August 26, 2002

FDEP's proposed methodology[1] for tracking marsh compliance with the phosphorus criterion at a given marsh station is stated as follows:

> *"Therefore, based on the compliance measurement methodology developed in Section 2 above, waters represented by a sampling station shall be deemed in compliance with the phosphorus criterion for a given year if* ***either*** *of the following conditions are satisfied:*

1) *The annual geometric mean of measured phosphorus concentrations for that station during that year does not exceed the 10 µg/L criterion demonstrated to be protective of the natural flora and fauna within the EPA, or*

2) *The annual geometric mean of measured phosphorus concentrations for that station during that year does not exceed the natural spatial and temporal variation captured by the 15 µg/L upper compliance limit* ***and*** *the arithmetic average of the annual geometric mean TP concentrations measured at that station during the five-year period encompassing that year and the preceding four years is maintained at or below (i.e., does not exceed) the 10 µg/L protective criterion."*

This algorithm would be applied annually to data from a marsh monitoring network to develop a "report card" that reflects the status of the system relative to the 10-ppb criterion.

This document evaluates the power of the test for identifying sites that are impacted to various degrees by elevated P concentrations. "Power" is defined as the probability of failing the test at an impacted site. Power depends upon (1) the parameters of the test (as defined above); (2) random variability (as reflected in year-to-year variance in the geometric mean) and (3) the extent of impact; i.e., the higher the average P concentration, the higher the probability of biological impact ("imbalance"), and the higher the probability of failing the test.

---

[1] FDEP, "Development of Proposed Phosphorus Criterion Compliance Measurement Methodology for the Everglades Protection Area", April 2002.

Power is evaluated by applying the test to a simulated 1000-year time series of yearly geometric means. The time series is specified by a hypothetical long-term arithmetic mean (LTM) and standard deviation. The LTM reflects the underlying statistical distribution of yearly geometric means. The 1-year and 5-year means fluctuate about the LTM. The yearly values are modeled by a lognormal distribution with a standard deviation of 0.24, as estimated by FDEP based upon monitoring data from reference sites in the Water Conservation Areas.

An example of a simulated time series for a hypothetical site with a long-term mean of 11 ppb is shown in Figure 1:

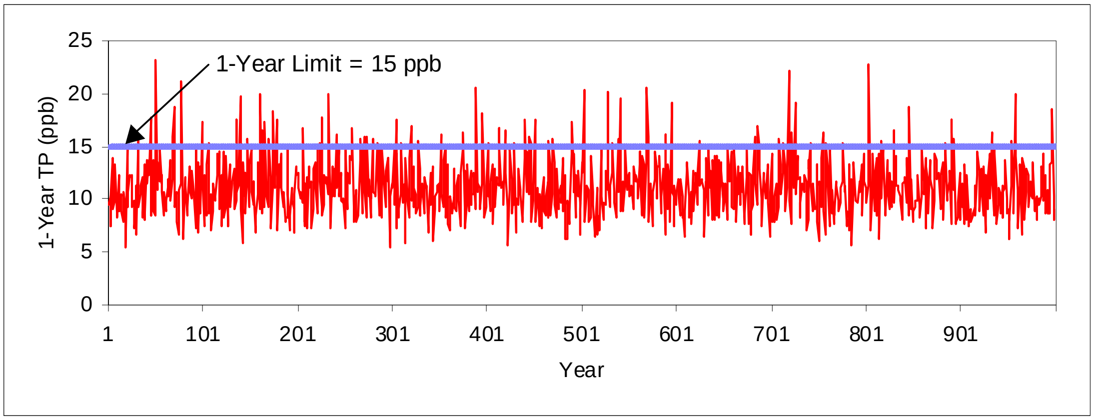


Figure 1 – Simulated 1-Year Geometric Means at a Site with a Long-Term Mean of 11 ppb

The dotted line shows the yearly limit proposed by FDEP (15 ppb). The hypothetical site fails the test when the yearly value is above the limit. This occurs in 9 % of the years. Since the LTM of this time series is above the 10-ppb threshold, we would expect that it would fail the compliance test at some frequency.

The test also requires comparison of the 5-year rolling mean of the yearly geometric means with a 10 ppb limit. By averaging the simulated yearly geometric means over successive 5-year intervals, a time series of 5-year means can be generated:

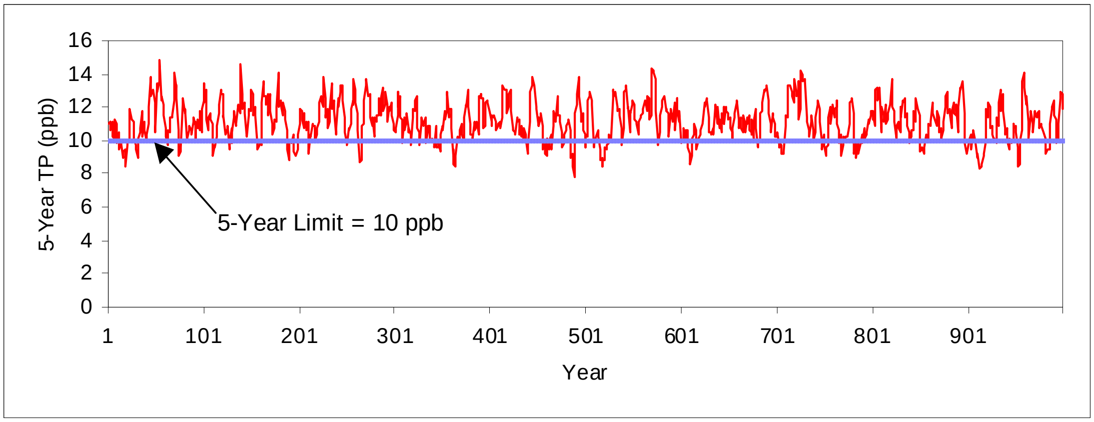


Figure 2 – Simulated 5-year Rolling Means at a Site with a Long-Term Mean of 11 ppb

The 5-year mean exceeds 10 ppb in 82% of the years. The site fails the test when 5-year mean exceeds 10 ppb, provided that the1-year value also exceeds 10 ppb. This occurs in 57% of the years. In 25% of the years, the 5-year mean exceeds 10 ppb, but the 1-year mean is less than 10 ppb, and the site passes the test.

When the 1-year and 5-year tests are applied simultaneously, the net failure rate is 57%. This indicates that there would be a 57% probability of failing the test in any year at a site with a long-term mean of 11 ppb. The net failure rate is controlled primarily by the 5-year test (vs. the 1-year test). The low power of the 1-year test partially reflects the fact that it was derived as the upper 95th percentile of the distribution of yearly geometric means at a site with a long-term geometric mean of 10 ppb. All other compliance tests used in tracking phosphorus levels in the Everglades (ENP inflow limits, Refuge marsh levels, EAA Regulatory Rule, C139 Regulatory Rule) use the upper 90th percentile. If the 90th percentile were used in this case, the 1-year limit would be 13.6 ppb instead of 15 ppb. The would increase the failure rate for the 1-year test from 9% to 17% at a site with a long-term mean of 11 ppb and increase the sensitivity to short-term spikes, but have little effect on the net average failure rate (57% to 58%). The latter would still be controlled by the 5-year test.

By repeating the above exercise for long-term means ranging from 5 to 25 ppb, we can derive the following relationship between the long-term mean and the expected failure rate at a given site:

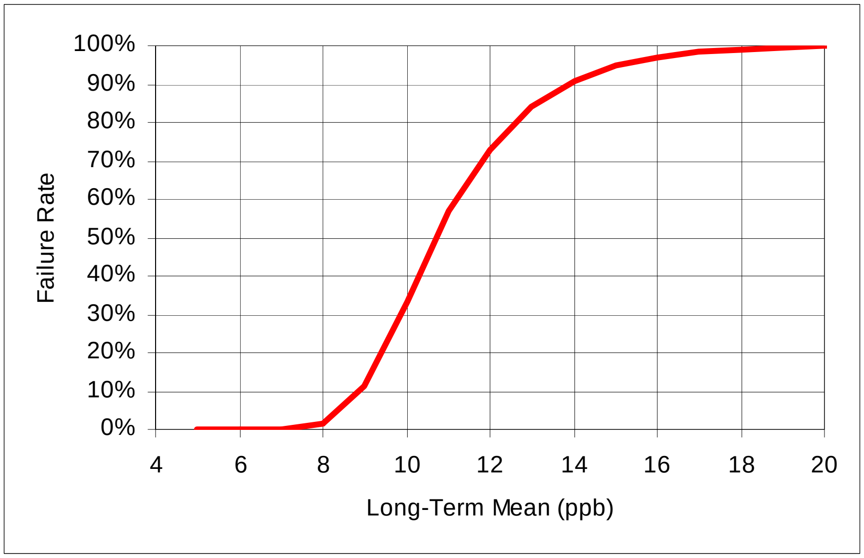


Figure 3 – Net Failure Rate vs. Long-Term Mean

The probability of failing the test increases with the long-term mean. At sites with long-term means of 8, 10, 12, and 14 ppb, predicted failure rates are 2%, 33%, 73%, and 91%, respectively.

The performance of the test can be further evaluated from a biological perspective by considering the fact that the distinction between “unimpacted” and “impacted” is not a sharp line. Transect monitoring data and supporting experimental research have demonstrated that the transition from reference to impacted sites occurs over a relatively narrow range of ~9 to ~14 ppb, depending upon biological indicator. The extent of impact can be expressed as the percent of the biological indices that show a significant departure from reference sites, as derived from FDEP’s change-point analyses.

Spatial gradients in mean phosphorus concentration & percent of indices impacted in WCA-2A are shown in Figure 4:

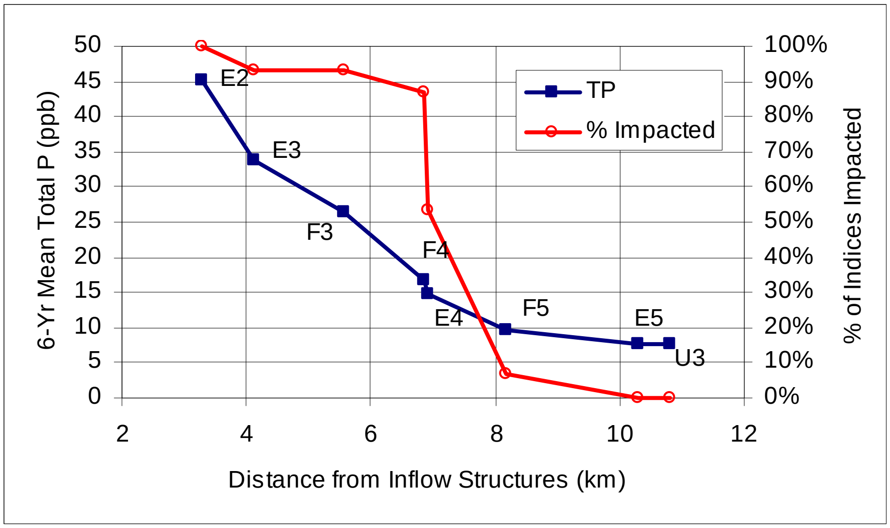


Figure 4 – Mean Total P & Percent of Biological Indices Impacted vs. Distance from Inflow Structures in WCA-2A

As provided by FDEP, the mean Total P values are the arithmetic means of the yearly geometric means for 1994-1999 at each of the labeled monitoring sites. As phosphorus concentrations increase from < 10 ppb to ~15 ppb between the E5/F5 sites and the F4/E4 sites, there is a sharp increase in the percent of indices impacted. This response is the primary basis for the 10 ppb criterion.

The two curves in Figure 4 can be combined to develop a relationship between mean concentration and percent of the indices that are impacted:

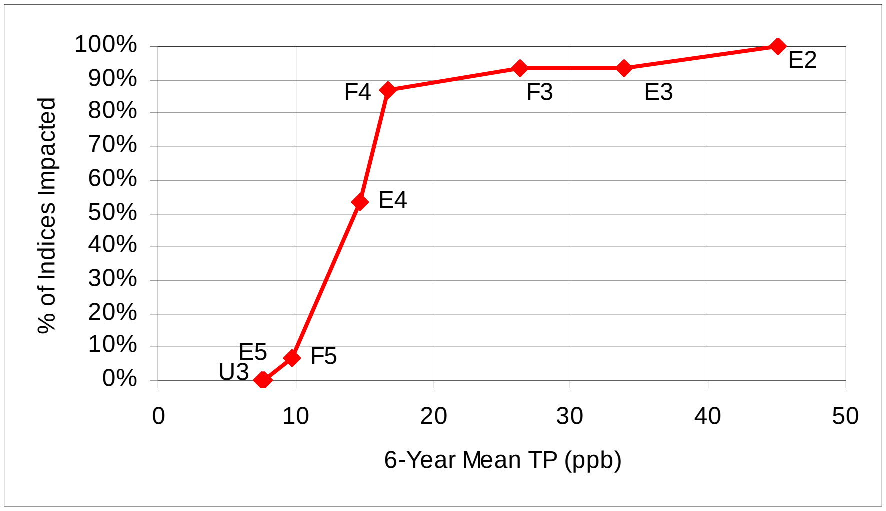


Figure 5 – Percent of Biological Indices Impacted vs. Mean Total P Concentration

Figure 5 does not place any particular weights in the various indices. As phosphorus concentrations increase above ~9 ppb, the more sensitive indices (i.e., minimum daily dissolved oxygen, calcareous periphyton mats, and macroinvertebrates) would be impacted first, followed by the less sensitive indices (mean daily dissolved oxygen, periphyton & macrophyte species, loss of open-water habitat).

In turn, Figures 2 & 5 can be combined to compute the percent of indices impacted during each 5-year interval in the simulated 1000-year time series, as illustrated in Figure 6 for a site with a long-term mean of 11 ppb:

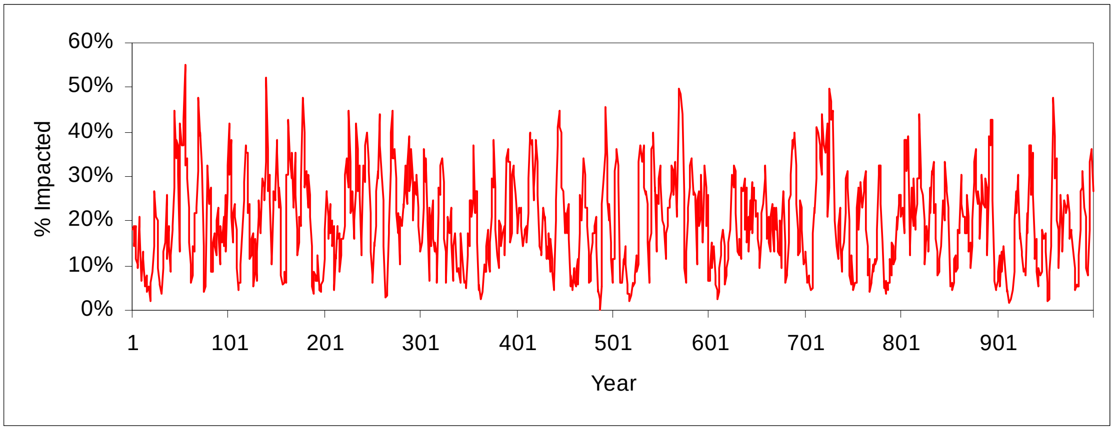


Figure 6 – Percent of Biological Indices Impacted for a Site with a Long-Term Mean Total P Concentration of 11 ppb

On the average, 20% of the indices would be impacted at this site. These would tend to include the more sensitive indices (i.e., minimum daily dissolved oxygen, calcareous mats, and macroinvertebrates). Percentages range from ~5% to ~50% in different 5-year intervals. The computation assumes that the indices respond to the 5-year-average Total P concentration. For simplicity, the 5-year and 6-year rolling-means are assumed to be equal.

It is likely that response time varies with biological index. For example, periphyton would be expected to respond more rapidly than macrophytes. The 5-year interval is consistent with the period of intensive research along the nutrient gradient in WCA-2A when paired water quality and biological measurements were made. Sufficient data are not available for modeling variations in response time as a function of biological index. If the framework were restructured assuming a 1-year response time, the maximum 1 year limit (15 ppb) would have to be reduced significantly to avoid significant impacts on sensitive indicators.

By mapping the biological response curve (Figure 5) onto the power curve (Figure 3), we can derive a relationship between the long-term-average percent of indices impacted at a given site and the probability of failing the compliance test in any year:

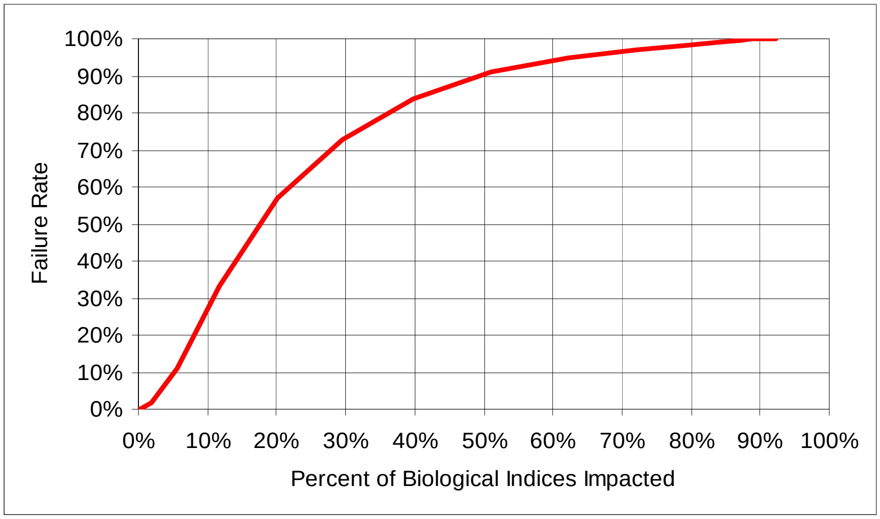


Figure 7 – Net Failure Rate vs. Long-Term-Average Percent of Indices Impacted

This curve reflects the combined effects of (a) the test design, (b) the expected temporal variability in marsh concentrations, and (c) the potential range of biological responses over a range of P concentrations. At sites with 0%, 10%, 30%, and 50% of the indices impacted, the expected failure rates would be 0%, 28%, 73%, and 91%, respectively.

The test specifies an automatic pass in years when the geometric mean is less than 10 ppb, regardless of the 5-year mean. Impacted sites with 5-year or long-term means above 10 ppb would be expected to have geometric means below 10 ppb in some years. As demonstrated above, the latter situation would occur at a frequency of 25% at a site with a long-term mean of 11 ppb. Passing the test in such years would allow impacted sites to go undetected, assuming that biological responses are driven by concentrations averaged over more than one year. The effect of the 1-year pass feature is demonstrated in Figure 8:

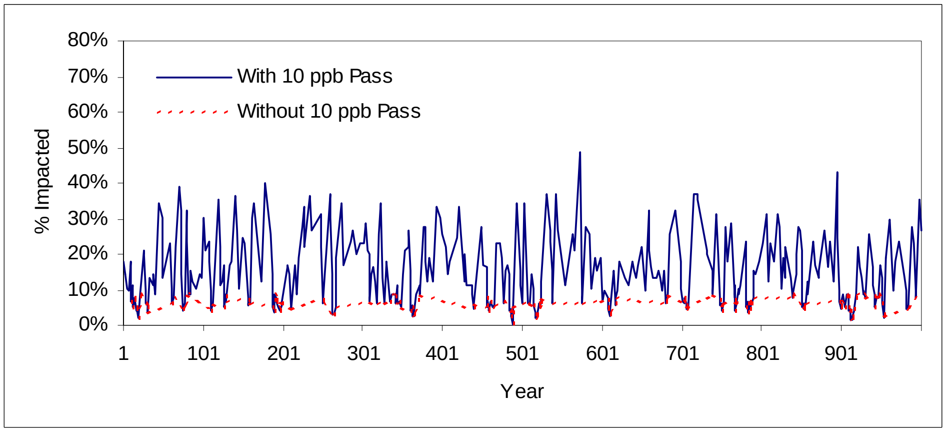
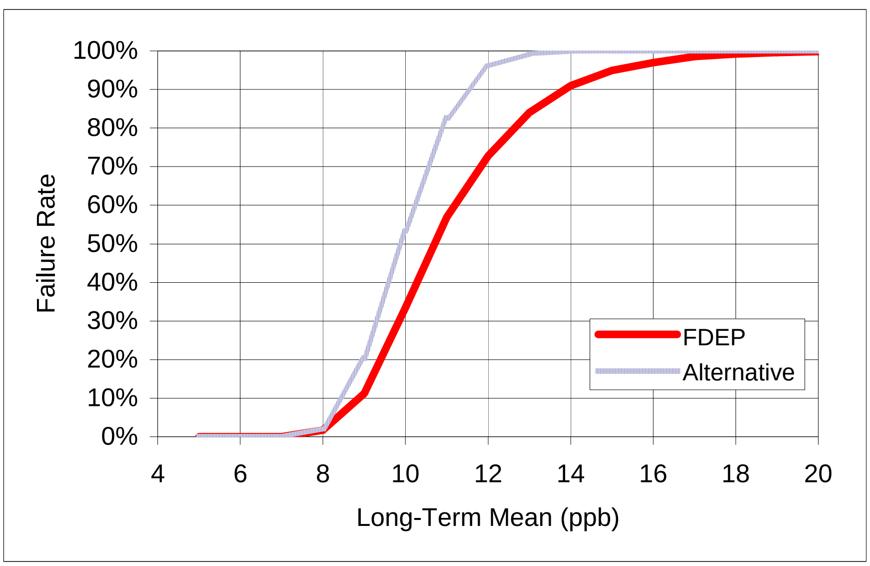


Figure 8 – Percent of Indices Impacted in 5-Year Intervals Passing the Test With & Without the Automatic 10 ppb Pass Feature for a Hypothetical Site with a Long-Term Mean of 11 ppb

With the proposed test design, ~5% to ~40% of the indices would be impacted in in years when the site passes the test. Eliminating the provision that the site passes the test in any year with the geometric mean is less than 10 ppb, the maximum impact would be limited to <10 %.

Power curves with and without the 10 ppb pass feature are shown in Figure 9:



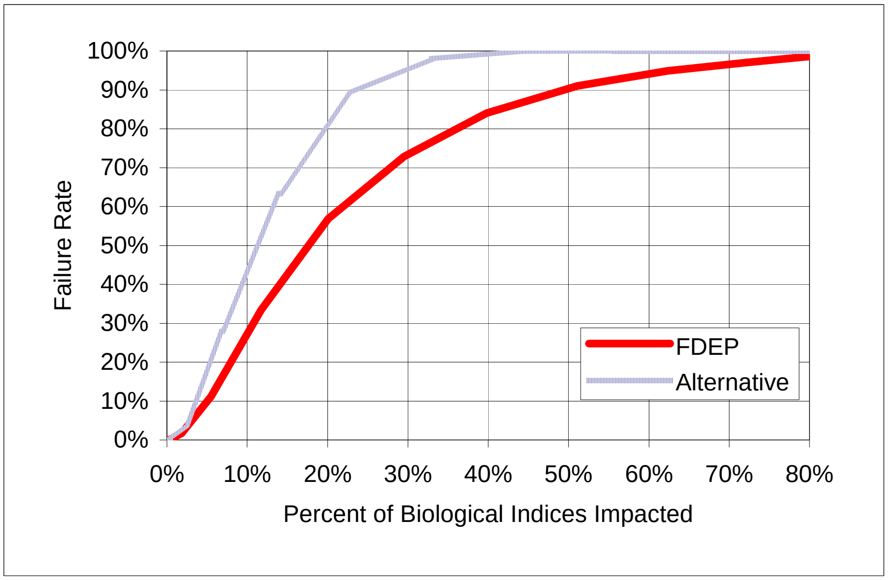


Figure 9 – Failure Rates with (FDEP) & without (Alternative) the 10 ppb Pass Feature

Eliminating the 10 ppb pass feature would increase power for detecting impacted stations and be consistent with the concept that impacts are driven more by 5-year average than by the 1-year average P concentration. It would also stabilize the pass/fail signal. That is, there would be less of a tendency for sites that have long-term means near the 10 ppb criterion to bounce in and out of compliance from one year to the next. Results of the modified test would tend to vary at 5-year intervals, but be relatively stable from one year to the next. This would provide a more stable signal for driving management decisions.

Either versions of the test would be reasonably effective for detecting impacted stations when applied on an annual basis, provided that marsh monitoring stations are located in areas that are likely to be impacted by external inputs (i.e. close to inflow points). The overall effectiveness of the program for implementing the criterion depend critically upon how the “report card” results are interpreted and linked to management decisions. These features have not been clearly defined.